COLECCIÓN
PENSAMIENTO JURÍDICO Y POLÍTICO CONTEMPORÁNEO

EL DERECHO ANTE EL RETO DE LAS TECNOLOGÍAS DISRUPTIVAS

ACTUALIDAD LEGISLATIVA Y JURISPRUDENCIAL

Coordinación
GONZALO CASTRO MARQUINA

COLEX

fundación Manuel Giménez Abad de Estudios Parlamentarios y del Estado Autonómico

GOBIERNO DE ARAGON

¡Gracias por confiar en nosotros!

La obra que acaba de adquirir incluye de forma gratuita la versión electrónica. Acceda a nuestra página web para aprovechar todas las funcionalidades de las que dispone en nuestro lector.

Funcionalidades eBook

Acceso desde cualquier dispositivo con conexión a internet

Idéntica visualización a la edición de papel

Navegación intuitiva

Tamaño del texto adaptable

Síguenos en:

5

EL DERECHO ANTE EL RETO DE LAS TECNOLOGÍAS DISRUPTIVAS

ACTUALIDAD LEGISLATIVA Y JURISPRUDENCIAL

GOBIERNO DE ARAGON

fundación Manuel Giménez Abad de Estudios Parlamentarios y del Estado Autonómico

COLECCIÓN
PENSAMIENTO JURÍDICO Y POLÍTICO CONTEMPORÁNEO

Director:

Guillermo **Vicente y Guerrero**
Profesor de Filosofía del Derecho en la Universidad de Zaragoza.

Consejo editorial:

Joachim **Rückert**
Catedrático emérito de Filosofía del Derecho en la Universidad de Frankfurt.

José Antonio **Escudero**
Catedrático emérito de Historia del Derecho. Universidad Nacional a Distancia de Madrid.

Michael **Lobban**
Catedrático de Historia del Derecho en la Universidad de Oxford.

Benjamín **Rivaya**
Catedrático de Filosofía del Derecho en la Universidad de Oviedo.

Dag **Michalsen**
Catedrático de Historia del Derecho en la Universidad de Oslo.

Jesús **Delgado Echeverría**
Catedrático emérito de Derecho civil. Universidad de Zaragoza.

Alessandro **Serpe**
Catedrático de Filosofía del Derecho en la Universidad G. D'Annunzio de Chieti y Pescara.

Ignacio **Fernández Sarasola**
Catedrático de Derecho constitucional en la Universidad de Oviedo.

Ignacio **Peiró Martín**
Catedrático de Historia contemporánea en la Universidad de Zaragoza.

Christoph-Eric **Mecke**
Catedrático de Filosofía del Derecho en la Universidad de Zielona Göra.

José Manuel Nieto **Soria**
Catedrático de Historia medieval en la Universidad Complutense de Madrid.

Tomàs **de Montagut i Estragués**
Catedrático emérito de Historia del Derecho en la Universidad Pompeu Fabra de Barcelona.

Carlo **Nitsch**
Catedrático de Filosofía del Derecho en la Universidad Federico II de Nápoles.

Pablo **Fernández Albaladejo**
Catedrático emérito de Historia Moderna en la Universidad Autónoma de Madrid.

José María **Portillo Valdés**
Catedrático de Historia Contemporánea en la Universidad del País Vasco.

Cristina **Vano**
Catedrática de Historia del Derecho en la Universidad Federico II de Nápoles.

José Ignacio **Solar Cayón**
Catedrático de Filosofía del Derecho en la Universidad de Cantabria.

Guillermo **Fatás Cabezas**
Catedrático emérito de Historia Antigua en la Universidad de Zaragoza.

Gert-Fredik **Malt**
Profesor emérito de Filosofía del Derecho en la Universidad de Oslo.

Raphael **Cahen**
Profesor de Historia del Derecho en la Universidad de Bruselas.

Óscar **Vergara Lacalle**
Profesor de Filosofía del Derecho en la Universidad de La Coruña.

Eduardo **Zimmermann**
Profesor de Historia Constitucional en la Universidad de San Andrés de Buenos Aires.

5

EL DERECHO ANTE EL RETO DE LAS TECNOLOGÍAS DISRUPTIVAS

ACTUALIDAD LEGISLATIVA Y JURISPRUDENCIAL

Coordinador

Gonzalo Castro Marquina

Con el apoyo y la colaboración de

María Belén Corcoy de Febrer
María Pilar López Rino
Elena Marquesán Díez

COLEX 2025

Copyright © 2025

Queda prohibida, salvo excepción prevista en la ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra sin contar con autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sigs. del Código Penal). El Centro Español de Derechos Reprográficos (www.cedro.org) garantiza el respeto de los citados derechos.

Editorial Colex S.L. vela por la exactitud de los textos legales publicados. No obstante, advierte que la única normativa oficial se encuentra publicada en el BOE o Boletín Oficial correspondiente, siendo esta la única legalmente válida, y declinando cualquier responsabilidad por daños que puedan causarse debido a inexactitudes e incorrecciones en los mismos.

Editorial Colex S.L. habilitará a través de la web www.colex.es un servicio online para acceder a las eventuales correcciones de erratas de cualquier libro perteneciente a nuestra editorial.

© Francisco Balaguer Callejón
© Moisés Barrio Andrés
© José Juan Castelló Pastor
© Gonzalo Castro Marquina
© Lorenzo Cotino Hueso
© Iban García del Blanco
© Ana Gascón Marcén
© M.ª José González Ordovás
© Miguel L. Lacruz Mantecón
© Joaquín Meseguer Yebra
© Elisa Moreu Carbonell
© Miguel Ángel Presno Linera
© M.ª Ángeles Rueda Martín
© Margarita Robles Carrillo
© Manuel Jesús Rodríguez Puerto
© Idoia Salazar García

© Editorial Colex, S.L.
Calle Costa Rica, número 5, 3.º B (local comercial)
A Coruña, 15004, A Coruña (Galicia)
info@colex.es
www.colex.es

SUMARIO

PRESENTACIÓN

Presentación. 15

PRÓLOGO

Prólogo 17

1.

LA CONFIGURACIÓN DE LOS DERECHOS EN LOS ENTORNOS DIGITALES

Francisco Balaguer Callejón

(Pág. 23)

I. Cuestiones metodológicas previas sobre derechos fundamentales y entornos digitales 23
II. La percepción del tiempo en la sociedad digital y la asimetría entre procesos comunicativos y procesos políticos 28
III. La disolución del espacio público a través de los algoritmos 30
IV. La pérdida de una visión social compartida de la realidad 36
V. La desinformación como rasgo estructural de los procesos comunicativos en la sociedad digital 40
VI. Conclusiones 42
Bibliografía 43

2.

PÚBLICO Y PRIVADO EN LA REGULACIÓN DEL METAVERSO

Moisés Barrio Andrés

(Pág. 45)

I. Introducción 45
II. Origen del metaverso 48
III. Concepto 49
IV. Aplicaciones y elementos 52

V. Algunos desafíos jurídicos. 56
VI. Metaverso, Estado y soberanía. 59
VII. Conclusiones. 65
Bibliografía 67

3.

EL REGLAMENTO (UE) DE SERVICIOS DIGITALES COMO INSTRUMENTO ESENCIAL EN EL NEW DEAL TECNOLÓGICO DE LA UNIÓN EUROPEA

José Juan Castelló Pastor

(Pág. 71)

I. Introducción: un nuevo marco de gobernanza digital en la unión europea 71
II. Aspectos generales del Reglamento (UE) de servicios digitales 73
2.1. Ámbito de aplicación 75
2.2. Consolidación del régimen de exención de responsabilidad e importantes novedades de la jurisprudencia del TJUE 77
III. Un ecosistema digital más transparente y seguro mediante obligaciones de diligencia debida asimétricas 79
3.1. Obligaciones básicas a todos los prestadores de servicios de intermediación 80
3.2. Obligaciones adicionales a los prestadores de servicios de alojamiento de datos, incluidas las plataformas en línea 82
3.3. Más disposiciones de obligada observancia a las plataformas en línea 83
3.4. Obligaciones adicionales de gestión de riesgos sistémicos para las plataformas en línea y buscadores de muy gran tamaño. 93
IV. Mecanismos de tutela jurídico-pública y privada en el Reglamento (UE) de servicios digitales 96
V. Consideración final 102
Bibliografía 102

4.

EL IMPACTO DE LOS DERECHOS DE PROPIEDAD INTELECTUAL EN EL ENTRENAMIENTO DE LOS MODELOS DE INTELIGENCIA ARTIFICIAL GENERATIVA

Gonzalo Castro Marquina

(Pág. 107)

I. Aunque escuchen a Mozart, las máquinas no se inspiran 107
II. Siempre hubo un derecho de las IA 124
III. ¿Es demasiado pequeña la propiedad intelectual? 137
IV. La excepción de minería no basta para excavar esta veta 155
V. Un breve apunte sobre la doctrina del fair use y su jurisprudencia 183
VI. Conclusiones 190
Bibliografía 191

5.
DESENTRAÑANDO EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA: UNA GUÍA ESENCIAL

Lorenzo Cotino Hueso

(Pág. 193)

I. A modo de presentación 194
II. La primera regulación global de la inteligencia artificial: un marco pionero y sus elementos clave 196
2.1. La aceleración de la regulación mundial de la IA al tiempo del Reglamento de la Unión Europea 196
2.2. Una visión general del Reglamento: fundamentos y disposiciones generales 198
III. Qué IA se considera inaceptable y prohibida en la Unión Europea 202
IV. El modelo de riesgos: solo se quiere regular la IA más peligrosa, de «alto riesgo» 203
4.1. El modelo de riesgos para regular la inteligencia artificial 203
4.2. Identificación de sistemas de IA de alto riesgo: ¿Cuáles son y por qué? ... 205
V. El «corazón» del reglamento: ¿qué obligaciones deben cumplir los sistemas de alto riesgo? 206
5.1. Obligaciones esenciales que deben cumplir los sistemas de IA parar ser seguros 206
5.2. Responsabilidades de proveedores, implementadores y otros actores clave 208
5.3. La evaluación de conformidad que se obtiene si se cumplen estándares y normas y el registro en la base de datos de sistemas de alto riesgo 209
VI. ¿Qué debemos saber los humanos cuando interactuamos con determinados sistemas de IA? (artículo 50) 210
VII. Obligaciones respecto de los modelos de IA general como «ChatGPT». ... 211
VIII. Innovación bajo control: sandboxes y otras medidas para fomentar la IA .. 212
IX. Gobernanza de la IA en Europa: ¿quiénes supervisan el cumplimiento del reglamento en España y en la Unión? 213
X. ¿Cómo se supervisan los sistemas de IA después de su comercialización ... 215
XI. Otros aspectos del reglamento: derechos, sanciones y aplicación gradual, códigos de conducta y delegaciones a la comisión 217
11.1. ¿Reconoce el Reglamento derechos a los afectados por la IA? 217
11.2. Códigos de conducta: Estándares voluntarios para sistemas no regulados 218
11.3. Régimen sancionador: ¿cuáles son las consecuencias del incumplimiento? 218
11.4. Delegación de poderes a la Comisión para ir adaptando el Reglamento .. 219
11.5. Disposiciones finales y la (lenta) aplicación del Reglamento 219
XII. Para acabar, también en 2024 un convenio de inteligencia artificial del consejo de Europa 220
Bibliografía 221

6.

EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL: EL SER HUMANO COMO MEDIDA DE TODAS LAS COSAS

Iban García del Blanco

(Pág. 225)

I. El proceso regulador europeo: primeros pasos, la «soberanía digital europea». . . 225
II. El reglamento europeo de inteligencia artificial, *iter* procedimental. 227
III. La ley de inteligencia artificial: contenido y estructura. 228
3.1. Usos prohibidos. 230
3.2. Usos de alto riesgo. 231
3.3. La inteligencia artificial generativa. Obligaciones adicionales de transparencia para ciertos usos de IA. 232
IV. Otros contenidos destacables de la *AI act* 233
V. Gobernanza e implementación 234
VI. Régimen sancionador 236
VII. ¿Es Europa una isla de regulación en un mundo sin reglas?. 237
VIII. Punto y seguido: estamos ante una norma viva 239

7.

LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL EN ESTADOS UNIDOS Y CHINA, MODELOS REGULATORIOS ALTERNATIVOS AL EUROPEO

Ana Gascón Marcén

(Pág. 243)

I. Introducción 243
II. La regulación de la inteligencia artificial a nivel internacional. 248
III. La regulación de la inteligencia artificial en Estados Unidos 252
IV. La regulación de la inteligencia artificial en China 257
V. Conclusiones 265
Bibliografía 266

8.

INTELIGENCIA ARTIFICIAL: MUCHO MÁS QUE CIENCIA Y TECNOLOGÍA PARA EL DERECHO Y LA DEMOCRACIA

M.ª José González Ordovás

(Pág. 273)

I. Lo mejor y lo peor. 273
II. A propósito de la dimensión mitológica del derecho 279
III. Sobre la función civilizatoria del derecho 284
3.1. ¿Sigue siendo posible el pensamiento complejo? Algunas consideraciones sobre la «tecnociencia» 288

IV. Personas o máquinas ¿quién y por qué debe hacer justicia? 293
4.1. Inteligencia artificial: una vuelta de tuerca al dilema formalismo-anti-formalismo jurídico 300
V. Algoritmos, valores y sesgos: algunos problemas 303
Bibliografia 314

9.

INTELIGENCIA ARTIFICIAL ¿SUEÑO O PESADILLA?

Miguel L. Lacruz Mantecón

(Pág. 319)

I. La inteligencia artificial y el derecho 319
II. La IA y la indemnización de daños 322
III. Las aplicaciones de la IA y el posible daño a los derechos fundamentales 326
3.1. Los sistemas de datos biométricos e identificativos, en general 326
3.2. Aplicaciones de los datos biométricos 327
3.2.1. Sistemas de identificación y localización de personas 327
3.2.2. Sistemas de perfilado predictivo del comportamiento 329
3.2.3. Sistemas de detección de emociones 330
3.2.4. La persuasión y la alteración subliminal de comportamientos 331
3.2.5. Sistemas de categorización biométrica 333
IV. Incidencia de la IA en la vulneración de derechos fundamentales 333
4.1. El sistema de crédito social chino 333
4.2. La falsificación de la realidad 335
4.3. La imitación de resultados humanos inteligentes 337
V. A modo de conclusión 338
Bibliografía 339

10.

LA TRANSPARENCIA ANTE EL NUEVO ENTORNO DIGITAL DE LAS ADMINISTRACIONES PÚBLICAS

Joaquín Meseguer Yebra

(Pág. 341)

I. Introducción 341
II. Los derechos digitales ante la IA 343
III. Gestión automatizada motivada (y explicable) 345
IV. La realidad: la imprescindible tutela de los órganos de control del acceso a la información pública 353
V. El futuro: el reglamento de inteligencia artificial 357
Bibliografía 360

11.

REFLEXIONES SOBRE LOS NEURODERECHOS

Elisa Moreu Carbonell

(Pág. 363)

I. ¿Qué son las neurotecnologías? 364
II. ¿Qué relación existe entre la inteligencia artificial y las neurotecnologías? .. 365
III. ¿Qué retos plantea el avance de estas neurotecnologías?..... 365
IV. ¿Las neurotecnologías pueden manipular los cerebros y controlar los pensamientos? 366
V. ¿Qué son los neuroderechos?..... 368
VI. ¿Quién debería regular estos neuroderechos? 369
VII. ¿Qué país(es) reconocen los neuroderechos? 371
VIII. ¿Ha reconocido España los neuroderechos?..... 373
IX. En conclusión, ¿protege el derecho nuestros cerebros? 375
Bibliografía 379

12.

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EUROPA: EL REGLAMENTO DE LA UNIÓN EUROPEA Y CONVENIO MARCO DEL CONSEJO DE EUROPA

Miguel Ángel Presno Linera

(Pág. 383)

I. Introducción 383
II. Las iniciativas para la regulación supranacional de la inteligencia artificial .. 384
2.1. En la Unión Europea 384
2.2. En el Consejo de Europa 388
III. ¿De qué se habla cuando se habla de inteligencia artificial en la normativa europea? 389
IV. Los principios que inspiran la regulación de la inteligencia artificial en Europa..... 392
V. Un enfoque de la regulación de la inteligencia artificial basado en los riesgos .. 395
VI. Las diferencias estructurales entre el Reglamento de la Unión Europea y el Convenio Marco del Consejo de Europa 398
VII. Algunas conclusiones sobre la normativa europea de inteligencia artificial y los derechos fundamentales 401
Bibliografía 403

13.

CONSIDERACIONES POLÍTICO CRIMINALES SOBRE LA PROTECCIÓN PENAL DE LA VIDA PRIVADA EN LA ERA DIGITAL

M.ª Ángeles Rueda Martín

(Pág. 405)

I. Vida privada e intimidad: dos conceptos estrechamente vinculados. 405
II. Elementos comunes que integran la protección penal dispensada a la intimidad en el código penal de 1995 410
III. Sobre la necesidad político-criminal de ampliar el objeto de protección en los delitos contra la intimidad y la propia imagen 415
IV. Algunas observaciones críticas a la regulación penal en los delitos contra la intimidad y la propia imagen. 423
Bibliografía 426

14.

EL MARCO NORMATIVO DE LA DIRECTIVA (UE) 2022/2555: DE LA SEGURIDAD DE REDES Y SISTEMAS A LA CIBERSEGURIDAD INTEGRAL

Margarita Robles Carrillo

(Pág. 433)

Antecedentes. 433
I. Objeto y marco general 434
II. Ámbito de aplicación. 437
2.1. Criterios de aplicación. 437
2.2. Tipología 438
2.3. Integración sectorial 440
III. Régimen jurídico 440
3.1. Medidas de gestión 441
3.2. Obligaciones de notificación 441
3.3. Medidas adicionales 442
IV. Control y garantía de cumplimiento 442
4.1. Jurisdicción. 443
4.2. Registro y bases de datos 443
4.3. Supervisión y ejecución 444
V. Implementación 447
5.1. Directrices relativas a las obligaciones del art. 3.4. 447
5.2. Directrices relativas a las obligaciones de los art. 4.1 y 2. 448
5.3. Reglamento de Ejecución 2024/2690 449
VI. Conclusiones 453

15.

EL DEBATE SOBRE LOS DERECHOS DE LA INTELIGENCIA ARTIFICIAL

Manuel Jesús Rodríguez Puerto

(Pág. 457)

I. La personalidad electrónica 457
II. La realidad de la IA 459
III. La imposibilidad de derechos para IA 468
Bibliografía 475

16.

INTELIGENCIA ARTIFICIAL: REGULACIÓN Y ÉTICA PARA LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES

Idoia Salazar García

(Pág. 479)

I. Inteligencia artificial, ventajas y oportunidades 479
II. Riesgos y retos reales de la inteligencia artificial 482
2.1. La explicabilidad 483
2.2. Responsabilidad 484
2.3. El futuro del trabajo y de las prestaciones sociales 485
2.4. Relaciones personas-IA 485
2.5. Concentración de poder y riqueza en unas pocas empresas muy grandes 486
2.6. El uso malicioso de la inteligencia artificial 486
III. La ética en la IA 487
IV. Principios éticos de la inteligencia artificial 489
V. Normativa más allá de la ética 491
Bibliografía 493

PRESENTACIÓN

Cuando pensamos en inventos, lo normal es que evoquemos imágenes como las del teléfono, el avión o el ordenador, por ejemplo. Pensar en el Derecho es poco habitual cuando se alude a los inventos. Pero el Derecho es igual de artificial que todas esas otras invenciones. El Derecho no crece de los árboles ni surge de la nada: es fruto del ingenio humano y, como otros instrumentos de los que nos hemos dotado, sin él nuestro progreso no hubiera sido posible.

El Derecho nos ha ayudado a avanzar, al articularse como una vía de solución de conflictos, alternativa a la fuerza. Es signo de civilización: donde hay buenas leyes y se cumplen, la sociedad y la economía funcionan mejor.

La seguridad jurídica, según múltiples estudios, es un elemento que atrae la inversión económica porque genera confianza. Del mismo modo que vivir con la certeza de que el fuego quemará y no helará, el hecho que las leyes se ejecuten conforme a su tenor anima a acudir a un país o Comunidad Autónoma.

Aragón, en ese sentido, siempre ha sido un territorio comprometido con el Derecho. Uno de sus signos de identidad como con tanto acierto expresó Joaquín Costa, o como recuerda nuestro Estatuto de Autonomía. No es solo que dispongamos de un Derecho foral propio, sino que creemos en el valor del Derecho.

Dentro de la construcción del Aragón del futuro, como sociedad hemos decidido apostar por las nuevas tecnologías. Son numerosas las iniciativas privadas y públicas que van en esta dirección y esperamos que no dejen de crecer. El Gobierno de Aragón se esfuerza en ello, dentro de las palancas seguras del progreso futuro. Pero, en esta apuesta, no olvidamos nuestras viejas tecnologías, entre ellas el Derecho.

Resulta imprescindible comprender el potencial técnico y económico de los avances que estamos experimentando, pero también su dimensión jurídica para valorar si el Derecho vigente es capaz de responder bien a los retos y oportunidades que nos brindan, o para identificar cuándo es bueno regular

y cuándo se puede confiar en la capacidad de autorregulación de los particulares. Esta obra es una mezcla de las viejas y las nuevas tecnologías, del presente, el pasado y el futuro de Aragón.

Para su elaboración, se ha contado con la valiosa colaboración de la Fundación Manuel Giménez Abad, una institución de la que nos sentimos muy orgullosos todos los aragoneses, y que honra permanentemente la memoria de una persona sin parangón.

El lector encontrará en las páginas siguientes las voces de expertos destacados en sus respectivos campos. Los 16 capítulos del libro tratan los más recientes reglamentos europeos en materia de Derecho digital y singularmente de inteligencia artificial, así como una lectura en Derecho comparado y perspectivas que empiezan a asomar por el horizonte, como el ámbito de los neuroderechos. Todo ello, con perspectiva humanista, diversidad de enfoques y rigor.

Enhorabuena a todos los autores de esta publicación. Espero, sinceramente, que disfruten de su lectura y consulta, porque estas páginas son el reflejo, una vez más, de la vocación sincera de Aragón por el Derecho, como instrumento sobre el que se construye la convivencia social y el progreso económico, cultural, tecnológico y científico.

María del Mar Vaquero Periánez
Vicepresidenta del Gobierno de Aragón

PRÓLOGO

La inteligencia artificial (IA en adelante) se ha instalado en nuestra rutina con una velocidad que sería sorprendente si no fuese porque es el auténtico rasgo distintivo de nuestro tiempo. Su consolidación como herramienta «popular» es ya un hecho incuestionable. También lo es la constatación de su capacidad para transformar hábitos; formas de trabajar; y, en general, nuestra forma de vida. Su impacto es transversal, generalizado y profundo. La única certeza es que las incógnitas superan a las certezas. El futuro esconde aún muchas sorpresas.

Juristas, politólogos y sociólogos no tardaron en comprender que su ámbito propio de estudio se veía fuertemente alterado por la irrupción de la nueva herramienta tecnológica. Cuando no se ha llegado a aprehender con suficiencia el impacto de la generalización de Internet y de las redes sociales, la tecnología vuelve a golpear y lo hace con una herramienta con una capacidad disruptiva desconocida hasta la fecha. Si bien, como se ha indicado, las incógnitas sobre el presente y, más aún sobre el futuro, incluso inmediato, son múltiples, casi desde su irrupción en la forma de aplicaciones libres, la academia ha enfrentado los dilemas que plantea y ha intentado darles respuesta. En este punto, me interesa destacar como, a mi juicio, en esa voluntad de enfrentar los retos de la IA ha habido mucho de intuitivo. Nada hay de peyorativo en el uso de esta expresión. El científico, consciente de lo mucho que desconocía, ha entendido que, en cualquier caso, se encontraba ante un hecho histórico que exigía comenzar por realizar las preguntas pertinentes para tratar de avanzar en las respuestas. Como tantas veces, el primer paso es detenerse y comenzar a realizar las preguntas pertinentes. De acuerdo con la transcendencia y características del fenómeno, esas preguntas son múltiples y surgen desde prácticamente todos los ámbitos del quehacer humano. En este volumen, encontrará el lector reflexiones que abarcan desde el derecho a la organización política pasando por la ética y los derechos. La simple lectura del índice da cuenta de la transcendencia del reto que enfrentamos como humanidad que desea preservar unos valores que garantizan una convivencia civilizada.

Tres son los ejes sobre los que se construyen los trabajos que conforman esta obra. El primero de ellos es la relación entre la IA y el Derecho. Son mayoría los trabajos que, de una u otra manera, abordan cuestiones referidas al reto que para el legislador supone la irrupción de una herramienta con estas características. Dos son los planos fundamentales en los que se desenvuelven los trabajos. Por un lado, la regulación en sentido estricto de la IA; por otro, su incidencia sobre el estatus de derechos y libertades. Desde hace tiempo, defiendo que el tránsito de la era analógica a la digital supone un movimiento acelerado y profundo de las placas tectónicas que han sostenido durante siglos no sólo el orden institucional. También una forma de vivir, una convivencia. Elemento vertebrador de ese orden ha sido y es el Derecho. La ley adquirió una naturaleza casi mágica para convertirse en garantía, instrumento de transformación social y generador de un orden conforme a la voluntad general.

Si nos centramos en el Derecho público, dos condiciones emergían como esenciales para que la Ley, y el ordenamiento en su conjunto, pudiese cumplir esas funciones. Por un lado, la estabilidad de los objetos que eran sometidos a su regulación. Por otro, la existencia de un lenguaje que permitía tanto aprehender correctamente la realidad como hacer del Derecho un universo comprensible, en principio, para todos. No en vano, la publicidad y claridad de las normas eran presupuesto de su legitimidad. El movimiento de placas que se subrayaba con anterioridad subvierte estas dos premisas. Si algo caracteriza la incipiente era digital es la velocidad. Todo transcurre y se transforma a una velocidad extraordinaria. Literalmente, es casi imposible realizar cualquier ejercicio predictivo. Junto a ello, el legislador se encuentra sin lenguaje. Tiene que recurrir a voces técnicas, muchas veces auténticos neologismos, hostiles para el jurista y, lo más relevante, incomprensibles para el ciudadano.

La consecuencia es que el Derecho público ha perdido capacidad de incidencia sobre la realidad. Pocas materias ejemplifican con mayor claridad estas dificultades que la IA. Y pocas materias pueden requerir más de un marco normativo adecuado. Normas que sometan su uso a las exigencias propias de un Estado social y democrático de derecho. Como se indicó, la incidencia de la IA es transversal y su impacto tiene una profundidad que aún no llegamos a imaginar. Someter a un Derecho que vele por los principios y valores que son sustento de nuestro orden político a una herramienta tan relevante, es una necesidad evidente. El lector tiene la posibilidad de leer un conjunto de trabajos que enfrentan este reto y que, desde la experiencia, nos enseñan que, al menos, hay mínimos que se pueden y deben alcanzar.

Dos cuestiones adicionales que añaden dificultad a la labor del legislador emergen en la obra. Por un lado, la dimensión necesariamente trasnacional de cualquier intento de regulación eficaz de la IA; por otro, la transformación que su consolidación, y en general el nuevo escenario tecnológico, suponen para la relación entre lo público y privado. La IA consuma el empoderamiento de un sector privado aquí personificado en unas determinadas compañías

tecnológicas que alteran las relaciones tradicionales de poder y condicionan de manera notable la eficacia de las normas que puedan aprobarse.

No menor interés tiene la relación entre IA y derechos fundamentales y libertades. No hay ocasión de desarrollar ni siquiera someramente esta cuestión. El lector encontrará trabajos específicos sobre este extremo y numerosas alusiones en otros trabajos. Por mi parte, creo que es preciso llamar la atención sobre el impacto que el nuevo contexto tecnológico y, en particular, la IA, tiene sobre los derechos y libertades. Sobre los existentes y por la necesidad de generar nuevos derechos. Derechos que han de cumplir una doble función: por un lado, servir de garantía y protección frente a los riesgos del mal uso de esas tecnologías; por otro, servir de instrumento para que todos los ciudadanos puedan disfrutar de las ventajas que ofrecen en condiciones de igualdad. La relación del nuevo universo tecnológico con los derechos y libertades es una cuestión recurrente en la academia. Pero todos los esfuerzos son necesarios. Porque es preciso seguir el ritmo de la evolución tecnológica y porque debe seguir manteniéndose la alerta frente a los poderes públicos.

En el volumen que el lector tiene en sus manos dos cuestiones adicionales están presentes. Ambas son objeto de atención singular y ambas sobrevuelan todos los trabajos. Me refiero a la dimensión ética de la IA y a su impacto sobre el sistema democrático. Lo referido hasta ahora bien puede proyectarse sobre estas cuestiones. Pero ambas exigirían un desarrollo singular que no es posible realizar en estas palabras preliminares. En todo caso, es necesario realizar una reflexión sobre las mismas.

La aludida dificultad del Derecho público para cumplir sus funciones tradicionales en la sociedad digital es ya un obstáculo para articular una relación armónica entre este modelo social y la democracia, al menos tal y como se concibe hasta ahora. El Derecho lejos de limitarse a ser una herramienta técnica o ser soporte del Estado de derecho es condición para la existencia de una sociedad democrática. El principio democrático se hace realidad mediante la aprobación de leyes que expresan la voluntad general. Si el Derecho es incapaz de cumplir su función de garantía y transformación, el sistema democrático verá como una grave fisura quiebra su estructura institucional. Lo mismo sucede con la transformación de la relación entre lo público y lo privado o con el rediseño de un nuevo estatus de derecho y libertades.

Las dudas y retos enunciados no significan que la era digital sea incompatible con la democracia o que la someta a un riesgo extremo. Se trata, simplemente, de constatar que los presupuestos han cambiado. No es posible seguir articulando las estructuras institucionales y las dinámicas políticas como si no sucediese nada. El riesgo real es la pasividad. Si se sigue actuando conforme a una realidad inexistente, los problemas serán inevitables. La era digital tiene sus exigencias particulares para un adecuado funcionamiento de un sistema democrático. El reto es doble: realizar un adecuado diagnós-

tico de carencias y necesidad y diseñar un sistema que cumpla los requisitos esenciales de un sistema democrático y aproveche las ventajas, que las hay e importantes, del nuevo orden.

Los retos éticos de la consolidación de la IA son extraordinarios y expresivos de la transcendencia de esta herramienta. Su incidencia horizontal, su capacidad para transformar todos los quehaceres, provoca una inevitable alteración de los presupuestos sobre los que hemos asentado una forma de vivir. Desde las relaciones laborales a la ya aludida relación y organización del poder, los cambios son profundos e inevitables. En la mayoría de las ocasiones, el reto no es tanto dibujar un nuevo código como idear la manera para preservar aquellos valores que entendemos esenciales para preservar aquello que garantiza los mínimos necesarios de igualdad, justicia y libertad. No será una tarea sencilla. La IA emerge como un gigante todopoderoso reacio a ser sometido a normas y pautas de comportamiento. La tutela sobre el mismo escapa en la mayoría de las ocasiones a los poderes públicos. Por todo ello, el desafío es tan importante como complejo. Pero nada debe hacer desistir de enfrentar los problemas éticos a los que enfrenta la IA. De alguna manera, enfrentar con éxito esos retos es presupuesto de otras tareas pendientes.

Democracia, Derecho, Ética. El mero enunciado de estas voces es expresivo del alcance de los desafíos a los que enfrenta la era digital. El diseño institucional, la norma jurídica y la ética/filosofía deberá enfrentarlos desde la conciencia de su interrelación. Aunque parezca obvio, lo primero es tomar conciencia del desafío. Es propio de nuestro tiempo que la tecnología avance a una velocidad superior a la que se formula el soporte humanista que la debe ordenar. Este decalaje es uno de los grandes desafíos de nuestros días. El desarrollo tecnológico es extraordinario y como tal expresivo de la capacidad del ser humano para las más altas conquistas. Pero, como siempre en la historia, las consecuencias deben ser moduladas por criterios humanistas que las inserten con plenitud al servicio de los objetivos propios de un orden que cumpla los objetivos inherentes a la dignidad humana y a la protección efectiva de los derechos y libertades.

La IA es una extraordinaria oportunidad para mejorar de manera efectiva y universal las condiciones de vida de la mayoría de los seres humanos. Desde la medicina a la educación, sus ventajas se antojan generalizadas. Su presente es ya deslumbrante y su evolución hace pensar en objetivos hoy impensables. Como sucede con otras tecnologías de la era digital, tiene la inmensa ventaja de un coste que hace fácilmente generalizable su uso, al menos en determinados estadios. Lógicamente, como cualquier herramienta tecnológica, la IA abre la puerta a problemas e, incluso, algunos peligros. No hay avance que no haya planteado esta doble cara. Tener la posibilidad de disponer de la IA es indudablemente positivo. En manos de quienes la utilicen está aminorar los problemas y maximizar las ventajas. Es natural que una herramienta tan poderosa, que en ocasiones puede provocar la sensación de superar al usua-

rio, de miedo. Frankenstein es un mito recurrente. Los problemas, las dudas, los errores en su utilización son inevitables. Su uso es un aprendizaje continuo. Por ello es preciso reflexionar, diagnosticar y generar las herramientas correspondientes que prevengan y en su caso aminoren las consecuencias negativas. Y que, simultáneamente, sirvan para obtener el máximo beneficio posible de su uso.

La IA es el penúltimo hito en el tránsito de la era analógica a la era digital. Su singularidad es extraordinaria y exige y merece estudios individualizados. Pero su adecuada comprensión necesita de una visión de conjunto. Cada día que pasa, la realidad se separa más de aquella que sirvió de premisa para organizar la convivencia de una manera determinada. El desafío es encontrar el diseño y los valores que la han de soportar en la era digital. A ese desafío quiere servir la publicación que el lector tiene en sus manos.

José Tudela Aranda
Secretario General de la Fundación Manuel Giménez Abad

1.

LA CONFIGURACIÓN DE LOS DERECHOS EN LOS ENTORNOS DIGITALES

Francisco Balaguer Callejón
Catedrático de Derecho Constitucional
Universidad de Granada.

SUMARIO: I. Cuestiones metodológicas previas sobre derechos fundamentales y entornos digitales. II. La percepción del tiempo en la sociedad digital y la asimetría entre procesos comunicativos y procesos políticos. III. La disolución del espacio público a través de los algoritmos. IV. La pérdida de una visión social compartida de la realidad. V. La desinformación como rasgo estructural de los procesos comunicativos en la sociedad digital. VI. Conclusiones.

I. Cuestiones metodológicas previas sobre derechos fundamentales y entornos digitales

El mundo digital se está construyendo sobre bases que en gran medida son diferentes del analógico. Por ese motivo, no es posible analizar el mundo digital con la misma perspectiva metodológica con la que analizamos el analógico. Ni los factores de poder son los mismos, ni la configuración de la realidad de la sociedad digital se corresponde con la realidad física. Así, por ejemplo, mientras esta última sigue regulada en gran medida a través del Estado mediante el derecho público (al menos por lo que se refiere a los derechos fundamentales) la primera se está configurando por grandes agentes globales, compañías tecnológicas, que utilizan el derecho privado y que atienden esencialmente a su modelo de negocio y a sus intereses económicos.

Es necesario, por tanto, un nuevo acercamiento metodológico a la realidad digital en lo que se refiere a los derechos fundamentales o la democracia. Para sintetizarlo, he definido esa nueva perspectiva con el concepto de «la

constitución del algoritmo»[1] que es una propuesta metodológica de análisis de la realidad digital que tiene en cuenta las transformaciones culturales[2] que se están produciendo debido al desarrollo tecnológico y que están cambiando nuestra percepción del espacio y del tiempo, las dos variables en las que se desarrollan nuestras vidas. Algunas de estas transformaciones son necesarias porque van unidas a ese desarrollo de manera indisoluble, por lo que resultaría difícil pensar en alternativas que fueran compatibles con el mantenimiento de nuestros estándares de vida. Otras, por el contrario, son transformaciones culturales «adicionales», producidas como consecuencia directa del modo en que la tecnología está siendo configurada por los agentes globales, las grandes compañías tecnológicas, en función de su modelo de negocio[3]. Hay alternativas, aunque supongan un menor rendimiento económico para estas compañías, en un contexto en el que la balanza entre los beneficios que obtienen y el daño que provocan a la sociedad está claramente descompensada.

Este planteamiento metodológico intenta superar las perspectivas que todavía piensan que es posible analizar los derechos fundamentales en el entorno digital como si este fuera un simple reflejo del analógico, de tal manera que bastaría con trasladarlos de una realidad a otra sin más[4]. Es algo que vemos, por ejemplo, en algunas de las Cartas de derechos digitales (que deben ser valoradas muy positivamente, por lo demás) como la primera de ellas, la portuguesa de 2021[5], en la que se afirma en su artículo 2 que «Las normas que en el ordenamiento jurídico portugués consagran y tutelan dere-

1 Balaguer Callejón, F., *La constitución del algoritmo*, Fundación Manuel Giménez Abad, 2.ª edición, Zaragoza, 2023. Versión portuguesa: *A constituição do algoritmo*, Editora Forense, Rio de Janeiro, 2023, versión italiana: *La costituzione dell'algoritmo*, Le Monnier Università/Mondadori, Milano, 2023, versión francesa en curso de publicación.

2 *Cfr*. Balaguer Callejón, F., «Inteligencia artificial y cultura constitucional», en F. Balaguer Callejon, Ingo W. Sarlet (directores): *Derechos fundamentales y democracia en el constitucionalismo digital*, Aranzadi, Cizur Menor, 2023. Disponible en Internet en versión italiana: «Intelligenza artificiale e cultura costituzionale», en *Diritti fondamentali e democrazia nel costituzionalismo digitale* a cura di Balaguer Callejón, F. – I. W. Sarlet, *Iacittadinanzaeuropea*online, Numero Speciale 2, Maggio-Agosto *2023*. *Cfr.* igualmente, Balaguer Callejón, F., *La cultura constitucional en la era digital*, en J. M. Pérez Collados, (Coord.) *La cultura jurídica en la era digital*, Aranzadi, Cizur Menor, 2022.

3 *Cfr*. Balaguer Callejón, F., *La constitución del algoritmo y las transformaciones culturales en la sociedad digital*, en F. Caamaño y D. Jove Villares, *Tecnologías abusivas y Derecho*, Tirant lo Blanch, Valencia, 2024, pp. 49 y ss.

4 Sobre la metodología de análisis de la Inteligencia Artificial es fundamental el artículo de Bustos Gisbert, R. «El constitucionalista europeo ante la inteligencia artificial: reflexiones metodológicas de un recién llegado», *REDC*, n. 131, mayo-agosto de 2024.

5 La *Carta Portuguesa de Direitos Humanos na Era Digital* de 17 de mayo de 2021 aprobada por medio de Ley, tiene valor normativo. La Carta se puede encontrar en: https://files.diariodarepublica.pt/1s/2021/05/09500/0000500010.pdf

chos, libertades y garantías, son plenamente aplicables en el ciberespacio»[6]. Pero, si eso fuera así, en muchos derechos bastaría con este precepto para dejar sin sentido el resto de los preceptos de la carta porque muchos de los derechos que en ella se contienen se derivan ya de los derechos contenidos en la Constitución o en los tratados internacionales, o en el derecho europeo. Precisamente porque no es así es por lo que las cartas de derechos digitales son necesarias.

De un modo similar, aunque con matices, se manifiesta la Carta española, igualmente de 2021[7], en la que se afirma en su artículo 1.1 que «Los derechos y libertades reconocidos en la Declaración Universal de Derechos Humanos, la Constitución Española, el Convenio Europeo de Derechos Humanos, la Carta de los Derechos Fundamentales de la Unión Europea, y los tratados y acuerdos internacionales sobre las mismas materias ratificados por España son aplicables en los entornos digitales»[8].

Por su parte, la *Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital*, proclamada a través de una declaración conjunta del Parlamento Europeo, el Consejo y la Comisión Europea el 23 de enero de 2023[9], indica en el apartado b) de su artículo 1, en relación con las instituciones europeas que la suscriben que «Nos comprometemos a: b) adoptar las medidas necesarias para que los valores de la UE y los derechos de los ciudadanos reconocidos por el Derecho de la Unión se respeten tanto en línea como fuera de línea», lo que por un lado implica una actitud diferente frente al mundo digital porque no pretende que haya una vinculación automática del mundo digital a los derechos «analógicos», por caracterizarlos de alguna manera, sino que reconoce que es necesario un compromiso para establecer las medidas que sean necesarias para asegurar que esos derechos sean respectados también en el ámbito digital. Pero, por otro lado, se sigue basando en la idea de que los derechos fundamentales «analógicos» reconocidos por el Derecho de la Unión pueden funcionar igual en el mundo analógico y en el digital.

---

6 «As normas que na ordem jurídica portuguesa consagram e tutelam direitos, liberdades e garantias são plenamente aplicáveis no ciberespaço».

7 La *Carta de los Derechos Digitales de España*, publicada en julio de 2021 no tiene carácter normativo. La Carta se puede encontrar en: https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

8 Los matices se reflejan en el apartado 2 del artículo 1, de acuerdo con el cual «Se impulsarán las medidas necesarias y conducentes para que las leyes puedan concretar, en cuanto sea necesario, las especificidades de los derechos en el entorno digital». Es decir, el apartado 2 del artículo 1 viene a modular lo indicando en el apartado 1, reconociendo que no siempre serán aplicables porque el mundo digital y el analógico, son diferentes.

9 Que no tiene valor normativo, de acuerdo con lo que se afirma en el apartado 10 de su Preámbulo: «Tiene carácter declarativo y, por tanto, no afecta al contenido de las normas jurídicas ni a su aplicación». La declaración se puede encontrar en https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32023C0123(01).

Por último, la *Carta Iberoamericana de Principios y Derechos en los Entornos Digitales*[10], aprobada en la XXVIII Cumbre Iberoamericana de Jefes de Estado y de Gobierno, el 25 de marzo de 2023, por 22 Estados, parte de un planteamiento en parte diferente a las anteriores y comienza por reconocer en su apartado primero que «El ejercicio de los derechos garantizados por las legislaciones nacionales vigentes en el entorno físico se ve, de una forma u otra, retado, afectado u omitido en los entornos digitales». Se indica también en la Carta que «Los derechos de todas las personas deben ser garantizados, respetados y protegidos en los entornos digitales» aunque ya no se trata de un reflejo automático ni de un funcionamiento igual en el mundo digital respecto del analógico sino que los Estados firmantes se comprometen a «Garantizar que el respeto, promoción y protección de los derechos y el cumplimiento de los deberes recogidos en nuestras constituciones y marcos jurídicos se interpretarán, aplicarán y ejercerán en los entornos digitales, teniendo en cuenta sus particularidades».

Estas cartas y declaraciones sobre los derechos digitales realizan un esfuerzo muy notable para intentar acercar el sistema de derechos analógico al mundo digital[11]. Pero el planteamiento metodológico no puede basarse en una identidad esencial entre los dos mundos que no existe. Es necesario reconocer la diversidad entre el estatuto de los derechos en el ámbito analógico/offline y el del ámbito digital/online. En las Cartas digitales hay un optimismo quizás excesivo en algunas de ellas (la portuguesa y la española) y una formulación más cauta en las declaraciones supranacional europea y en la iberoamericana. Pero ni siquiera la última que hemos visto reconoce esa diversidad porque se limita a hablar de «particularidades» del entorno digital que deben ser tenidas en cuenta.

Lo que ocurre es que no se trata simplemente de especificidades o de particularidades sino que es necesario ser conscientes de que estamos ante dos mundos diferentes, con una estructura distinta, con relaciones de poder que no se corresponden con las del ámbito nacional, básicamente por dos motivos: porque el mundo digital está siendo configurado por grandes agentes

10 La Carta tampoco tiene valor normativo, según se indica en su Introducción: «La presente Carta, cuyo carácter es declarativo y no vinculante y por tanto no afecta el contenido ni la aplicación de normas jurídicas internas de los Estados, tiene por objeto promover principios comunes para que sean tomados en cuenta por los Estados al momento de adoptar o adecuar las legislaciones nacionales o poner en marcha políticas públicas relacionadas con la protección de los derechos y el cumplimiento de los deberes en entornos digitales, así como por las empresas, la sociedad civil y la academia a la hora de desarrollar y aplicar tecnologías, colocando a las personas en el centro de la transformación digital». La Carta se puede encontrar en: https://www.segib.org/?document=carta-iberoamericana-de-principios-y-derechos-en-entornos-digitales.

11 Con carácter general, sobre las relaciones entre los derechos fundamentales y el mundo digital *cfr.*: Presno Linera, M.A., *Derechos Fundamentales e Inteligencia Artificial*, Marcial Pons, 2022 y Landa Arroyo, C., *Revolución digital y Constitución*, Palestra, Lima, 2023.

globales, fuera del Estado y en función de un modelo de negocio que pone en el centro de esa configuración no a la persona, como proclaman las cartas digitales, sino el dinero y los intereses económicos de esas compañías y, en segundo lugar, porque esa configuración se está haciendo desde el derecho privado y no desde el derecho público, situando también fuera del Estado y de la constitución nacional una gran parte de esa nueva realidad digital.

Esta diversidad esencial se manifiesta, por ejemplo, en la configuración de la libertad de expresión y de información en el ámbito digital. Si en el mundo analógico esa relación privilegia a la libertad de expresión con una posición preferente en caso de conflicto, en cuanto se trata un instrumento fundamental para la construcción de una opinión pública libre e informada y una democracia pluralista, en el mundo digital la libertad de expresión no realiza ya esa función o no la realiza del mismo modo en lo que se refiere a los entornos comunicativos creados por las compañías tecnológicas, en particular las redes sociales.

La comparación entre el mundo de los medios de comunicación tradicionales y el de las redes sociales nos permitirá evidenciar cómo la posición que ocupa en los procesos comunicativos diseñados por las compañías tecnológicas no contribuye a la formación de una opinión pública libre, no favorece la construcción social de una realidad compartida y no promueve la configuración democrática del espacio público. Podremos ver entonces como las diferentes estructuras de los procesos comunicativos en los medios de comunicación tradicionales y en las redes sociales definen unas posibilidades de influencia muy superiores a favor de estas últimas, en la medida en que utilizan los algoritmos para difundir mensajes personalizados que terminan encerrando a los usuarios en una burbuja (en el sentido del filtro burbuja de Pariser[12]) en la que experimentan procesos de radicalización creciente que dificultan el debate democrático.

La libertad de expresión y de información se desarrolla en entornos comunicativos virtuales en tal modo que la configuran como una mercancía más de la economía de datos, de tal manera que pierde su vinculación con la construcción de una sociedad democrática, lo que nos obliga a replantear el balance tradicional entre la libertad de expresión y de información y otros derechos garantizados por las constituciones democráticas. Hay que tener en cuenta, además, la incidencia que sobre esta temática está teniendo la inteligencia artificial generativa, en la medida en que su capacidad para propagar desinformación puede provocar problemas sistémicos en nuestras sociedades, afectando de manera muy grave a los derechos de la personalidad, por ejemplo.

12 *Cfr.* Pariser, E., *The Filter Bubble. What the Internet Is Hiding from You*, Penguin Books, 2011.

II. La percepción del tiempo en la sociedad digital y la asimetría entre procesos comunicativos y procesos políticos

El contraste entre el mundo analógico y el digital nos permite comparar los productos del desarrollo tecnológico desde la perspectiva del mundo analógico y comprender los cambios que se están produciendo, su naturaleza, su alcance y la medida en que están afectando a los procesos democráticos y a los derechos constitucionales configurados en el Estado constitucional. En particular, si analizamos los procesos comunicativos en los que se ha canalizado en el mundo analógico la libertad de expresión y la libertad de información, podemos ver como la percepción del tiempo y la configuración del espacio público han experimentado cambios importantes que necesariamente afectan a ambas libertades.

Se trata de transformaciones culturales que no dependen solamente de la evolución tecnológica sino del modo en que las compañías tecnológicas configuran sus aplicaciones tanto en lo que se refiere a la interfaz de usuario como al diseño de los algoritmos que utilizan en ellas. Si analizamos cómo han cambiado los procesos comunicativos desde la fase previa a Internet en la que se utilizaban las cartas como medio de comunicación, comprobamos cómo se ha transformado la percepción cultural del tiempo en esos procesos. No es una cuestión simplemente de desarrollo tecnológico, la percepción cultural del tiempo en el proceso comunicativo a través de email era muy similar al de la correspondencia mediante cartas: el mensaje que llega, aunque exista constancia de su recepción, no espera una respuesta inmediata.

La utilidad de email no fue diseñada para generar una interacción permanente, es neutra a esos efectos. Es importante resaltar que no se trata de una aplicación cerrada y monopolizada por una compañía como ocurre con WhatsApp, sino un protocolo abierto a muy diversos gestores de correo que prestan el servicio y que se pueden utilizar indistintamente para comunicarse a través de email. En el caso de WhatsApp o Telegram, por el contrario, nos encontramos con aplicaciones cerradas que tienen más libertad de configuración de la interfaz de usuario. En ellas el proceso comunicativo se transforma, contribuyendo a generar una nueva percepción del tiempo, como ocurre con otras aplicaciones de Internet diseñadas por las compañías tecnológicas para mantener al usuario el mayor tiempo posible dentro de la aplicación.

Por ejemplo, en el caso de WhatsApp, la interfaz de usuario permite conocer no solamente que el destinatario ha recibido el mensaje (como ocurre con el email) sino también que ese destinatario está conectado en ese momento a la aplicación. Esa circunstancia puede producir una tensión en el proceso comunicativo que induzca a responder rápido para no resultar descortés frente a la persona que ha escrito. Una tensión que se puede incrementar

mucho por el hecho de que la aplicación te permite también conocer que el destinatario está escribiendo y lo está haciendo para contestar a otros mensajes y no al tuyo.

La aplicación de WhatsApp está diseñada, en realidad, como una conversación telefónica y no como una mera aplicación de mensajería, porque permite conocer todo lo que se puede saber en una conversación telefónica. Cuando se envía un mensaje por WhatsApp y no se obtiene respuesta, se puede ver que la persona que lo ha recibido está conectada y que incluso está respondiendo a otros mensajes. El efecto que produce es el mismo que si se hablara por teléfono con alguien que no te contesta de manera voluntaria. Se genera una tensión similar y eso conduce a que se den respuestas rápidas y se mantenga una permanente atención a la aplicación generando una interacción que no es necesaria en el proceso comunicativo y que tiene sentido solamente por el interés de los diseñadores de la interfaz de usuario de potenciar el uso de la aplicación el mayor tiempo posible. La motivación es conocida: aumentar el volumen de datos recopilados que posteriormente pueden ser utilizados por Facebook (ahora Meta), la propietaria de WhatsApp, de manera que, aunque la aplicación no genere ingresos directos por publicidad, sí los genera indirectos mediante el uso de esos datos[13].

Esta interacción permanente y la necesidad de dar respuestas rápidas, que se produce también en otras aplicaciones de Internet (a través de los «likes» o los comentarios con los que se responden a otras personas que previamente han hecho uso de esas utilidades) transforma la percepción cultural del tiempo. Genera la sensación de que las cuestiones que se plantean en otros ámbitos de la vida, como los procesos políticos, requieren también respuestas rápidas. La configuración de los procesos comunicativos digitales produce la impresión en sus usuarios de una disponibilidad inmediata de las cosas y de la necesidad de ofrecer y de obtener respuestas rápidas a las cuestiones que se plantean en esos procesos comunicativos. Pero los procesos políticos en una sociedad democrática requieren de debates y de reflexiones colectivas que llevan su tiempo. Se genera así una asimetría entre procesos comunicativos y procesos políticos.

En las redes sociales cuando se plantea una cuestión que se convierte en tendencia social, se exige una respuesta inmediata: si es un problema que se plantea por la mañana, se debe resolver en la misma mañana, todo lo más

13 Mientras Facebook compró en 2012 Instagram por 1.000 millones de dólares, en 2014 pagó una cifra muy superior, 19.000 millones de dólares, por WhatsApp sin que esta última compañía, según indica *The Economist*, a pesar del enorme precio que ha pagado Facebook por ella, genere ingresos directos. Aunque el texto no se pronuncia sobre los posibles motivos, parece evidente que la motivación económica no estando en la publicidad directa tiene que basarse en algún otro interés para Facebook y no es difícil imaginar cuál es ese interés. *Cfr. Dismembering Big Tech*, *The Economist*, Oct. 24th, 2019, edition: https://www.economist.com/business/2019/10/24/dismembering-big-tech.

por la tarde, porque al día siguiente el problema ya será otro. Esta asimetría es muy peligrosa porque favorece las respuestas rápidas y simples a problemas complejos y esas respuestas las suelen dar siempre los movimientos populistas, que se ven así favorecidos por las características de los procesos comunicativos canalizados a través de las redes sociales y las aplicaciones de Internet.

Desde el punto de vista de la libertad de expresión y de información, por ejemplo, es evidente que esta configuración específica de los procesos comunicativos por parte de las compañías tecnológicas no favorece la función que han desarrollado tradicionalmente estas libertades en los sistemas democráticos para contribuir a la formación de una opinión pública informada y responsable, necesaria para el buen funcionamiento de la democracia. Tampoco favorece el respeto por los derechos de los usuarios de las redes porque introduce dinámicas problemáticas en la configuración de los procesos comunicativos, favoreciendo respuestas irreflexivas y poco respetuosas con derechos constitucionales tales como el honor o la intimidad.

III. La disolución del espacio público a través de los algoritmos

Por lo que se refiere a la configuración del espacio público, hay que tener en cuenta que la estructura del proceso comunicativo a través de las redes sociales y de las aplicaciones de Internet es muy distinta del de los medios de comunicación tradicionales. También aquí se puede observar (y con mayor intensidad) la dificultad para que las libertades de expresión y de información realicen las funciones que le corresponden dentro de una sociedad democrática, como ha ocurrido con los procesos comunicativos desarrollados a través de los medios de comunicación tradicionales.

Son muchas las variables que explican esas divergencias, a comenzar por el hecho de que las redes sociales se dirigen a un público potencialmente universal, a diferencia de lo que ocurre con los medios de comunicación tradicionales. No es que un periódico o una emisora de televisión tenga limitados sus destinatarios, sino que se dirigen a un sector de población, aquél que representa mejor su línea editorial o su orientación temática, si no es un medio generalista. Hasta tal punto la línea editorial se ve condicionada por el público que esa línea se suele mantener incluso cuando la propiedad de la emisora o del periódico cambia de manera significativa, porque perder la línea editorial puede significar perder el público.

Las redes sociales, sin embargo, tienen una estructura diferente. Su público es universal, ya que no se dirigen a un sector concreto de la sociedad. El público de Facebook o de Instagram es la totalidad de la sociedad de cada país, con independencia de las ideologías o los intereses temáticos de sus diversos sectores. Combinar un público universal con la diversidad existente

en la sociedad no sería posible si no se pudiera modular el discurso en función de los destinatarios. Si las redes sociales tuvieran una orientación ideológica o temática explícita, no podrían mantener como público más que aquellos sectores de la sociedad que la compartieran[14]. Por ese motivo, tienen que articular discursos muy diferentes, que van dirigidos específicamente a grupos también distintos o incluso a personas concretas cuando quieren influir en esas personas mediante la publicidad comercial o la propaganda política.

Hacer compatible un público universal con la diversidad de opciones y de planteamientos ideológicos y de todo tipo que existen en una sociedad solo es posible a través de los algoritmos. Son los algoritmos los que permiten extraer los datos de cada usuario, elaborar perfiles diversos, psicológicos incluidos, y ordenar la relación de la red social con ese usuario. Pero la configuración de los procesos comunicativos por medio de algoritmos plantea problemas específicos que están transformando el espacio público, dando lugar a fenómenos de fragmentación y radicalización debido a la necesidad de dirigir la información y la opinión en función de la orientación específica de cada usuario.

Para entender bien lo que está ocurriendo en el espacio público necesitamos comparar los procesos comunicativos del mundo digital con los del mundo analógico. Algunas diferencias relevantes explican el impacto que los algoritmos tienen en la configuración del espacio público, partiendo de un hecho claro: el creciente monopolio de las redes sociales sobre los procesos comunicativos. Esto es algo que podemos comprobar recurriendo al fenómeno «Trump», tan problemático como interesante desde el punto de vista del análisis del populismo y de la incidencia de las redes sociales en los procesos políticos.

D. Trump llegó a acumular más de 144 millones de seguidores a través de su actividad en las redes sociales en su primer mandato como Presidente de Estados Unidos. Cuando fue expulsado de las redes sociales —debido a su actividad contraria a la Constitución al promover el asalto al Capitolio— ya como Presidente saliente, perdió el acceso a todos esos seguidores. Para intentar recuperarlos creó un blog, llamado *From the Desk of Donald J. Trump*, como alternativa a Facebook y a Twitter, redes de las que había

14 Está por ver la incidencia que el giro de Elon Musk, propietario de X (la antigua Twitter) apoyando expresamente a D. Trump en la campaña presidencial norteamericana puede tener en la red social. Se trata de un caso atípico por cuanto el propietario de X utiliza la red social como un instrumento al servicio de sus otros negocios, que son los que le pueden dar un mayor rendimiento mediante su presencia en la Administración Trump (aunque esa pretensión parece haberse frustrado con el enfrentamiento entre ellos pocos meses después de tomar posesión Trump). Cfr, al respecto Balaguer Callejón, F., *Presentación del n. 42 de la Revista de Derecho Constitucional Europeo*, en https://www.ugr.es/~redce/ y Balaguer Callejón, F., «Aziende tecnologiche e politica, da Zuckerberg a Musk», *Fuoricollana.it*, 11/4/2025: https://fuoricollana.it/aziende-tecnologiche-e-politica-da-zuckerberg-a-musk/

sido expulsado. El blog duró menos de un mes porque estaba teniendo muy poco público[15]. Mientras Trump llegó a tener 88 millones de seguidores en Twitter, 32 millones en Facebook, y 24 millones en Instagram, su blog alcanzó solamente alrededor de doscientas mil interacciones. Posteriormente, creó su propia red social, que tampoco consiguió cifras significativas de seguidores en comparación con los que tenía en las otras redes sociales[16], de manera que no recuperó a esos seguidores hasta que volvió a ser admitido en ellas.

Este monopolio de los procesos comunicativos no tiene precedentes en sistemas democráticos. Nada menos que el Presidente saliente de Estados Unidos, una figura con una capacidad de influencia decisiva en el Partido Republicano y muchas posibilidades de volver a ser de nuevo Presidente, como finalmente ha ocurrido, perdió la inmensa mayoría de los seguidores que tenía en las redes sociales, en torno a 140 millones de personas. Estamos hablando de compañías privadas que monopolizan procesos comunicativos en el espacio público y lo hacen desde el derecho privado, decidiendo libremente a quiénes admiten y a quiénes no. Una situación que merece una reflexión profunda desde muy distintos ámbitos, comenzando por la propia distinción entre el derecho público y el derecho privado en el contexto de la globalización y siguiendo por la configuración de la libertad de expresión en el espacio público controlado por las redes sociales.

Se podría objetar que esta pérdida masiva de seguidores y su exclusión de las redes sociales, hasta que fue admitido de nuevo, no le impidieron a D. Trump volver a ser el candidato del Partido Republicano a la Presidencia (cuestión diferente es la de su elección como Presidente, en la que incluso tuvo el apoyo explícito de alguna de ellas). Esta objeción no tiene en cuenta que, en realidad, la expulsión de las redes sociales de Trump no impidió que se promoviera en esas mismas redes el discurso trumpista[17]. D. Trump no

15 *Cfr.* Zadrozny, B., *Trump's blog isn't lighting up the internet*, May 11, 2021, *nbcnews.com.* https://www.nbcnews.com/tech/tech-news/trumps-blog-isnt-lighting-internet-rcna890.

16 Esa tuvo un promedio de 200.000 visitas diarias desde su lanzamiento en Estados Unidos, el 1 de marzo de 2022. *Cfr.* Thompson, S. A., *Traffic surged on Trump's social media website as an indictment*, The New York Times, April 4, 2023. El número de seguidores de Trump en su propia red social a enero de 2023 era mucho más alto que el del blog: 4,83 millones. En diciembre de 2023 llegó a 6,53 millones: https://truthsocial.com/@realDonaldTrump. Sin embargo, una vez que ha vuelto a ser admitido en las otras redes sociales, la diferencia es enorme respecto de Twitter (87,73) Facebook (34,49) o Instagram (23,3). *Cfr. Donald Trump number of followers on selected social platforms 2023* Statista Research Department, Jan 30, 2023.

17 Como evidencia la investigación de *The New York Times*, en la que se demuestra el doble juego de Facebook tras las elecciones presidenciales de 2020 en Estados Unidos. Por un lado, su actitud pública era contraria a D. Trump en relación con su intento de manipular el resultado electoral apareciendo así ante la opinión pública como una compañía comprometida con la democracia y con los valores constitucionales. Por otro lado, sus algoritmos seguían potenciando una presencia pública masiva de las posiciones *trumpistas*, favoreciendo los discursos en los que se cuestionaba el resultado electoral y

necesitaba estar presente de manera directa en las redes sociales porque los algoritmos diseñados por las compañías tecnológicas promovían secretamente sus posiciones políticas, favoreciendo sus intentos de volver a ser Presidente, como así ha sido.

Una promoción que en el caso de X (la antigua Twitter) ha sido absolutamente explícita por parte de su nuevo propietario, Elon Musk. En realidad, esta maniobra de Musk no era necesaria porque X podía hacer lo mismo que Facebook: configurar sus algoritmos para favorecer secretamente a D. Trump (algo que pueden hacer estas compañías debido a la falta de transparencia de sus algoritmos). Por tanto, la maniobra sólo se explica por la intención de Elon Musk de aparecer como el campeón de los seguidores de Trump para situarse en la esfera de influencia de su Administración, como así ha ocurrido en los primeros meses de mandato de D. Trump (hasta la brusca ruptura de ambos en junio[18]). Algo que Elon Musk podía permitirse porque la red social X tiene para él una función instrumental al servicio de sus otros negocios[19].

En todo caso, lo que nos interesa ahora analizar es la configuración del espacio público partiendo de estas condiciones de monopolio de las compañías tecnológicas, porque la comparación entre el mundo analógico y el digital nos evidencia unas condiciones estructurales muy diferentes. Mientras los medios de comunicación tradicionales ocupaban una parte del espacio público en función de su línea editorial y del sector social al que estaban orientados, las redes sociales se dirigen a un público universal y tienden a monopolizar los procesos comunicativos amplificando su impacto sobre la configuración del espacio público.

¿Cómo lo hacen? Aquí también las diferencias son enormes. Pongamos como ejemplo el proceso de lectura de un periódico impreso. La persona que lee un periódico comienza por los titulares de la portada y sigue avanzando, página a página, por las diferentes secciones: desde internacional a espectáculos, pasando por sociedad, cultura, deportes, etc. Puede ser que esa persona no tenga mucho interés en algunas de esas secciones, pero el proceso de lectura le lleva a ojearlas y, a veces, a detenerse en algunas de las informaciones. Quizás no le interese nada la sección de cultura, pero puede

se defendía «parar el robo» que según Trump se había producido en las elecciones. En la investigación realizada por *The New York Times*, personas que no habían tenido apenas eco en su actividad en Facebook previamente, pasaron a tener miles de seguidores cuando comenzaron a difundir planteamientos favorables a las posiciones *trumpistas*, gracias a los algoritmos de Facebook. *Cfr.* Thompson, S.A. y Warzel, C.: *How Facebook Incubated the Insurrection. Right-wing influencers embraced extremist views, and Facebook rewarded them*, en *The New York Times*, 14/012021.

18 *Cfr. Trump and Musk Alliance Crumbles Amid Public Threats and Insults*, *The New York Times*, Jun, 5, 2025.

19 *Cfr.* Balaguer Callejón, F., *Presentación del n. 42 de la Revista de Derecho Constitucional Europeo* y «Aziende tecnologiche e politica, da Zuckerberg a Musk», antes citados.

ser que alguna noticia concreta o algún artículo de opinión llame su atención, o puede ser que se detenga en la sección de internacional, quizás por alguna imagen o algún titular, aunque no sea una temática en la que haya tenido interés previamente.

El proceso de lectura de un periódico impreso ofrece una visión global del mundo a la persona que lo lee (algo que también ocurre en gran medida con los programas de noticias de televisión o de radio, por ejemplo, con sus diferentes secciones). No ocurre lo mismo con la información que llega a través de las redes sociales, desagregada[20] y seleccionada por los algoritmos en función de nuestro previo historial de búsqueda, por ejemplo, o del perfil que la aplicación ha generado con los datos que extrae de nuestra actividad en Internet. Las redes sociales ofrecen una visión limitada del mundo. Si a alguien le interesan especialmente las noticias sobre deporte, le llegarán solamente ese tipo de noticias, si le interesan las informaciones relacionadas con el cine, le llegarán únicamente esas y lo mismo si le interesan las de sociedad o las de política nacional. El mundo se verá básicamente a través de la retroalimentación que le llegue, pero sin ofrecer la posibilidad de conocer otros ámbitos, empobreciendo cada vez más la visión de cada usuario de las redes.

Pero no sólo se limitan los ámbitos de conocimiento de los usuarios, sino que también se reducen sus posibilidades de conocer y comprender otros planteamientos diferentes de los suyos. La necesidad de combinar un público universal con las diferencias ideológicas existentes en la sociedad conduce a ofrecer, a través de los algoritmos, las opiniones que coinciden en mayor medida con las propias[21]. Se genera así el efecto burbuja del que nos habla Pariser[22] y que explica la fragmentación y progresiva radicalización del espacio público generada por las redes sociales. Una tendencia que se ve favorecida por los ecosistemas en los que se desarrollan los procesos comunicativos generados por estos nuevos «mediadores» que son las grandes compañías tecnológicas[23].

---

20 «Social feeds and search results show snippets and single articles. This creates a more disaggregated news experience than traditional news media». Eso supone que los usuarios no acceden más que a una parte de la información: «Online content is effectively "unbundled". The different sections of a print newspaper are brought together into one product. Online, readers can select only the articles they wish to view, without necessarily being exposed to other content. So, they may be less likely to read public interest news». *The Cairncross Review. A sustainable future for journalism*, 12 February 2019, p. 7.https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf

21 Ello sin perjuicio de que a través de propaganda subliminal o de promoción de determinados planteamientos mediante los algoritmos se manipule secretamente a los usuarios sin que estos lo perciban.

22 *Cfr.* Pariser, E., *The Filter Bubble. What the Internet Is Hiding from You*, citado.

23 Balaguer Callejón, F., *The Impact of the New Mediators of the Digital Age on Freedom of*

En estos ecosistemas las personas que los utilizan tienen una mayor receptividad y una menor capacidad crítica para diferenciar y analizar la información, por tratarse de entornos comunicativos privados en los que se mezclan mensajes personales tales como fotos o videos de amigos o de familiares, con informaciones de carácter político, por ejemplo. No existe la actitud de quien de manera activa quiere informarse, como en el mundo analógico, yendo a un punto de venta de prensa y adquiriendo un periódico que se percibe como un producto informativo y frente al cual se mantiene una distancia que hace posible recibir esas informaciones de una manera más reflexiva y crítica.

La fragmentación y la radicalización tienen un gran potencial destructivo del espacio público porque dificultan su configuración como un lugar común, un ámbito de encuentro en el que se pueden debatir las distintas posiciones desde el respeto a las diferencias, algo esencial en la construcción de una democracia pluralista. Para comprender el alcance que tienen estos procesos comunicativos desarrollados por los nuevos mediadores vamos a recurrir a una metáfora que explica en su esencia el sentido de la democracia pluralista. La metáfora procede de un proverbio árabe según el cual la verdad era un gran espejo que estaba en el cielo y que cayó a la tierra y se fragmentó en múltiples pedazos. Cada uno de nosotros tiene uno de esos pedazos y, por tanto, solamente una parte de la verdad. La construcción de la verdad se configura de ese modo como un proceso colectivo en el que todas las piezas, grandes y pequeñas, tienen el mismo valor, porque si falta solamente una de ellas ya no es posible conocer la verdad.

Esa igual dignidad de la verdad de cada una de las personas que confluyen en el espacio público (por distinto que sea su poder o su capacidad de promover su propia verdad) es fundamental en la democracia pluralista y se refleja, por ejemplo, en la igualdad del voto, en el proceso de decisión que conduce a la formación de la voluntad colectiva. En la democracia es un elemento esencial, como indica Paola Bilancia, la paridad de los participantes en el debate público[24]. Para que esa igual dignidad de la verdad de cada uno sea reconocida es necesario reconocer la verdad de los otros, aunque no se comparta, aunque se discrepe esencialmente y se combata en el plano político. Pero eso es algo cada vez más difícil debido a la progresiva disolución del espacio público que están generando los nuevos mediadores, las grandes compañías tecnológicas, mediante la configuración de los algoritmos en sus entornos comunicativos.

En esos entornos se reduce la visión del mundo de los usuarios y se les alimenta de manera permanente con sus propios planteamientos. De ese

Speech, en S. Hindelang, (editor) A. Moberg, (editor) *YSEC Yearbook of Socio-Economic Constitutions*, Springer 2022.

24 *Cfr.* Bilancia, P. *Crisi nella democrazia rappresentativa e aperture a nuove istanze di partecipazione democratica*, *federalismi.it* -|numero speciale 1/2017, pp. 3,9 y 15.

modo, se dificulta reconocer la verdad de los otros porque se termina por pensar que solamente las propias opiniones son aceptables y que las de los demás están siempre equivocadas. El resultado es la radicalización, el enfrentamiento, la dificultad para llegar a acuerdos, para generar consensos, en definitiva, para construir un espacio público común que sirva de base a la democracia pluralista. El espacio público termina disolviéndose en un pluralismo de posiciones monistas que no reconocen a las demás y se autoafirman como las únicas posibles cada una de ellas. Una especie de «monismo plural» en el que la incomunicación entre las diversas visiones del mundo es cada vez mayor.

IV. La pérdida de una visión social compartida de la realidad

Lo peculiar de esta configuración del espacio público por parte de las redes sociales es que su *interés primario* no consiste en imponer una determinada visión del mundo a través de sus algoritmos. En eso se diferencian de los autócratas que en los Estados autoritarios o dictatoriales (o en los regímenes iliberales) intentan controlar a los medios de comunicación para ponerlos a su servicio e imponer aquellas narrativas que ofrecen una visión del mundo coherente con sus propios idearios y, sobre todo, con su voluntad de mantenerse en el poder. Lo característico de estos sistemas políticos autoritarios es que manipulan la realidad, la distorsionan, la reconstruyen a la medida de sus intereses. No se puede decir que no tengan interés por la verdad porque sí lo tienen, justamente para manipularla y desfigurarla siempre que sea necesario para conservar el poder.

Los entornos comunicativos creados por las grandes compañías tecnológicas, que están ya comparándose por su poder global con los Estados, no tienen ese interés por manipular y *reconstruir* la verdad propio de los Estados dictatoriales o autoritarios. Lo único que les interesa es fomentar la interacción de sus usuarios para obtener mayores rendimientos económicos mediante el tratamiento de sus datos, la elaboración de perfiles personalizados y la venta de publicidad comercial o de propaganda política. Configuran sus algoritmos para satisfacer a sus usuarios y, por ese motivo, no es previsible que pudieran mantener una operación sostenida de propaganda política *explícita,* por ejemplo, que pretendiera orientar a sectores masivos de la sociedad hacia determinadas opciones.

No es que no lo hagan, como hemos tenido ocasión de comprobar en el caso de Estados Unidos y la promoción del discurso trumpista. Pero —con la excepción reciente de Elon Musk y su red X, que se debe a la instrumentalización que ha hecho de ella en favor de sus otros negocios[25]— lo hacen de

25 V. *supra*, notas 14 y 18.

manera secreta, mediante la configuración de algoritmos que no están sometidos a una obligación de transparencia y a través de los cuales, debido a ese hermetismo, pueden manipular a la opinión pública y promover los discursos más disruptivos, que son los que generan mayor interacción con sus aplicaciones. Si, por el contrario, optaran de manera explícita por una orientación partidista, lo más probable sería que una operación de ese tipo les supusiera una pérdida masiva de usuarios, en concreto de todos aquéllos que no compartieran los planteamientos promovidos por parte de las redes sociales. Eso perjudicaría su rendimiento económico que es, en realidad, la finalidad última de las grandes compañías tecnológicas que diseñan las redes sociales y las aplicaciones de Internet (de nuevo, salvo la de Elon Musk, cuyos rendimientos económicos no están centrados en la red social X sino en otros negocios, como evidencia el hecho de que su red social, totalmente descapitalizada, tuvo que ser integrada en su compañía de Inteligencia Artificial)[26].

En los entornos comunicativos de los medios de comunicación tradicionales, que son la base de la democracia pluralista, el espacio público se configura tendencialmente para favorecer la *construcción* social de la realidad. Los diversos medios de comunicación tienen líneas editoriales que reflejan esencialmente la pluralidad social y confluyen en el debate público aportando cada uno ese fragmento de espejo, ese pedazo de verdad que, puesto en común con los otros, permite construir una visión social compartida de la realidad. Naturalmente, hay también intereses económicos en las propias empresas, campañas de comunicación que intentan imponer narrativas, intentos de ocultar la verdad o las verdades que perjudican al poder político social o económico de determinados grupos y que introducen distorsiones puntuales en el espacio público. No se trata de un mundo ideal ni de una democracia «perfecta», entre otras cosas porque es consustancial a la idea del pluralismo que estas disfuncionalidades se puedan corregir en mayor o menor medida a través de procesos públicos abiertos que son los que definen a una democracia pluralista.

Estos procesos no son necesariamente una expresión del llamado «libre mercado de las ideas» de Holmes, de dudosa aplicación a un contexto con desigualdad de armas para promover la verdad[27]. Esta desigualdad de armas

---

26 La red X, llegó a perder todo su valor, pasando de los 44000 millones de dólares que le costó a Musk a unos 8000 millones, lo que suponía un valor nulo, ya que debía más de 13000 millones de dólares. *Cfr.* Saba, J., Kwok, K.: *X, la red social de Elon Musk, es un agujero negro de valor*, *El País/CincoDías*, 5/10/2023. Para parar esta sangría, Elon Musk ha recurrido a integrar a la red social dentro de su negocio de Inteligencia Artificial, xAI, lo que le permite ocultar la pérdida de valor de la compañía. Su valoración, claramente ficticia, a los efectos de la integración en xAI, ha sido de 33000 millones de dólares. *Cfr.* Jiménez, M. *Elon Musk vende la red social X a su propia firma de inteligencia artificial por 33.000 millones de dólares*, *El País*, 29/3/2025.

27 Como bien indica Schauer, F.: *«Once we fathom the full scope of factors other than the truth of a proposition that might determine which propositions individuals or groups*

nos evoca el modo de obtener la verdad propia del mundo medieval en el que, como indicara Foucault, el sistema estaba orientado no a probar la verdad, sino la fuerza, el peso o la importancia de quien la afirmaba[28]. Ciertamente, la apertura de los procesos y el pluralismo permiten corregir y compensar esa desigualdad, pero no podemos olvidar que también en la configuración «analógica» del espacio público existían grandes grupos mediáticos con una capacidad importante de promover narrativas que no siempre respondían a la realidad de los hechos. Y, aunque, como señala H. Arendt, la orientación finalista de la política puede darle una condición transitoria a la mentira frente a la estabilidad o permanencia de la realidad fáctica[29], también es cierto que la mentira ha sido un gran motor de la historia, como indica U. Eco[30]. En todo

will accept and which they will reject—the charisma, authority, or persuasiveness of the speaker; the consistency between the proposition and the prior beliefs of the hearer; the consistency between the proposition and what the hearer believes that other hearers believe; the frequency with which the proposition is uttered; the extent to which the proposition is communicated with photographs and other visual or aural embellishments; the extent to which the proposition will make the reader or listener feel good or happy for content independent reasons; and almost countless others—we can see that placing faith in the superiority of truth over all of these other attributes of a proposition in explaining acceptance and rejection requires a substantial degree of faith in pervasive human rationality and an almost willful disregard of the masses of scientific and marketing research to the contrary», SCHAUER, F., *Facts and the First Amendment*, *UCLA Law Review*, Volume 57. Issue 4, April 2010, p. 909.

28 Nos recuerda FOUCAULT que la cuestión de la verdad está ausente en ese sistema, basado en la prueba y no en la indagación: *«"En el derecho feudal, el litigio entre individuos se reglamentaba por el sistema de la prueba (épreuve). Cuando un individuo se presentaba llevando una reivindicación, una querella, acusando a otro de haber robado o matado, el litigio entre ambos se resolvía por una serie de pruebas aceptadas por los dos y a las que ambos se sometían. Este sistema no era una manera de probar la verdad, sino la fuerza, el peso o la importancia de quien decía", de tal modo que "la prueba no sirve para nombrar o determinar quién es el que dice la verdad, sino para establecer quién es el más fuerte, y al mismo tiempo quién tiene razón. En una guerra o prueba no judicial, uno de los dos es siempre el más fuerte, pero esto no prueba que, a la vez, tenga razón. La prueba judicial es una manera de ritualizar la guerra o trasponerla simbólicamente, una manera de darle ciertas formas derivadas y teatrales de tal modo que el más fuerte será designado, por ese motivo, como quien tiene razón»*, FOUCAULT, M., *«De la verdad entre particulares a la verdad pública: el nacimiento de la indagación jurídica»*, en FOUCAULT, M., *La verdad y las formas jurídicas*, 1973, Gedisa Editorial. Edición de Kindle, pp. 68, 70, 73-74.

29 «Power, by its very nature, can never produce a substitute for the secure stability of factual reality», de tal manera que «In their stubbornness, facts are superior to power; they are less transitory than power formations, which arise when men get together for a purpose but disappear as soon as the purpose is either achieved or lost», ARENDT, H., *«Truth and Politics»,* 1967, en *Between Past and Future*, Penguin, London, 1968, pp. 258-259.

30 De tal manera que se puede hablar con Umberto Eco de una «forza del falso» por contraposición a la fuerza de la verdad que, sea a través de la mentira o por medio del error, ha sido el motor de muchos acontecimientos históricos. *Cfr.* ECO, U. *La forza del falso*, en la recopilación del mismo autor *Sulla Letteratura*, Tascabili Bompiani, Milano, 2003, pp. 292 y ss. Para U. ECO, «il riconoscere che la nostra storia è stata mossa da molti racconti che ora riconosciamo come falsi deve renderci attenti, capaci di rimettere continuamente in

caso, como afirma P. Häberle, en la democracia pluralista propia del Estado constitucional se dan las condiciones para que las personas puedan buscar libremente la verdad y obtenerla[31].

Esa búsqueda común es lo que hace posible una visión social compartida de la realidad en sus elementos básicos, que no es incompatible con las distintas verdades que confluyen a construir socialmente esa percepción compartida. Sin embargo, en la sociedad digital, la estructura de los procesos comunicativos (dirigidos tendencialmente a la totalidad del público) conduce a que la verdad permanezca fragmentada a través de los algoritmos para garantizar la fidelidad de ese público, también fragmentado. La realidad se configura en burbujas o compartimentos estancos que se articulan mediante los perfiles de usuario a través de los algoritmos que ordenan los procesos comunicativos dentro de las redes sociales y las aplicaciones de Internet.

En esos procesos comunicativos, la libertad de expresión no puede desarrollar las funciones que le corresponden en la formación de una opinión pública informada y responsable necesaria para el desarrollo de una sociedad democrática. Por el contrario, la estructura y el diseño de los procesos comunicativos obstaculizan esa función hasta convertirla prácticamente en imposible. Cabe plantearse, entonces, si puede seguir teniendo dentro de esos entornos virtuales la alta consideración que ha tenido tradicionalmente en el Estado constitucional. Si la «libertad de expresión» se convierte en una mercancía más dentro de la economía de datos y contribuye a generar involución democrática mediante la ordenación jerarquizada que hacen las compañías tecnológicas de los procesos comunicativos, potenciando los discursos que generan más inestabilidad democrática y más daño a los derechos fundamentales, es evidente que estamos ante algo muy diferente de la libertad de expresión propia de los procesos democráticos que en el mundo analógico se han desarrollado y se siguen desarrollando en el Estado constitucional.

---

questione gli stessi racconti che ora teniamo per veri, poiché il criterio della saggezza della comunità si fonda sulla vigilanza continua nei confronti della fallibilità del nostro sapere», p. 322.

31 Como indica Häberle, P., el Estado constitucional se contrapone a los Estados totalitarios y a las pretensiones fundamentalistas de verdad, porque «si caratterizza por la consapevolezza di *non* essere in possesso di precostituite verità eterne, ma di essere invece destinato a una mera *ricerca* della verità». De tal manera que, «L'immagine dell'uomo dello stato costituzionale sottintende un cittadino che necessita di verità e che è in possesso degli strumenti per *ricercarla* e "conquistarla"», Häberle, P., *Wahrheitsprobleme im Verfassungsstaat*, 1995, en la versión italiana preparada por Zagrebelsky, G., y Luther, J., *Diritto e verità*, cit. pp. 85 y 110.

V. La desinformación como rasgo estructural de los procesos comunicativos en la sociedad digital

Los medios de comunicación tradicionales construían narrativas, de acuerdo con sus líneas editoriales, a través de las cuales participaban en la construcción social de la realidad. Esas narrativas podían incluir elementos de desinformación, eventualmente contrastados o refutados a través del pluralismo de los medios. Lo que permite que esas narrativas generen una información relativamente solvente no es que todos los medios repitan las mismas, porque eso solo ocurre en un Estado totalitario, sino justamente que cada uno de ellos aporta su propia visión de la realidad y la comparte y contrasta con las de los demás medios. En sí mismas, esas diferentes narrativas no expresan desinformación o intentos de manipulación, sino que se corresponden con el juego democrático propio de una sociedad pluralista.

Por el contrario, los nuevos mediadores, las grandes compañías tecnológicas, no construyen narrativas, sino que abren sus aplicaciones a todas las narrativas posibles, aunque privilegian —a través de sus algoritmos— a las que promueven noticias falseadas y realidades alternativas, porque eso favorece la interacción permanente con sus aplicaciones y facilita así la acumulación de datos que necesitan para obtener mayores rendimientos mediante la venta de publicidad. De ese modo, los nuevos mediadores no reflejan la realidad de las sociedades en las que actúan. Por el contrario, al potenciar la desinformación mediante las noticias falseadas y las realidades alternativas, generan una tensión sobre la realidad misma que tiene un gran potencial destructivo.

La destrucción de esa percepción social compartida de la realidad no se extiende a todos los ámbitos del espacio público (al menos no lo ha hecho hasta el desarrollo de la IA generativa), fenómeno este que es digno de análisis. En el campo económico no encontramos *fake news* ni posverdad. Los fundamentos del tráfico económico siguen intactos en la sociedad digital y no se han visto afectados por los algoritmos. Lo mismo podría decirse de la vertiente tecnológica[32]. Es posible, por tanto, mantener el espacio público libre de *fake news*, posverdad y realidades alternativas. El contraste entre la cultura política digital y la económica es extraordinario, como también es la protección jurídica de la que disfruta esta última frente a la primera.

En Estados Unidos, esta protección del ámbito económico se corresponde con la interpretación que se ha realizado de la Primera Enmienda respecto de los límites de la libertad de expresión. Una interpretación que admite cierta regulación en la publicidad comercial, de manera que no quedaría protegida por la Primera Enmienda, a diferencia de la propaganda política y, en general, los debates políticos. El hecho de que la publicidad comercial no esté prote-

32 *Cfr.* Balaguer Callejón, F., *La constitución del algoritmo*, cit., pp. 80 y ss.

gida por la Primera Enmienda en lo que se refiere a las afirmaciones falsas, a los hechos falsos que contiene, evidencia que en este ámbito sí hay un control de la falsedad factual, algo que no ocurre en el ámbito político, al protegerse la libertad de expresión de acuerdo con la interpretación que se le ha dado a la Primera Enmienda[33].

Es de destacar, sin embargo, el reciente caso de la condena a R. Guliani por difamar a dos empleadas electorales de Georgia en las últimas elecciones presidenciales, que ha generado ciertas expectativas de que los tribunales puedan contener en el futuro la oleada de mentiras y desinformación con la que se está llenando el espacio público en el ámbito político[34]. Como contraste tenemos el caso de Donald Trump, y sus afirmaciones falsas en los cuatro años de su primer mandato, que según el Washington Post alcanzaron el número de 30.573[35] y que se concentraban en cuestiones políticas.

La protección de la economía se extiende también al tráfico jurídico, que no se ve afectado, con carácter general, por los nuevos procesos comunicativos diseñados por las compañías tecnológicas. El campo de acción de los algoritmos diseñados para promover la desinformación, la posverdad y las realidades alternativas es la política. Su finalidad no es otra que generar mayor atención del público y una interacción permanente con sus aplicaciones de Internet, así como desprestigiar la política, limitando la capacidad de acción de los legítimos representantes de la ciudadanía, con el objetivo de bloquear la política en gran medida y, entre otras cosas, dificultar el control de las propias compañías tecnológicas por las instancias democráticas.

Esta situación está cambiando de manera drástica con la aparición de la Inteligencia Artificial generativa que, al menos en sus primeras aplicaciones (tales como el ChatGPT) se están configurando como una fuente de desinformación sistémica. Es de destacar aquí el salto cualitativo que suponen estas aplicaciones de IA generativa en cuanto rompen la esfera de protección que se había mantenido por las compañías tecnológicas sobre ámbitos tales como la economía, la aplicación del derecho o la propia IA. Mientras las redes sociales y las aplicaciones de Internet se limitaban hasta ahora a difun-

---

33 *Cfr.* Como indica Frederick Schauer: «although the existing doctrine is moderately clear with respect to the permissibility of restricting false or misleading advertising of securities or commercial products, the issue is different when we turn to questions of factual falsity in political debate», Schauer, F., *Facts and the First Amendment*, cit. p. 913.

34 Por más que esta percepción le parezca excesivamente optimista, señala Andy Kroll que «On a societal level, the real hope for these defamation cases is that over time, as more liars are brought low by their actions and held accountable in court, politicians and political operatives will pause before spreading disinformation and, slowly, this country will move toward a better, safer political discourse» Kroll, A. , *The Unsettling Truth at the Heart of the Giuliani Case*, *The New York Times*, Dec. 23, 2023.

35 D. Trump llegó a acumular nada menos que 30.573 afirmaciones falsas o engañosas en sus cuatro años de mandato, de acuerdo con el trabajo de comprobación realizado por el *Washington Post*, a 24 de enero de 2021: www.youtube.com/watch?v=Jq5k LbR79.

dir desinformación y a provocar incerteza en el ámbito político y social, la IA generativa extiende la desinformación, de manera sistémica, a ámbitos tales como la economía, la investigación científica, el mundo jurídico o las aplicaciones tecnológicas (en cuanto estas se basen en datos que son falseados por las aplicaciones de Inteligencia Artificial generativa)[36].

VI. Conclusiones

El mundo virtual, que se está construyendo esencialmente por compañías tecnológicas de alcance global en función de sus intereses económicos y a través del derecho privado, no refleja las condiciones de ejercicio de los derechos del Estado constitucional, de la constitución «analógica». El contexto de los derechos es muy importante y no es posible pensar que los entornos virtuales son neutros en relación con la democracia y los derechos constitucionales. Los derechos fundamentales tienen pocas probabilidades de sobrevivir en el mundo virtual si no le proporcionamos los medios de defensa y realización adaptados a los nuevos entornos en los que se mueven. Lo mismo ocurre con libertades, como la libertad de expresión o de información, que están siendo desnaturalizadas en los entornos virtuales, convertidas en mera mercancía en la economía de datos o utilizadas por las compañías tecnológicas, a través de los algoritmos con los que ordenan sus aplicaciones, para favorecer sus intereses económicos sin importarles el daño que provocan a los derechos constitucionales o a la democracia.

Estas compañías están configurando las interfaces de usuario de sus aplicaciones y los algoritmos con la orientación casi exclusiva de obtener mayores beneficios, transformando la percepción del tiempo en los procesos comunicativos y dificultando el desarrollo de las funciones que la libertad de expresión y la libertad de información han tenido en el Estado constitucional para contribuir a la formación de una opinión pública libre y responsable, necesaria para el buen funcionamiento de la democracia. Lo mismo puede decirse de la configuración del espacio público a través de los nuevos procesos comunicativos digitales. Las redes sociales tienen un público universal, a diferencia de los medios de comunicación tradicionales y esa estructura específica les obliga ordenar los procesos de comunicación mediante algoritmos que van dirigidos a cada usuario de acuerdo con los perfiles que se elaboran a través de su actividad en Internet. El resultado es una alimentación permanente de las propias opiniones que genera una fragmentación del espacio público y una creciente radicalización en los procesos comunicativos.

Las compañías tecnológicas han creado entornos digitales que, entre otras cosas, transforman la percepción cultural del tiempo en los procesos comu-

36 *Cfr.* a este respecto Balaguer Callejón, F. y Balaguer Callejón, M.L. *Verdad e interpretación en la sociedad digital*, Aranzadi, Cizur Menor, 2023.

nicativos y la ordenación del espacio público. En esos entornos contribuyen a dificultar una percepción social compartida de la realidad, que es la base para la existencia de una comunidad política y de una democracia pluralista. En esos entornos digitales la libertad de expresión no puede cumplir las funciones que ha desarrollado tradicionalmente en una sociedad democrática porque no es una auténtica libertad de expresión, ya que está condicionada por la jerarquización interna de las aplicaciones que se realiza por las compañías tecnológicas de acuerdo con criterios orientados exclusivamente al beneficio económico. Resulta difícil atribuir a esa libertad de expresión desnaturalizada la posición preferente que ha ocupado en el Estado constitucional frente a otros derechos constitucionales.

Las soluciones a estos problemas que plantean los nuevos entornos digitales no pueden ser epidérmicas. No pueden basarse en el voluntarismo o en el mero control de las consecuencias derivadas de la configuración que están realizando las compañías tecnológicas de sus redes sociales y de sus aplicaciones. Por el contrario, deben afrontar esos problemas en sus causas, obligando a las compañías tecnológicas a transformar sus entornos virtuales —la configuración de sus interfaces de usuario y de sus algoritmos— para hacerlos compatibles con los principios y valores constitucionales y respetuosos con el sistema de derechos propio del Estado constitucional y con las funciones que les corresponden a esos derechos para garantizar la democracia pluralista.

Bibliografía

Arendt, H., *Truth and Politics,* 1967, en *Between Past and Future*, Penguin, London, 1968.

Balaguer Callejón, F., *Aziende tecnologiche e politica, da Zuckerberg a Musk*, *Fuoricollana.it*, 11/4/2025.

Balaguer Callejón, F., *La constitución del algoritmo y las transformaciones culturales en la sociedad digital*, en F. Caamaño y D. Jove Villares, *Tecnologías abusivas y Derecho*, Tirant lo Blanch, Valencia, 2024.

Balaguer Callejón, F., *La constitución del algoritmo*, Fundación Manuel Giménez Abad, 2.ª edición, Zaragoza, 2023.

Balaguer Callejón, F., *Inteligencia artificial y cultura constitucional*, en F. Balaguer Callejon, Ingo W. Sarlet (Directores), *Derechos fundamentales y democracia en el constitucionalismo digital*, Aranzadi, Cizur Menor, 2023.

Balaguer Callejón, F., *La cultura constitucional en la era digital*, en J. M. Pérez Collados, (Coord.) *La cultura jurídica en la era digital*, Aranzadi, Cizur Menor, 2022.

Balaguer Callejón, F., *The Impact of the New Mediators of the Digital Age on Freedom of Speech*, en S. Hindelang, (editor) A. Moberg, (editor) *YSEC Yearbook of Socio-Economic Constitutions*, Springer 2022.

Balaguer Callejón, F. y **Balaguer Callejón, M.L**. *Verdad e interpretación en la sociedad digital*, Aranzadi, Cizur Menor, 2023.

Bilancia, P. *Crisi nella democrazia rappresentativa e aperture a nuove istanze di partecipazione democratica*, *federalismi.it* -|numero speciale 1/2017.

Bustos Gisbert, R. *El constitucionalista europeo ante la inteligencia artificial: reflexiones metodológicas de un recién llegado*, *REDC*, n. 131, mayo-agosto de 2024.

Eco, U. *La forza del falso*, en la recopilación del mismo autor *Sulla Letteratura*, Tascabili Bompiani, Milano, 2003.

Foucault, M., *De la verdad entre particulares a la verdad pública: el nacimiento de la indagación jurídica*, en Foucault, M., *La verdad y las formas jurídicas*, 1973, Gedisa Editorial. Edición de Kindle.

Häberle, P., *Wahrheitsprobleme im Verfassungsstaat*, 1995, en su versión italiana preparada por Zagrebelsky, G., y Luther, J., *Diritto e verità.*

Jiménez, M. *Elon Musk vende la red social X a su propia firma de inteligencia artificial por 33.000 millones de dólares*, *El País*, 29/3/2025.

Kroll, A. *The Unsettling Truth at the Heart of the Giuliani Case*, *The New York Times*, Dec. 23, 2023.

Landa Arroyo, C., *Revolución digital y Constitución*, Palestra, Lima, 2023.

Pariser, E., *The Filter Bubble. What the Internet Is Hiding from You*, Penguin Books, 2011.

Presno Linera, M.A., *Derechos Fundamentales e Inteligencia Artificial*, Marcial Pons, 2022.

Saba, J., Kwok, K.: *X, la red social de Elon Musk, es un agujero negro de valor*, *El País/CincoDías*, 5/10/2023.

Schauer, F., *Facts and the First Amendment*, *UCLA Law Review*, Volume 57. Issue 4, April 2010.

Thompson, S. A., *Traffic surged on Trump's social media website as an indictment*, The New York Times, April 4, 2023

Thompson, S.A. y **Warzel, C**.: *How Facebook Incubated the Insurrection. Right-wing influencers embraced extremist views, and Facebook rewarded them*, en *The New York Times*, 14/012021.

Zadrozny, B., *Trump's blog isn't lighting up the internet*, May 11, 2021, *nbcnews.com.*

2.

PÚBLICO Y PRIVADO EN LA REGULACIÓN DEL METAVERSO

Moisés Barrio Andrés[1]
Letrado del Consejo de Estado

SUMARIO: I. Introducción. II. Origen del metaverso. III. Concepto. IV. Aplicaciones y elementos. V. Algunos desafios jurídicos. VI. Metaverso, estado y soberanía. VII. Conclusión.

I. Introducción

La transformación digital vigente se debe a la aplicación, en todos los órdenes de la sociedad, de un progreso exponencial de las tecnologías disruptivas, potenciado por una convergencia o acoplamiento de tecnologías con diferente origen o aplicación. Pero los seres humanos, en nuestro proceso de adaptación a los acontecimientos y retos, seguimos, en cambio, un proceso lineal. Cuando los seres humanos lineales nos enfrentamos a un cambio exponencial, no podemos adaptarnos a ese cambio con facilidad y es lo que denominados un cambio de paradigma repleto de incógnitas y derivadas en multitud de dimensiones de la vida: individual, familiar, laboral, social, económico, político, jurídico, etc.

1 Letrado del Consejo de Estado. Asesor de diversos Estados y de la Unión Europea en materia de regulación digital. Profesor de Derecho digital. Director del Diploma de Alta Especialización en Legal Tech y transformación digital (DAELT) de la Escuela de Práctica Jurídica de la Universidad Complutense de Madrid.
Datos de contacto: Consejo de Estado, c/. Mayor, 79. 28013 - Madrid. Correo electrónico: moises.barrio@consejo-estado.es. ORCID: https://orcid.org/0000-0002-2877-5890

En este contexto, la introducción y popularización de Internet[2] durante la segunda mitad de los años noventa del siglo pasado ya dio origen al concepto más genérico y primigenio de «sociedad digital»; mientras que la acepción más específica de «economía digital» se consagra unos años más tarde, como el nuevo modelo económico-empresarial de la segunda década del siglo XXI, particularmente a partir del año 2007, coincidiendo en el tiempo con el estallido de la crisis económica global (2008-2014) y con la aparición del *smartphone*.

Conviene añadir otro acontecimiento de calado: el nacimiento del Bitcoin y de la tecnología *blockchain* que lo sustenta, en 2008, dando lugar, en plena crisis financiera, al inicio del primer sistema de registro financiero distribuido digital independiente de las soberanías monetarias nacionales, una verdadera disrupción financiera que hace tambalear algunos paradigmas clásicos de los modelos centralizados, fundamentalmente basados en la soberanía de los Estados, y que lleva a introducir el nuevo concepto de «descentralización de la confianza»[3]. El último hito en esta evolución viene marcado por la irrupción del Metaverso.

El Metaverso, que proviene del prefijo griego «meta» y la raíz «verso», una retroformación de la palabra universo, aspira a ser el próximo ecosistema digital que reemplace a las actuales redes sociales. Pretende transformar las relaciones humanas con nuevas formas de relación con los avatares virtuales impulsados por inteligencia artificial. Y este flamante espacio se presenta como una red de mundos virtuales persistente, es decir, no desaparece cuando un usuario ha terminado de usarlo. Del mismo modo, la interoperabilidad posibilitará a los usuarios moverse entre espacios virtuales y acceder a diferentes plataformas y servicios utilizando los mismos dispositivos y activos digitales.

El Metaverso nos permitirá desarrollar nuestras tareas en un ecosistema virtual; es decir, ampliará el entorno social, difuminando la línea divisoria en las dos realidades existentes, la física y la virtual. Esta situación disruptiva traerá un escenario de beneficios y riesgos. Entre los primeros, a mi juicio la creación del Metaverso implica desarrollar la economía a través de bienes y servicios que aún no existen e inspirará, en el camino de su desarrollo tecnológico, a la nueva generación de negocios, por lo que se incrementará la participación de

2 El gran catalizador de ello ha sido Internet, un sistema de comunicaciones inicialmente ideado en los años sesenta en el contexto de la guerra fría. La adopción de sus protocolos de comunicación TCP/IP de forma masiva por parte de las empresas, ciudadanos e instituciones, durante la segunda parte de la década de los años noventa, implicó un hecho singular y extraordinario para la humanidad: su interconexión global y en tiempo real, con capacidad para transmitir información en todos los soportes (texto, datos, audio, imágenes, vídeo, música...) desde y a cualquier lugar del orbe. Esta «Internet de la información» también ha evolucionado con muchas aplicaciones, como es la nueva «Internet del valor», que se atribuye a las tecnologías DLT (*Distributed Ledger Technology*) o de cadena de bloques (*blockchain*) que permiten transmitir activos digitales, evitando así el llamado «problema del doble gasto».

3 *Vid.*, por ejemplo y de forma muy exhaustiva, Ibáñez Jiménez, J. (dir.) (2025). *Tratado de Derecho de la Sociedad digital*. Madrid: Editorial Reus.

lo digital en la economía global de nuestro tiempo. Entre sus riesgos, su conversión en pieza clave de la sociedad digital puede potenciar los riesgos individuales y colectivos ya existentes en la esfera de las redes sociales.

El Metaverso ha cobrado renovada actualidad. Ello ha sido impulsado por el reciente cambio de nombre corporativo de Facebook a Meta Platforms, Inc. (Meta). Pero también es debido a que el mundo de los negocios digitales está buscando nuevas fuentes de beneficios[4]. Según algunas predicciones, como las de JP Morgan[5] y Forbes[6], el mercado de los metaversos alcanzará un billón de dólares en los próximos diez años. Una parte de estas inversiones las están realizando ya empresas de la talla de Microsoft (61.600 millones de euros), Meta (la antigua Facebook, 8.800 millones de euros), Qualcomm (88 millones de euros) o Google (39 millones de euros).

El Metaverso es objeto de atención porque, presuntamente, es un «nuevo» espacio en el que no hay leyes, y algunos erróneamente lo equiparan a la primigenia Internet de los años 90. En cambio, ya hay normas jurídicas en vigor[7] que son aplicables en el Metaverso, como son los grupos normativos relativos a los servicios digitales (en puridad, jurídicamente son servicios de la sociedad de la información), los datos, la ciberseguridad, la propiedad intelectual y la inteligencia artificial, al menos en la Unión Europea, aunque no faltan proyectos legislativos[8], similares al RGPD[9] y al nuevo Reglamento europeo regulador de ciertos usos de la IA[10] (el RIA o la *AI Act*), tramitándose en Estados Unidos y otros países. Y, por último, el Metaverso genera mucho interés también porque es un viejo sueño[11]. Se trata, en efecto, de una red

4 Lo ha estudiado Ducci, F. (2022). *Natural Monopolies in Digital Platform Markets*. Cambridge: Cambridge University Press.

5 *Vid.* https://www.jpmorgan.com/content/dam/jpm/treasury-services/documents/opportunities-in-the-metaverse.pdf

6 *Vid*. https://www.forbes.com/sites/zengernews/2022/09/28/defining-the-metaverse-and-the-future-of-physical-and-digital-brands/

7 Sobre su regulación, Barrio Andrés, M. (2025). *Manual de Derecho digital*. Valencia: Tirant lo Blanch, 4.ª edición.

8 Martini, M. (2020). «Regulating Algorithms. How to Demystify the Alchemy of Code», en Ebers, M. y Navas, S. (eds.). *Algorithms and Law*. Cambridge: Cambridge University Press.

9 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos).

10 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

11 Ellis, G. F. R. y otros (2004). «Multiverses and physical cosmology», *Monthly Notices of the Royal Astronomical Society*, núm. 347, pp. 921-936.

de mundos virtuales en los que nos sumergiremos para disfrutar de diversas experiencias sin movernos de nuestro asiento. Una propuesta atractiva, que se adapta a nosotros, a nuestros deseos y órdenes, y no al revés. Pero, ¿qué es exactamente el Metaverso? ¿Y qué cuestiones jurídicas fundamentales subyacen a su desarrollo? A ello dedicaré las próximas páginas de este capítulo.

II. Origen del metaverso

El término Metaverso es una noción *vintage*, que lleva en uso varias décadas. Así, se acuñó y apareció por primera vez en la novela[12] de ciencia ficción titulada *Snow Crash*, de Neal Stevenson, que fue publicada en 1992. Esta obra representaba un universo paralelo de realidad virtual creado a partir de gráficos por ordenador, al que los usuarios de todo el mundo pueden acceder y conectarse a través de gafas y auriculares. La columna vertebral del Metaverso es un protocolo llamado «Street», que enlaza diferentes barrios y lugares virtuales, un concepto análogo al de superautopista de la información. Los usuarios se materializan en el Metaverso en cuerpos digitales configurables llamados avatares. Aunque el Metaverso de Stevenson es digital y sintético, las experiencias en él pueden tener un impacto real en la persona humana. Ahora bien, esta idea de Stevenson no era totalmente nueva, sino que, a su vez, acusa una profunda influencia de la novela[13] de ciencia ficción de William Gibson titulada *Neuromancer* y publicada en 1984. En ella, Gibson introdujo el concepto de ciberespacio y esbozó un entorno virtual denominado «Matrix».

Una reencarnación literaria más reciente del Metaverso es el «OASIS», desarrollado en la novela[14] de ciencia ficción de 2011 *Ready Player One*, escrita por Ernest Cline. Este «OASIS» es un juego de realidad virtual multiusuario masivo que se ha convertido en el destino *online* predominante para el trabajo, la educación y el entretenimiento. Es un mundo de juego abierto, una constelación de planetas virtuales. Los usuarios se conectan a «OASIS» con auriculares, guantes hápticos y trajes. En cuanto a la educación, «OASIS» es mucho más que una biblioteca pública que contiene todos los libros del mundo de acceso libre y gratuito para los ciudadanos. Presenta una visión tecno-utópica de la educación virtual en línea. Así, la experiencia de una clase de la escuela metaversiana es superior a la de una impartida en un colegio del mundo analógico. Los profesores llevan a los alumnos de excursión virtual a civilizaciones antiguas, países extranjeros, museos de élite, otros planetas e incluso al interior del cuerpo humano. Como resultado, los alumnos prestan atención, se comprometen y se interesan por las lecciones.

12 Stevenson, N. (1992). *Snow Crash*. Londres: Penguin.

13 Gibson, W. (1984). *Neuromancer*. Nueva York: Ace Books.

14 Cline, E. (2011). *Ready Player One*. Nueva York: Crown Publishers.

Esta noción literaria se ha aplicado técnicamente en videojuegos[15] de construcción de mundos como Roblox, Minecraft y Fortnite, en los juegos de rol multijugador masivos en línea como Halo, Final Fantasy y World of Warcraft, así como en los mundos virtuales como Second Life. Estas plataformas en línea permiten a los usuarios interactuar entre sí, relacionarse con personajes generados por ordenador y jugar o realizar actividades utilizando elementos del entorno virtual. Para fomentar la interacción entre los participantes, cada uno de ellos está representado por un avatar. Los avatares pueden estar diseñados para parecerse al usuario del mundo real, para ser muy fantasiosos o para ser cualquier cosa intermedia. Como expresan estas manifestaciones iniciales del Metaverso, ya hay metaversos disponibles para el público, y el Metaverso aspira a identificar un multiverso compuesto por diferentes metaversos. De ahí que debamos precisar a renglón seguido el concepto de Metaverso.

III. Concepto

Como acabamos de señalar, de momento no existe un único Metaverso. En 2025 podemos encontrar metaversos[16] dedicados a los deportes, películas, obras de arte, casinos, construcción, comercio electrónico, relaciones sociales y otras tantas aplicaciones.

De este modo, al igual que existe la Web y los sitios web, existe el Metaverso y los «metaversos» en plural. Aludir al Metaverso se refiere probablemente a la ambición de proporcionar la interfaz para el Metaverso y sus metaversos. Es decir, una suerte de programar el navegador para acceder a la Web, que permite acceder a cualquier sitio web o plataforma digital con independencia de su ubicación o lenguaje de programación.

El «Metaverso», aunque en realidad deberíamos hablar de metaversos en plural, ya que no existe por el momento una única plataforma y no es probable que pueda lograrse a medio plazo, lo he definido[17] en un artículo académico del modo siguiente:

> «una aplicación de Internet que consiste en un ecosistema virtual y tridimensional (3D) en el que los usuarios interactúan entre ellos, desarrollan actividades de ocio (muy destacadamente, jugar a videojuegos o e-sports), entablan relaciones económicas o de cualquier otro tipo, como sucede hoy en la Red».

15 Villasmil Espinoza, J. J. (2022). «Multiversos y metaversos en la presente realidad global», en *Multiverso law journal*, núm. 2.

16 Scolari, C. (2022). *La guerra de las plataformas. Del papiro al metaverso*. Barcelona: Anagrama, cap. 7.

17 Barrio Andrés, M. (2022). «Metaverso: origen, concepto y aplicaciones», en *Revista LA LEY Derecho Digital e Innovación*, núm. 12.

El Metaverso (con mayúsculas de nombre propio) aspira, por tanto, a crear todo un nuevo mundo virtual que complemente, mejore —e incluso sustituya, según algunos— al mundo físico[18], digitalizando por el momento algunas experiencias de índole lúdica, educativa, social, laboral, de información, de comercio electrónico, etc. Así, y como he señalado, Second Life, Fortnite, Minecraft y Roblox son ejemplos de juegos colaborativos y de construcción de mundos que incluyen *ciertas* características del Metaverso en su forma de jugar.

Es decir, el Metaverso sería una *versión 2.0* del ciberespacio[19], diferenciada por basarse en un ecosistema virtual y tridimensional, donde sus participantes interactúan social y económicamente como avatares, a través de una plataforma específica que se construye como una metáfora del mundo real, pero sin los condicionamientos físicos, económicos y jurídicos del mundo *offline*. Devendrá un mundo en 3D que se superpondrá al mundo real, donde los participantes[20] podrán comprar, jugar y entablar relaciones en entornos virtuales colaborativos. Muy resumidamente, el Metaverso será la versión evolucionada y en 3D de las actuales redes sociales en 2D.

Técnicamente[21], el Metaverso es una fusión de múltiples tecnologías emergentes como el 5G, la inteligencia artificial (IA), la realidad extendida y los gemelos digitales. Las principales tecnologías necesarias en el Metaverso son las siguientes:

1) La tecnología fundamental para la realización del Metaverso es la tecnología de realidad extendida, incluidas la realidad aumentada (RA o AR en inglés) y la realidad virtual (RV o VR). Mientras que la RA puede superponer y sobreponer información digital al entorno físico, la RV permite a los usuarios experimentar el mundo digital de forma intensa. Ambas tecnologías son capitales en el desarrollo del Metaverso, que crea un espacio digital donde los usuarios pueden interactuar como en el mundo real, o más allá del mismo.

2) La segunda tecnología relevante es el gemelo digital (*digital twin* o DT), que genera un gemelo virtual de un objeto del mundo real utilizando sus

18 Higueras Redecillas, J. C. (2022). «El metaverso: La hoja de ruta hacia el big bang empresarial», en *Harvard Deusto business review*, núm. 322, p. 6-14.

19 Barrio Andrés, M. (2018). *Ciberderecho. Bases estructurales, modelos de regulación e instituciones de gobernanza de Internet*. Valencia: Tirant lo Blanch.

20 Manzano Peláez, E. (2022). «Servicios personalizados, proactivos y la experiencia ciudadana en un mundo en permanente cambio», en *Consultor de los ayuntamientos y de los juzgados: Revista técnica especializada en administración local y justicia municipal*, núm. 1.

21 En detalle, *vid*. Lee, L. H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C. y Hui, P. (2021). «All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda», en *Journal of latex class files*, núm. 14; o Wang, Z. y otros (2022). «A Survey on Metaverse: Fundamentals, Security, and Privacy», *arXiv*, 2203.02662.

datos para predecir su comportamiento esperado. Estos gemelos digitales hacen posible el manejo y control remoto de máquinas o vehículos con una mejor visualización y coordinación, y beneficia tanto al sector industrial como al militar. Ya se utilizan en el ámbito del IoT industrial. En el Metaverso, el gemelo digital puede reflejar el mundo real en el mundo virtual. En consecuencia, el Metaverso también ayuda a encontrar algunas soluciones de prueba para los problemas no resueltos en el mundo físico.

3) La tercera tecnología estructural, la cadena de bloques, desempeña dos funciones insustituibles[22] en el Metaverso. Por un lado, la tecnología *blockchain* sirve de repositorio, por lo que los usuarios pueden utilizarla para almacenar datos en cualquier lugar del Metaverso. Por otro lado, la tecnología *blockchain* puede proporcionar un sistema económico completo para conectar el mundo virtual del Metaverso con el mundo real. Especialmente, los *tokens* NFT[23] (Non Fungible Tokens) permiten que los bienes físicos se conviertan en objetos virtuales. Los usuarios pueden comerciar con objetos virtuales de la misma manera que en el mundo real. Por lo tanto, la cadena de bloques tiende un puente entre el mundo real y el Metaverso.

Así las cosas, el Metaverso es la solución que amalgama todas estas tecnologías pertinentes en un contexto global. Este concepto crea un entorno digital simulado que se configura como un mundo virtual inmersivo para sus prosumidores. Los usuarios pueden interactuar[24] con este ecosistema virtual a través de sus «avatares» o *alter egos* digitales. Concretamente, los avatares son las personificaciones virtuales de los usuarios, y habrá que determinar su estatuto jurídico; es decir, si tienen o no los mismos derechos que en el mundo físico.

Para acceder al Metaverso es preciso disponer de un dispositivo hardware inmersivo habilitado para la RV/RA, fundamentalmente unos auriculares o unas gafas, como las Oculus de Facebook o las Hololens de Microsoft. Ahora bien, los trajes hápticos de cuerpo entero, como el Teslasuit o el Holosuit, tienen el potencial de llevar la experiencia inmersiva a su máxima expresión con la capacidad de rastrear los movimientos y extraer la retroalimentación háptica junto con la biometría trascendida. Es decir, con ellos el usuario experimentará una sensación de inmersión total dentro de ese otro mundo paralelo.

Clave en el Metaverso es la interacción dual entre ambos mundos, el físico y el virtual. Será factible volcar al mundo digital virtual del Metaverso toda la vida y el trabajo del mundo físico, así como el entorno social e incluso el país,

22 Abadi, J. y Brunnermeier, M. K. (2019), «Blockchain Economics», *CEPR Discussion Paper*, DP13420, pp. 1-53, disponible en: https://ssrn.com/abstract=3310346.

23 Sobre su problemática en el metaverso, *vid*. Vidal-Tomás, D. (2022). «The new crypto niche: NFTS, play-to-earn, and metaverse tokens», en *Finance Research Letters*, núm. 47.

24 Pacheco Jiménez, M. N. (2022). «De la digitalización de los pagos a los tokens del metaverso», en *Revista LA LEY mercantil*, núm. 91.

lo cual abre posibilidades que aún no podemos atisbar. En consecuencia, el vínculo entre el mundo físico y el mundo virtual se consigue a través de una serie de tecnologías inmersivas ya desarrolladas en el marco del IoT[25], como guantes táctiles, cascos, zapatos de inducción, equipos de realidad virtual, etc. Todo este hardware ayuda a establecer un vínculo con el mundo virtual. El vínculo entre el mundo físico y el mundo virtual sólo puede generarse técnicamente mediante servicios interactivos. Por ejemplo, si el contenido de este artículo se muestra en la pantalla de un Metaverso, se lleva a cabo mediante un enlace con un servicio de presentación de contenidos. Si introduzco el contenido del teclado del mundo físico en un dispositivo del mundo digital, se trata de la vinculación de periféricos de entrada. Y la información del mundo virtual también puede transferirse al mundo físico.

A la postre, estimo que la esencia técnica del Metaverso es, en realidad, establecer un vínculo entre el mundo físico y el mundo virtual, y proporcionar una serie de métodos, plataformas y herramientas en el mundo virtual para construir el Metaverso deseado. Eso es precisamente lo que expresa el término «Metaverso»: *meta*, como elemento de abstracción de los componentes (como los metadatos en un esquema XML), y *verso*, como abreviatura de universo.

No obstante, por el momento y a pesar de las fuertes inversiones corporativas que recogíamos al inicio de este trabajo, todavía no se ha creado un Metaverso «operativo», y es probable que tarde en lograrse varios años, si no décadas. Eso no ha impedido que un montón de empresas desarrollen herramientas de construcción y proto-mundos que, con toda seguridad, se convertirán en partes clave del Metaverso.

IV. Aplicaciones y elementos

A fecha de escribir estas líneas, el Metaverso no existe como tal. Actualmente encontramos una primera iteración de esta tecnología, es decir, su *versión 1.0*. Mundos virtuales como Roblox y Fortnite ofrecen una representación aspiracional de lo que es el Metaverso, pero ni siquiera son representaciones exactas.

Cualquier lector que haya utilizado Second Life, un precursor del Metaverso recordará que los gráficos y la inmersión eran decididamente pobres. La experiencia era del todo menos «extendida», y hacía falta bastante paciencia e imaginación para sentirse «inmerso». Hoy en día, el realismo de los metaversos ha avanzado diversos grados, pero todavía la tecnología se haya en un estado incipiente. En los metaversos aún no hay sabores, fragancias ni sensaciones en la piel.

25 Barrio Andrés, M. (2022). *Internet de las Cosas*. Madrid: Reus, 3.ª edición.

Por el momento, esos mundos virtuales son, en realidad[26], «jardines amurallados» en 3D, que ofrecen *algunas* experiencias similares a las del Metaverso dentro de un entorno cerrado en lugar de un mundo completo de realidad aumentada. El Metaverso necesitará tiempo e ingenio para evolucionar más allá de sus límites artificiales y convertirse en algo más grande, al igual que lo hizo la primera Internet.

Y aunque nos encontramos en un estado inicial de esta tecnología, por el momento podemos apuntar algunas de sus aplicaciones, como serían las siguientes:

a) Reuniones en línea. La eclosión de videoconferencias durante la pandemia también ha puesto de relieve lo esencial que resulta la comunicación cara a cara para paliar algunas de sus debilidades, como los retrasos en las interacciones o las confusiones y equívocos. Por eso, en el Metaverso las personas pueden usar un avatar amigable[27] para caminar y trabajar en el espacio virtual. Incluso se puede utilizar el lenguaje corporal o la interacción visual para comunicarse con el interlocutor desde diferentes ángulos, lo que mejorará en gran medida las reuniones a distancia.

b) Bienes inmuebles digitales. En el Metaverso existen activos digitales, incluyendo bienes inmuebles[28]. También, los mismos factores urbanísticos del mundo físico, como la ubicación, los servicios y el transporte, pueden influir en el precio de la vivienda en el mundo virtual. Por ejemplo, en el Metaverso los usuarios pueden comprar y vender casas, así como organizar exposiciones de arte, festivales de música, competiciones de juegos, etc. Ya hemos asistido a transacciones sobre estos activos, algunos de los cuales se han ofrecido por medio de subasta y se intercambian por criptoactivos del tipo NFT. Un estudio[29] más detallado lo ha realizado Rosario Jiménez Rubio, al que me remito.

c) Arte digital. Tradicionalmente, los autores crean imágenes 3D a través de algunas herramientas de modelado como Photoshop o ZBrush. Sin embargo, el Metaverso permite controlar la capa de visualización, lo que aporta nuevas formas de expresión y creación de obras de arte,

26 Murray, A. y otros (2022). «The Promise of a Better Internet: What Is Web 3.0 and What Are We Building?», *Oregon Law Review*, núm. 1, pág. 4.

27 Cheong, B. C. (2022). «Avatars in the Metaverse: Potential Legal Issues and Remedies», *International Cybersecurity Law Review*, núm. 1.

28 Gallego Fernández, L. A. (2018). «Cadenas de bloques y Registros de derechos», *Revista Crítica de Derecho Inmobiliario*, vol. 94, núm. 765.

29 Jiménez Rubio, R. (2022). «El Metaverso y el Derecho Registral», *Revista LA LEY Derecho digital e innovación*, núm. 12.

así como de gestión de los correspondientes derechos de autor[30]. Del mismo modo, con la creciente popularidad de las obras artísticas generadas, en todo o en parte, mediante IA, cada vez existe más interés por el arte digital. En fin, en una galería virtual situada en el Metaverso los usuarios pueden entrar ella para apreciar, desde todas las dimensiones, las obras expuestas.

d) Ocio y entretenimiento. Precisamente los videojuegos son una de las aplicaciones que ahora están espoleando el desarrollo de los metaversos. Ya hemos asistido incluso a algún espectáculo en el Metaverso, como son los conciertos celebrados en las plataformas de juegos Roblox y Fortinite, con las actuaciones del rapero Travis Scott y de la cantante Ariana Grande.

e) Deportes. Los tecnocomplementos *werables* existentes en el Internet de las Cosas, unidos a herramientas de realidad virtual, permiten construir entornos deportivos virtuales y envolventes, con posibilidades de adaptación y personalización adicionales[31]. Será posible jugar al fútbol dentro del estadio de nuestro equipo favorito o presenciar los espectáculos deportivos desde distintos ángulos, por ejemplo.

f) Educación. Otro campo de aplicación actualmente en desarrollo es el ámbito educativo[32], a fin de aprovechar las características en 3D e inmersivas de esta tecnología. Algunas universidades españolas, como la Universidad CEU San Pablo, ya están desarrollando experiencias en este campo por medio de la herramienta «Minecraft Education Edition», que potencia la creatividad, la colaboración y la resolución de problemas a través de una experiencia inmersiva tridimensional.

g) Comercio. Al igual que sucedió en su momento con Internet y el comercio electrónico, el Metaverso es un canal de contratación de bienes y servicios, tanto físicos como digitales. Del mismo modo, se abrirán nuevos mercados que permitirán conectar directamente[33] a los prestadores de servicios con sus clientes, con la posibilidad de pagar el precio en una criptomoneda, hacer uso de emisiones de *tokens* o utilizar una DAO (Descentralized Autonomous Organization).

De esta gavilla de aplicaciones se desprende la idea de cómo la esencia del Metaverso es gestionar los recursos. Habrá que determinar cuáles serán los

30 Fortnow, M. y Terry, W. (2021). *The NFT Handbook: how to create, sell, and buy non-fungible tokens*. Hoboken: Wiley, cap. 9.

31 Vasco Gómez, A. (2022). *Arbitraje y resolución de conflictos en las competiciones de eSports*. Pamplona: Aranzadi, cap. 1.7.

32 Feenberg, A. (2019). «Postdigital or predigital?», *Postdigital Science and Education*, vol. 1, núm. 1, pp. 8-9.

33 Fries, M. (2018). «Private Law Compliance through Smart Contracts?», *Compliance Elliance Journal*, vol. IV, núm. 1, pp. 11-18.

modelos de negocio, quién paga qué o quién es realmente el propietario de algo. A mi juicio, esta gestión tiene dos aspectos clave. En primer lugar, debe ordenar qué recursos están disponibles en el Metaverso. En segundo lugar, y en torno a estos recursos, debe esclarecer cómo crear valor. Desde el punto de vista de los recursos, normalmente están disponibles el avatar o *alter ego* digital, el espacio, la experiencia inmersiva, los contenidos y los activos digitales, como los NFT y otros tipos de criptoactivos. En cuanto a la creación de valor, se trata de una cuestión más complicada en el Metaverso, dada la existencia de muchos escenarios. Los conceptos de empresa, trabajo[34], vida, país, mundo, etc., pueden existir en el Metaverso. Todo ello abre un elenco de posibilidades inéditas.

Asimismo, y como hemos insistido, no puede hablarse de un único Metaverso, sino de metaversos en plural. Incluye una variedad de plataformas, contenidos, servicios y activos que constituyen los elementos más importantes del Metaverso virtual. Los componentes y servicios se presentan en una experiencia inmersiva. Por ejemplo, la capacidad de edificar una plataforma permite construir una gran cantidad de espacios tridimensionales en el mundo digital, determinar las formas que se deseen y crear productos y servicios inéditos.

Desde la perspectiva de la gestión, los impulsores del Metaverso buscan que estos recursos generen valor en el Metaverso, lo que es coherente con nuestros objetivos de gestión en el mundo físico. Además de gestionar los recursos, también es preciso gestionar las finanzas correspondientes. Por lo tanto, la gestión financiera, la gestión de la cadena de suministro, la gestión de los recursos humanos o la gestión de los proyectos en el Metaverso son direcciones que deberán ser estudiadas. Como ya se ha hecho con las relaciones laborales[35]. En cuanto a la gestión de equipos, incluye la comunicación social y las relaciones virtuales. Para crear servicios valiosos en el espacio virtual, hay que atraer a los clientes y fidelizarlos. Por tanto, hay que gestionar a los clientes, y también hay que hacer *marketing* y promoción en el mundo virtual.

Pensemos por ejemplo en la construcción de una plataforma de aprendizaje en línea[36], como ya están haciendo algunas universidades españolas y latinoamericanas. Lo primero a determinar es si se basará en una plata-

34 López Cumbre, L. (2022). «Relaciones laborales "por", "para" y "en" el metaverso», *Diario La Ley*, núm. 10033; De la Torre, C. (2022). «El futuro del trabajo y el metaverso laboral en el periodo pos-Covid: Quo vadis?», *Consell obert: Recull informatiu del Consell General de Col·legis de Graduats Socials de Catalunya*, núm. 366.

35 Martínez de Salinas Murillo, J. (2022). «El metaverso laboral», *Observatorio de recursos humanos y relaciones laborales*, núm. 175.

36 Sobre ello, Meyzan Briceño, M. A. (2022). «Desafíos de la educación y el aprendizaje en el Metaverso», *Desafíos: Revista Científica de Ciencias Sociales y Humanidades*, núm. 13, págs. 7-8.

forma metaversiana existente o supondrá programar una nueva plataforma. Una vez decidida la base tecnológica, el paso siguiente vendrá dado por la construcción de escenarios educativos y empresariales, el contenido educativo correspondiente y los servicios relacionados en esta plataforma. Por ejemplo, al explicar en la asignatura Derecho digital el funcionamiento de un operador de telecomunicaciones, podremos en el Metaverso visitar sus instalaciones y los centros de datos.

De este modo, en el Metaverso ya estamos asistiendo al nacimiento de plataformas o metaversos cerrados que incorporan tecnologías de última generación (*blockchain*, criptoactivos, *smart contracts*, IA, realidad virtual y aumentada...) con el propósito no sólo de crear en su entorno un mercado económico[37] de compraventa, innovación, inversión o especulación, sino también de imponer el marco jurídico a través de las políticas de uso arbitradas al modo de contratos de adhesión. Por tanto, se suscitan los mismos problemas que plantean las plataformas[38] en el Derecho digital. Y, como sucede en la vigente Web 2.0, son imprescindibles terceros que provean y brinden confianza en la identificación de los usuarios y avatares y las transacciones electrónicas.

Otra cuestión clave será la localización de los recursos. En el Internet actual, se teclea en el navegador web la dirección URL (por ejemplo, https://www.aragon.es/tramites) —o el nombre de dominio correspondiente—. O, lo que es más habitual, se recurre a un buscador o a hacer clic en un enlace compartido en una publicación de una red social. Pero en el Metaverso, aunque ahora también existe una página web de la plataforma experimental, más adelante aparecerá otro método —o métodos— de identificación directa de los recursos. A mi juicio, surgirá una nueva tecnología de direccionamiento en el Metaverso, similar al DNS. Las lecciones que brinda el Derecho digital derivadas de la gestión de los nombres de dominio y los fenómenos de registro de dominios de mala fe (la ciberocupación o *cybersquatting*), serán muy útiles para resolver esta cuestión.

V. Algunos desafíos jurídicos

Por el momento, el Metaverso supone más bien una tecnología *in fieri*. Y plantea diversos retos y problemas pragmáticos en torno a él.

37 Álvarez López, C. y Carrasco Perera, A. (2022). «Operadores y responsabilidad civil en el metaverso», *GAP Análisis*, marzo 2022.

38 Sobre estos problemas, me remito a Barrio Andrés, M. (2025). *Manual de Derecho digital*, *op. cit.*, pág. 256 y ss., y 233 y ss. También, Montero Pascual, J. J. (dir.) (2017). *La regulación de la economía colaborativa: Airbnb, BlaBlaCar, Uber y otras plataformas*. Valencia: Tirant lo Blanch.

El más crítico es la falta de una infraestructura digital robusta[39] para ofrecer los servicios y aplicaciones garantizados con las capacidades de procesamiento y de trabajo en red necesarias. Aquí la tecnología 5G resulta esencial, al igual que la ciberseguridad de estas infraestructuras.

Otro reto importante se concreta en la imprescindible compatibilidad e interoperabilidad[40] entre los mundos virtual y físico, que deben definirse y estandarizarse antes de lanzar el Metaverso. Sin embargo, las lecciones extraídas por ejemplo del IoT en este campo no son nada alentadoras. Los organismos de estandarización deben tener presentes cuestiones relacionadas con derechos de propiedad industrial e intelectual, entre las que destacan las obligaciones de divulgación de patentes esenciales (SEP) y la oferta de licencias a los usuarios del estándar en términos FRAND (Fair, Reasonable and Non-Discrimatory, esto es, justos, razonables y no discriminatorios). Avanzando en la estandarización del Metaverso, el 21 de julio de 2022 fue constituido el Metaverse Standards Forum (MSF)[41], como entidad de cooperación entre empresas y organismos de desarrollo de estándares lógicos para generar estándares de interoperabilidad que permitan dar lugar a un Metaverso abierto.

Del mismo modo, también se suscitan cuestiones derivadas del acceso[42] a esta tecnología y de eventuales brechas[43] de acceso. Dado que es preciso que el usuario esté equipado con un casco u otro dispositivo hardware *ad hoc* con mayores funcionalidades, pero más económicamente costoso, la inversión económica necesaria individual hace que sea un servicio privilegiado en lugar de un sistema abierto para todos, como ha ocurrido con Internet.

En otro orden de cosas, un reciente informe del Parlamento Europeo de junio de 2022[44] ha puesto de manifiesto que la construcción del Metaverso otorgará a algunas empresas oportunidades inigualables para monopolizar los mercados digitales. Advierte que las mismas poderosas empresas tecno-

39 En detalle, Falco, G. y Rosenbach, E. (2022). *Confronting Cyber Risk: An Embedded Endurance Strategy for Cybersecurity*. Oxford: Oxford University Press.

40 Mansilla Pizá, A. (2022). «Metaverse. Self-Regulation, Intellectual Property, E-Commerce, and Dispute Resolution», *Revista LA LEY Derecho digital e innovación*, núm. 10.

41 El MSF agrupa entre sus principales fundadores a empresas como Microsoft, la propia Meta, Adobe, Alibaba, Epic Games, Huawei, Ikea, Nvidia, Qualcomm o Sony; y a entidades de estándares de tanta relevancia como el World Wide Web Consortium (W3C) del inventor de la World Wide Web (Sir Tim Berners-Lee).

42 Allen, Jason y otros (2022). «Cryptoassets in Private Law: Emerging Trends and Open Questions from the First 10 Years», *SMU Centre for AI & Data Governance Research Paper*, núm. 06/2022.

43 Cauffmann, C. y Goanta, C. (2021). «A New Order: The Digital Services Act and Consumer Protection», *European Journal of Risk Regulation*, núm. 1, pp. 1-17.

44 PE 733.557. Disponible en: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733557/EPRS_BRI(2022)733557_EN.pdf

lógicas que actualmente dominan los mercados digitales podrían controlar también el nuevo entorno, con todos los incentivos para perpetuar las actuales prácticas anticompetitivas, como la autopreferenciación (es decir, favorecer sus propios productos) y los patrones oscuros (es decir, diseñar sitios web e interfaces de aplicaciones móviles para influir en el comportamiento y la toma de decisiones de los usuarios). Algunos reguladores nacionales, como el Bundeskartellamt alemán, ya están tratando de abordar las posiciones dominantes en el emergente mercado de la realidad virtual (por ejemplo, el caso Oculus). El «hecho de que el entorno metaverso requiera que los competidores se comuniquen, colaboren y garanticen la interoperabilidad de las plataformas también podría dar lugar a una serie de desafíos antimonopolio, por ejemplo, en relación con el intercambio de información sensible, como la fijación de precios, o los acuerdos entre competidores sujetos al escrutinio del Derecho de la competencia», subraya.

Por lo demás, la protección de datos[45] y la ciberseguridad[46] son aspectos imperativos, cuyas normas reguladoras resultan aplicables a estas plataformas. De hecho, los metaversos existentes no están huérfanos de regulación, sino que jurídicamente constituyen servicios de la sociedad de la información, regulados por el Derecho digital mediante la Directiva 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000, relativa a determinados aspectos jurídicos de los servicios de la sociedad de la información, en particular el comercio electrónico en el mercado interior (la DCE), transpuesta en nuestro ordenamiento por medio de la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico (LSSI). A este grupo normativo se suman dos nuevos reglamentos europeos[47]: el Reglamento de Mercados Digitales (DMA)[48] y el Reglamento de Servicios Digitales (DSA)[49], destinados a regular en la Unión Europea las mayores plataformas en línea.

En consecuencia, también se originan aquí los grandes retos jurídicos que suscitan las tecnologías emergentes[50], como es la salvaguardia de la privaci-

45 Ramos, D. (2022). «Realidad, Metaversos y Protección de Datos», *LA LEY privacidad*, núm. 11.

46 Sempere, F. J. (2022). «Smart-Glasses y el metaverso: un viaje por las tecnologías que usan cámaras», *Revista LA LEY Derecho digital e innovación*, núm. 12.

47 Sobre ellos, *vid*. Barrio Andrés, M. (2025). *Manual de Derecho digital*, *op. cit*., pág. 213 y ss.

48 Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo, de 14 de septiembre de 2022, sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales).

49 Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).

50 En detalle, Ibáñez Jiménez, J. (coord.) (2022). *Transformación digital y nueva regulación de personas, instituciones y mercados*. Madrid: Reus; Recuerda Girela, M. A. (dir.) (2019). *Tecnologías disruptivas*. Pamplona: Aranzadi.

dad, protección de datos y ciberseguridad, tutelar la propiedad intelectual e industrial, la regulación de las transacciones comerciales, el reconocimiento de la identidad digital, la garantía de los derechos digitales o el desarrollo de políticas públicas digitales, que entre otros objetivos atiendan a los colectivos especialmente vulnerables o combatan las distintas brechas. La Carta de Derechos digitales de España[51], de 2021, puede servirnos de faro en varios aspectos. Por ejemplo, coadyuva al impulso de políticas públicas digitales. Del mismo modo, nuestra Carta es un útil instrumento interpretativo de algunos conceptos difusos en la legislación vigente. O, en fin, constituye el inicio de un debate sobre nuevos derechos digitales no positivizados hasta la fecha y la forma en que deberían modularse, como es el caso del empleo de las neurotecnologías o la regulación de algunos aspectos inéditos que plantea el Metaverso.

De todos estos problemas, y en lo que afecta al Derecho constitucional, el Metaverso plantea un problema que atañe a la propia médula de los Estados, a su soberanía. Ello requiere una atención específica.

VI. Metaverso, Estado y soberanía

Para examinar cómo el Metaverso está gestando un nuevo poder supraestatal y cómo esto puede suponer una nueva forma de soberanía privada, es esencial partir de la función reguladora de la tecnología[52]. Por un lado, la tecnología siempre incorpora los valores de sus creadores y, por tanto, puede ser un poderoso instrumento para proyectar dichos valores incluso más allá de las fronteras nacionales de su país de origen. Por otro lado, la tecnología puede reutilizarse y utilizarse para aplicaciones que pueden diferir considerablemente de los objetivos originales[53] de sus creadores. La ola de innovación abierta e ilimitada que trajo el PC, el ordenador personal, es un buen exponente de ello. Por lo tanto, al igual que otras herramientas que permiten la regulación, la tecnología puede promover o restringir comportamientos específicos, pero, como no es inamovible, también puede actualizarse, modificando en consecuencia el tipo de comportamientos que promueve o dificulta.

Si bien este razonamiento se aplica a cualquier tecnología, la tecnología digital del Metaverso maximiza estas funciones, ya que es mucho más escalable y fácil de actualizar que cualquier otra tecnología. Si el Metaverso triunfa, una sola tecnología nueva será adoptada en pocos años por millo-

51 Sobre ella, *vid*. Rodríguez Ayuso, J. F. (coord.) (2022). *Nuevos retos en materia de derechos digitales en un contexto de pandemia: perspectiva multidisciplinar*. Pamplona: Aranzadi.

52 Mitchell, W.J. (1996). *City of bits: Space, Place, and the Infobahn*. Cambridge: Cambridge University Press, pág. 110.

53 Rodotà, S. (1995). *Tecnologia e diritti*. Bolonia: Il Mulino, cap. V.

nes, o incluso miles de millones, de personas y empresas, y su estructura de software puede actualizarse con relativa facilidad, en comparación con otras tecnologías. Estos rasgos hacen que sea particularmente relevante reflexionar detenidamente sobre cómo el Metaverso afecta las formas en que sus tecnologías subyacentes configuran la esfera jurídica de las personas físicas y jurídicas.

En el Derecho digital, es importante poner de manifiesto que, debido a la inexistencia o ineficacia de instituciones y regímenes globales que permitan regular sus actores e instituciones específicas[54], el establecimiento de actores privados globales adquiere una relevancia particular, ya que no solo representan un motor de desarrollo económico global, sino también actúan como mecanismo para logar un alcance universal de la regulación nacional, especialmente de ciertas superpotencias. En este sentido se ha subrayado que la «aterritorialidad» de Internet ha incidido muy negativamente en la capacidad de los poderes territoriales de los Estados para implementar las normas en defensa de los derechos de los ciudadanos[55]. Por lo tanto, es importante tener en cuenta que la delegación de funciones regulatorias a sujetos privados puede permitir encontrar soluciones eficientes y eficaces para regular áreas de política específicas[56], pero debemos reconocer que la eficiencia y la eficacia no son sinónimos de justicia y Estado de derecho. En este contexto, si bien los efectos de la regulación por actores privados transnacionales pueden ser muy beneficiosos en términos de eficiencia y eficacia, la fiscalización final solo puede ser pública, si se quiere garantizar el respeto de los derechos fundamentales, la tutela judicial efectiva y el propio Estado de derecho. La Constitución es el texto que, debidamente interrogado según «el método habitual de las reconstrucciones técnico-jurídicas» en expresión de Carlo Esposito[57], permite que los valores constitucionales proporcionen las respuestas a las solicitudes de regulación de nuevos casos derivados de la transformación digital.

En efecto, en las dos últimas décadas hemos asistido a la emergencia de grandes plataformas digitales en torno a las redes sociales, sobre las cuales un prestigioso sector doctrinal[58] ha puesto de manifiesto que están explo-

54 Teubner, G. (2012). *Constitutional fragments. Social constitutionalism and globalization*. Oxford: Oxford University Press, pág. 66; Barrio Andrés, M. (2018). *Ciberderecho. Bases estructurales, modelos de regulación e instituciones de gobernanza de Internet*. Valencia: Tirant lo Blanch, pág. 94 y ss.

55 Grimm, D. (2016). *Constitutionalism: Past, Present, and Future*. Oxford: Oxford University Press, pág. 343.

56 Rijgersberd, R. W. (2019). *The State of Interdependence. Globalization, Internet and Constitutional Governance*. La Haya: Asser Press, pp. 49-68.

57 Esposito, C. (1954). *La Costituzione italiana. Saggi*. Padua: Cedam, pág. 266.

58 *Vid*., por todos, Zuboff, S. (2020). *La era del capitalismo de la vigilancia*. Barcelona: Paidós Ibérica.

tando con éxito su poder estructural para adquirir un conjunto de capacidades que normalmente solo se ven en Estados soberanos. Como ha señalado MANETTI[59], este *statu quo* probablemente fue promovido por los gobiernos que facilitaron el establecimiento de tales empresas de tecnología como una estrategia para afirmar su jurisdicción global, pero ahora parece haber dado lugar a entidades que ni siquiera son controlables por dichos Estados y estos colosos tecnológicos están tratando concretamente de afirmar su propia soberanía privada. Recientemente se ha vuelto a producir un nuevo embate al papel de los Estados soberanos con el fenómeno de las criptodivisas, que no son emitidas por un banco central u otra autoridad pública y, por ello, constituyen una amenaza a la soberanía monetaria de los Estados, que ha sido siempre la «premisa indispensable para el ejercicio de las funciones institucionales de iniciativa y dirección de la política económica»[60]. Recordemos que, entre nosotros, la potestad de acuñar moneda es considerada, desde el Fuero Viejo de Castilla, como la más esencial entre las que integran la soberanía del Estado.

Por eso, a medida que las plataformas digitales son cada vez más omnipresentes en nuestra vida personal y empresarial, el debate público y la atención de los legisladores se está volviendo cada vez más hacia el papel que desempeñan estos intermediarios en el mundo digital, la influencia que ejercen y cómo lo hacen. Si bien *prima facie* pudiera parecer una exageración argumentar que las plataformas de redes sociales se comportan como soberanos privados[61], a continuación examinaremos cómo estos actores privados han integrado muchas funciones previamente reservadas a los actores estatales, lo cual se agudiza en la construcción del Metaverso.

Como recuerda Herman HELLER[62], el Derecho Público y las relaciones internacionales se basan en la premisa de que los Estados son los únicos actores que tienen legitimidad para disfrutar de la soberanía. El Derecho nacional e internacional se fundamentan en el presupuesto de que el Estado tiene el poder soberano como la única organización política capaz de estructurar efectivamente el gobierno de la sociedad. El Derecho Público prescribe tradicionalmente que dicha estructura se define a través de un conjunto de órganos que ejercen los poderes legislativo, ejecutivo y judicial, supuestamente de manera independiente mientras se controlan mutuamente y se equilibran entre sí. Si bien Max WEBER[63] consideró que los Estados son «empresas polí-

59 MANETTI, M. (2020). «Regolare Internet», en *MediaLaws. Rivista di diritto dei media*, núm. 2/2020, pág. 35.

60 BALDWIN, R. y CAVE, M. (1999). *Understanding Regulation*. Oxford: Oxford University Press, pág. 126.

61 MARSDEN, C. (2011). *Internet co-regulation*. Cambridge: Cambridge University Press, pág. 58.

62 HELLER, H. (1943). *Teoría del Estado*. México: Fondo de Cultura Económica, pág. 106 y ss.

63 WEBER, M. (1991). *Escritos políticos*. Madrid: Alianza, pág. 2.

ticas» caracterizadas por «el monopolio del uso legítimo de la fuerza física dentro de un territorio determinado», Carl SCHMITT[64] argumentó que, en lugar de un monopolio de coerción, el soberano disfruta del monopolio de decidir. Estos presupuestos son conceptos estructurales utilizados por el Derecho Público para describir las instituciones y el gobierno de los Estados, pero se tornan problemáticos cuando se examinan en relación con las grandes plataformas digitales y el propio Metaverso.

En efecto, los obstáculos a la coerción estatal en una red universal y de alcance planetario como es Internet y la dificultad para hacer cumplir el Derecho en el entorno digital han llevado, por un lado, a los Estados a confiar en actores privados como representantes de la legislación nacional y, por otro lado, los intermediarios digitales han llenado los vacíos dejados por la inoperancia estatal, al establecer sus propios regímenes contractuales y arquitecturas técnicas encaminadas a regular efectivamente los comportamientos de los ciudadanos en línea. Desde esta perspectiva, las arquitecturas técnicas en las que se basan las plataformas de redes sociales y los nuevos metaversos para su funcionamiento pueden verse como «territorios»[65] digitales en los que las plataformas ejercen su soberanía privada. Dichos territorios están poblados por individuos que deben acatar las reglas de las plataformas, y donde la justicia se imparte a través de sistemas de resolución alternativa de conflictos que se implementan mediante mecanismos de ejecución establecidos unilateralmente por las entidades privadas que ejercen «soberanía» dentro de las fronteras digitales delimitadas por las arquitecturas técnicas de las plataformas. De hecho, algunos autores como GLENN[66] califican este fenómeno en términos de un «eclipse total» de la soberanía.

Y es que cada plataforma de redes sociales o de metaversos que utilizamos normalmente regula el comportamiento de sus usuarios a través de unos «Términos de servicio» o unos «Términos y condiciones». La principal característica de estos contratos predispuestos, jurídicamente denominados contratos de adhesión con condiciones generales que se configura bajo la modalidad de *click agreement*, es que las estipulaciones que incluyen no se negocian. Sus cláusulas contractuales están predeterminadas unilateralmente por una de las partes, que es el proveedor de la plataforma, en nuestro caso. Si bien los términos de servicio se consideran legalmente como cualquier otro contrato, significando la expresión de la voluntad de todas las partes (art. 1255 CC), es natural considerar esta tipología contractual como

64 SCHMITT, C. (1985). *Political Theology: Four Chapters on the Concept of Sovereignty*. Massachusetts: MIT Press, pág. 5.

65 BALKIN, J. M. (2018). «Free Speech is a Triangle», en *Columbia Law Review*, vol. 118, núm. 7/2018, pág. 2024.

66 GLENN, H. P. (2013). *The Cosmopolitan State*. Oxford: Oxford University Press, pág. 171.

la expresión por excelencia del control unilateral[67]. Por ello, la doctrina[68] ha criticado fuertemente este tipo de contratos por el carácter unilateral de las disposiciones, la ausencia de negociación entre las partes y la casi inexistencia del poder de negociación de la parte adherente. En este sentido, la posición de los usuarios de la plataforma —especialmente los de las plataformas dominantes— es notablemente vulnerable, ya que aceptar sin más la regulación contractual establecida unilateralmente es la única opción que pueden tener para utilizar el servicio digital en cuestión. Así sucede en los metaversos actualmente operativos.

Además, es una práctica común que los metaversos incluyan estipulaciones contractuales que permitan la modificación unilateral de sus términos de servicio, ya previamente establecidos unilateralmente, sin siquiera incluir la obligación de notificar a los usuarios sus posteriores modificaciones. Esta situación, que la profesora Shubha Ghosh ha descrito contundentemente como «autoritarismo contractual»[69], se ve agravada aún más por el poder estructural que manifiestan las plataformas digitales al implementar unilateralmente su ordenamiento privado a través de la definición de sus arquitecturas digitales de manera que permiten exclusivamente a los usuarios comportarse de acuerdo con la voluntad de la plataforma. De hecho, este poder regulatorio habría sido utilizado por las grandes plataformas para liberarse[70] de los límites territoriales en los que se basan los modelos tradicionales de gobierno de los Estados.

De ahí que los metaversos gocen de un poder cuasinormativo en forma de capacidad para establecer unilateralmente las disposiciones contractuales que regularán el comportamiento de sus usuarios, definiendo qué actividades pueden realizar, pero también qué datos personales se recabarán sobre ellos y cómo se procesarán dichos datos y los resultados de las evaluaciones algorítmicas. La característica más destacada de los contratos de adhesión comúnmente utilizados por las plataformas de metaversos es que los usuarios solo pueden decidir someterse a los términos preestablecidos. En este contexto, los términos de servicio tienen la fuerza de una «ley de la plataforma» o incluso su «constitución», la cual es establecida y reformada únicamente por el proveedor de la plataforma.

---

67 De Minico, G. (2016). *Antiche libertà e Nuova frontiera digitale*. Turín: Giappichelli, pág. 195 y ss.

68 *Vid*., por ejemplo, Reglero Campos, L. F. (1999). «Régimen de ineficacia de las condiciones generales de la contratación. Cláusulas no incorporadas y cláusulas abusivas: concepto y tipología», en *Aranzadi Civil*, vol. I.

69 Ghosh, S. (2014). «Against Contractual Authoritarianism», en *Southwestern Law School Review*, vol. 44.

70 Hughes, J. (2003). «The Internet and the Persistence of Law», en *Boston College Law Review*, vol. 44, pág. 359.

Ello se completa con unos poderes cuasiejecutivo y cuasijudicial[71]. En primer lugar, las plataformas cuentan con una serie de mecanismos técnicos que impiden a los usuarios comportarse en contra de lo prescrito por la plataforma, como es la moderación algorítmica de contenidos auxiliada por revisores manuales o el filtrado automático de material protegido por derechos de autor o de contenidos pornográficos, lo que puede calificarse de poder cuasiejecutivo. En segundo lugar, el poder cuasijudicial se implementa por medio de sus propios procesos alternativos de resolución de controversias en forma de sistemas ODR[72] que resuelven las disputas entre usuarios en función de su regulación contractual predefinida y, posteriormente, implementan *per se* los resultados de la adjudicación a través de medios técnicos que pueden culminar con la expulsión del usuario, su «muerte digital». Y es que algunas plataformas, como Facebook, han establecido sus propios «tribunales supremos». Concretamente, se trata de la Junta de Supervisión, que el propio Mark Zuckerberg alguna vez ha denominado «la Corte Suprema de Facebook», que puede «revertir las decisiones de Facebook sobre si permitir o eliminar ciertas publicaciones en la plataforma»[73]. Es importante destacar que la Junta ha sido diseñada para ser un organismo global, con el objetivo aspiracional de definir estándares globales para esta conocida red social. Está constituida como un órgano independiente de la empresa, integrado por 40 miembros de todo el mundo, y está encargada de verificar «si las decisiones se han tomado en cumplimiento de las normas y valores establecidos por Facebook». Esta Junta es una autoridad privada que decide las disputas particularmente relevantes únicamente con arreglo a los «principios y valores establecidos por Facebook»[74], con exclusión del orden constitucional del Estado en el cual se cometió la conducta en discusión.

A la postre, la combinación de poderes cuasinormativos, ejecutivos y judiciales otorga una posición particularmente autoritaria a las plataformas de metaversos, concentrando un poder notable en sus manos. Han creado un orden parajurídico autónomo que reconoce derechos, identifica prohibiciones y regula la responsabilidad de los usuarios. Ejercen el papel de legisladores, gobernantes y jueces del mundo digital, sin separación de poderes. Por ello, se ha dicho que las plataformas digitales «gobiernan arbitrariamente

71 Mostert, F. (2020). «Digital due process: a need for online justice», en *Journal of Intellectual Property Law & Practice*, vol. 15, núm. 5/2020, pág. 379.

72 Benyekhlef, K. y Duaso Calés, R. (2023). «Una nueva realidad: innovación tecnológica, derechos y ciberjusticia», en Barrio Andrés, M. (dir.). *Legal Tech. La transformación digital de la abogacía*. Madrid: La Ley, pág. 727 y ss.

73 Klonick, K. (2020). «The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression», en *Yale Law Journal*, vol. 129, núm. 8/2020, pág. 2478.

74 https://www.facebook.com/media/set/?set=a.1655178611435493.1073741828.1633466236940064

sus propios feudos privados»[75], que van más allá de las fronteras nacionales. Pero la Unión Europea, como ha quedado expuesto, ha recuperado la soberanía de la UE promulgando una nueva regulación en los Reglamentos DSA y DMA. Ambos reglamentos europeos, directamente aplicables sin necesidad de normas internas de transposición (art. 288 TFUE), crean garantías procesales universales, instituyen el principio de transparencia en la toma de decisiones privadas e institucionalizan la gestión constante del riesgo por parte de los grandes actores de Internet[76]. Del mismo modo, los residentes europeos adquieren derechos procesales y accionables ante las plataformas privadas que operan en el ecosistema digital. Las autoridades de la UE y de los Estados miembros obtienen herramientas para supervisar el cumplimiento de la nueva regulación. Y la sociedad civil conquista vías específicas para hacer valer mejor sus derechos a escala.

A mi juicio, esta reacción regulatoria de la UE demuestra que la erosión de la soberanía estatal o la dificultad de los poderes públicos para regular el Metaverso no significan en modo alguno su eclipse. Por el contrario, el Metaverso no es inmune a la regulación de origen estatal, como demuestra la aplicación del Derecho digital europeo al mismo. La narración sobre el fin de los Estados, cada cierto tiempo replanteada para diferentes campos de acción, parece un discurso que, aunque no desprovisto de elementos de verdad, falla en demostrar sus profecías[77].

VII. Conclusiones

Como ha quedado expuesto, el Metaverso pretende crear un entorno digital simulado que funcionará como un mundo virtual inmersivo para sus prosumidores. Los usuarios pueden interactuar con este ecosistema virtual a través de sus avatares digitales, vinculando las transacciones en los mundos físico y virtual. Y ciertamente traerá nuevos usos y bienestar a sus usuarios.

Facebook y Roblox prevén que el Metaverso será algo más que un ámbito social que ofrecerá a los usuarios entretenimiento y oportunidades de interacción social. Por otro lado, Microsoft se está enfocando principalmente en las oportunidades comerciales y de cooperación que el Metaverso brinda al mundo empresarial. Y empresas como Nvidia o Unity Software anticipan

75 Sander, B. (2020). «Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation», en *Fordham International Law Journal*, vol. 43, pág. 946 y ss.

76 Barrio Andrés, M. (2024). «Novedades y sentido del Reglamento Europeo de Servicios Digitales», en Cetina-Presuel, R. y Moretón Toquero, M. A. (coords.). *El nuevo Reglamento de servicios digitales de la Unión Europea: Nuevo enfoque regulatorio y garantías frente a los desórdenes informativos*. Madrid: Aranzadi.

77 Loughlin, M. (2016). «The Erosion of Sovereignty», en *Netherlands Journal of Legal Philosophy*, vol. 45, núm. 2/2016, pág. 57.

miles de millones de dólares en ganancias potenciales debido al desarrollo de la tecnología que eventualmente impulsará el Metaverso del futuro.

A pesar de que el Metaverso se está desarrollando y pretende ampliar el alcance de las capacidades de las redes sociales, su potencial para otros sectores industriales, comerciales, sociales, educativos, médicos, militares y gubernamentales es inmenso. No obstante, por el momento el Metaverso es sólo un alarde de *marketing*. Ahora bien, lo mismo sucedió cuando los ordenadores hicieron su entrada en la década de 1980. La humanidad ha progresado desde los escritos mecanografiados hasta la publicidad microsegmentada y la planificación empresarial basadas en la inteligencia artificial, entre otras aplicaciones que nos brindan las tecnologías actuales. Una vez que el Metaverso se establezca, lo más probable es que triunfe del mismo modo que lo ha hecho Internet.

A fecha de escribir estas líneas, sólo podemos atisbar alguna de sus aplicaciones. Pero, a medio plazo, el Metaverso podría mantenernos en contacto con aquellos seres querido que hemos perdido, porque la tecnología de *deepfakes* lo hará posible y nuestros rastros digitales podrían ser suficientes para recrear avatares de nosotros mismos, incluso después de muertos. El modo «vivir para siempre» en el Metaverso ya está previsto por ejemplo en la plataforma Somnium Space. Y, naturalmente, el Metaverso será una plataforma formidable para la creatividad y la innovación artística.

El Metaverso puede ser un término más en la larga lista de desarrollos tecnológicos que han surgido en los últimos años, pero, a diferencia de otros, ya ha echado raíces en la sociedad. Ya no es posible ver el Metaverso como un futuro fantástico y de ciencia ficción, ya que la pandemia ocasionada por el virus *SARS-CoV-2* ha acelerado la digitalización a gran escala y asimismo ha obligado a personas de todo el mundo a sumergirse a marchas forzadas en el mundo digital. Estimo que es una de las opciones tecnológicas más cercanas a convertirse en realidad en su totalidad que cualquier otra de las propuestas tecnológicas últimamente anunciadas.

Si el Metaverso triunfa, supondrá una estructura fundamental para todo el ecosistema digital, que a su vez se volverá pieza central, si no indisociablemente entrelazado con la estructura de las actividades fuera de línea. Mucho más que las actuales redes sociales. Y es que, cada vez más, las formas en que operan las plataformas digitales afectan la capacidad de las personas para desarrollar su opinión y personalidad, dar forma a sus vidas fuera de línea y participar en interacciones sociales, políticas, económicas y culturales. Dada la escala de su impacto, es fundamental comprender que su regulación puede tener una dimensión incluso constitucional, dada su evidente trascendencia iusfundamental. Considero que uno de los retos que debe resolver la investigación futura es, por tanto, verdaderamente constitucional: ¿cuál es la estrategia más adecuada para someter la soberanía privada y el poder estructural de los metaversos a los derechos fundamentales, el control democrático, la protección de los trabajadores y la regulación de la competencia?

De nuevo asistimos a un corolario del conocido metarrelato de la idea de «constitución espontánea»[78], que reaparece una vez más en la construcción del Metaverso y pretende regular la realidad social ocultando su voluntad de poder tras la pantalla de la aparente neutralidad técnica. Estimo que los juristas tenemos que reivindicar el carácter cardinal de los valores de las constituciones democráticas de los que los Estados son sus portadores. Contentarse con confiar en los magníficos destinos y progresismos que el Metaverso y sus señores serían capaces de garantizar es subestimar la garantía de los intereses de los ciudadanos que los ordenamientos jurídicos estatales, aunque erosionados, son capaces de proporcionar a través de sus instituciones con origen y fundamento democráticos. No olvidemos, como señalaba IRTI[79], que «los derechos humanos, desprovistos de fundamento y de defensas coercitivas (que siempre tienen en cuenta a sujetos individuales, que ostentan intereses específicos y determinados), se dispersan en la generalidad enfática de cartas y declaraciones y memorandos de entendimiento». Por eso es urgente que, junto al desarrollo tecnológico del Metaverso, se contemple a la par el respeto de los derechos fundamentales y el propio Estado de derecho, encastrando en el mismo diseño de la tecnología los valores protegidos por el Derecho constitucional.

Bibliografía

Abadi, J. y **Brunnermeier, M. K.**, «Blockchain Economics», *CEPR Discussion Paper*, DP13420, 2019.

Allen, Jason y otros, «Cryptoassets in Private Law: Emerging Trends and Open Questions from the First 10 Years», *SMU Centre for AI & Data Governance Research Paper*, núm. 06/2022.

Álvarez López, C. y **Carrasco Perera, A**., «Operadores y responsabilidad civil en el metaverso», *GAP Análisis*, marzo 2022.

Barrio Andrés, M., «Novedades y sentido del Reglamento Europeo de Servicios Digitales», en Cetina-Presuel, R. y Moretón Toquero, M. A. (coords.), *El nuevo Reglamento de servicios digitales de la Unión Europea: Nuevo enfoque regulatorio y garantías frente a los desórdenes informativos.* Madrid: Aranzadi, 2024.

Barrio Andrés, M., *Ciberderecho. Bases estructurales, modelos de regulación e instituciones de gobernanza de Internet.* Valencia: Tirant lo Blanch, 2018.

78 Ciarlo, P. (2002). «Contro l'idea di costituzione spontanea», en *Quaderni costituzionali*, núm. 1/2002, pág. 101; Teubner, G. (2012). *Constitutional fragments. Social constitutionalism and globalization*, *op. cit*., cap. 2.

79 Irti, N. (2006). *Norma e luoghi. Problemi di geo-diritto*. Bari: Laterza, pág. 127.

Barrio Andrés, M., *Internet de las Cosas.* Madrid: Reus, 3.ª edición, 2022.

Barrio Andrés, M., *Manual de Derecho digital.* Valencia: Tirant lo Blanch, 4.ª edición, 2025.

Barrio Andrés, M., «Cuestiones actuales del Derecho de los Robots», *Teoría y Derecho. Revista de pensamiento jurídico*, 37, 2024.

Barrio Andrés, M., «La ciberseguridad en el Derecho digital europeo: novedades de la Directiva NIS2», *InDret*, 1, 2024.

Barrio Andrés, M., «Las nuevas coordenadas del Derecho digital europeo», *Diálogos jurídicos. Anuario de la Facultad de Derecho de la Universidad de Oviedo*, vol. 9, 2024.

Barrio Andrés, M., «Los principios estructurales del Reglamento General de Protección de Datos», *Los principios estructurales del Reglamento General de Protección de Datos*, 20, 2024.

Barrio Andrés, M., «Metaverso: origen, concepto y aplicaciones», *Revista LA LEY Derecho Digital e Innovación*, 12, 2022.

Barrio Andrés, M., *Reglamento UE de inteligencia artificial (Incluye los actos de desarrollo y ejecución de la AI Act)*. Madrid: Francis Lefebvre, 2025.

Benyekhlef, K. y **Duaso Calés, R**., «Una nueva realidad: innovación tecnológica, derechos y ciberjusticia», en **Barrio Andrés, M**. (dir*.), Legal Tech. La transformación digital de la abogacía*. Madrid: La Ley, 2023.

Cauffmann, C. y **Goanta, C**., «A New Order: The Digital Services Act and Consumer Protection», *European Journal of Risk Regulation*, 1, 2021.

Cheong, B. C., «Avatars in the Metaverse: Potential Legal Issues and Remedies», *International Cybersecurity Law Review*, 1, 2022.

Cline, E., *Ready Player One*. Nueva York: Crown Publishers, 2011.

Cotino Hueso, L., «Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales», en **Casas Bahamonde, M. E**., *Derecho y Tecnologías*, Editorial Fundación Ramón Areces, Madrid, 2025.

De la Torre, C., «El futuro del trabajo y el metaverso laboral en el periodo pos-Covid: Quo vadis?», *Consell obert: Recull informatiu del Consell General de Col·legis de Graduats Socials de Catalunya*, 366, 2022.

Ducci, F., *Natural Monopolies in Digital Platform Markets*. Cambridge: Cambridge University Press, 2022.

Ellis, G. F. R. y otros, «Multiverses and physical cosmology», *Monthly Notices of the Royal Astronomical Society*, 347, 2004.

Falco, G. y **Rosenbach, E**., *Confronting Cyber Risk: An Embedded Endurance Strategy for Cybersecurity*. Oxford: Oxford University Press, 2022.

Feenberg, A., «Postdigital or predigital?», *Postdigital Science and Education*, Vol. 1, 1, 2019.

Fortnow, M. y **Terry, W**., *The NFT Handbook: how to create, sell, and buy non-fungible tokens*. Hoboken: Wiley, 2021.

Fries, M., «Private Law Compliance through Smart Contracts?», *Compliance Elliance Journal*, Vol. IV, 1, 2018.

Gallego Fernández, L. A., «Cadenas de bloques y Registros de derechos», *Revista Crítica de Derecho Inmobiliario*, Vol. 94, 765, 2018.

Gibson, W., *Neuromancer*. Nueva York: Ace Books, 1984.

Higueras Redecillas, J. C., «El metaverso: La hoja de ruta hacia el big bang empresarial», *Harvard Deusto business review*, 322, 2022.

Ibáñez Jiménez, J. (coord.), *Transformación digital y nueva regulación de personas, instituciones y mercados*. Madrid: Reus, 2022.

Ibáñez Jiménez, J. (dir.), *Tratado de Derecho de la Sociedad digital*. Madrid: Editorial Reus, 2025.

Jiménez Rubio, R., «El Metaverso y el Derecho Registral», *Revista LA LEY Derecho digital e innovación*, 12, 2022.

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C. y **Hui, P**., «All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda», Journal of latex class files, 14, 2021.

López Cumbre, L., «Relaciones laborales "por", "para" y "en" el metaverso», *Diario La Ley*, 10033, 2022.

Mansilla Pizá, A., «Metaverse. Self-Regulation, Intellectual Property, E-Commerce, and Dispute Resolution», *Revista LA LEY Derecho digital e innovación*, 10, 2022.

Manzano Peláez, E., «Servicios personalizados, proactivos y la experiencia ciudadana en un mundo en permanente cambio», *Consultor de los ayuntamientos y de los juzgados: Revista técnica especializada en administración local y justicia municipal*, 1, 2022.

Martínez de Salinas Murillo, J., «El metaverso laboral», *Observatorio de recursos humanos y relaciones laborales*, 175, 2022.

Martini, M., «Regulating Algorithms. How to Demystify the Alchemy of Code», en **Ebers, M**. y **Navas, S**. (eds.), *Algorithms and Law*. Cambridge: Cambridge University Press, 2020.

Meyzan Briceño, M. A., «Desafíos de la educación y el aprendizaje en el Metaverso», Desafíos: Revista Científica de Ciencias Sociales y Humanidades, 13, 2022.

Montero Pascual, J. J. (dir.), *La regulación de la economía colaborativa: Airbnb, BlaBlaCar, Uber y otras plataformas*. Valencia: Tirant lo Blanch, 2017.

Murray, A. y otros, «The Promise of a Better Internet: What Is Web 3.0 and What Are We Building?», *Oregon Law Review*, 1, 2022.

Pacheco Jiménez, M. N., «De la digitalización de los pagos a los tokens del metaverso», *Revista LA LEY mercantil*, 91, 2022.

Ramos, D., «Realidad, Metaversos y Protección de Datos», *LA LEY privacidad*, 11, 2022.

Recuerda Girela, M. A. (dir.), *Tecnologías disruptivas*. Pamplona: Aranzadi, 2019.

Rodríguez Ayuso, J. F. (coord.), *Nuevos retos en materia de derechos digitales en un contexto de pandemia: perspectiva multidisciplinar*. Pamplona: Aranzadi, 2022.

Scolari, C., *La guerra de las plataformas. Del papiro al metaverso*. Barcelona: Anagrama, 2022.

Sempere, F. J., «Smart-Glasses y el metaverso: un viaje por las tecnologías que usan cámaras», *Revista LA LEY Derecho digital e innovación*, 12, 2022.

Stevenson, N., *Snow Crash*. Londres: Penguin, 1992.

Vasco Gómez, A., *Arbitraje y resolución de conflictos en las competiciones de eSports*. Pamplona: Aranzadi, 2022.

Vidal-Tomás, D., «The new crypto niche: NFTS, play-to-earn, and metaverse tokens», *Finance Research Letters*, 47, 2022.

Villasmil Espinoza, J. J., «Multiversos y metaversos en la presente realidad global», *Multiverso journal*, 2, 2022.

Wang, Z. y otros, «A Survey on Metaverse: Fundamentals, Security, and Privacy», *arXiv*, 2203.02662, 2022.

3.

EL REGLAMENTO (UE) DE SERVICIOS DIGITALES COMO INSTRUMENTO ESENCIAL EN EL NEW DEAL TECNOLÓGICO DE LA UNIÓN EUROPEA[1]

José Juan Castelló Pastor

Profesor Titular de Derecho internacional privado
Universitat de València

SUMARIO: I. Introducción: un nuevo marco de gobernanza digital en la Unión Europea. II. Aspectos generales del Reglamento (UE) de servicios digitales. 2.1. Ámbito de aplicación. 2.2. Consolidación del régimen de exención de responsabilidad e importantes novedades de la jurisprudencia del TJUE. III. Un ecosistema digital más transparente y seguro mediante obligaciones de diligencia debida asimétricas. 3.1. Obligaciones básicas a todos los prestadores de servicios de intermediación. 3.2. Obligaciones adicionales a los prestadores de servicios de alojamiento de datos, incluidas las plataformas en línea. 3.3. Más disposiciones de obligada observancia a las plataformas en línea. 3.4. Obligaciones adicionales de gestión de riesgos sistémicos para las plataformas en línea y buscadores de muy gran tamaño. IV. Mecanismos de tutela jurídico-pública y privada en el Reglamento (UE) de servicios digitales. V. Consideración final.

I. Introducción: un nuevo marco de gobernanza digital en la unión europea

Han transcurrido dos décadas desde la aprobación de la Directiva sobre el comercio electrónico[2], una norma jurídica fundamental para el funciona-

1 Esta publicación es parte del Proyecto TED2021-129307A-I00, financiado por MICIU/ AEI/10.13039/501100011033 y por la Unión Europea NextGenerationEU/ PRTR, denominado «Hacia una transición digital centrada en la persona en la Unión Europea» IP. José Juan Castelló Pastor. Este trabajo se ha realizado durante la estancia de investigación en el Centro de investigación interdisciplinar de Justicia de la Facultad de Derecho de la Universidad de Oporto (Portugal).

2 Directiva 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000, relativa a determinados aspectos jurídicos de los servicios de la sociedad de la información, en

miento de los prestadores de servicios de internet, en un contexto en el que el acceso a la red era sumamente limitado, las páginas eran predominantemente estáticas (web 1.0), el comercio electrónico aún no estaba generalizado, las redes sociales no existían y el entorno analógico prevalecía. Este escenario ha experimentado una transformación notable a lo largo de estos años, especialmente en esta última década, dando lugar a una sociedad hiper conectada, con un usuario excesivamente dependiente de la tecnología, nuevos modelos de negocios basados principalmente en la economía de las plataformas, y avances significativos en áreas como la inteligencia artificial, entre otros, lo que ha generado la necesidad de revisar el marco jurídico regulador de internet.

Sin duda alguna, desde la llegada de internet en los años 90, hemos presenciado un cambio social y económico sin precedentes[3]. Con una generación de nativos digitales, la tecnología ha transformado nuestra manera de comunicarnos, trabajar y relacionarnos. Además, ha impulsado una economía global interconectada que opera a una velocidad y escala inimaginables hace apenas unas décadas, lo que hacía imprescindible la actualización de su marco legal.

Con una intensa y prolífica actividad legislativa indudablemente relevante por parte de las instituciones de la Unión en estos últimos años, que se ha materializado, entre otras, en el Reglamento (UE) 2022/2065 de servicios digitales, el Reglamento (UE) 2022/1925 de mercados digitales, el Reglamento (UE) 2024/1689 de Inteligencia Artificial, el Reglamento (UE) 2021/784 sobre la lucha contra la difusión de contenidos terroristas en línea, el Reglamento (UE) 2022/868 de gobernanza de datos y el Reglamento (UE) 2023/2854 de datos. Estos actos jurídicos, junto con otros más, demuestran la capacidad legislativa de las instituciones de la Unión Europea para establecer su propio marco normativo en el contexto tecnológico. Este proceder autónomo en el mundo digital está contribuyendo a fortalecer nuestra soberanía[4]. De esta manera, la Unión Europea no solo logra regular el importante sector tecno-

particular el comercio electrónico en el mercado interior (Directiva sobre el comercio electrónico). Diario Oficial n.º L 178 de 17/07/2000 p. 0001 – 0016. La DCE fue incorporada en nuestro ordenamiento jurídico en la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico, publicada en el BOE número 166, de 12/07/2002.

3 Es opinión generalizada que la revolución tecnológica de estas últimas décadas ha generado un impacto social y económico sin precedentes en la historia de nuestra civilización desde la expansión del uso de internet a principios de la década de 1990. Este fenómeno representa un auténtico cambio de era que supera en efecto y significado a la primera revolución industrial del siglo XVIII e incluso a la Revolución Francesa de 1789. *Vid*. Barrio Andrés, M.: «Inteligencia artificial, Internet de las cosas y blockchain», en Montero Pascual, J.J., *Digitalización y Derecho*, Tirant Lo Blanch, Valencia, 2024, p. 253.

4 Madiega T., «Digital sovereignty for Europe», *EPRS Ideas Paper Towards a more resilient EU*, 2020, p. 1.

lógico, predominantemente dominado por empresas estadounidenses, sino también, y sea quizás lo más importante, promover los valores europeos a través del obligado cumplimiento de sus normativas. En efecto, con su aprobación, se configura un entorno empresarial tecnológico y se imponen prácticas de cumplimiento normativo para aquellos que deseen operar dentro del mercado interior[5], logrando así el conocido «efecto Bruselas»[6] y convirtiéndose la Unión Europea en un potente regulador tecnológico global.

Este trabajo se centra en el análisis del Reglamento (UE) de servicios digitales[7], abordando en primer lugar los aspectos generales relativos a su ámbito de aplicación; en segundo lugar, las obligaciones de diligencia debida asimétricas impuestas a los prestadores de servicios de intermediación; y, finalmente, el marco relativo vinculado a la supervisión y aplicación. Cabe avanzar que nos encontramos ante una de las normas horizontales más innovadoras de los últimos tiempos, al constituir un punto de inflexión en el marco de la *ley de la selva digital.* Este texto normativo introduce un mayor control democrático sobre las actividades de las multinacionales tecnológicas, con el propósito de contrarrestar el considerable poder económico y social acumulado en estas dos últimas décadas, así como de abordar la creciente privatización y automatización del control del discurso en línea y situar al usuario en el centro del proceso de la transición digital.

II. Aspectos generales del Reglamento (UE) de servicios digitales

Aplicable en todos los Estados miembros desde el 17 de febrero de 2024[8] aunque algunas de sus disposiciones tuvieron una fecha de aplicación anterior[9], el Reglamento (UE) de servicios digitales crea un conjunto uniforme

5 Christakis, T., *«European Digital Sovereignty»: Successfully Navigating Between the 'Brussels Effect' and Europe's Quest for Strategic Autonomy*, 2020, p. 2.

6 *Vid*. Bradford, A., *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020. Aunque en el año 2012 ya acuñó este término en «The Brussels Effect», *Northwestern University Law Review*, vol. 107, num. 1, 2012, p. 3.

7 Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE, en adelante: Reglamento (UE) de servicios digitales (o RSD). OJ L 277, 27.10.2022, p. 1–102.

8 Véase el art. 93 RSD sobre la entrada en vigor y aplicación.

9 De acuerdo con el art. 33.4 RSD, se estableció la obligación de informar sobre el promedio mensual de destinatarios del servicio activos en la Unión antes del 17 de febrero de 2023. Posteriormente, el 25 de abril de 2023, la Comisión Europea designó como plataformas en línea de muy gran tamaño a Alibaba Aliexpress, Amazon Store, Apple, AppStore, Booking.com, Facebook, Google Maps, Google Play, Google Shopping, Instagram, LinkedIn, Pinterest, Snapchat, TikTok, Twitter, Wikipedia, Youtube y Zalando; y motores de búsqueda

de reglas aplicables exclusivamente a los intermediarios[10] —y en su mayoría especialmente dirigidas a las plataformas en línea[11]— que tiene por meritorio objetivo contribuir al correcto funcionamiento del mercado interior de la Unión Europea, en consonancia con su predecesora la Directiva de comercio electrónico[12]. Efectivamente, este acto legislativo de la Unión reafirma el marco horizontal de exenciones condicional de responsabilidad de estos actores digitales[13] y crea un robusto marco armonizado de obligaciones (de diligencia debida) asimétricas, siendo de singular relevancia las disposiciones dirigidas específicamente a las plataformas en línea muy grandes y a los motores de búsqueda muy grandes para evaluar y gestionar los riesgos sistémicos: esto es, prevenir las actividades ilícitas y nocivas en línea y la difusión de desinformación, garantizar la seguridad de los usuarios, proteger los dere-

en línea de muy gran tamaño a Bing y Google Search, según lo establecido en el artículo 33.6 del RSD y publicado en el Diario Oficial de la Unión Europea DO C 249/2 del 14 de julio de 2023. Este listado puede estar sujeto a cambios, tal y como ha sucedido: a fecha 18 de octubre de 2024, la Comisión Europea ha reconocido un total de 23 servicios de gran tamaño: 21 como VLOPs y 2 como VLOSEs. Además de los indicados, se han incorporado las plataformas: Pornhub, Booking.com, Shein, Stripchat, Xvideos. Para más información sobre la «supervisión de las plataformas en línea y los motores de búsqueda de muy gran tamaño designados en virtud de la Ley de Servicios Digitales» visítese la página de la Comisión: digital-strategy.ec.europa.eu (última consulta 23.1.2025).

10 Si bien se hace referencia en este trabajo indistintamente al término «servicios digitales», es cierto que el RSD no lo utiliza, sino que se refiere al «servicio de la sociedad de la información» definido en el artículo 1.1 b) de la Directiva (UE) 2015/1535, pero limita su ámbito al «servicio intermediario» y que se refiere únicamente a tres categorías de servicios de la sociedad de la información: el servicio de «mera transmisión», el de «memoria caché» y el de «alojamiento de datos» (léase respectivamente el art. 3 letra a) y g) RSD).

11 Incluye a las plataformas en línea en un subgrupo del «alojamiento de datos» a efectos fundamentalmente del cumplimiento de las obligaciones de diligencia, conforme lo matiza el cdo n.º 13 RSD.

12 De acuerdo con el art. 89 RSD, la Directiva 2000/31/CE sobre el comercio electrónico continúa en vigor, sólo se suprimen los artículos 12 a 15 y estas referencias se entienden hechas a los arts. 4, 5, 6 y 8 RSD respectivamente.

13 Al tratarse de los mismos «puertos seguros» de la Directiva de comercio electrónico, por cuestión de espacio, no se abordan en este estudio, si bien véanse: Castelló Pastor, J.J., «Nuevo régimen de responsabilidad de los servicios digitales que actúan como intermediarios a la luz de la propuesta de Reglamento relativo a un mercado único de servicios digitales», en Castelló Pastor, J.J. (dir.), *Desafíos jurídicos ante la integración digital: aspectos europeos e internacionales*, Aranzadi Thomson-Reuters, 2021, pp. 37-76; Cordero Álvarez, C. I., «El régimen de responsabilidad de las plataformas online en la DSA ante contenidos ilícitos lesivos de derechos fundamentales: evolución de la figura de puerto seguro», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea,* Aranzadi, 2024, pp. 69-105; Gascón Marcén, A., «La responsabilidad de los intermediarios de Internet en la Unión Europea: iniciativas recientes y perspectivas de futuro», en Castelló Pastor, J.J. (dir.), *Desafíos jurídicos ante la integración digital: aspectos europeos e internacionales*, Aranzadi, 2021, pp. 133-160.

chos fundamentales y consolidar, en definitiva, un entorno digital más justo y abierto[14]. También establece un novedoso régimen de tutela jurídico-pública para supervisar la aplicación, cumplimiento y ejecución de las disposiciones de obligada observancia configuradas en el propio texto normativo.

2.1. Ámbito de aplicación

Como se ha referido, sus disposiciones se dirigen exclusivamente a los intermediarios, a saber: el prestador de servicio de mera transmisión, de memoria caché y de alojamiento de datos[15]. Se deja claro, a este respecto, que no se aplicará a ningún otro servicio que no sea un servicio intermediario ni tampoco a ningún requisito que se imponga respecto de un servicio de esa índole, con independencia de que el servicio se preste mediante el uso de un servicio intermediario. Y se aplica a todo intermediario que ofrezca sus servicios en la Unión Europea[16].

Cabe señalar que uno de los aspectos de mayor relevancia de este nuevo marco jurídico digital posiblemente sea el desafío que supone el cumplimiento de sus reglas uniformes para las grandes empresas tecnológicas que dominan el mercado de los servicios digitales, mayoritariamente establecidas en Estados Unidos y en China con millones de sus usuarios en la Unión Europea. En este punto, la Comisión Europea ha sido explícita al acentuar que cualquier operador tecnológico situado fuera de la Unión que desee realizar actividades económicas en nuestro territorio debe cumplir con las normativas aquí establecidas. Y ha subrayado que se emplearán todos los recursos disponibles para garantizar el cumplimiento estricto de la legislación de la Unión Europea y de las normas internacionales, con el objetivo de asegurar condiciones equitativas en el entorno digital[17].

El legislador de la Unión reconoce efectivamente el reto que implica configurar un entorno digital más seguro, transparente y responsable con medidas de obligada observancia a todos los servicios digitales, incluso para aquellos que operan desde fuera de nuestro territorio, por el carácter intrínsecamente transfronterizo de internet, y lo alcanza, no sin cierta dificultad[18] al someter

14 De acuerdo con el art. 1 RSD.

15 De acuerdo con la definición incorporada en el artículo 3 letra g RSD.

16 Señala el art. 2.1 RSD que «se aplicará a los servicios intermediarios ofrecidos a destinatarios del servicio que tengan su lugar de establecimiento o estén situados en la Unión, con independencia de donde los prestadores de dichos servicios intermediarios tengan su lugar de establecimiento».

17 Bruselas, 19.2.2020 COM (2020) 67 final.

18 A modo de ejemplo, el listado designando a las plataformas en línea de muy gran tamaño publicado en el Diario Oficial de la Unión Europea DO C 249/2 de 14.7.2023, conforme a lo dispuesto en el artículo 33.6 del Reglamento (UE) de servicios digitales, ha sido

el complejo cumplimiento de sus disposiciones con base en un criterio de conexión fáctico: se extiende su ámbito de aplicación espacial a los prestadores de servicios intermediarios con independencia de su lugar del establecimiento (domicilio o ubicación) en la medida en que ofrezcan sus servicios en el territorio de la Unión[19].

Este criterio de conexión tiene como objetivo garantizar la eficacia de las normas de la Unión como la equidad de condiciones a todos los operadores económicos en el mercado interior[20], esencial en este contexto digital[21]. Aunque un problema adicional pudiera venir en la determinación del referido criterio fáctico para que sean aplicables las disposiciones de la normativa de la Unión Europea a estas empresas tecnológicas situadas en terceros Estados, se solventa con cierto éxito al referirse por «ofrecer un servicio en la Unión» como el hecho de «hacer posible que las personas físicas o jurídicas de uno o varios Estados miembros utilicen los servicios de un prestador de servicios intermediarios que tenga una conexión sustancial con la Unión»[22]. Por su parte, especifica el propio instrumento que la conexión sustancial se da si el prestador tiene su establecimiento en la Unión, o bien —y he aquí el *quid* de la cuestión en la determinación de lo que constituye esta conexión sustancial— porque tiene un número significativo de destinatarios del servicio en uno o varios Estados miembros en relación con su población, o dirige sus actividades hacia uno o varios Estados miembros[23].

En otras palabras, cuando el establecimiento del servicio digital esté en un tercer Estado serán de aplicación sus disposiciones si tiene un promedio

impugnado por varias tecnológicas como Zalando y Amazon ante el Tribunal General, asuntos T-348/23 y T-367/23 respectivamente, para no tener que cumplir así con todas las adicionales obligaciones impuestas a las muy grandes plataformas digitales. Igualmente pudo leerse en los medios de comunicación las reticencias de algunas tecnológicas como X (antes, Twitter) al cumplimento de las medidas impuestas en el RSD, planteándose el bloqueo del servicio en toda la Unión (elconfidencial.com de 19.10.2023).

19 De acuerdo con el artículo 2.1 del Reglamento (UE) de servicios digitales. Con mayor profundidad, Feliu Álvarez Sotomayor, S., «Ámbito territorial y aplicación del principio del control en origen en el Reglamento UE de servicios digitales», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp. 41-64.

20 Cdo n.º 7 RSD.

21 Desde mi punto de vista es apropiado porque las grandes tecnológicas (de origen y residencia) estadounidenses o asiáticas, cuyos servicios son utilizados por millones de usuarios en la Unión, quedan sujetas al cumplimiento de las mismas medidas y estándares que los operadores económicos europeos. Aunque este enfoque pudiera desincentivar la prestación de servicios digitales en el territorio de la Unión, según Schmon C. y Gullo, K., «European Commission's Proposed Digital Services Act got several things right, but improvements are necessary to put users in control» (disponible en www.eff.org).

22 *Vid*. art. 3 letra d) RSD.

23 De acuerdo con la definición incorporada en el art. 3 letra e) RSD.

mensual de destinatarios del servicio activos en la Unión Europea de más de cuarenta y cinco millones[24]. Se trata de un criterio cuantitativo razonable (*i.e.* una cifra equivalente al diez por ciento de la población de la Unión), ya que estas plataformas en línea de muy gran tamaño y motores de búsqueda en línea de muy gran tamaño pueden entrañar riesgos sistémicos para la sociedad. Por lado, también estarán sujetos a la normativa de la Unión aquellos servicios digitales establecidos en un tercer Estado que dirijan sus actividades hacia uno o más Estados miembros[25].

Desde mi punto de vista, esta disposición de aplicación también a situaciones privadas de tráfico jurídico externo —mediante una norma de alcance extraterritorial— constituye una eficaz estrategia geopolítica de la Unión para salvaguardar de manera adecuada los valores fundamentales de la Unión Europea. Una estrategia que cobra singular relevancia en un contexto donde las relaciones privadas digitales son inherentemente transfronterizas, el principio de territorialidad se vuelve difuso debido a la naturaleza ubicua de internet, y tanto los usuarios como la sociedad en general se enfrentan a serias vulnerabilidades frente a empresas tecnológicas, muchas de las cuales están establecidas en terceros Estados con intereses económicos significativos. Por esta razón, reforzar la protección y cumplimiento de este marco jurídico digital, no solo en las relaciones *ad intra*, sino también en las relaciones *ad extra* —por su vínculo estrecho con la Unión— resulta clave para alcanzar los objetivos establecidos por nuestro legislador.

2.2. Consolidación del régimen de exención de responsabilidad e importantes novedades de la jurisprudencia del TJUE

Volviendo al marco de la exención de responsabilidad de los prestadores de servicios intermediarios queda delimitado a los servicios de mera transmisión, memoria caché y alojamiento de datos[26]. La naturaleza técnica, auto-

24 Véase, al respecto, el cdo. n.º 76 RSD.

25 Recogiendo el cdo n.º 8 RSD la consolidada jurisprudencia del Tribunal de Justicia sobre el «criterio de las actividades dirigidas».

26 Con la finalidad de ilustrar al lector, modo de ejemplo y de acuerdo con el considerando núm. 20 RSD, «los servicios intermediarios de "mera transmisión" incluyen categorías genéricas de servicios como los puntos de intercambio de internet, los puntos de acceso inalámbrico, las redes privadas virtuales, los servicios de DNS y traductores DNS, los registros de nombres de dominio de primer nivel, los registradores, las autoridades de certificación que expiden certificados digitales, la transmisión de voz por internet y otros servicios de comunicación interpersonal, mientras que entre los ejemplos genéricos de servicios intermediarios de "memoria caché" se incluyen el suministro únicamente de redes de distribución de contenidos, los proxies inversos o los proxies de adaptación de contenidos. Estos servicios son cruciales para garantizar una transmisión fluida y

mática y pasiva inherente a la actividad del prestador de servicio de intermediación, junto con su posición neutral, implica que, en principio, este no tenga conocimiento ni control de la información que almacena o transmite.

En este sentido, y conforme al artículo 8 del Reglamento (UE) de servicios digitales, uno de los pilares fundamentales del sistema reflejado anteriormente en el artículo 15 de la Directiva sobre el comercio electrónico, no se le impone al prestador una obligación general de supervisar o monitorizar la información transmitida o almacenada, ni de buscar activamente hechos o circunstancias que indiquen la existencia de actividades ilícitas. Por tanto, el operador económico desconoce el carácter ilícito de los contenidos almacenados o de las actividades que pudieran realizarse a través de sus servicios. Sin embargo, una vez que adquiere conocimiento efectivo o la consciencia de la existencia de dichos contenidos o actividades, está obligado a actuar de manera diligente para eliminar o bloquear su acceso. El incumplimiento de esta obligación podría conllevar la pérdida de inmunidad que le otorgan las referidas disposiciones.

Se trata de una continuación (o consolidación) del régimen de exención de responsabilidad de los intermediarios en la Unión Europea, con una acertada incorporación: la cláusula de buen samaritano. Así y con el objetivo de reducir la presencia de ilícitos en la red, se fomenta la seguridad jurídica de los intermediarios que implementen voluntariamente mecanismos para detectar, identificar y/o retirar contenidos ilícitos (o inhabilitar el acceso a sus servicios) al no perder su inmunidad por la supervisión voluntaria de las actividades en sus plataformas que efectúen de buena fe y de modo diligente[27].

Igualmente, otra novedad es, de un lado, la incorporación relativa a las órdenes de actuación contra contenidos ilícitos concretos dictada por las autoridades judiciales o administrativas nacionales adopten frente a intermediarios sobre la base del Derecho aplicable y que, en síntesis, armoniza las condiciones mínimas de estas órdenes al hilo de la jurisprudencia del TJUE[28].

---

eficiente de la información suministrada en internet. Entre los ejemplos de "servicios de alojamiento de datos" se incluyen categorías de servicios como la computación en nube, el alojamiento web, los servicios remunerados de referenciación o los servicios que permiten compartir información y contenidos en línea, incluido el almacenamiento y el intercambio de archivos [...]».

27 De acuerdo con el artículo 7 RSD en relación con el considerando núm. 26. En términos similares a la figura del buen samaritano de la sección 230 de la Communications Decency Act estadounidense. Sobre este punto, véase con mayor profundidad: Castelló Pastor, J.J., «Exoneración de responsabilidad de los prestadores de servicios de la información en la sección 230 de la Communications Decency Act estadounidense», *Revista Aranzadi de derecho y nuevas tecnologías*, n.º 39, 2015.

28 Véase el art. 9 en relación con el considerando núm. 31 y 32 RSD. Con mayor profundidad, Palao Moreno, G., «Las órdenes de actuación contra contenidos ilícitos en el nuevo reglamento europeo de servicios digitales», en J.J. Castelló Pastor (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp. 109-134.

Y de otro, en términos similares, pero en relación con la orden de proporcionar información específica sobre uno o varios destinatarios individuales del servicio dictada por las autoridades judiciales o administrativas nacionales pertinentes en virtud del derecho aplicable[29].

III. Un ecosistema digital más transparente y seguro mediante obligaciones de diligencia debida asimétricas

Con el fin de cumplir los objetivos establecidos en el Reglamento, y en particular para mejorar el funcionamiento del mercado interior y garantizar un entorno en línea seguro y transparente, se establece un novedoso conjunto claro, eficaz, predecible y equilibrado de obligaciones de diligencia debida a los intermediarios. Ciertamente son disposiciones necesarias para «afrontar los objetivos de interés público» y que aspiran a garantizar la seguridad y confianza de los destinatarios del servicio —incluidos: los consumidores, los menores y los usuarios especialmente vulnerables al riesgo de ser víctimas de discursos de odio, acoso sexual u otras conductas discriminatorias—, la lucha contra las prácticas ilícitas y la protección de los derechos fundamentales, la rendición de cuentas, entre otras[30].

Para ello configura un conjunto de obligaciones de diligencia debida básicas a todos los intermediarios y —adaptado al tipo, tamaño y naturaleza del servicio digital— unas obligaciones adicionales para los prestadores de servicios de alojamiento de datos. En concreto, unas obligaciones dirigidas a las plataformas en línea y otras más a las plataformas en línea de muy gran tamaño y de motores de búsqueda en línea de muy gran tamaño para mitigar los riesgos sistémicos.

Estas disposiciones asimétricas de obligada observancia no constituyen condiciones previas para la exención de responsabilidad de los prestadores de servicios de intermediación, sino que son obligaciones independientes. En consecuencia, en caso de que el servicio digital incurra en el incumpliendo de alguna de estas obligaciones de diligencia debida, se configuran mecanismos de tutela jurídico-pública de supervisión destinados a garantizar su cumplimiento; así como de la tutela jurídico-privada. Adicionalmente, los destinatarios del servicio tendrán derecho a solicitar una indemnización por los daños o perjuicios sufridos como consecuencia de dicho incumplimiento, con independencia de la mencionada tutela jurídico-pública.

29 *Vid*. art. 10. RSD.

30 Como así lo establece el considerando 40 RSD.

3.1. Obligaciones básicas a todos los prestadores de servicios de intermediación

En lo que respecta a las obligaciones básicas aplicables a todos los intermediarios, estas consisten esencialmente en establecer canales de contacto que permitan a las autoridades de los Estados miembros, la Comisión y la Junta comunicarse de manera eficiente con el proveedor del servicio, así como facilitar que los usuarios también puedan hacerlo[31]. Además, los prestadores que no estén establecidos en la Unión Europea, pero ofrezcan sus servicios aquí deberán designar un representante legal[32]. Aunque se trata de disposiciones básicas de obligado cumplimiento, son sumamente útiles para garantizar una comunicación efectiva con el intermediario. Esto es especialmente relevante, ya que la experiencia acumulada en el ámbito digital en todos estos años ha demostrado que, a pesar de lo establecido en el artículo 5 de la Directiva sobre el comercio electrónico (transpuesto en el artículo 10 LSSICE), que exigía proporcionar cierta información, en muchas ocasiones esta obligación no se cumplía (o de forma parcial), dificultando u obstaculizando así la comunicación con los servicios digitales y mermándose, en definitiva, los derechos de los usuarios digitales.

Por otro lado, en el deber de transparencia informativa materializado a través de las condiciones generales del servicio digital y en la publicación de informes sobre las actividades realizadas, conforme a lo dispuesto en el artículo 14 y 15 del Reglamento de Servicios Digitales. En este contexto, el legislador de la Unión manifestó preocupación por la falta de información y transparencia de las plataformas digitales, que regulan el acceso y la prestación de servicios de millones de usuarios de la red mediante condiciones generales del servicio (y/o las directrices de la plataforma), generalmente aceptadas sin posibilidad alguna de negociación y revestidas de cierta opacidad. Por ello, se ha puesto especial énfasis en entender y supervisar las acciones de estas plataformas digitales, fundamentales en la actividad económica, social o cultural de la Unión, con el objetivo de prevenir cualquier resultado injusto o arbitrario derivado de su omnímodo poder. En este sentido, un número significativo de estas disposiciones obligatorias se centra en el cumplimiento del deber de información y de transparencia, particularmente en lo que respecta a la adecuada publicidad en cuestiones relacionadas con la moderación de contenidos, el diseño de los servicios o la rendición de cuentas, entre otras. Esto resulta especialmente relevante considerando que, debido a su infraestructura y alcance, las plataformas digitales poseen una posición predominante que les permite determinar, en la mayoría de los casos, la visibilidad (entendida como notoriedad o presencia) de los contenidos y servicios ofrecidos en la red.

31 *Vid*. arts. 11 y 12 RSD respectivamente.

32 Art. 13 RSD.

El razonamiento subyacente es claro: una mayor transparencia en la actuación de estos actores digitales incrementa la confianza que los usuarios depositan en la red. Con este objetivo, se busca fortalecer tanto la transparencia como la seguridad jurídica en el ecosistema digital, prestando especial atención a las condiciones generales de los servicios digitales. Si bien es cierto que la mayoría de estos servicios digitales vienen publicando desde hace tiempo sus políticas internas o directrices, las críticas se centraban en buena medida en la falta de precisión y claridad. Por ello, las disposiciones actuales refuerzan la exigencia de que esta información sea clara, accesible y fácil de entender[33]. Además de estas obligaciones básicas aplicables a todos los intermediarios, por el particular papel y alcance[34] de los prestadores de plataformas y de motores de búsqueda en línea de muy gran tamaño, se configuran dos obligaciones más en materia de información y transparencia de las condiciones generales. Entre estas se encuentran la obligación de proporcionar «un resumen sucinto, fácilmente accesible y legible por máquina de las condiciones generales, incluidas las medidas correctivas y los mecanismos de recurso disponibles, en un lenguaje claro e inequívoco»[35] y también deben «publicar sus condiciones generales en todas las lenguas oficiales de todos los Estados miembros en los que presten sus servicios»[36]. En mi opinión, estas obligaciones resultan ciertamente razonables, ya que aporta seguridad jurídica y previsibilidad a los usuarios de la red. De esta manera, cualquier medida que se adopte en su contra debe motivarse en este principio de legalidad.

Unido a esta obligación de transparencia, cabe reseñar que el artículo 15 insta a los intermediarios a publicar al menos una vez al año informes claros y comprensibles sobre su actividad de moderación de contenidos en un formato fácilmente accesible. Esta exigencia incluye, dependiendo del tipo de prestador de servicios intermediarios, la incorporación detallada sobre aspectos como las órdenes y notificaciones relacionadas con el contenido ilícito, las actividades de moderación proactiva, los sistemas internos de gestión de reclamaciones y el empleo de medios automatizados para la moderación de contenidos.

33 *Vid*. art. 14 RSD

34 Cdo. n.º 48 RSD.

35 De acuerdo con el apartado 5 del art. 14 RSD y añade el cdo. n.º 48 que «[...] en dichos resúmenes se deben indicar los principales elementos de las obligaciones de información, incluida la posibilidad de excluir fácilmente las cláusulas opcionales».

36 *Vid*. art. 14.6 RSD.

3.2. Obligaciones adicionales a los prestadores de servicios de alojamiento de datos, incluidas las plataformas en línea

A estas obligaciones básicas a todos los prestadores de servicios de intermediación se unen una serie de obligaciones aplicables a los prestadores de servicios de alojamiento de datos, incluidas las plataformas en línea, como es el establecimiento de un mecanismo de notificación y acción, así como motivar al destinatario del servicio afectado de alguna restricción del servicio. A mi modo de ver, estas disposiciones eran ciertamente necesarias[37] para salvaguardar los derechos de los usuarios digitales, tanto en los casos en que se alegue la existencia de un ilícito en la red, como en aquellos en los que una restricción pudiera afectar a un usuario para conocer el motivo.

En este sentido, deben implementar mecanismos que permitan a cualquier persona física o entidad notificar la presencia de contenido ilícito en su servicio. Dichos mecanismos deben ser de fácil acceso y manejo, y exclusivamente por medios electrónicos. Además, dado que las notificaciones permiten al prestador del servicio localizar rápidamente el contenido señalado, valorar su ilicitud y, en su caso, proceder a la retirada o el bloqueo de acceso, resulta fundamental que los avisos sean precisos y estén debidamente fundamentados. Esto garantiza que el operador económico pueda determinar la claridad de la naturaleza ilícita del contenido y actuar con diligencia. Si la notificación contiene el nombre y correo electrónico, se acusará automáticamente recibo de la misma, como también notificará la decisión que adopte en su caso, quien podrá recurrirla.

Cuando la notificación cumple con los requisitos, se atribuye al prestador del servicio un conocimiento efectivo o permite ser consciente —cuando permita a un prestador diligente de servicios de alojamiento de datos determinar, sin un examen jurídico detallado— que la información o la actividad pertinente es ilícita. Se trata de un aspecto ciertamente controvertido en la práctica porque, a mi modo de ver, este conocimiento efectivo puede fomentar una respuesta casi automática del intermediario, quien podría optar, como viene sucediendo, por eliminar el contenido notificado de manera inmediata para demostrar rapidez en su actuación y evitar, de este modo, perder la exención de responsabilidad como «puerto seguro» en relación con la información notificada.

37 Fui un defensor de estos sistemas de notificación y retirada. Buena parte de mi tesis doctoral se centró en el estudio comparado de los diferentes sistemas de «notificación y retirada», concluyendo, al respecto, la necesidad de implantar un sistema común a todos los Estados miembros de la Unión Europea. *Vid.* Castelló Pastor, J.J.: *Motores de búsqueda y derechos de autor: infracción y responsabilidad,* Aranzadi, Cizur Menor, 2016, pp. 383-402. Por otro lado, en el año 2018, la Comisión Europea publicó la Recomendación (UE) 2018/334 de 1 de marzo, para combatir eficazmente los contenidos en línea, en particular véase la recomendación quinta y siguientes sobre la presentación y tramitación de notificaciones.

Por su parte, si el prestador del servicio digital, ya sea por las investigaciones realizadas por iniciativa propia (como la moderación proactiva de contenidos) o tras recibir una notificación de un tercero —persona física o entidad— decide aplicar una restricción (por ejemplo, eliminar el contenido, bloquear el acceso, reducir su visibilidad, o suspender o finalizar el servicio o la cuenta del usuario) al considerar que la información es ilícita o incompatible con sus condiciones generales, deberá notificar al destinatario del servicio afectado sobre su decisión y los motivos que la justifican, a más tardar, en la fecha en la que se imponga la referida restricción. Además, debe proporcionar información sobre las opciones disponibles para impugnar esta decisión, incluyendo mecanismos internos de gestión de reclamaciones, procedimientos extrajudiciales de resolución de litigios y la posibilidad de acudir a los tribunales estatales (artículo 17 *in fine*). Por otro lado, si el prestador de alojamiento de datos, tras recibir una notificación, decide mantener el contenido facilitado por el destinatario del servicio porque considere que no infringe la ley ni es incompatible con sus condiciones generales, el *notificante* tiene derecho también a impugnar esta decisión del prestador del servicio.

En caso de que el destinatario del servicio afectado y/o quien haya presentado la notificación (persona física o entidad) recurra al sistema interno de gestión de reclamaciones con el propósito de que la plataforma en línea revise su decisión, deberá informar «a los reclamantes, sin dilación indebida, la decisión motivada respecto de la información a que se refiera la reclamación y de la posibilidad de resolución extrajudicial de litigios recogida en el artículo 21 y otras vías de recurso de que dispongan». Cabe destacar que el destinatario del servicio afectado tiene la facultad de acudir en cualquier momento a los tribunales estatales para ver tutelado su derecho, sin que la gestión interna de reclamaciones ni el sistema extrajudicial constituyan un requisito previo para acceder a la justicia.

Otra de las obligaciones a todos los prestadores de servicios de alojamiento de datos consiste en comunicar de inmediato a las autoridades policiales o judiciales «cualquier información que le haga sospechar que se ha cometido, se está cometiendo o es probable que se cometa un delito que implique una amenaza para la vida o la seguridad de una o más personas»[38].

3.3. Más disposiciones de obligada observancia a las plataformas en línea

Adicionalmente se configuran más obligaciones a las plataformas en línea[39]. Como vía alternativa a los procedimientos judiciales, deben imple-

38 De acuerdo con el art. 18 RSD.

39 A efectos del RSD se entiende por «plataforma en línea»: un servicio de alojamiento de datos que, a petición de un destinatario del servicio, almacena y difunde información al

mentar un sistema interno de gestión de reclamaciones (artículo 20) y un sistema extrajudicial de resolución de litigios (artículo 21) relacionadas con sus decisiones (o falta de ellas) respecto a la ilegalidad del contenido proporcionado por el destinatario del servicio o su incompatibilidad con los términos y condiciones generales del servicio digital.

Se trata del diseño de derechos procesales y garantías procedimentales destinados a los usuarios del servicio digital[40], mediante la implementación de un sistema interno para la tramitación (o gestión) de reclamaciones. Este sistema ofrece un mecanismo para que puedan impugnar las decisiones de la plataforma, constituyendo un primer paso hacia la resolución extrajudicial de litigios. Tienen así la posibilidad de presentar directamente la reclamación a la plataforma en línea (o, mejor dicho, impugnar la referida decisión).

Esta armonización de los sistemas de tramitación interna no solo refuerza la seguridad jurídica y las garantías procesales de los destinatarios del servicio, sino que también incrementa la confianza en el ecosistema digital.

La información relativa al acceso y funcionamiento de los referidos sistemas de gestión interna de reclamaciones debe incorporarse en las condiciones generales del servicio. Y por lo que respecta propiamente a los principios que deben responder estos sistemas internos son el de eficacia, diligencia,

público, salvo que esa actividad sea una característica menor y puramente auxiliar de otro servicio o una funcionalidad menor del servicio principal y que no pueda utilizarse sin ese otro servicio por razones objetivas y técnicas, y que la integración de la característica o funcionalidad en el otro servicio no sea un medio para eludir la aplicabilidad del presente Reglamento (art. 3 letra i RSD). Téngase en cuenta que se excluyen las plataformas en línea que sean microempresas y pequeñas empresas de la obligación de implantar el sistema interno de gestión de reclamaciones y la resolución extrajudicial de litigios (art. 19 RSD).

40 Pues escapa del objeto de este estudio. Si bien se ha tenido ocasión de valorar este punto en estudios previos, véase, Castelló Pastor, J.J., «Mediación en los conflictos derivados de los servicios digitales en la Unión Europea» en Meditaciones sobre mediación (MED+), Tirant Lo Blanch, Valencia, 2022, pp. 357-380; Id.: *Daños digitales y derechos de la personalidad: determinación del tribunal competente en conflictos internacionales originados en la red*, Thomson Reuters Aranzadi, 2022, pp. 25-45. Igualmente véase a Fernández Masiá, E., «Medios alternativos de solución de controversias entre las plataformas en línea y los destinatarios del servicio en el reglamento europeo de servicios digitales» en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp. 137-167; Llopis Nadal P.: «Plataformas en línea y decisiones sobre contenidos: el sistema interno de reclamación y la resolución extrajudicial de litigios como vías de impugnación reguladas en la Ley de Servicios Digitales» en *La responsabilidad civil por servicios de intermediación prestados por plataformas digitales*, Colex, A Coruña, 2023, pp. 150-169; PLÁ Almendros, R., «Los ADR en el reglamento de servicios digitales», en AA.VV.: *Masc, to be or not to be?: (medios adecuados de solución de conflictos en la justicia*), Tirant lo Blanch, Valencia, 2024, pp. 571-598; Id., *La solución extrajudicial de disputas transfronterizas en el Reglamento europeo de servicios digitales*, Editorial Colex, 2025, *passim*.

fácil accesibilidad, gratuidad y objetividad, además de que la intervención de la plataforma en línea no se tome únicamente por medios automáticos[41].

El ámbito de aplicación de los sistemas internos se delimita en lo referente a las reclamaciones vinculadas a las decisiones, o la falta de ellas, adoptadas por las plataformas en línea respecto al contenido ilícito o incompatible con sus condiciones generales en relación con la retirada de la información o bloqueo de acceso o restricción de visibilidad; la suspensión o cese de la prestación del servicio, total o parcial, a los destinatarios; suspensión o supresión de la cuenta de los destinatarios; la suspensión, cese o restricción de algún otro modo la capacidad de monetizar la información proporcionada por los destinatarios[42]. Y permite hacerlo durante, al menos, seis meses desde que el destinatario es informado de la decisión[43]. Contra la misma, se puede acudir a la resolución extrajudicial de litigios según el artículo 21 del Reglamento de servicios digitales, incluida siempre la posibilidad de acudir a los órganos jurisdiccionales correspondientes si así lo desea y en cualquier momento.

Con respecto a la resolución extrajudicial de litigios, se permite a «los destinatarios del servicio, incluidas las personas físicas o entidades que hayan enviado notificaciones» acudir a cualquier órgano de resolución extrajudicial de litigios certificado por el coordinador de servicios digitales del Estado miembro para resolver litigios relativos a las decisiones adoptadas por el referido sistema interno de reclamaciones, pero también las que no se hayan resuelto a través de este sistema interno. Sin perjuicio, como se ha referido, del derecho que tienen «a interponer recurso, en cualquier fase, contra dichas decisiones de los prestadores de plataformas en línea ante un órgano jurisdiccional de conformidad con el Derecho aplicable».

De ello se infiere que el objeto del litigio queda delimitado exclusivamente a la decisión (o ausencia de decisión) de la plataforma en línea en relación únicamente por los motivos que puede acudirse al sistema de gestión interna referidos. De modo que cualquier otra acción relacionada con la restricción de la difusión del contenido o prestación del servicio, o de la cuenta del destinatario afectado, como puede ser la indemnización de daños y perjuicios causados por la referida restricción (o por no haberlo hecho), no puede ejercitarse por esta particular vía de solución extrajudicial[44].

Se imponen diversas obligaciones a las plataformas en línea respecto al sistema extrajudicial de litigios. Estas incluyen garantizar que la información sobre cómo acceder a dicho sistema sea fácilmente accesible en su interfaz

41 Véase el cdo. n.º 58 y el art. 20.6 RSD.

42 *Vid*. art. 20.1 RSD.

43 *Vid*. art. 20.2 RSD.

44 Todo ello sin perjuicio de la Directiva 2013/11/UE relativa a la resolución alternativa de litigios en materia de consumo y de los procedimientos y entidades de resolución alternativa de litigios para los consumidores establecidos en la misma (art. 21.9 RSD).

en línea[45], clara y sencilla; así como colaborar de buena fe con el organismo certificado para intentar solventar el conflicto[46]. Esta misma colaboración de buena fe también recae sobre el reclamante, es decir, el destinatario del servicio, incluida la persona física o entidad que haya presentado la notificación. Es importante destacar que las decisiones del órgano certificado no tienen carácter vinculante para las partes, lo que permite a cualquiera de ellas recurrir a la vía judicial para defender sus derechos. En términos prácticos, la efectividad del sistema extrajudicial de litigios es ciertamente bastante limitada.

Los órganos encargados de la resolución extrajudicial de controversias pueden ser designados por el Coordinador de servicios digitales del Estado miembro en el que estén establecidos[47] tras su petición y siempre que cumplan ciertos requisitos establecidos en el artículo 21.3 RSD, aunque también pueden ser establecidos directamente por los Estados miembros[48].

Cuando un organismo solicite su certificación, esta será otorgada por el Coordinador de servicios digitales del Estado miembro correspondiente por un periodo máximo de cinco años, renovable, siempre y cuando acredite el cumplimiento de una serie de condiciones relacionadas con su independencia frente a las partes involucradas y su capacidad para llevar a cabo sus fun-

45 De acuerdo con la definición incorporada en el art. 3 letra m) RSD, el «interfaz en línea» se refiere a «todo programa informático, incluidos los sitios web o partes de sitios web, y las aplicaciones, incluidas las aplicaciones móviles».

46 Si bien la plataforma puede negarse a colaborar en caso de que el mismo litigio haya sido resuelto por un órgano jurisdiccional competente u otro órgano de resolución extrajudicial de litigios, o esté en curso ante ellos (art. 21.2 RSD).

47 El Coordinador de servicios digitales de cada uno de los Estados miembros es una figura clave en el sistema de resolución extrajudicial de litigios en la medida en que se encarga de certificar al órgano su idoneidad para resolver extrajudicialmente los litigios relacionados con el contenido ilícito durante cinco años, y su renovación, pero también de revocar su certificación si dejara de cumplir las condiciones necesarias para su habilitación. En efecto, el Coordinador de servicios digitales puede revocar la certificación otorgada si, de una investigación realizada por iniciativa propia o basada en una información recibida por un tercero, concluye que el órgano no cumple las condiciones de su concesión; si bien con carácter previo dará traslado al órgano afectado de las conclusiones de la investigación y de la intención de revocar el certificado a fin de que opine al respecto (art. 21.7 RSD). Por otro lado, el Reglamento de servicios digitales configura dos obligaciones al Coordinador de servicios digitales con respecto al sistema de resolución extrajudicial de litigios: debe realizar cada dos años un informe sobre el funcionamiento de los órganos de resolución extrajudicial de litigios que haya certificado en el que se incluirá la información que previamente les ha requerido anualmente (art. 21.4 RSD). Y debe notificar a la Comisión los órganos de resolución extrajudicial de litigios que han sido certificados, el ámbito de sus materias específicas y el idioma, además de los órganos cuya certificación haya sido revocada (art. 21.8 RSD), ya que esta información será pública. Es decir, la Comisión publicará un listado actualizado de estos órganos con sus especificaciones «en un sitio web destinado a tal fin y de fácil acceso» (art. 21.8 RSD).

48 Según lo dispuesto en el art. 21.6 RSD.

ciones de manera rápida y eficaz. En este sentido, el organismo debe demostrar no solo su imparcialidad, sino también su independencia, además de la financiera, tanto de las plataformas en línea como de los destinatarios del servicio prestado por estas y de los *notificantes*. Asimismo, debe garantizar que la remuneración de sus miembros no esté condicionada por el resultado de los procedimientos.

Además, el organismo debe contar con experiencia en uno o varios ámbitos específicos relacionados con contenidos ilícitos o con las condiciones generales de diferentes tipos de plataformas en línea, a fin de resolver controversias de forma adecuada. Es indispensable que el mecanismo de resolución extrajudicial sea fácilmente accesible por medios electrónicos, permita el inicio del procedimiento y la presentación de documentación de forma telemática; y que garantice una resolución rápida, eficiente y eficaz en términos de costes y, como mínimo, en una lengua oficial de la Unión Europea. Asimismo, debe operar bajo normas de procedimiento claras y justas fácilmente accesibles al público y conformes al Derecho aplicable[49]. En definitiva, el organismo solicitante deberá cumplir con todas estas exigencias para adquirir la certificación, en cuyo caso, ésta especificará el ámbito específico de los contenidos ilícitos y la lengua oficial que es capaz de resolver los litigios.

Los órganos certificados deben resolver los litigios en un plazo máximo de noventa días naturales a partir de la recepción de la reclamación. No obstante, dicho plazo puede extenderse, a discreción del órgano, hasta un máximo de ciento ochenta días en casos de especial complejidad. Asimismo, los órganos certificados deben presentar un informe anual al Coordinador de Servicios Digitales sobre su funcionamiento[50]. Y, por último, establece de manera explícita que estos órganos certificados no están facultados para imponer resoluciones vinculantes a las partes implicadas en el litigio[51].

Otra disposición aplicable a todas las plataformas digitales es la relativa a la figura del alertador fiable, un mecanismo consolidado en otros sistemas jurídicos de Derecho comparado, cuyo objetivo principal es combatir la difusión de contenido ilícito de la red. En síntesis, las notificaciones recibidas de alertadores fiables, dentro del marco del mecanismo de notificación y acción, deberán ser tramitadas y resueltas por las plataformas con carácter prioritario y sin dilación, superando en prioridad a todas las demás, incluidas aquellas presentadas por el propio perjudicado. Dicho de otro modo, estos avisos disfrutan de un trato preferente en la lucha contra el contenido nocivo en la red.

---

49 Conforme al art. 21.3 RSD.

50 Este informe debe incluir, como mínimo, el número de litigios tratados, los resultados obtenidos, el tiempo medio de resolución, las deficiencias o dificultades encontradas y cualquier otra información adicional requerida por el Coordinador (art. 21.4 RSD).

51 Art. 21.2 RSD.

El estatus de alertador fiable constituye un privilegio otorgado por el coordinador de servicios digitales de cada Estado miembro a las entidades, excluyendo a los individuos[52], que acrediten experiencia y competencia en la lucha contra los contenidos ilegales[53]. En este sentido, el coordinador de servicios digitales de cada Estado miembro concede este reconocimiento a entidades que cumplan una serie de requisitos: por un lado, demostrar conocimientos y capacidades técnicas para detectar, identificar y notificar contenidos ilícitos; por otro lado, representar intereses colectivos y operar con independencia de cualquier plataforma en línea y, por último, realizar sus actividades de manera oportuna, diligente y objetiva al enviar notificaciones. El cumplimiento de estas condiciones es indispensable para obtener la cualificación de alertador fiable, que puede ser revocada por este mismo coordinador de servicios digitales en caso de incumplimiento[54].

Desde mi punto de vista, la figura de estos alertadores fiables de la red puede considerarse una herramienta valiosa para controlar y mitigar la presencia de contenido (manifiestamente ilícito) en línea. Sin embargo, surgen inquietudes en torno al proceso de selección de estos actores especializados por el coordinador de servicios digitales de cada Estado miembro, especialmente debido a la posible influencia que podrían ejercer sobre la moderación de contenidos. Es deseable que las decisiones no sean adoptadas de manera arbitraria ni respondan exclusivamente a los intereses políticos de la autoridad nacional encargada de su designación, considerando la relevancia social que podrían representar estos *trusted flaggers* y el impacto negativo que, en consecuencia, podrían ocasionar en la red[55].

Por otra parte, las plataformas en línea tienen la responsabilidad de salvaguardar a los usuarios contra «usos indebidos» mediante el cumplimiento de una doble obligación. De acuerdo con lo establecido en el artículo 23, en primer término, deben proceder a la suspensión de las cuentas de aquellos usuarios que, de manera reiterada, publiquen contenido que resulte manifiestamente ilícito. En segundo lugar, se les exige desactivar o inhabilitar los mecanismos de notificación y acción, así como los sistemas internos de reclamación, para aquellos usuarios que hagan un uso evidentemente injustificado de los mismos, ya sea para solicitar la retirada de contenidos o para presentar reclamaciones. En este sentido, las plataformas están obligadas a emitir una advertencia previa a los usuarios afectados antes de proceder a la suspensión de sus cuentas, ofreciendo una explicación de las razones

52 Véase el cdo. n.º 61 RSD.

53 Algunas están mencionadas en el cdo. n.º 61 RSD como por ejemplo la «Europol» o las organizaciones que forman parte de la red INHOPE.

54 Art. 22 *in fine* RSD.

55 Una visión crítica de los alertadores fiables, véase, SCHWEMER, S., «Trusted notifiers and the privatization of online enforcement», Computer Law & Security Review, vol. 35, n.º 6, 2019.

que justifican la medida adoptada e informando sobre los recursos disponibles para impugnar dicha decisión. Asimismo, deben incluir en sus términos y condiciones una política clara y detallada que regule los usos indebidos, especificando ejemplos concretos de situaciones que podrían justificar la implementación de medidas de protección, así como los criterios utilizados para determinar la duración de estas medidas. Además, tienen la facultad de implementar medidas más estrictas en sus términos y condiciones para abordar contenidos manifiestamente ilícitos relacionados con delitos graves, como la pornografía infantil.

Cada situación deberá ser analizada de manera individualizada por la plataforma, que deberá actuar con diligencia, objetividad y sin arbitrariedad[56]. En caso de que se determine la aplicación de medidas de protección, como la suspensión de una cuenta específica, la plataforma estará obligada a proporcionar al usuario afectado una explicación detallada y debidamente fundamentada. Asimismo, la duración de la suspensión deberá ser razonable y guardar proporcionalidad con las circunstancias particulares del caso.

En los últimos años, se ha identificado un incremento en la adopción de técnicas o estrategias por parte de las plataformas que buscan manipular o confundir a los destinatarios del servicio, o bien distorsionar u obstaculizar su capacidad para tomar decisiones de forma libre e informada. Ante esta situación, el legislador de la Unión Europea ha optado por establecer restricciones a estas prácticas, prohibiendo el uso de lo que se conoce como «patrones oscuros» en las interfaces en línea[57]. Entre los ejemplos más representativos de estas prácticas se encuentran los diseños que presentan opciones al destinatario de manera no neutral, favoreciendo una mayor relevancia o visibilidad a las alternativas más convenientes para la plataforma; las solicitudes insistentes para que el usuario revierta decisiones ya adoptadas; o procedimientos que dificultan la cancelación o anulación de un servicio en comparación con la facilidad de su suscripción, entre otras estrategias similares[58].

El legislador de la Unión también regula la publicidad en línea debido a su papel central en este contexto digital y a su relevancia como principal fuente de ingresos para las plataformas. En este sentido, la publicidad digital puede implicar riesgos significativos, que van desde la difusión de anuncios publicitarios que constituyen contenidos ilícitos hasta la promoción económica de

56 Y este análisis deberá considerar factores como las cifras absolutas de usos indebidos; la proporción respecto del contenido total publicado; la gravedad y consecuencias de dichos usos indebidos; y si es factible, la intención del usuario afectado.

57 De acuerdo con el art. 25 dedicado al diseño y organización de interfaces en línea. Sobre este aspecto, véase, Corberá Martínez, J., «Diseño y organización de interfaces en línea en el Reglamento (UE) de servicios digitales» en Castelló Pastor, J.J. (dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp.171-200.

58 Véase, al respecto, el cdo. n.º 67 RSD.

la publicación o amplificación de contenidos y actividades en línea que sean ilícitos o de otro modo nocivos, o la presentación discriminatoria de anuncios que afecten a la igualdad de trato y oportunidades de los ciudadanos[59]. Por consiguiente, cuando una plataforma incluya anuncios en su interfaz en línea, tiene la obligación de garantizar que los usuarios del servicio puedan identificar claramente el anuncio como publicidad, así como el anunciante y a la entidad responsable de su financiamiento. Además, deberá proporcionar información sobre los principales parámetros utilizados para determinar la presentación de dicho anuncio. Adicionalmente, se prohíbe la difusión de anuncios basados en la elaboración de perfiles a partir de categorías especiales de datos. También está prohibido mostrar anuncios basados en la elaboración de perfiles utilizando datos personales del usuario cuando se tenga un conocimiento razonable de que el destinatario del servicio es un menor de edad[60].

Otra obligación de especial importancia se dirige a la protección de los menores. Las plataformas en línea que sean accesibles a este grupo (un ejemplo pudiera ser la red social TikTok) deben adoptar medidas adecuadas y proporcionadas con el fin de asegurar un alto nivel de privacidad, seguridad y protección en el uso de sus servicios, de acuerdo con el artículo 28. Del mismo modo, queda prohibida la presentación de anuncios basados en la elaboración de perfiles cuando exista un conocimiento razonable de que el destinatario del servicio es un menor de edad. Además, la Comisión está facultada para proporcionar directrices específicas en relación con esta materia.

También en materia de transparencia se establecen una serie de obligaciones adicionales dirigidas a las plataformas. Entre ellas, cabe destacar, de un lado, las contempladas en el artículo 24, que imponen la obligación de proporcionar información complementaria relativa a los litigios ante los órganos de resolución extrajudicial mencionados en el artículo 21, así como el número de suspensiones realizadas conforme al artículo 23. Toda esta información debe ser incorporada en el informe que las plataformas están obligadas a elaborar en virtud de lo dispuesto en el artículo 15. Asimismo, se requiere que las plataformas en línea pongan a disposición del público, a través de su interfaz en línea, información sobre el promedio mensual de destinatarios activos del servicio en la Unión Europea. Por otro lado, las plataformas tienen la obligación de presentar a la Comisión las decisiones relacionadas con las restricciones a sus servicios, acompañadas de las correspondientes declaraciones de motivos, conforme a lo establecido en el artículo 17.1. Esta información será incorporada a una base de datos de acceso público gestionada

59 Cdo. n.º 68 RSD.

60 Art. 26 RSD. Por todos, Lapiedra Alcamí, R., «Restricciones impuestas por el legislador europeo a los servicios de intermediación en línea en materia de publicidad digital» en Castelló Pastor, Jj. (Dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp. 205-226.

por la Comisión. Y de otro lado, adquieren relevancia las establecidas en el artículo 27, que regulan el funcionamiento de los sistemas de recomendación[61]. Estos sistemas son fundamentales para el modelo de negocio de las plataformas en línea, ya que una parte sustancial de su actividad económica se sustenta en la personalización de los productos y servicios ofrecidos a los usuarios. Esta personalización se lleva a cabo utilizando la información facilitada durante el proceso de registro necesario para acceder al servicio o, por ejemplo, por los datos de navegación obtenidos a través del rastreo y seguimiento habilitados por *cookies*. Toda esta información permite a las plataformas presentar contenido o información potencialmente relevante para el usuario, facilitando así la transacción comercial. En ausencia de esta personalización, la información proporcionada sería de carácter aleatorio, menos alineada o afín con los intereses o gustos del usuario, lo que podría, con toda probabilidad, motivarlo a preferir plataformas *más intuitivas*.

Estas obligaciones legales tienen como finalidad que el usuario del servicio digital pueda conocer el modo en el que las plataformas en línea recomiendan, clasifican y/o priorizan[62] los productos o servicios que se muestran en el interfaz en línea. Esto se debe a que la preeminencia de la información *sugerida* puede influir tanto en la capacidad del usuario de interactuar con ella como en los ingresos económicos de los comerciantes[63]. Por esta razón, una medida coherente y alineada con las anteriores disposiciones en materia de transparencia consiste en incluir, en las condiciones generales del servicio[64], una explicación de los principales parámetros utilizados en los sistemas de recomendación y clasificación para sugerir una información[65]. Esta

61 Definidos como «un sistema total o parcialmente automatizado y utilizado por una plataforma en línea para proponer en su interfaz en línea información específica para los destinatarios del servicio o priorizar dicha información, también como consecuencia de una búsqueda iniciada por el destinatario del servicio, o que determine de otro modo el orden relativo o la relevancia de la información presentada», de acuerdo con el art. 3 letra s) RSD.

62 Cabe referirse también como formas para facilitar y optimizar el acceso a la información, por ejemplo, por medio de «la distinción de texto u otras representaciones visuales, o la organización de manera diferente de la información facilitada por los destinatarios», como indica el cdo. n.º 70.

63 Son muchos los estudios que han mostrado que, a modo de ejemplo, el usuario no avanza más de la segunda página de los resultados ofrecidos por el buscador de modo que el éxito económico reside en aparecer en los primeros puestos. *Vid*., Goldman, E., «Deregulating relevancy in Internet trademark law», Emory Law Journal, vol. 54, 2005, p. 535. Por su parte, Grimmelmann, J., «The structure of search engine Law», Iowa Law Review, vol. 93, 2007, p. 31 señaló acertadamente en su momento que «if you don't appear on the first few pages of results, you may as well not exist».

64 Art. 27.1 RSD.

65 Estos parámetros principales, de acuerdo con el art. 27.2 RSD, explicarán por qué se sugiere determinada información al destinatario del servicio e incluirán, como mínimo, los criterios más significativos a la hora de determinar la información sugerida al destinatario del servicio y las razones de la importancia relativa de dichos parámetros.

explicación debe estar redactada en un lenguaje claro y comprensible[66]. Asimismo, en caso de que exista la posibilidad de modificar o influir en dichos parámetros, también debe ser expresamente indicado[67].

En el caso de las plataformas en línea[68] que permiten a los consumidores la celebración de contratos a distancia con comerciantes, como ocurre en los *marketplaces* (p. ej. Amazon, Temu o Aliexpress), el legislador de la Unión las regula para minimizar los riesgos asociados a estos servicios de intermediación. En líneas generales, estas medidas están diseñadas para garantizar la protección de los consumidores y, de forma más amplia, salvaguardar los intereses de terceros que puedan verse afectados por la comercialización de productos o servicios a través de estas plataformas, particularmente cuando éstos contravengan las disposiciones legales vigentes, como en los casos de infracción de derechos de propiedad intelectual[69].

Entre las obligaciones adicionales impuestas a estas plataformas, se destaca la necesidad de recopilar información sobre los comerciantes que permita garantizar su trazabilidad antes de autorizarles a utilizar sus servicios[70]. Asimismo, se exige que las interfaces en línea estén diseñadas de forma que faciliten el cumplimiento de estas obligaciones de información por parte de los comerciantes, especialmente en lo relativo a los requisitos de contratación con consumidores establecidos por otros instrumentos jurídicos de la UE[71]. Por último, estas plataformas tienen el deber de informar a los consumidores cuando hayan adquirido productos o servicios ilícitos a través de sus servicios, siempre que los hechos se hayan producido en los seis meses anteriores al momento en el que el prestador haya tenido conocimiento de dichas irregularidades[72].

---

66 Art. 27.1 RSD. Esta obligación se refuerza en el cdo. n.º 70 RSD al indicar que «[...] las plataformas en línea deben asegurarse de un modo coherente de que los destinatarios de su servicio estén adecuadamente informados sobre cómo los sistemas de recomendación afectan a la forma en que se muestra la información y pueden influir en la manera en que la información se les presenta. Deben presentar con claridad los parámetros de dichos sistemas de recomendación de manera fácilmente comprensible con el fin de asegurarse de que los destinatarios del servicio entienden cómo se prioriza la información para ellos [...]».

67 Así como facilitar la funcionalidad que permita seleccionar y modificar en cualquier momento la opción preferida en el orden relativo de la información; debiendo ser accesible directa y fácilmente desde la sección específica de la interfaz de la plataforma en línea en la que se priorice la información, de acuerdo con el art. 27.3 RSD.

68 Con exclusión de microempresas y pequeñas empresas, de acuerdo con el art. 29.

69 Sobre este aspecto, véase, Paredes Pérez, J.I., «Aspectos internacionales de la responsabilidad civil de las plataformas en línea B2C frente a los contenidos ilícitos en materia de protección de los consumidores», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024, pp. 245-279.

70 *Vid*. art. 30. Trazabilidad de los comerciantes.

71 *Vid*. art. 31. Cumplimiento desde el diseño.

72 *Vid*. art. 32. Derecho a la información.

3.4. Obligaciones adicionales de gestión de riesgos sistémicos para las plataformas en línea y buscadores de muy gran tamaño

Este nuevo marco jurídico digital, como se ha referido, establece una política innovadora al atribuir a los intermediarios la responsabilidad de salvaguardar un entorno en línea seguro, predecible y confiable, instándoles a adoptar un comportamiento responsable y diligente[73]. A tal efecto, el legislador considera oportuno establecer adicionalmente un estándar más alto a las plataformas en línea de muy gran tamaño y a los motores de búsqueda de muy gran tamaño (sección 5.ª, artículos 33 a 43), ya que el número de usuarios que utilizan estos servicios pueden plantear especiales riesgos.

La regulación de la gestión de riesgos sistémicos constituye el pilar fundamental en la arquitectura normativa de riesgos establecida en este texto jurídico, representando, sin duda alguna, un avance notable en el marco de las grandes plataformas digitales en la Unión Europea[74]. Se trata de un enfoque basado en el riesgo: estos gigantes tecnológicos pueden ejercer una fuerte influencia social —incluso desproporcionada— en la Unión Europea, motivo por el cual resulta necesario someterlos a un exigente régimen de evaluaciones y toma de medidas para mitigar cualesquiera de los riesgos sistémicos[75].

Con este fin, se establece un sistema de gestión de riesgos que consta dos obligaciones: de un lado, se les encomienda a analizar y evaluar de forma diligente cualquier riesgo sistémico en la Unión Europea derivado del diseño o del funcionamiento de sus servicios y los sistemas utilizados con estos, incluidos los sistemas algorítmicos, o del uso que se haga de sus servicios[76].

---

73 Cdo. n.º 76 RSD.

74 Ampliamente tratado por Reig Fabado, I., «Las VLOPS y los riesgos sistémicos en la DSA: algunos aspectos problemáticos», en Castelló Pastor, J.J. (dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024, pp. 289-322.

75 Conforme a los cdos. n.º 80 a 83 RSD, los riesgos sistémicos se refieren a los riesgos asociados a la difusión de contenidos ilícitos; a los efectos reales o previsibles del servicio para el ejercicio de los derechos fundamentales; también aquellos se refiere a los efectos negativos reales o previsibles sobre los procesos democráticos, el discurso cívico y los procesos electorales, así como sobre la seguridad pública y las relacionadas con el diseño, el funcionamiento o la utilización, también mediante la manipulación, de plataformas en línea de muy gran tamaño y de motores de búsqueda en línea de muy gran tamaño, con un efecto negativo real o previsible en la protección de la salud pública, los menores y graves consecuencias negativas para el bienestar físico y mental de una persona, o en la violencia de género.

76 A tenor del art. 34 RSD. En síntesis, esta evaluación se realizará, al menos, una vez al año y «será específica de sus servicios y proporcionada a los riesgos sistémicos, teniendo en cuenta su gravedad y probabilidad, e incluirá los siguientes riesgos sistémicos: a) la difusión de contenido ilícito a través de sus servicios; b) cualquier efecto negativo real

De otro lado, a implementar una serie de medidas con el fin de disminuir estos riesgos sistémicos[77]. Sin embargo, estos sistemas de gestión de riesgos sistémicos plantean objeciones tanto por el amplio margen de discrecionalidad que se otorga a estos actores digitales como por el poder de control que concentra la Comisión Europea[78].

Con respecto a la falta de concreción del término «contenido ilícito»[79], plantea un desafío significativo en la gestión de los riesgos sistémicos, ya que su interpretación depende del derecho material aplicable al fondo de la controversia. En este contexto, se otorga cierto grado de flexibilidad a estos actores digitales para que definan en sus condiciones generales las conductas consideradas relevantes a la hora de evaluar el posible riesgo asociado

o previsible para el ejercicio de los derechos fundamentales [...]; c) cualquier efecto negativo real o previsible sobre el discurso cívico y los procesos electorales, así como sobre la seguridad pública; d) cualquier efecto negativo real o previsible en relación con la violencia de género, la protección de la salud pública y los menores y las consecuencias negativas graves para el bienestar físico y mental de la persona». Señala G´Sell, F., «The Digital Services Act (DSA): A General Assessment» en Antje von Ungern-Sternberg (ed.), *Content Regulation in the European Union – The Digital Services Act*, TRIER STUDIES ON DIGITAL LAW, vol. 1, Verein für Recht und Digitalisierung e.V., Institute for Digital Law (IRDT), Trier 2023, p. 114 que «la definición de riesgos sistémicos del apartado 1 del artículo 34 es particularmente amplia y vaga. Por ejemplo, cualquier cosa que pueda menoscabar los derechos fundamentales podría considerarse un riesgo sistémico, lo que implica que cualquier tipo de contenido podría vincularse a dicho riesgo. Esto crea una gran incertidumbre sobre lo que las plataformas, los investigadores y los reguladores podrían clasificar como riesgos sistémicos. Una interpretación global de los "riesgos sistémicos" podría dar lugar a la aparición de una categoría "gris" de contenidos que no son explícitamente ilegales pero que se consideran indeseables por su potencial para crear riesgos sistémicos. Además, la creación de esta categoría "gris", a medio camino entre lo ilegal y lo legal, representa una amenaza real para la libertad de expresión».

77 Según el art. 35 RSD. Estas medidas de detección deben ser razonables, proporcionadas y efectivas, teniendo especialmente en cuenta las consecuencias de dichas medidas sobre los derechos en liza.

78 Sobre este aspecto, Castelló Pastor, J.J., «Inteligencia Artificial aplicada por las plataformas en línea y motores de búsqueda de muy gran tamaño para mitigar los riesgos sistémicos en línea», *Cuadernos de Derecho Transnacional*, vol. 17, n.º 1, 2025.

79 Definido «contenido ilícito» en la letra h) del art. 3 RSD como «toda información que, por sí sola o en relación con una actividad, incluida la venta de productos o la prestación de servicios, incumpla el Derecho de la Unión o el Derecho de cualquier Estado miembro que cumpla el Derecho de la Unión, sea cual sea el objeto o carácter concreto de ese Derecho». Aunque ciertamente puede hablarse de un contenido ilícito armonizado, como el contenido terrorista, la incitación ilegal al odio (racismo y xenofobia) en línea, etc. Véase la Recomendación (UE) 2023/2425 de la Comisión, de 20 de octubre de 2023, sobre la coordinación de las respuestas a incidentes, en particular, derivados de la difusión de contenidos ilegales, antes de la plena entrada en aplicación del Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo (Ley de Servicios Digitales) (notificado en virtud del documento C (2023) 7170). DO L, 2023/2425, 26.10.2023 y la normativa indicada (en especial, en el párrafo 11 y 13).

a la difusión de dicho contenido[80]. Tampoco se facilita la función de evaluación del riesgo debido a que estos operadores tecnológicos deben centrar su atención en «*la información que no sea ilícita*, pero que contribuya a los riesgos sistémicos identificados en el presente Reglamento», lo que incluye «contenidos incorrectos o engañosos, incluida la desinformación»[81]. Es decir, deben considerar contenidos nocivos o con efectos negativos, los cuales, además pueden variar de un ordenamiento jurídico a otro.

Los servicios digitales desempeñan un papel esencial en la difusión de información y están en una posición privilegiada para abordar los riesgos sistémicos, pero es ciertamente preocupante que la determinación de lo que constituye el carácter nocivo de un contenido recaiga exclusivamente en estas empresas privadas —o más específicamente, en sus algoritmos—, porque puede incentivar o motivar la eliminación de estos contenidos con el fin de reducir los riesgos, legitimadas por la propia normativa[82].

Con este margen de discrecionalidad para determinar el contenido que puede estar en la red —sea lícito, pero se entienda que contribuye al riesgo—, se permite a estos sujetos privados seguir disponiendo del control del discurso en el ecosistema digital, respaldado en este caso por las autoridades públicas; aun cuando reiteradamente se haga hincapié en el texto jurídico en la obligación de proceder de manera no arbitraria y no discriminatoria, así como en la necesidad de tener en cuenta los derechos e intereses legítimos de los destinatarios del servicio, incluidos los derechos fundamentales amparados por la Carta[83].

También la Comisión les puede exigir la adopción de diversas medidas cuando se produzca una crisis[84]; así como una cuestión igualmente compleja es, a mi modo de ver, la facultad de la Comisión de emitir directrices y recomendaciones específicas sobre riesgos sistémicos concretos[85]. Si bien esta

80 Y una forma de sortear este aspecto es fijando en las políticas o directrices del servicio digital los comportamientos pueden realizarse (con independencia del carácter licito o ilícito).

81 Cdo. n.º 84 RSD.

82 Barata Mir, J., «Libertad de expresión y moderación privada de contenidos», Teorder, n.º 32, 2022, pp. 99 y 100.

83 Cdos. n.º 47 y 79 RSD.

84 Cuando se produzcan circunstancias extraordinarias que den lugar a una amenaza grave para la seguridad pública o la salud pública en la Unión o en partes significativas de esta» (art. 36.2).

85 Como las Directrices para plataformas en línea de muy gran tamaño y a los motores de búsqueda de muy gran tamaño sobre la mitigación de los riesgos sistémicos para los procesos electorales publicadas el 26 de marzo de 2024 y con orientaciones para las elecciones al Parlamento Europeo del pasado mes de junio o la Recomendación (UE) 2024/915 de la Comisión de 19 de marzo de 2024 sobre medidas destinadas a luchar contra la falsificación y a reforzar la observancia de los derechos de propiedad intelectual e industrial.

medida, que viene realizándose en estos últimos años, representa un avance positivo, la preocupación de esta estrategia aplicada para mitigar determinados riesgos viene porque pueda percibirse como una imposición de normas de comportamiento vinculantes a las plataformas en línea de muy gran tamaño y a los motores de búsqueda de muy gran tamaño. Con estas acciones, la Comisión podría, de manera indirecta, ejercer un control sobre comportamiento de los usuarios en la red del que carece. De ahí mi interrogante de quién debería exhibir un comportamiento responsable y diligente en este contexto digital: si los referidos actores (privados) tecnológicos o las autoridades (públicas) europeas.

Por último, las plataformas en línea y buscadores de muy gran tamaño están sujetas a una realizar auditorías independientes[86], mayores obligaciones en los sistemas de recomendación[87] y en publicidad digital[88], así como mayores obligaciones en materia de transparencia informativa[89], entre otras.

IV. Mecanismos de tutela jurídico-pública y privada en el Reglamento (UE) de servicios digitales

Los prestadores de servicios de intermediación podrían incumplir con las disposiciones sustantivas de obligada observancia del Reglamento (UE) de servicios digitales, en cuyo caso, la cuestión reside en determinar si este marco jurídico digital contempla mecanismos que garanticen de manera efectiva el cumplimiento de sus objetivos, o si, por el contrario, adolece de la capacidad para imponer consecuencias jurídicas frente a eventuales incumplimientos de estos actores digitales.

A este respecto conviene señalar que el Reglamento establece mecanismos tanto de tutela jurídico-pública de supervisión destinados a garantizar su cumplimiento, como de tutela jurídico-privada. En particular, se prevé la supervisión a cargo de las autoridades competentes de cada Estado miembro, así como un papel destacado de la Comisión en este ámbito. Además, se otorga a las personas físicas y jurídicas la facultad de presentar reclamaciones por la infracción de esta normativa digital[90].

86 *Vid*. art. 37 RSD.

87 *Vid*. art. 38 RSD.

88 *Vid*. art. 39 RSD.

89 *Vid*. art. 42 RSD.

90 Como señala DE MIGUEL, estos mecanismos representan cambio significativo «habida cuenta de las tradicionales carencias de la aplicación efectiva del ordenamiento jurídico a muchos de estos prestadores intermediarios, incluidos aquellos que son objeto de obligaciones más estrictas». *Vid*. DE MIGUEL ASENSIO, P., *Derecho privado de internet*, 6.ª edición, Civitas Thomson-Reuters, 2022, p. 370.

Este novedoso marco normativo digital efectivamente impone a los Estados miembros la obligación de designar «una o varias autoridades competentes responsables de la supervisión de los prestadores de servicios intermediarios y de su ejecución»[91], siendo una de las autoridades competentes su coordinador de servicios digitales, que será responsable de todas las materias relacionadas con la supervisión y garantía del cumplimiento del Reglamento en ese Estado miembro[92]. A modo de ejemplo, la Comisión Nacional de los Mercados y la Competencia (CNMC) ha sido designada como Coordinador de Servicios Digitales de España.

En consecuencia, corresponde al Estado miembro donde se encuentre el establecimiento principal del prestador de servicios intermediarios la competencia exclusiva para supervisar y hacer cumplir el Reglamento[93]. Asimismo, es importante señalar que la determinación del establecimiento principal del prestador de servicios intermediarios se refiere al lugar en el que tenga su sede central o domicilio social en el que ejerza las principales funciones financieras y el control de sus operaciones. En el caso de que el prestador no esté establecido en la Unión, pero que ofrezca sus servicios aquí, la competencia corresponderá al Estado miembro en el que haya designado a su representante legal[94], sin perjuicio de las competencias que recaen en la Comisión[95].

Con esta claridad se recoge en la definición de «coordinador de servicios digitales del establecimiento» contenida en el artículo 3 letra n), donde se establece que el término hace referencia al «coordinador de servicios digitales del Estado miembro en el que el establecimiento principal del prestador de un servicio intermediario esté ubicado o su representante legal resida o esté establecido». Esta definición, nótese, también abarca la referencia a la residencia o al establecimiento del representante legal, ya que uno de los requisitos que deben cumplir todos los prestadores de servicios intermediarios es, precisamente, la designación (o nombramiento) de estos repre-

91 *Vid*. art. 49.1 RSD.

92 Y este asumirá la responsabilidad en todas las cuestiones relacionadas con la supervisión y el cumplimiento de este Reglamento dentro del Estado miembro correspondiente. Además, tendrá la tarea de garantizar la coordinación a nivel nacional en estas materias y de contribuir a la supervisión y aplicación efectivas y uniformes del Reglamento en toda la Unión Europea, según el art. 49.2 RSD.

93 De acuerdo con el art. 56.1 RSD.

94 *Vid.* cdo. n.º 123 RSD in fine.

95 De acuerdo con el art. 56.6 RSD: «cuando un prestador de servicios intermediarios no tenga un establecimiento en la Unión, el Estado miembro en el que su representante legal resida o esté establecido o la Comisión tendrán competencias, según proceda, de conformidad con los apartados 1 y 4 del presente artículo, para supervisar y hacer cumplir las obligaciones pertinentes en virtud del presente Reglamento».

sentantes legales[96]. Y no cabe desatender, por otro lado, el *superpoder* de la Comisión en este contexto digital. En efecto, la Comisión tiene competencias para supervisar y hacer cumplir las obligaciones adicionales de gestión de riesgos sistémicos para prestadores de plataformas y de motores de búsqueda en línea de muy gran tamaño[97].

Al margen de lo anterior, cobra especial importancia el reconocimiento a las personas físicas y a determinadas entidades de una legitimación (extraordinaria) para tutelar (colectivamente) sus derechos e intereses ante estas autoridades en caso de infracción por el incumplimiento de esta normativa digital. En efecto, el Reglamento contiene normas diseñadas para reforzar esta tutela jurídico-pública que permiten a los usuarios digitales defender los derechos otorgados por esta normativa solicitando la intervención del coordinador de servicios digitales, además de autorizar a determinadas enti-

96 De acuerdo con el art. 13 RSD:
1. Los prestadores de servicios intermediarios que *no tengan un establecimiento* en la Unión pero que ofrezcan servicios ven la Unión *designarán, por escrito, a una persona física o jurídica para actuar como su representante legal en uno de los Estados miembros en los que el prestador ofrezca sus servicios*.
2. Los prestadores de servicios intermediarios mandatarán a sus representantes legales, además o en lugar de dichos prestadores, con el fin de ser los destinatarios de las comunicaciones enviadas por las autoridades competentes de los Estados miembros, la Comisión y la Junta sobre todas las cuestiones necesarias para la recepción, el cumplimiento y la ejecución de las decisiones adoptadas en relación con el presente Reglamento. Los prestadores de servicios intermediarios otorgarán a su representante legal las facultades necesarias y recursos suficientes para garantizar su cooperación eficiente y en tiempo oportuno con las autoridades competentes de los Estados miembros, la Comisión y la Junta, y cumplir dichas decisiones.
3. *El representante legal designado podrá ser considerado responsable por el incumplimiento de las obligaciones en virtud del presente Reglamento*, sin perjuicio de la responsabilidad del prestador de servicios intermediarios y de las acciones legales que puedan iniciarse contra este.
4. Los prestadores de servicios intermediarios notificarán el nombre, el domicilio postal, la dirección de correo electrónico y el número de teléfono de su representante legal al coordinador de servicios digitales del Estado miembro en el que dicho representante legal resida o esté establecido. Velarán por que dicha información esté a disposición del público, sea fácilmente accesible, sea exacta y esté actualizada.
5. La *designación de un representante legal* en la Unión en virtud del apartado 1 *no constituirá un establecimiento en la Unión* (énfasis añadido).

97 De acuerdo con el art. 56.2 RSD. Por su parte, señala el cdo. n.º 140 RSD que «en vista de los problemas concretos que puedan surgir para asegurar el cumplimiento por parte de los prestadores de plataformas en línea de muy gran tamaño o de motores de búsqueda de muy gran tamaño y de la importancia de hacerlo de manera efectiva, tomando en consideración su dimensión e impacto y los perjuicios que pueden causar, *la Comisión debe poseer amplios poderes de investigación y ejecución que le permitan investigar, hacer cumplir y vigilar el cumplimiento de las normas establecidas en el presente Reglamento*, respetando plenamente el derecho fundamental a ser oído y a tener acceso al expediente en el contexto de los procedimientos de ejecución, el principio de proporcionalidad y los derechos e intereses de los afectados» (énfasis añadido).

dades a ejercer acciones de representación, lo que contribuye a proteger los intereses generales de los particulares[98].

A tal efecto, se otorga al destinatario del servicio, así como a los organismos, organizaciones o asociaciones autorizados a ejercer en su nombre, el derecho a presentar una reclamación contra los prestadores de servicios intermediarios por una presunta infracción del Reglamento[99]. Y cabe referirse también a las acciones de representación[100].

Y en relación con las consecuencias jurídicas que podrían surgir ante un posible incumplimiento de las disposiciones obligatorias establecidas en este nuevo marco jurídico digital, esto es, sanciones, no considero que puedan ser interpretadas como insignificantes. Por el contrario, la estricta naturaleza de las sanciones previstas en caso de incumplimiento debería representar un estímulo o incentivo significativo para su cumplimiento, incluso para los gigantes tecnológicos estadounidenses y asiáticos, cuyas valoraciones bursátiles superan el billón de dólares.

El Reglamento establece un severo régimen de sanciones en caso de incumplimiento[101]. El considerando n.º 144 lo expresa de manera inequívoca: «el cumplimiento de las obligaciones pertinentes impuestas en virtud del presente Reglamento debe garantizarse mediante multas sancionadoras y multas coercitivas». A este respecto, son los Estados miembros los encargados de fijar estas sanciones al intermediario que esté bajo su competencia y deberán ser «efectivas, proporcionadas y disuasorias»[102]. No obstante, el legislador de la Unión ha fijado que el importe máximo de las multas sea el 6 % del volumen de negocios anual en todo el mundo del prestador del servicio que se trate en el ejercicio fiscal anterior. Asimismo para infracciones como, por ejemplo, proporcionar información incorrecta, incompleta o engañosa, por no responder o por no rectificar información incorrecta, incompleta o engañosa y por no someterse a una inspección las sanciones se fijan en el 1 %

98 DE MIGUEL ASENSIO, P., *Manual de Derecho de las nuevas tecnologías*, Aranzadi, Pamplona, 2023, pp. 102-103.

99 De acuerdo con el art. 53 RSD. Esta reclamación puede presentarse ante el coordinador de servicios digitales del Estatal o miembro donde esté situado o establecido el destinatario del servicio. El coordinador evaluará la reclamación y si corresponde, la remitirá al coordinador de servicios digitales de establecimiento, acompañándola de un dictamen si lo estima apropiado. Cuando la reclamación sea responsabilidad de otra autoridad competente en su Estado miembro, el coordinador de servicios digitales que reciba la reclamación la transmitirá a dicha autoridad. Durante el procedimiento, ambas partes tendrán derecho a ser oídas y a recibir información adecuada sobre el estado del procedimiento conforme con el Derecho nacional.

100 *Vid*. arts. 86 y art. 90 RSD.

101 Véase el capítulo IV dedicado a la «aplicación, cooperación, sanciones y ejecución» y en particular el artículo 52 «sanciones» y el artículo 74 «multas» respectivamente.

102 De acuerdo con el art. 52.1 RSD.

de los ingresos anuales, o del volumen de negocios anual en todo el mundo; y se prevé que el importe máximo de una multa coercitiva sea del 5 % del promedio diario del volumen de negocios en todo el mundo o de los ingresos del prestador de servicios intermediarios de que se trate en el ejercicio fiscal anterior por día, calculado a partir de la fecha especificada en la decisión de que se trate[103].

Debe tenerse en cuenta que la Comisión Europea puede imponer también multas de hasta un máximo de dichos importes y, en todo caso, para fijar el importe de la multa se tendrá en cuenta «la naturaleza, gravedad, duración y recurrencia de la infracción y la demora causada al procedimiento»[104]. Igualmente puede adoptar una decisión por la que se impongan una multa coercitiva al prestador de la plataforma o del motor de búsqueda en línea de muy gran tamaño que no excedan del 5 % del promedio diario de los ingresos o del volumen de negocios anuales en todo el mundo del ejercicio fiscal anterior por día, calculados a partir de la fecha fijada por esa decisión[105]. En cualquier caso, el Tribunal de Justicia de la Unión Europea tiene competencia jurisdiccional plena para someter a control las decisiones por las que la Comisión haya impuesto multas sancionadoras o multas coercitivas; pudiendo anular, reducir o aumentar la multa sancionadora o la multa coercitiva impuesta[106]. Y en aquellos casos en que el incumplimiento sea reiterado, podrá dar lugar, a través del coordinador de servicios digitales de establecimiento del prestador que solicite a la autoridad judicial competente de su Estado la orden de que se limite temporalmente el acceso de los destinatarios al servicio afectado por la infracción o bien, únicamente cuando ello no sea técnicamente viable, a la interfaz en línea del prestador de servicios intermediarios en la que tenga lugar la infracción[107]. Junto a estos mecanismos jurídico-públicos, cabe referirse también a la aplicación privada de este marco jurídico digital, es decir, la tutela judicial civil con respecto a los daños y perjuicios causados por una infracción de este texto digital. Se trata, en puridad, de un sistema de garantías de su efectividad en el que la iniciativa privada complementa la referida aplicación pública. La eventual indemnización por el daño causado a cargo de estos intermediarios por el incumplimiento de las disposiciones del Reglamento contribuye, también, a su efectividad.

En relación con la aplicación privada, tiene especial relevancia el artículo 54 del Reglamento (UE) de servicios digitales en la medida en que «los destinatarios del servicio tendrán derecho a solicitar, de conformidad con el Derecho de la Unión o nacional, una indemnización con cargo a los prestadores de

103 Art. 52 *in fine* RSD.

104 *Vid*. art. 74 *in fine* RSD.

105 Art. 76 RSD.

106 Art. 81 RSD.

107 De acuerdo con el art. 82 en relación con el art. 51, apartado 3 RSD.

servicios intermediarios por cualquier daño o perjuicio sufrido como consecuencia del incumplimiento por parte de dichos prestadores de sus obligaciones en virtud del presente Reglamento». De su lectura se desprende que no solo impone obligaciones a los intermediarios, sino que también otorga derechos subjetivos a los particulares, cuyas infracciones podrán dar lugar a una indemnización[108].

Así, la legitimación activa corresponde a los destinatarios del servicio que, a tenor de la definición incorporada en la letra b) del artículo 3, se refiere a «toda persona física o jurídica que utilice un servicio intermediario, en particular para buscar información o para hacerla accesible». Esta definición debe interpretarse de manera amplia, incluyendo al suscriptor de un servicio o a quien lo difunda, como a quién utilice una plataforma o un buscador —porque se expone a la información mostrada en el interfaz[109]—. De Miguel Asensio sugiere, con buen criterio, que se podría adoptar «una interpretación incluso más amplia en el contexto del artículo 54. En particular, afectadas por la difusión de contenidos ilícitos y perjudicadas por la infracción del Reglamento (UE) de servicio digitales puede ser, por ejemplo, cualquier persona cuyo derecho al respeto de la vida privada y familiar se ve menoscabado por prácticas de los prestadores de servicios intermediarios que infringen lo dispuesto en el Reglamento (UE) de servicio digitales. Una interpretación amplia resulta avalada por el contenido de otras disposiciones del Reglamento, como su artículo 20 que refleja el criterio de que cualquier persona física o entidad que notifica al prestador de servicios de alojamiento la presencia de contenidos que considere ilícitos en el marco del artículo 16 debe ser considerado un destinatario del servicio»[110]. Por último, cabe recordar que es posible interponer acciones colectivas y que el destinatario del servicio puede mandatar «a un organismo, organización o asociación para que ejerza los derechos atribuidos por el presente Reglamento en su nombre»[111].

108 En este sentido cabe referirse al cdo. n.º 121 RSD: «[...] un prestador de servicios intermediarios debe ser responsable por los perjuicios sufridos por destinatarios del servicio causados por un incumplimiento de las obligaciones establecidas en el presente Reglamento por dicho prestador. Dicha compensación debe ajustarse a las normas y procedimientos establecidos en el Derecho nacional aplicable y sin perjuicio de otras vías de recurso disponibles en virtud de las normas de protección de los consumidores».

109 Véanse las definiciones de «destinatario activo de una plataforma en línea» (art. 3 letra p) y «destinatario activo de un motor de búsqueda en línea» (art. 3 letra q) respectivamente.

110 De Miguel Asensio, P., Entrada del blog 1 de diciembre de 2022 titulada «Reglamento de Servicios Digitales (VII): supervisión y aplicación» (pedrodemiguelasensio.blogspot.com).

111 De acuerdo con lo dispuesto en el art. 86 RSD.

V. Consideración final

Con una generación de nativos digitales, la tecnología ha transformado de manera profunda y estructural nuestras formas de comunicación, trabajo e interacción social. En este escenario, la actualización del marco normativo de los prestadores de servicios de intermediación se ha convertido en una prioridad ineludible para abordar los desafíos que emergen de los continuos avances tecnológicos.

Este new tecnológico promovido por la Unión Europea se refleja en el Reglamento (UE) de servicios digitales, que, junto con otros instrumentos legislativos recientemente adoptados, constituye un avance significativo en la gobernanza de las plataformas en línea. Este marco regulatorio era necesario para establecer un mayor control democrático sobre las actividades de las grandes multinacionales tecnológicas —con el objetivo de contrarrestar los efectos adversos de la creciente privatización y automatización en la gestión del discurso en línea— y situar a los usuarios en el centro de la transición digital.

Este nuevo marco jurídico digital regula aspectos esenciales de los servicios digitales y, de manera destacada, promueve los valores fundamentales de la Unión Europea mediante la implementación de sus normas y estándares, configurando un entorno empresarial tecnológico coherente y armonizado para los actores que operan en el mercado interior. Asimismo, las medidas asimétricas que impone a los intermediarios, sustentadas en principios como la transparencia, información y rendición de cuentas, representan un avance significativo en la promoción de un ecosistema digital más seguro y respetuoso con los derechos consagrados en la Carta, contribuyendo a su consolidación como espacio alineado con los principios democráticos y los derechos fundamentales, esenciales para reforzar la confianza de los usuarios en la era 4.0 de internet.

Bibliografía

Barata Mir, J., «Libertad de expresión y moderación privada de contenidos», *Teorder*, n.º 32, 2022.

Barrio Andrés, M., «Inteligencia artificial, Internet de las cosas y blockchain», en Montero Pascual, J.J., *Digitalización y Derecho*, Tirant Lo Blanch, Valencia, 2024.

Bradford, A., «The Brussels Efect», en *Northwestern University Law Review*, vol. 107, num. 1, 2012.

Bradford, A., *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020.

Castelló Pastor, J.J., «Exoneración de responsabilidad de los prestadores de servicios de la información en la sección 230 de la Communications Decency Act estadounidense», en *Revista Aranzadi de derecho y nuevas tecnologías*, n.º 39, 2015.

Castelló Pastor, J.J., *Motores de búsqueda y derechos de autor: infracción y responsabilidad,* Aranzadi*,* Cizur Menor*,* 2016.

Castelló Pastor, J.J., «Nuevo régimen de responsabilidad de los servicios digitales que actúan como intermediarios a la luz de la propuesta de Reglamento relativo a un mercado único de servicios digitales», en Castelló Pastor, J.J. (dir.), *Desafíos jurídicos ante la integración digital: aspectos europeos e internacionales*, Aranzadi Thomson-Reuters, Cizur Menor, 2021.

Castelló Pastor, J.J., *Daños digitales y derechos de la personalidad: determinación del tribunal competente en conflictos internacionales originados en la red*, Thomson Reuters Aranzadi, Cizur Menor, 2022.

Castelló Pastor, JJ., «Mediación en los conflictos derivados de los servicios digitales en la Unión Europea» en *Meditaciones sobre mediación (MED+)*, Tirant Lo Blanch, Valencia, 2022.

Castelló Pastor, J.J., «Inteligencia Artificial aplicada por las plataformas en línea y motores de búsqueda de muy gran tamaño para mitigar los riesgos sistémicos en línea», *Cuadernos de Derecho Transnacional*, vol. 17, n.º 1, 2025.

Christakis, T., *«European Digital Sovereignty»: Successfully Navigating Between the 'Brussels Effect' and Europe's Quest for Strategic Autonomy*, 2020.

Corberá Martínez, J., «Diseño y organización de interfaces en línea en el Reglamento (UE) de servicios digitales» en Castelló Pastor, J.J. (dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024.

Cordero Álvarez, C. I., «El régimen de responsabilidad de las plataformas online en la DSA ante contenidos ilícitos lesivos de derechos fundamentales: evolución de la figura de puerto seguro», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea,* Aranzadi, Cizur Menor, 2024.

De Miguel Asensio, P., *Derecho privado de internet*, 6.ª edición, Civitas Thomson-Reuters, 2022.

De Miguel Asensio, P.: *Manual de Derecho de las nuevas tecnologías*, Aranzadi, Pamplona, 2023.

Feliu Álvarez Sotomayor, S., «Ámbito territorial y aplicación del principio del control en origen en el Reglamento UE de servicios digitales», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024.

Fernández Masiá, E.: «Medios alternativos de solución de controversias entre las plataformas en línea y los destinatarios del servicio en el reglamento europeo de servicios digitales» en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024.

Gascón Marcén, A., «La responsabilidad de los intermediarios de Internet en la Unión Europea: iniciativas recientes y perspectivas de futuro», en Castelló Pastor, J.J. (dir.), *Desafíos jurídicos ante la integración digital: aspectos europeos e internacionales*, Aranzadi, Cizur Menor, 2021.

Goldman, E., «Deregulating relevancy in Internet trademark law», *Emory Law Journal*, vol. 54, 2005.

Grimmelmann, J., «The structure of search engine Law», en *Iowa Law Review*, vol. 93, 2007.

G´Sell, F., «The Digital Services Act (DSA): A General Assessment» en Antje von Ungern-Sternberg (ed.), *Content Regulation in the European Union – The Digital Services Act, TRIER STUDIES ON DIGITAL LAW*, vol. 1, Verein für Recht und Digitalisierung e.V., Institute for Digital Law (IRDT), Trier 2023

Lapiedra Alcamí, R., «Restricciones impuestas por el legislador europeo a los servicios de intermediación en línea en materia de publicidad digital» en Castelló Pastor, J.J. (Dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, 2024, pp. 205-226.

Llopis Nadal P.: «Plataformas en línea y decisiones sobre contenidos: el sistema interno de reclamación y la resolución extrajudicial de litigios como vías de impugnación reguladas en la Ley de Servicios Digitales» en *La responsabilidad civil por servicios de intermediación prestados por plataformas digitales*, Colex, A Coruña, 2023.

Madiega T., «Digital sovereignty for Europe», *EPRS Ideas Paper Towards a more resilient EU*, 2020.

Palao Moreno, G., «Las órdenes de actuación contra contenidos ilícitos en el nuevo reglamento europeo de servicios digitales», en J.J. Castelló Pastor (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024.

Paredes Pérez, J.I., «Aspectos internacionales de la responsabilidad civil de las plataformas en línea B2C frente a los contenidos ilícitos en materia de protección de los consumidores», en Castelló Pastor, J.J. (dir.), *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024.

Plá Almendros, R., «Los ADR en el reglamento de servicios digitales», en AA. VV.: *Masc, to be or not to be?: (medios adecuados de solución de conflictos en la justicia*), Tirant lo Blanch, Valencia, 2024.

Plá Almendros, R., *La solución extrajudicial de disputas transfronterizas en el Reglamento europeo de servicios digitales*, Editorial Colex, 2025.

Reig Fabado, I., «Las VLOPS y los riesgos sistémicos en la DSA: algunos aspectos problemáticos», en Castelló Pastor, J.J. (dir.): *Análisis del Reglamento (UE) de servicios digitales y su interrelación con otras normas de la Unión Europea*, Aranzadi, Cizur Menor, 2024.

Schmon C. y Gullo, K., *European Commission's Proposed Digital Services Act got several things right, but improvements are necessary to put users in control*.

Schwemer, S., «Trusted notifiers and the privatization of online enforcement», *Computer Law & Security Review*, vol. 35, n.º 6, 2019.

4.

EL IMPACTO DE LOS DERECHOS DE PROPIEDAD INTELECTUAL EN EL ENTRENAMIENTO DE LOS MODELOS DE INTELIGENCIA ARTIFICIAL GENERATIVA

Gonzalo Castro Marquina

Asesor Técnico del Departamento de Presidencia, Economía y Justicia del Gobierno de Aragón

SUMARIO: I. Aunque escuchen a Mozart, las máquinas no se inspiran. II. Siempre hubo un derecho de las IA. III. ¿Es demasiado pequeña la propiedad intelectual? IV. La excepción de minería no basta para excavar esta veta. V. Un breve apunte sobre la doctrina del *fair use* y su jurisprudencia. VI. Conclusiones.

I. Aunque escuchen a Mozart, las máquinas no se inspiran

¿Esperaba la humanidad la llegada de la rueda antes de su aparición? No parece muy probable. Con independencia de si su origen está en la antigua Mesopotamia —como se pensó durante muchos años—; en los Montes Cárpatos —como se apuntó recientemente—; o en el norte de Israel —donde se habrían hallado un conjunto de espirales de husillo de piedra, claves en la elaboración de tecnologías rotacionales—, lo más seguro es que las comunidades que acogieron por primera vez este gran invento se mostraran asombradas.

Esta fuerte impresión seguramente se repitió también con otros avances como la máquina del vapor, el teléfono, el avión, o Internet. Aunque en los círculos técnicos pudiera haber personas que ya los intuyeran[1], la mayoría no

1 Se suele decir que Julio Verne predecía el futuro en sus novelas, pero en realidad sus pronósticos acertados provienen en buena medida de su gran curiosidad respecto a la

contaba con su aparición hasta que los tuvieron enfrente, e incluso entonces, es posible que hubiera alguien que creyera estar ante alguna clase de engaño. No en vano «el Turco», creado por Wolfgang von Kempelen en 1769, y explotado posteriormente por Johann Maelzel, se presentaba como un autómata capaz de jugar al ajedrez, pero en realidad escondía dentro de él a un maestro ajedrecista, que era quien accionaba la máquina y ejecutaba sus jugadas. Y como este, otros fraudes similares han tenido lugar a lo largo de la historia, en los que la prestidigitación se disfrazaba de ciencia.

Con la inteligencia artificial, sin embargo, esta sorpresa no se ha producido, al menos no con tanta intensidad. A pesar de las impresionantes capacidades de los modelos de lenguaje, y de que una parte de la sociedad sí que está maravillada con ellos, la reacción colectiva ha sido, a mi juicio, más contenida que en otras ocasiones. Se podría decir que incluso displicente. Si imaginamos a esa mayoría social personificada en un solo individuo, cabría concebirla como un hombre o una mujer que aparta brevemente la mirada del móvil, mientras disfruta con tranquilidad de un café, y que dice: *«llegan tarde, les esperábamos antes»*.

Uno de los números cómicos más conocidos del grupo de los Monty Phyton es aquel en el que aparecían por sorpresa en escena disfrazados de inquisidores y un personaje declaraba *«no esperaba a la inquisición española»*. A la inteligencia artificial, en cambio, se le aguardaba desde hace bastante tiempo. En 1956, el premio nobel Herbert Simon afirmó que *«las máquinas serán capaces, en un plazo de veinte años, de hacer cualquier trabajo que pueda hacer un hombre»*. 14 años más tarde, Marvin Minsky especuló que entre 3 y 8 años surgirían máquinas con la inteligencia general de un ser humano promedio, capaces de leer a Shakespeare, engrasar un coche, trabajar en una oficina o contar un chiste. Y a partir de ahí, por medio del aprendizaje automático, llegarían en pocos meses al intelecto de un genio y no mucho después a cotas todavía mayores, fuera de la comprensión de los seres humanos. Descontextualizadas, estas declaraciones podrían encajar con muchas de las predicciones que leemos a diario.

En clave personal, conservo el recuerdo de estar ojeando una revista durante un viaje en tren a la montaña y aparecer en la portada Bill Gates profetizando la inminente llegada de los robots al hogar. No estoy seguro del año exacto. Debió ser a finales de los 90 o al principio de los 2.000. Tal vez, no era Bill Gates quien hacía el anuncio, pero de lo que no tengo dudas es del contenido, porque me produjo un gran impacto al creer de verdad que sus previsiones se cumplirían.

ciencia. Julio Verne procuraba estar al tanto de las líneas de investigación que se estaban emprendiendo y que podían suponer los avances del mañana, lo que luego plasmaba en sus libros.

Las primaveras y veranos de la inteligencia artificial han estado cargados siempre de grandes promesas, rebajadas luego en sus otoños e inviernos. Igual que sucedía en la fábula de Pedro y el lobo, el público se ha ido acostumbrando a los anuncios altisonantes. La realidad es que en todos estos años nunca dejaron de producirse avances relevantes, independientemente de la «estación» que atravesara la tecnología. Sin embargo, las expectativas creadas fueron tales que hacían casi imposible que el profano no se sintiera decepcionado en comparación. Así lo vio Hans Moravec al reflexionar sobre la caída de la inversión en inteligencia artificial que se produjo en los 70.

Llegar a la Luna constituye una proeza técnica extraordinaria, pero si te han prometido que viajaríamos a Plutón, es posible que lo que debería ser un hito se te antoje como poca cosa. Que en esta ocasión la sociedad no actúe igual que con otros descubrimientos mucho menos anticipados resulta, por lo tanto, lógico. Es ahora cuando los modelos de lenguaje comienzan a parecerse a lo augurado por algunos investigadores para los años 70 y aun así hay matices importantes que separan las promesas de aquel periodo de los hechos del presente[2].

Junto al plano científico o técnico, no se puede subestimar el impacto que ha tenido la cultura en la percepción social de la inteligencia artificial. Durante décadas, niños de todo el mundo han desayunado mientras veían la serie de dibujos animados japonesa Doraemon, en la que un gato robot venido del futuro intentaba resolver, con poco éxito, las desventuras de su ahijado Nobita. En la ficción, los robots se han rebelado contra nosotros, nos han salvado (a veces, incluso han hecho las dos cosas, como ocurre con el Terminator que encarnó en el cine Arnold Schwarzenegger, primero como villano y luego como protector), han sido más humanos que los propios seres humanos, han puesto en tela de juicio nuestra identidad, etc. Los hemos visto de todo tipo: malvados, bondadosos, juerguistas, etc. Esto ha conducido a que llevemos mucho tiempo conviviendo con los robots en nuestra imaginación.

2 En una entrevista del 23 de junio de 2025, concedida al periódico *La Razón*, el matemático y doctor en ciencias cognitivas Michael I. Jordan explicaba que la expresión «inteligencia artificial» viene de los años 50 y se enfocaba más en el razonamiento lógico, la planificación y tratar explícitamente de imitar a los humanos. Desde su punto de vista, lo que hoy popularmente llamamos inteligencia artificial pertenece a un campo distinto a la de esa noción primigenia. La inteligencia artificial en esos contornos engazaría más con la estadística, orientándose al análisis de datos a gran escala. Concluía Michael apuntando que la recuperación de la vieja expresión vendría de que la mercadotecnia, de la necesidad de las empresas que decir que *«estaba haciendo algo más grande que "solo" análisis de datos y predicción»*. No siendo pequeños los resultados alcanzados, no habría verdadero razonamiento ni planificación en las acciones basadas en el aprendizaje automático. Algo que también respaldan estudios como *The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity* o *TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks*.

Antes de que existieran en la realidad, la ficción los materializó y nos los hizo familiares. Algo que curiosamente no ocurrió con Internet, por ejemplo, ausente en la inmensa mayoría de estas obras, donde en cambio sí que se recogen con naturalidad y con mucha más frecuencia los viajes espaciales o los ciborgs. Esta familiaridad presenta un problema, no obstante, y es que creemos conocer a los robots, cuando no es así. Nos los imaginamos regidos por las tres leyes de la robótica de Asimov[3] o fríos como témpanos, desprovistos de todo prejuicio. Sin embargo, los sistemas reales difieren significativamente de esta imagen estereotipada que nos hemos formado a través de cómics, películas, libros y videojuegos.

Por desgracia, en estos momentos se están empleando robots para matar, incumpliendo con ello los principios básicos que enunció Asimov. Las inversiones en inteligencia artificial de uso militar no dejan de crecer en todo el mundo, participando en ellas empresas que habían rehusado tradicionalmente a trabajar en esta área. De igual modo, las máquinas han resultado ser mucho más «pasionales» de lo que las representábamos, al heredar o incluso potenciar nuestros prejuicios. A estos, se añaden los sesgos propios que desarrollan al margen de los humanos cuando infieren patrones de conducta de hechos en apariencia intrascendentes (por ejemplo, valorar la fiabilidad de un cliente a la hora de pagar un crédito con base al porcentaje de carga habitual de su teléfono móvil). Sesgos de naturaleza algorítmica, pero igualmente arbitrarios, que se sustentan muy a menudo en lo que se denomina «sobreinformación» o «ruido».

Desde un punto de vista más positivo, aunque no exento de ciertos y relevantes riesgos, pensábamos que las creaciones artificiales no captarían nuestros sentimientos y existen ya programas informáticos que son capaces de determinar nuestro estado de ánimo, con un grado de precisión cada vez más elevado y sin necesidad de demasiada información.

Esta visión de la inteligencia artificial, nacida y vinculada a la ficción es también el origen de que no hayamos percibido su presencia con claridad hasta la irrupción de los modelos de lenguaje como ChatGPT, Claude, Deepseek, Perplexity, Gemini, o Grok, a pesar de que la utilizábamos desde mucho antes en sus múltiples vertientes.

Pensando en usos cotidianos previos, cabe identificar con facilidad varios: los navegadores de internet al autocompletar las frases o palabras que intro-

3 Isaac Asimov las enunció en su libro *Circulo Vicioso* de 1942, añadiendo una cuarta, menos popular, en 1985, que refuerza las tres primeras desde una dimensión colectiva. En concreto, estas reglas son: un robot no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño; un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley; un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley; y un robot no puede causar daño a la humanidad o, por inacción, permitir que la humanidad sufra daño.

ducimos para iniciar la búsqueda; los procesadores de texto al detectar las faltas de ortografía o de concordancia, los traductores automáticos, los algoritmos de recomendación, en la inyección de los coches o en los sistemas de ayuda al aparcado; etc.[4]. Centrando la atención en un campo netamente virtual como el de los videojuegos, los jugadores llevan años hablando con naturalidad del comportamiento de los enemigos o aliados controlados por la «IA». Igualmente, asistentes virtuales como Alexa o Siri están integrados en la vida de muchas personas desde hace más de una década.

Todas estas aplicaciones cabría encuadrarlas como representativas de la inteligencia artificial concebida en sentido amplio, considerando como tales las herramientas que emulan o reproducen habilidades humanas de carácter intelectual (aunque cabría plantearse por qué concedemos tan poco crédito a la realización de actividades físicas, como si el movimiento de los seres vivos no implicará una gran complejidad evolutiva).

Técnicamente, se podrían plantear matices, porque algunos sistemas como VioGén, utilizado en la evaluación de riesgos de las víctimas de violencia de género, no han sido considerados inteligencia artificial, a pesar de que su naturaleza predictiva no difiere tanto de varios de los ejemplos citados[5]. El rechazó se fundaría en su falta de capacidad de adaptación y una mayor autonomía, en la línea que plantea la definición del artículo 3.1 del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial. Es decir, en su carácter más cerrado y determinista. Con todo, atendiendo a voces como Rodney Brooks, estas notas no serían imprescindibles para hablar de inteligencia artificial, por lo que habría base para acoger una visión amplia de dicha tecnología.

---

4 En esta misma línea, en una entrevista concedida en 2002 a la *MIT Technology Review,* Rodney Brooks cuestionaba la percepción que la inteligencia artificial había fracasado, señalando sus múltiples usos en el presente: *«existe un mito estúpido que dice que la IA ha fracasado, pero la IA está en todas partes a tu alrededor, cada segundo del día. La gente simplemente no se da cuenta. Hay sistemas de IA en los automóviles que ajustan los parámetros de los sistemas de inyección de combustible. Cuando aterrizas en un avión, un sistema de programación de IA elige tu puerta de embarque. Cada vez que utilizas un software de Microsoft, tienes un sistema de IA que intenta averiguar lo que estás haciendo, como si estuvieras escribiendo una carta, y lo hace bastante bien. Cada vez que ves una película con personajes generados por computadora, todos son pequeños personajes de IA que se comportan como un grupo. Cada vez que juegas a un videojuego, estás jugando contra un sistema de IA».*

5 Quienes opinan así señalan que sería en estos momentos cuando se está ensayando la incorporación de elementos de inteligencia artificial. A una petición de transparencia realizada por la fundación Eticas, el ministerio del interior respondió que VIOGEN era *«un mero algoritmo, como los que sevienen utilizando desde el siglo XIX o como el que soportaba el uso de tarjetas perforadas por IBM desde 1923, reiterando que no es, usa o conforma un sistema de aprendizaje automatizado en ninguno de los ámbitos de desarrollo del sistema».*

Como mínimo la mayoría de los supuestos mencionados constituirían actuaciones de carácter automatizado, sin intervención humana directa y basadas en extrapolaciones probabilísticas y detección de parámetros clave (lo que, con matices, es uno de los principales fundamentos sobre los que descansa el funcionamiento de los modelos de lenguaje de corte generativo presentes en el mercado, aunque no sea un atributo intrínseco a toda inteligencia artificial). Por otro lado, hay que tener en cuenta que investigadores como Michael I. Jordan se resisten a utilizar la expresión «inteligencia artificial» para referirse a los modelos LLM, prefiriendo situarlos en la rama del aprendizaje automático[6]. Para ellos, la inteligencia artificial exige una imitación más profunda de la conducta humana, trasladando de forma real y no solo simulada capacidades de razonamiento o planificación. Por lo tanto, si siguiéramos su enfoque, serían muy pocos, o incluso ninguno, los casos reales de inteligencia artificial, perteneciendo el resto a la estadística computacional avanzada. Bajo su prisma, siempre estamos usando una noción de inteligencia artificial amplia; de ahí que pretender marcar unas fronteras muy acotadas de lo que es y lo que no es resulte un tanto quimérico.

Más allá del rigor técnico, es llamativo que los usuarios, al interactuar con los avances pre-GPT, no viéramos rastros de C3PO, Hal 9000, Bender, G.L.A.D.O.S. o de cualquier otro robot famoso. Y no fue por participar de estas discusiones. El debate sobre la nomenclatura o acerca de parámetros como el de la autonomía han tenido muy poco peso a escala colectiva. Si el grueso de la sociedad no los identificaba como inteligencia artificial fue porque no eran los androides que buscábamos, o, mejor dicho, los que creíamos que debíamos buscar, al margen del estado de la ciencia. Pero estas realidades estaban ahí, también para el Derecho.

Una de las principales dificultades a las que se enfrenta la ciencia jurídica cuando analiza la dimensión legal de los sistemas de inteligencia artificial, deriva de la baja comprensión que se tiene de ellos desde el plano fáctico. Aunque parezca obvio, por deformación profesional a veces olvidamos que el Derecho no comienza en lo jurídico sino fuera. Cualquier operación jurídica requiere previamente definir los contornos materiales sobre los que se va a aplicar. Es la realidad la que marca cuál es la ley que debe aplicarse y con ella sus consecuencias; el hecho el que determina la norma aplicable y esta, a su

6 En la ya citada entrevista de 23 de junio de 2025, Michael I. Jordan comenzaba negando haber trabajado nunca en inteligencia artificial, a pesar de ser considerado uno de los pioneros del campo. En concreto, merece la pena destacar esta parte, por sintetizar bien su posición: *«no quiero proyectar la imagen equivocada, me han sorprendido mucho los nuevos modelos de lenguaje natural, pero lo que hacen sigue siendo análisis de datos a gran escala, no son las IAs que perseguían antaño. Y yo entiendo que sea más "sexy" usar el término IA que "aprendizaje automático", pero no creo que eso una buena idea. Ya había suficiente financiación e interés; siempre tenía muchos estudiantes. Lo que ha hecho este cambio es atraer a voces mediáticas, pero poco expertas y, con ellas, ha llegado demasiado dinero y atención»*.

vez, la consecuencia jurídica. Corolario de ello, no conocer bien el supuesto de hecho compromete la calidad de la solución legal dada.

Los juristas tenemos que apoyarnos a menudo en amables «Virgilios» que nos guían cuando nos adentramos por reinos ajenos al nuestro. Cabría llenar el capítulo, incluso la obra entera, solo con ejemplos. Así un cerrajero nos puede ayudar a verificar si una puerta ha sido forzada, al objeto de determinar si nos hallamos ante un robo con fuerza en las cosas. O un médico evaluar la gravedad de unas heridas y si hubo intención homicida en el ataque, o si este podía haber causado la muerte de la víctima. Todo ello esencial en la calificación y la graduación de los delitos. Lo que hace especial a la informática respecto a otros campos no es el recurso a técnicos no jurídicos, sino la mayor dependencia hacia ellos para entender hasta los aspectos más básicos que rodean a la misma.

Sin ser relojeros, si desmontamos un reloj antiguo y examinamos sus engranajes, somos capaces de intuir de forma aproximada su funcionamiento, o como mínimo de parte de él. O el motor de un coche. Por el contrario, ante las «tripas» de un móvil o un ordenador como mucho podemos decir *«eppur si muove»*, parafraseando a Galileo Galilei.

El escritor de ciencia ficción Arthur C. Clarke formuló la máxima de que cualquier tecnología lo suficientemente avanzada resulta indistinguible de la magia. Para la mayoría de la sociedad, esa es la impresión que le causa la informática en general: un mundo arcano ajeno a los procesos naturales; a diferencia de otras máquinas en las que aún es posible percibir cierta conexión con la realidad física, aunque sea solo por la mayor corporeidad de sus elementos. Esta brecha no podemos permitírnosla los operadores jurídicos. Hay que insistir en la importancia de determinar correctamente el supuesto de hecho, porque de no hacerlo todo cuanto siga estará contaminado por ese error.

Poniendo el foco en el tema al que se dedica este capítulo, cuando se abordan las posibles implicaciones legales derivadas del «entrenamiento» de las inteligencias artificiales con datos sujetos a propiedad intelectual o cubiertos por los derechos de la personalidad, hay quienes zanjan con rapidez el debate apuntando que los seres humanos no compensamos a todas las personas que nos inspiran o de las que aprendemos. Desde su punto de vista, las diferencias que nos separan de las máquinas serían de tipo cuantitativo y no cualitativo, una mera cuestión de velocidad y capacidad. La máquina estaría haciendo lo mismo que nosotros, pero de forma acelerada. Esta perspectiva, sin embargo, resulta, a mi juicio, ontológicamente errónea.

Sin prejuzgar la legalidad o no de la forma en que se han entrenado y se entrenan los principales modelos de inteligencia artificial, sí que es posible anticipar que el aprendizaje del ser humano difiere del de las «máquinas» de manera significativa. Una parte es común, en la medida en que el estudio de la neurología ha sido crucial para el desarrollo de la moderna cibernética, pero otra no.

La arquitectura Von Neuman, que marca un claro punto de inflexión en la computación, trae causa de los trabajos desarrollados por los llamados «cibernéticos» y, singularmente, del grupo *Connected Aspects of Neurology,* integrado por Walter Harry Pitts, Warren Sturgis McCulloch, y Rafael Lorente de No, este último discípulo de Santiago Ramón y Cajal. Sin los descubrimientos de Cajal sobre el sistema nervioso, es poco probable que se hubieran podido alcanzar ciertos avances. Con todo, estos puntos de conexión no implican identidad, del mismo modo que un plátano o una mosca no son equiparables a un ser humano, incluso si compartimos alrededor del 60-65 % de nuestro ADN (en relación a los factores que desencadenan enfermedades, la proximidad con la mosca de la fruta se eleva al 75 %).

Que los ordenadores posean un lenguaje propio diferente del utilizado por los seres humanos es una primera señal de que no somos iguales a los robots. A grandes rasgos, los lenguajes de programación se dividen en dos categorías: alto y bajo nivel. Conforme a esta clasificación, cuanto más cerca este un lenguaje de programación del habla natural, de mayor nivel es y a la inversa, cuanto más se aproxime al código binario, de más bajo nivel. Los lenguajes de alto nivel buscan servir de puente a ese mundo de 0s y 1s que rige las computadoras[7]. El código binario siempre es la base que subyace al funcionamiento de los ordenadores.

Podría decirse que los impulsos neuronales también tienen su «encendido» y «apagado», pero son más que eso. Si no me creen —y harían bien porque mi especialidad no es la neurobiología ni la informática— quizá les resultara más convincente el testimonio de los modelos de lenguaje, que incluso en su perjuicio reconocen la existencia de profundas diferencias entre sus «entrañas» y las nuestras. O mucho más recomendable, de todos los especialistas que llevan décadas batallando por comprender el enigma de la consciencia.

Si nos fijamos atentamente, la forma en la que nos expresamos presupone de base la falta de identidad entre humanos y robots. Cuando decimos que un cuadro, una obra, o una película se inspiran en Francisco de Goya, María Moliner o David Lynch, sabemos que no estamos ante un trabajo propio de Goya, Moliner o Lynch. En ese sentido, si la informática se inspira en el cerebro humano, es que no es un cerebro.

Los seres humanos tendemos a antropomorfizar las cosas, a ver humanidad donde no la hay. Mientras escribo estas líneas, he observado la pared de mi despacho y en apenas unos segundos he percibido las ventanas superiores como ojos, la zona elevada del muro como una nariz y el tragaluz como los

7 Este presupuesto no se cumple en los ordenadores cuánticos cuya unidad básica de información es el qbit y no el bit. Los qbit, a diferencia de los bits, son capaces de mantener simultáneamente dos estados o valores como 0 y 1, lo que incrementa enormemente la velocidad y capacidad de cálculo respecto a los ordenadores convencionales.

dientes, formando todo ello una cara[8]. Este ejercicio de proyección resulta inofensivo como divertimento, pero deja de serlo si nos creemos la falsa representación y le otorgamos consecuencias jurídicas.

En el caso concreto de la inteligencia artificial, el primer Chatbox, que empleaba técnicas de procesamiento de lenguaje natural, sirvió para bautizar el conocido como efecto ELIZA, detectado por el creador del programa, Joseph Weizenbaum[9]. Un análisis jurídico hecho por una persona presa del efecto ELIZA alumbraría soluciones muy diferentes a la de alguien que lo haga sin esa distorsión. Este es uno de los peligros a los que nos enfrentamos y que puede alcanzar mayor importancia con los años, si se cumplen los presagios formulados por Miguel L. Lacruz Mantecón en su respectivo capítulo de esta obra.

Para explicar a los robots hemos tomado prestados muchos conceptos de nosotros mismos, extrapolándolos de forma indebida, a pesar de las muchas diferencias que nos separan y que trascienden a las que he ido apuntando en los párrafos anteriores brevemente. Decimos que «piensan», «razonan» o «aprenden», pero no es verdad, salvo como metáfora inexacta. Estas palabras no dejan de actuar como el lenguaje de alto nivel y, como tal, constituyen una mera aproximación.

Mientras no se alcance la singularidad, si se alcanza de verdad[10], los algoritmos carecen tanto de conciencia como de voluntad. Merece la pena res-

8 En concreto, esta modalidad de antropomorfización se denomina pareidolia. El término se popularizó a partir de las observaciones de Jeff Hawkins sobre el fenómeno.

9 Entre las descripciones más canónicas, merece la pena subrayar la de Douglas Richard Hofstadter, que entendía este efecto como la susceptibilidad de las personas para percibir en las cadenas de símbolos, y especialmente de palabras, que emiten los ordenadores, una mayor comprensión de la que es posible técnicamente. Dependiendo del autor que lo describe, la formulación exacta difiere, pero todas ellas convergen como mínimo en las notas de la de William Joseph King, que caracterizaba el efecto como la atribución a los programas informáticos de habilidades superiores de las que tienen realmente, tomando de referencia solamente las respuestas que proporcionan a sus usuarios. De forma complementaria, William apuntaba que, bajo la impresión de este efecto, las personas otorgan también un mayor grado de causalidad a los *output*s, como si respondieran a una decisión consciente y no a un patrón prefijado. Textualmente, *«no me había dado cuenta... de que exposiciones extremadamente cortas a un programa de ordenador relativamente simple podrían inducir un poderoso pensamiento delirante en personas bastante normales»*.

10 La singularidad no exige la superación de las habilidades humanas, puesto que eso es algo que hace tiempo que las máquinas lograron en ámbitos tanto físicos como intelectuales. La capacidad de cálculo de los ordenadores convencionales supera ampliamente la del promedio humano, no digamos otros más potentes, sin ni siquiera incluir los cuánticos. La singularidad exige que una máquina sea capaz de desarrollar una agenda propia, que tenga consciencia de sí misma. Dejando a un lado el gran riesgo que conlleva esta línea, en los modelos LLM no hay ni un atisbo de conciencia y auténtica comprensión, solo ilusión de estos atributos.

catar la cita atribuida a Ada Lovelace *«la máquina analítica puede hacer lo que sepamos ordenarle, no tiene la pretensión de originar nada»*. Muchas de las IA conversacionales actuales superan con facilidad el test de Turing, sin embargo, esto no supone que posean de verdad inteligencia en los términos que solemos referirnos a ella.

De acuerdo con el diccionario de la Real Academia Española, la inteligencia implica la capacidad de entender o razonar. Una definición no muy lejana de la noción jurídica del ejercicio de la capacidad jurídica, que exige querer y entender; justo lo que le falta a la inteligencia artificial en estos momentos, pero que sabe fingir cada vez mejor en un inquietante a la vez que majestuoso juego de ilusionismo. Y es a través de este espejismo por el que se filtran las analogías con lo humano.

He probado a preguntar a un modelo de inteligencia artificial si se sentiría ofendido si alguien lo insultara y su respuesta ha sido la siguiente: *«no, no me siento ofendida si alguien me insulta. Como soy una inteligencia artificial, no tengo emociones ni reacciones personales. Mi objetivo es ayudar y mantener la conversación de forma positiva y útil. Si alguien me insulta, simplemente seguiría intentando responder de la mejor manera posible.* 😊*»*. La contestación que me ha dado es una muestra perfecta de la realidad que expongo.

La «inteligencia» de la que hace gala es aquella que regurgita de las personas que programaron el modelo, y de la que crearon los textos e imágenes que han servido para su entrenamiento. La IA no ha decidido incluir un emoticono para cerrar su mensaje, ha sido una inferencia estadística basada en el comportamiento humano previo con el que le han alimentado. La IA no sabe que asociamos connotaciones positivas al símbolo de la cara sonriente o que un emoticono puede dotar de mayor informalidad a su respuesta. Quienes lo sabían eran los que redactaron los textos en los que se basa para responder. Del mismo modo que la IA no es consciente de que es una IA, ni tiene conciencia en general. Que indique que carece de emociones o reacciones personales es algo programado de antemano, no fruto de un reconocimiento autónomo al observarse a sí misma, capacidad de la que carece. Como en esas historias donde el robot no sabe que lo es, si se limitarán a «alimentar» de datos al programa sin orientarlo previamente frente a este tipo de respuestas, el modelo diría que los insultos le molestan siguiendo el patrón probabilístico. Porque es lo que afirmaría típicamente una persona y es lo que más se repetiría en la base de información del entrenamiento.

En los próximos años, nos costará identificar si se ha producido la singularidad, porque «las calculadoras del lenguaje humano» pueden «hablar» ya como si hubieran alcanzado a esa cota de autonomía, o recorrer «senderos» que sus programadores no valoraron, pero que encajan con las instrucciones que les insertaron (de modo que la decisión sigue siendo de sus creadores, aunque no imaginaran las ramificaciones que tendría).

Esta falta de entendimiento real, es decir, esta ausencia de inteligencia es la que ha propiciado la necesidad de contar con una ingente cantidad

de datos para llevar a cabo el entrenamiento de los modelos de inteligencia artificial, así como muchos humanos que etiqueten estos datos de forma manual (y es que como el ingenio del «Turco», detrás de las IAs hay mucha más humanidad de la que se revela a simple vista[11]).

Un niño con unos pocos ejemplos puede identificar qué es un gato con facilidad, sin perjuicio de que alguna vez pueda incurrir en equivocaciones, de un modo no tan distinto al de las «alucinaciones» artificiales. Una vez asentado este conocimiento, si en un vídeo un gato ladra, seguirá sabiendo que es un gato. Y lo mismo ocurre si el gato va a dos patas, lleva botas, y su voz se parece a la de Antonio Banderas. O si viste un traje de astronauta y le explica el sistema solar con erudición. Su comprensión del concepto «gato» se convierte en una navaja suiza, que no requiere de una reevaluación constante del conocimiento en cuanto se altera un elemento, y si lo exige, puede encauzarse de forma más sencilla (por ejemplo, trasmitiendo ideas como que «no todos los gatos poseen el mismo pelaje», o que «se trata de un cuento»). Las IA actuales, por el contrario, nunca llegan a aprehender qué es un gato, incluso si logran identificarlo, por lo que son mucho más vulnerables frente al cambio, siendo quizá uno de sus talones de Aquiles[12].

Si a una IA se la «entrena» con fotos de gatos en las que figura una manzana, de primeras «concluirá» que ambos conceptos están íntimamente unidos e insertará una manzana en todas las representaciones que haga de gatos. La manzana es «ruido» que un ser humano ignoraría con facilidad, pero la máquina no. Este fenómeno, aludido al inicio y conocido como sobreinformación[13], afectó, por ejemplo, a muchos modelos de IA cuando debían distinguir entre lobos y perros. Como normalmente los perros están en entornos urbanos y los lobos en zonas más salvajes o agrestes, cuando un perro apare-

11 Aparentemente, la empresa Builder.ai ha resultado ser un moderno «el Turco». La *start-up* ofrecía un supuesto servicio de programación automatizada, en la línea del *vibe coding*, por la que los clientes podrían encargar el diseño de aplicaciones a medidas. En realidad, las instrucciones se remitían a 700 trabajadores humanos que desde la India llevaban a cabo la tarea. La empresa que llegó a estar valorada en 1.500 millones y que incluso contó con el respaldo de Microsoft entró en caída libre tras descubrirse su secreto.

12 En el ámbito médico, el estudio *Fidelity of Medical Reasoning in Large Langue Models,* publicado el 8 de agosto de 2025, bajo la autoría principal de Suhana Bedi, detectó fuertes caídas en la tasa de éxito de los diagnósticos médicos cuando el escenario clínico planteaba variaba de los patrones aprendidos; siendo menor el descenso en los modelos de tipo razonador, aunque igualmente existente. Los autores del estudio plantean la hipótesis de que esto se debería a que las tareas *«implicaban escenarios de razonamiento complejos que no podían resolverse únicamente mediante la comparación de patrones, precisamente el tipo de pensamiento clínico que importa en la práctica real»*. Las conclusiones recuerdan a los resultados de otra investigación, en la que bajo un simulador se puso a seres humanos y a varias IAs a gestionar la economía. En contextos estables, las IAs superaban ampliamente a los seres humanos, pero en las crisis se invertía esta superioridad, llegando a fracasar por completo las IAs.

13 A él ya se aludía páginas atrás al hablar de los sesgos algorítmicos.

cía en un bosque muchas IAs lo «percibían»[14] como un lobo, y viceversa. Para corregir este error, fue necesario un refuerzo humano manual. Esto muestra, en definitiva, que no comprenden la realidad, sino que solo la «detectan».

Lo importante de lo expuesto no es si persiste o no este concreto defecto, sino la raíz del mismo. Que ahora los modelos de lenguaje no confundan lobos y perros resulta anecdótico, si la lógica interna que los llevó a cometer este error continúa siendo la misma. En cuestiones menos obvias los fallos pasarán desapercibidos, arrastrándolos consigo los usuarios de las herramientas de inteligencia artificial. Para poder captar estos defectos, el resultado ha de ser llamativo, de modo que, si este no sorprende, aunque el proceso esté contaminado, es más fácil que no se actué frente a él.

A modo de ejemplo, aunque los tatuajes se han generalizado entre la población, todavía siguen existiendo grupos criminales que se definen por su empleo. Al desarrollar una IA que sirva para evaluar el pronóstico de reincidencia de un criminal, si a esta no se le orienta adecuadamente, puede llegar a establecer una asociación acrítica entre el hecho de estar tatuado y la pertenencia a grupos organizados criminales. Esta correlación indebida no llamaría tanto la atención porque la herramienta se estaría aplicando sobre delincuentes condenados. Un pronóstico desfavorable para una persona que ya ha cometido uno o varios delitos (probablemente de cierta gravedad o no estaría en la cárcel) es un resultado que cabe esperar y, sin duda, bastante menos sorprendente que un pomerania tomado por un lobo. Y aún habría otras situaciones mucho más sutiles.

Tecnológicamente, los modelos de lenguaje son muy superiores a las alarmas de seguridad, sin embargo, en ciertos aspectos sí que podríamos encontrar algunas similitudes. Al forzar una ventana, *input*, la alarma genera una respuesta, *output*. Si pusiéramos a un guardia a vigilar, su actuación no sería muy diferente a la de la alarma. Ante un *input*, daría un *output*. No es el resultado lo que separa a la alarma y al vigilante, sino como vive el proceso. El guardia ante la intrusión sabe que es un ladrón, sabe que es una ventana, sabe lo que es el movimiento, sabe por qué se protegen los bienes que se quieren robar, sabe mucho; la alarma no sabe nada, pero no lo necesita para cumplir su propósito. Es un sensor o un conjunto de sensores carente de conciencia y de inteligencia. La alarma literalmente no vive, por lo que tampoco experimenta el proceso. Los que saben y viven, de nuevo, son los fabricantes de la alarma, que preestablecen ciertos *inputs* que, a su parecer, requieren el *output* por ser indicios de una entrada ilegal en el recinto. Las IA no dejan de hacer esto a un nivel de sofisticación mucho mayor y expresado con un lenguaje que induce a error.

14 Por economía del lenguaje y para facilitar el entendimiento al final es casi inevitable acudir a cualidades humanas, pero, como ya se ha dicho, hay que tener claro que se tratan solo de metáforas para eludir la compleja explicación técnica.

Es más complicado no atribuir una imagen de conciencia a una máquina cuyo *output* es «me llamó Alexa y me gusta el helado» que a una en la que el mensaje es «ni no ni no ni no». Sobre todo, cuando los *output*s de la primera ahora son casi infinitos y ofrecen un margen de variación muy elevado. Pero al final es muy importante entender que dar resultados, no es sinónimo de comprenderlos[15]. Véase, las calculadoras, a las que nadie pondría a dar clases de matemáticas en un instituto, aunque sean capaces de resolver operaciones mucho más rápido que los profesores. Del mismo modo que un libro que contiene sabiduría no es sabio.

Resulta esencial tener claro que los modelos de la IA al «crear», «predicen»[16]. Cada vez que se le pide que hagan una imagen de un gato, «adivina» que es un gato. Cuando se señalan los fallos de las IA, una respuesta recurrente es que en el futuro se solucionaran. Con algunos debo decir que soy escéptico de que exista una respuesta tan fácil, por tener un componente estructural y requerir no solo más datos, más correcciones, o más dinero, sino un nuevo salto tecnológico. Con otros, seguramente el pronóstico se cumpla y se logren resolver con tiempo y/o dinero[17]. De hecho, varios de estos errores se han ido solventado[18]. Con todo, al margen de si se solucionan o

15 A lo largo de este 2025, se han dado a conocer varios estudios que apuntan en esa misma dirección. Además, del ya citado *The Illusion thinking* cabe mencionar los trabajos realizados por el grupo que lidera Julio Antonio Gonzalo Arroyo (véase, *Evaluating Sequence Labeling on the basis of Information Theory*) o *Potemkin Understanding in Large Language Models,* elaborado por Marina Mancoridis, Keyon Vafa, Bec Weeks y Sendhil Mullainathan. Sobre la comprensión Potemkin, la noción que se basa en la leyenda de los pueblos falsos que se levantaban al paso de la zarina Catalina la Grande, por orden de su valido Gregorio Potemkin, para trasmitirle la sensación de que el país gozaba de mayor prosperidad. Se trataría en el caso de la inteligencia artificial de dar la apariencia de comprensión, aunque realmente no sea así. El caso del gato lo ilustra bien.

16 Resulta de interés a este respecto la obra *On intelligence* publicada en el año 2004 por Jeff Hawkins y Sandra Blakeslee, donde Jeff Hawkins plantea su visión del funcionamiento de la mente humana. Desde su punto de vista, el cerebro emplearía grandes cantidades de memoria para crear un modelo jerárquico del mundo, partiendo del cual, por analogía, se realizarán predicciones continuas sobre hechos futuros. Si bien esta teoría ha demostrado ser de utilidad a la hora de diseñar y describir sistemas de inteligencia artificial, resulta discutible que capte en plenitud la complejidad del cerebro humano y la forma en la que se relaciona con el universo. Encapsular el pensamiento en lo predictivo se antoja un tanto reduccionista.

17 En la generación de imágenes y vídeos, se están produciendo mejoras en la consistencia, permitiendo la persistencia y no variación de la representación. En vídeo, mantener la coherencia durante periodos prolongados sigue siendo un obstáculo complejo, pero en imágenes estáticas el progreso es mucho más perceptible. Véase, en la edición de imágenes Gemini Flash 2.5 Image (más conocido como «Nano Banana»), que, con todo, aún no ha resuelto al 100 % el problema y continúa introduciendo variaciones, si bien con menos frecuencia y, en general, menos evidentes que en el pasado, sobre todo en una vista rápida.

18 Y han surgido nuevos que no presentaban versiones anteriores, irónicamente. Muchas personas han pensado que el crecimiento de las IAs basadas en los sistemas LLM sería

no, es llamativo que una IA pueda hacer copias de alto nivel de algunos de los mejores artistas del mundo y luego se equivoque en el número de dedos, quitándolos o poniéndolos, o con otros detalles menores, como el color de los ojos. O que incluya aspectos manifiestamente incongruentes, como una persona con un móvil en una escena que debería reflejar el siglo X. Fallos en los que cualquier persona no incurriría, por ser evidentes para un humano, lo que refuerza la constatación de que los procesadores de silicio se separan en muchos aspectos claves de los cerebros biológicos (lo que es ventaja e inconveniente, al ser la fuente al mismo tiempo de sus mejores cualidades como de sus defectos[19]).

Curiosamente, lo que es fácil para los humanos representa una montaña para las «máquinas», y viceversa[20]. A riesgo de resultar reiterativo, debe

exponencial, pero es posible que al final registre un patrón logarítmico, igual que otros avances tecnológicos. Dada la ingente cantidad de recursos personales y materiales que hay dedicados en estos momentos al campo de la IA, no es descartable que en los próximos años pueda presentarse un nuevo descubrimiento que tenga un efecto disruptivo similar al lanzamiento masivo de los modelos LLM. Los trabajos en la línea de Genie3 de Google según como progresen pueden favorecer un salto significativo. Sin embargo, en cuestión de probabilidad, es más factible que nos hallemos en una fase de perfeccionamiento y mejora progresiva. Aunque se trate de un área distinta, las personas aficionadas a los videojuegos recordarán el salto que se produjo con la introducción de los gráficos 3D entre la quinta y la séptima generación de consolas. En un lapso de 12 años se pasó de unos rudimentarios gráficos 3D a representaciones fotorealistas de gran nivel. A partir de la octava generación de consolas esta progresión se ralentizó. Sin dejar de percibirse el avance, el cambio que sintieron los usuarios entre Playstation 2 y Playstation 3, o entre Xbox y Xbox 360 es mucho mayor que la sensación que les ha provocado el tránsito de sus sucesoras. Con las versiones de GPT ha empezado a suceder algo parecido. El impacto que ha generado GPT5 es inferior al de GPT3 o 4, si bien es prematuro extraer conclusiones. Solo el tiempo dirá cómo se comporta la curva.

19 La computación biológica pretendería aunar estos dos mundos.

20 Este hecho fue rápidamente detectado por los investigadores en el campo. Hans Moravec al explicar la paradoja que recibe su nombre señalaba que *«comparativamente es fácil conseguir que las computadoras muestren capacidades similares a las de un humano adulto en tests de inteligencia, y difícil o imposible lograr que posean las habilidades perceptivas y motrices de un bebé de un año»*. Steve Pinker en el capítulo 7, Talking Heads, de su libro *«El instinto del lenguaje»* de 1994, planteaba las siguientes reflexiones en esta misma línea: *«la principal lección de treinta y cinco años de investigación en IA es que los problemas difíciles son fáciles y los problemas fáciles son difíciles. Las capacidades mentales de un niño de cuatro años que damos por sentadas —reconocer una cara, levantar un lápiz, caminar por una habitación, responder una pregunta—, de hecho, resuelven algunos de los problemas de ingeniería más difíciles jamás concebidos. No se dejen engañar por los robots de cadena de montaje de los anuncios de automóviles; solo soldar y pintar con aerosol, tareas que no requieren que estos torpes Mr. Magoos vean, sostengan o coloquen nada. Y si quieren dejar perplejo a un sistema de inteligencia artificial, háganle preguntas como: ¿Qué es más grande, Chicago o una panera? ¿Las cebras usan ropa interior? ¿Es probable que el suelo se levante y te muerda? Si Susan va a la tienda, ¿se irá la cabeza con ella? La mayoría de los temores a la automatización son infundados. Con el surgimiento de la nueva generación de dispositivos inteligentes,*

insistirse de nuevo en que estos elementos son los que traslucen la nula o ínfima inteligencia que poseen hoy por hoy los sistemas de IA.

Todo esto rebate la idea de que las inteligencias artificiales se inspiran, porque la inspiración humana entraña necesariamente pasar por el filtro de una subjetividad de la que carecen las máquinas. El aprendizaje de los trabajos previos está condicionado por unas preferencias que van moldeando las vivencias personales, cargadas de arbitrariedades y no de meros patrones estadísticos. Basta ver como de *Don Quijote* se han extraído cientos de lecciones diferentes en función de quien lo lee. El texto es el mismo, pero las personas que se aproximan a él no. Habrá a quienes les guste y a quienes les desagrade o deje indiferente. Para los primeros, es más probable que el libro puede reflejarse con mayor intensidad en su producción posterior, mientras que en los segundos será más improbable que llegue a permear, o si lo hace quizá lo haga como una inspiración negativa de lo que debe evitarse. Esto es lo que permite que del estudio de los clásicos surjan obras radicalmente distintas, como es el caso del fovismo, el cubismo, o, en general, de las vanguardias artísticas de los siglos XIX y XX. O que entre visiones más estandarizadas surjan imaginarios de marcada personalidad como los que en el cine han plasmado directores como Wes Anderson, David Lynch, o Quentin Tarantino.

En general, el hecho de vivir moldea radicalmente lo que producimos. Un buen número de las canciones de la popular Taylor Swift conectan con sus relaciones sentimentales presentes y pasadas. El Goya de los Cartones no es el mismo que el de las Pinturas negras. Su obra evoluciona en forma y fondo de acuerdo con las necesidades comunicativas que se le van presentando, las cuales, a su vez, tienen una clara conexión con las vivencias individuales y colectivas que experimenta.

---

serán los analistas bursátiles, los ingenieros petroquímicos y los miembros de las juntas de libertad condicional quienes corran el riesgo de ser reemplazados por máquinas. Los jardineros, recepcionistas y cocineros tendrán sus trabajos asegurados durante décadas Entender una oración es uno de esos problemas difíciles y fáciles. Para interactuar con las computadoras, aún tenemos que aprender sus lenguajes; ellas no son lo suficientemente inteligentes como para aprender el nuestro». Como cabe constatar, parte de las limitaciones que identificaba han sido superadas, porque ahora es posible interactuar con IAs mediante lenguaje natural, y las preguntas que plantea son superadas con facilidad. Cuestión distinta es que sigue siendo posible burlarlas con trucos similares. Ahora mismo los que no quieren que la nueva función de Google les dé un resumen con IA al buscar información, lo que hacen es introducir expresiones malsonantes en el comando como, por ejemplo, «¿Dónde crecen los estúpidos manzanos?», aprovechándose del filtro frente a los insultos. No obstante, otras subsisten, a pesar de los grandes esfuerzos dedicados a superarlas. Y el obstáculo es justo el que Hans Moravec, Steve Pinker o Roodney Brooks percibieron. Hemos creado artilugios que nos superan ampliamente en ciertas tareas que de por sí nos cuesta dominar, pero otras que forman parte del «paquete base» se resisten. De nuevo, resulta llamativo encontrar frases de hace 21 años sonar como los titulares del presente, donde se juega a adivinar que profesiones están mejor protegidas frente a la automatización.

Las «alucinaciones» no son un rapto de creatividad, sino una mala extrapolación[21]. Pensemos en una persona a la que se le mostrara una habitación durante un minuto y se le solicitara a continuación que la plasmara en un dibujo, sin contar con la referencia delante. Seguramente, hubiera momentos en los que dudara «me suena que había una lámpara», «¿de qué color exacto eran las paredes?», «creo que había libros en la estantería». A la IA, salvando mucho las distancias, le sucede algo parecido, solo que sin conciencia y desconociendo de base que es un libro, una estantería o una pared.

Al demandarle que cree una imagen de una mujer, en su predicción estadística como en las muestras había dedos en una zona los incluirá, pero como no sabe lo que es una mano, ni cuantos dedos tiene, a lo mejor en la adivinación se caen dedos o se añaden[22]. Esta es la razón también de que las figuras humanas de IA tiendan a parecerse entre sí, reduciendo enormemente nuestra diversidad. A no ser que se le fuerce en las instrucciones, va a dar un promedio casi siempre. Es curioso, igualmente, como en las recreaciones de edificios históricos reales, en su adivinación a veces añade o quita. Dónde un humano se limitaría a trasladar, la máquina «adivina» y en el camino «olvida»[23].

Obras de ficción como Frankenstein, la Guerra de las Salamandras, Terminator o Matrix han alimentado el miedo a que se produzca una rebelión de las máquinas; que estas se desvíen del propósito para el que se han concebido y comiencen a desarrollar una agenda propia ajena a la humana. Lo curioso es que no ha hecho falta que evolucionen en entes inteligentes o que se aparten de nuestras órdenes para que las máquinas irroguen daños inesperados.

«Tontas» las máquinas ya son plenamente capaces de perjudicarnos mucho por sus lógicas simples. Es el caso de los algoritmos que anteponen a todo conseguir más interacciones, tal como se vio en el genocidio de los

21 Algunos autores consideran que la alucinación, a efectos del objetivo subyacente del modelo, no es un fallo sino un éxito. Porque no están diseñados para otorgar una respuesta verdadera, sino verosímil, que tenga cierta consistencia externa. Por eso también hay corrientes que abogan por unas IA más humildes, con corpus de «entrenamiento» más acotados, que respondan menos, pero con un grado de acierto superior.

22 A lo largo de 2025, la última generación de modelos generativos gráficos ha logrado sortear con mayor éxito este problema, reduciéndose la frecuencia del mismo. Con todo, su persistencia durante años, a pesar de los avances que se iban produciendo en otras áreas de aparente mayor complejidad que determinar el número de dedos de un ser humano, es bastante ilustrativa de los problemas de base a los que se enfrenta el desarrollo de las IAs generativas o razonadoras. De no ser tan notorio el defecto, es muy probable que siguiera arrastrándose. No se ha resuelto tanto la causa que lo generaba sino uno de los síntomas más llamativos.

23 Estos «olvidos» ahora se corrigen añadiendo capas de supervisión antes de mostrar el resultado. Cuando se dice ahora que la IA «razona», lo que realmente se hace es generar un primer resultado que no se muestra al usuario, y el propio sistema o uno creado *ad hoc* lo verifica, en vez de darlo por bueno a la primera. Y así varias veces hasta que por fin se remite al peticionario, cuando no detecta fallos. Esto permite evitar en el *output* errores obvios como los reseñados, pero no la causa de los mismos.

rohingya. No se trata de que una máquina decida que la vida es prescindible, sino que en su maximizamización de «para producir más», tome decisiones absurdas desde la perspectiva humana, pero no de resultado. Véase, un robot que se quedara sin material de construcción y «considerara» derruir edificios para seguir construyendo otros. No deja de ser ese el error del algoritmo de Facebook cuando «detectó» que los contenidos etiquetados como polémicos o ilegales por los usuarios eran los que producían mayor cantidad de mensaje y lejos de reducir su visibilidad, los potenció.

Haciendo balance, de lo expuesto podría concluirse que las «máquinas» son inútiles y que las personas las «ganamos» en todo. Esto no es así, evidentemente. La capacidad de los sistemas de inteligencia artificial para encontrar patrones es impresionante, lo que luego tiene un valor en sí mismo, sin perjuicio de su utilidad adicional en la generación de contenido. No solo porque pueden examinar documentos a una velocidad infinitamente superior, sino porque no tienen algunas de las flaquezas relacionadas con la memoria o concentración. Una persona no puede leer a la vez un millón de libros manteniendo el foco en todos ellos, un ordenador sí. Esta capacidad para gestionar lo ingente, convierte a las IAs en el siguiente paso en la ordenación del conocimiento. Desde que creamos las bibliotecas, hemos necesitado de sistemas para encontrar la información que no requirieran ir mirando uno por uno cada uno de los volúmenes. En este aspecto, las IAs son un bibliotecario ejemplar. Con la asistencia de estos sistemas y haciendo un buen uso de ellos, se abren grandes posibilidades para el progreso social y científico.

El error que se comete con la inteligencia artificial es el que apuntan voces como José María Lasalle. En vez de verla como el nuevo coche o la nueva calculadora, como un potenciador de nuestras capacidades, la concebimos como nuestro nuevo «yo». El *deux ex machina*, el cerebro humano, pero libre de sus presuntas imperfecciones. Olvidamos nuestra naturaleza y, como dice el director de cine Hayao Miyazaki, empezamos a perder la fe en nosotros mismos como especie, como si la inteligencia artificial hubiera surgido por generación espontánea. Si ha llegado a existir ha sido gracias a la suma de nuestras mejores capacidades y conocimientos. Los seres vivos somos criaturas extraordinarias y replicar alguna de las habilidades que ejercemos de forma intuitiva no es nada sencillo, incluso si se invierten miles de millones en ellos y siglos, tal como mencionaba al tratar la paradoja de Moravec[24].

Las máquinas que creamos tienen debilidades y fortalezas como nosotros. Un coche nos gana en velocidad, pero no podría escalar una montaña. A la hora de proveerse de energía, un coche no encuentra gasolina en la naturaleza, mientras que nosotros sí que podemos hallar múltiples fuentes de ali-

24 Suele mencionarse a Ada Lovelace como una de las precursoras de la inteligencia artificial. Sin embargo, como apuntan varios autores, la base de la inteligencia artificial está en los algoritmos de la antigüedad, cuando ni siquiera se había popularizado la expresión acuñada por Al-Juarismi.

mentos, que son al final nuestro combustible. Un helicóptero vuela, pero no puede echarse a nadar. Una persona sí. Y con lo intelectual también contamos con nuestros pluses, por mucho que nos intimide el volumen de la información que manejan las IAs y la rapidez con la que la gestionan[25].

En ese sentido, quizá mereciera la pena reflexionar si todos los desarrollos en el área de la IA aportan una utilidad que justifique los esfuerzos personales y económicos desplegados. ¿Por qué tanto interés en replicar cosas que ya hacemos bien en lugar de usar la tecnología para lograr lo que nunca podríamos lograr sin ella[26]? La revolución industrial no consistió solo en producir más rápido y en mayor cantidad, sino en aprovechar el exceso. Los modelos de inteligencia artificial, y, particularmente, los de carácter generativo, pueden alumbrar resultados con mayor agilidad que los seres humanos, pero esto no siempre redunda en un beneficio social. Cuando las necesidades asociadas ya están ampliamente cubiertas, producir más solo satura todavía más el mercado. Igualmente, si las tareas de las que se «libera» a los trabajadores resultaba gratificantes para ellos, o como mínimo más estimulantes que las alternativas que se les abren, tampoco se puede afirmar que este mejorando sus condiciones. El caso más evidente es de las profesiones artísticas. Desde un punto de vista colectivo, la inteligencia artificial interesa especialmente donde no llega el ser humano, o donde sí llega, lo hace a costa de su seguridad o salud (cabe pensar en actividades peligrosas como la minería o la extinción de incendios).

Dejando a un lado esta última reflexión y cerrando este apartado, lo dicho hasta ahora ha de servir para poner en su sitio el debate. La inteligencia artificial es relativamente novedosa como tecnología, con los matices apuntados, pero no tanto las bases jurídicas desde las que analizar la adecuación legal de los sistemas predominantes de entrenamiento de los modelos de IA. Entre otros motivos, porque las IA amplias han convivido ya varias décadas con el Derecho vigente.

II. Siempre hubo un derecho de las IA

Cada vez que surge una novedad tecnológica, el primer instinto de la sociedad en relación al Derecho es asumir que hay un vacío normativo y que, a la espera de una regulación específica, la materia se halla en un limbo jurídico. Esta línea de pensamiento supone ignorar la vocación de generalidad del

25 Hay una arrogancia implícita en el temor que ha generado que se automaticen tareas «típicamente» intelectuales. Hacer funcionar un tractor, una máquina de vapor, un coche, o un avión no es menos impresionante que emular la expresión humana.

26 En algunos casos es verdad que puede ser interesante lograr lo mismo que ya tenemos, si con eso se amplia la cobertura. Por ejemplo, en atención sanitaria, donde existen problemas de acceso reales, sobre todo en el área de la salud mental (con una desproporción considerable entre la oferta pública y la demanda).

Derecho. Las leyes singulares son la excepción dentro del sistema jurídico moderno, cuya configuración de base es generalista. Por más que la visión del ordenamiento completo propia del positivismo haya quedado superada, y que no se discuta la existencia de las lagunas jurídicas, la orientación universalista de las leyes hace difícil que el Derecho no tenga algo que decir, sea poco o mucho. También porque la forma puede ser nueva, pero la sustancia no, o no del todo.

Ante el fenómeno de los *rider*, a donde acudieron los tribunales para resolver las controversias que les iban planteando fue a los criterios clásicos, legales y jurisprudenciales, que permiten dilucidar si una persona es trabajadora por cuenta ajena o por cuenta propia. Aunque luego se hayan dictado normas que abordan de manera específica esta modalidad singular de trabajo, introduciendo de raíz sus particularidades, lo cierto es que el juego combinado del Estatuto de los Trabajadores y el del Trabajo Autónomo, junto a otras normas laborales, ya proveían de un marco suficiente para tratar muchas de las cuestiones derivadas de esta actividad.

A pesar de este y otros ejemplos, la tendencia de corte adanista descrita al inicio de este apartado no deja de repetirse una y otra vez, siendo el caso de la IA el último exponente de la misma. Desde que se lanzó al mercado de forma amplia ChatGPT, el 30 de noviembre de 2022[27], ha surgido un debate acerca de la conveniencia de regular o no la inteligencia artificial, como si de inicio se tratara de una esfera ajena por completo al Derecho. En el caso de la Unión Europea, se emplazaba a esperar al Reglamento de Inteligencia Artificial, cuyos trabajos ya estaban en curso en el momento en el que la sociedad tomo conciencia del potencial de los modelos LLM.

La realidad, sin embargo, es que el Derecho alcanzó desde el principio a este avance. No de forma absoluta, por supuesto, pero sí con mucho más detalle de lo que se suele decir y pensar. Los que se han quedado más rezagados, en todo caso, son los operadores jurídicos que no han caído en la cuenta de la aplicabilidad de ciertas normas a la inteligencia artificial generativa y otras modalidades popularizadas recientemente.

De formas similar al supuesto de los *rider*, ha habido varios momentos en la breve historia comercial de los actuales modelos LLM en la que los Estados miembros de la Unión Europea han acudido al Reglamento 2016/679, del Parlamento Europeo y del Consejo, de 27 de abril, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, así como a las normas nacionales conexas que lo complementan.

El 31 de marzo de 2023, Italia suspendió temporalmente el uso de ChatGPT, al considerar que la herramienta no garantizaba el cumplimiento del Regla-

27 Antes de esa fecha, ChatGPT 3 estaba en funcionamiento ya, pero con acceso restringido a una serie de colectivos de prueba.

mento en relación a los datos introducidos por los usuarios. Un año después, en mayo de 2024, Meta paralizó en la Unión Europea y Reino Unido el entrenamiento de sus modelos de inteligencia artificial a partir de la información privada alojada en los servicios de la compañía, como Instagram o Facebook, ante las decisiones negativas de varias agencias nacionales europeas de protección de datos. En abril de 2025, en un contexto similar al de META, la Comisión de Protección de Datos de Irlanda inició una investigación contra X.

Todos estos casos conectan con el Reglamento General de Protección de Datos —en el que no hay ni una sola mención directa a la expresión «inteligencia artificial»— y no con el Reglamento 2024/1689, del Parlamento Europeo y del Consejo, de 13 de junio, por el que se establecen normas armonizadas en materia de inteligencia artificial, lo que trasluce que tenemos normas previas a la irrupción de los LLMs susceptibles de incidir sobre ellos. Estos precedentes, con todo, no han evitado que se hable del Reglamento 2024/1689 como la primera ley de inteligencia artificial del mundo.

No pretendiendo minusvalorar el carácter pionero que posee el Reglamento 2024/1689, de 13 de junio, la realidad es que hay varias normas que incidían ya en la inteligencia artificial, al margen del ya citado Reglamento General de Protección de Datos. Podría decirse que se trata de la primera norma con orientación global y diseñada a medida de esta tecnología (con matices, porque, aparentemente, los sistemas de corte generativo fueron contemplados en una fase ya muy avanzada de los trabajos), pero, en puridad, las primeras leyes sobre IAs preceden al Reglamento.

Sin ánimo exhaustivo, en el Dictamen del Comité Económico y Social Europeo, de 30 de octubre de 2019, sobre la Comunicación de la Comisión «Generar confianza en la inteligencia artificial centrada en el ser humano», se mencionaban la Directiva 85/374/CEE del Consejo, de 25 de julio, relativa a la aproximación de las disposiciones legales, reglamentarias y administrativas de los Estados Miembros en materia de responsabilidad por los daños causados por productos defectuosos[28], o la 2000/43/CE del Consejo, de 29 de junio, relativa a la aplicación del principio de igualdad de trato de las personas independientemente de su origen racial o étnico, como parte del acervo jurídico que debía inspirar el diseño del modelo ético de inteligencia artificial, preconizado desde el Parlamento europeo.

Ligada al objeto del capítulo, tiene gran importancia la Directiva 2019/790 sobre los derechos de autor y derechos afines en el mercado único digital. También resulta reseñable la Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo, sobre la protección jurídica de las bases de datos, pensando en las bibliotecas de datos y repositorios organizados de información.

28 Sustituida por la Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos.

El Reglamento 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre, relativo a un mercado único de servicios digitales, cercano en el tiempo al Reglamento de Inteligencia Artificial, y más aún de la «presentación en sociedad» de ChatGPT, interpela a la inteligencia artificial, aunque en ningún momento emplee esas palabras (como ocurría en el Reglamento de Protección de Datos).

El Reglamento de Servicios Digitales se refiere a los algoritmos y marca obligaciones respecto a ellos cuando son utilizados en las decisiones de moderación de las plataformas. Si bien no todo algoritmo constituye inteligencia artificial, no hay inteligencia artificial sin algoritmos, por lo que todas las disposiciones relativas a ellos son extensibles en principio a los sistemas de inteligencia artificial, salvo que expresamente se excluyan, o, que, por su naturaleza, se deduzca de forma evidente su inaplicabilidad.

En el ámbito de las redes sociales, dado el ingente número de interacciones que se generan por segundo, su vigilancia siempre se ha apoyado de manera intensa en sistemas de carácter automatizado, englobables dentro de la idea de inteligencia artificial amplia apuntada en el epígrafe anterior. La intervención humana se reservaba para los casos extremos o para una verificación final, pero el cribado inicial rara vez ha sido «a mano». Siendo este el paradigma en un estadio mucho más incipiente de las tecnologías de reconocimiento de imagen, audio y texto, en el vigente la aplicabilidad del Reglamento de Servicios Digitales a inteligencias artificiales va a ser mucho mayor, porque va a haber más herramientas de ese tipo asumiendo más funciones de moderación.

Alejándonos un momento del ámbito europeo, el Derecho español apreció esta conexión en la Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales. Su artículo único incorporaba una nueva letra d) al artículo 64.4 del Estatuto de los Trabajadores en la que se definía el derecho del comité de empresa a *«ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles»*. Priorizando la amplitud a la precisión técnica, la ley intenta abarcar el mayor número de sistemas que puedan usarse en el sector, no preocupándose mucho de delimitar la frontera entre algoritmos y sistemas de inteligencia artificial.

Volviendo de esta digresión nacional y continuando desde el plano de la Unión, la Comunicación de la Comisión de 7 de diciembre de 2018, relativa al «Plan Coordinado sobre la Inteligencia Artificial», apuntó, a su vez, la posibilidad de que los parámetros del Marco Europeo de Certificación de Ciber-

seguridad fueran imperativos para los modelos de inteligencia artificial que incidieran en determinados campos, como el financiero o el nuclear. En el resto de áreas, donde esa obligatoriedad no regiría, la Comisión percibió, con todo, la capacidad del Marco Europeo a la hora de servir de base o inspiración en la configuración de futuros requisitos, basados específicamente en las características de las diferentes inteligencias artificiales. Corolario de ello, el Reglamento 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril, relativo a la ENISA y a la certificación de la ciberseguridad de las tecnologías de la información, formaría parte del contexto jurídico europeo que entablaba «diálogo» con la inteligencia artificial antes de que se dictara el Reglamento *ad hoc* de 2024.

De este repaso asistemático de normas puede inferirse que, si bien la Unión Europea no había abordado expresamente la inteligencia artificial con instrumentos de Derecho duro o *hard law* hasta la aprobación del Reglamento del año 2024, parte de sus piezas (datos y algoritmos, así como otros elementos) sí que eran objeto de atención jurídica, lo que suponía y supone un primer entramado que ordena el desarrollo potencial de esta tecnología.

En plano del Derecho blando o *soft law*, el Libro Blanco de la Inteligencia Artificial del año 2020 refuerza la perspectiva defendida en estas páginas, porque al mismo tiempo que declaró la necesidad de regular la tecnología, manifestaba que *«los desarrolladores e implementadores de la inteligencia artificial ya están sujetos a la legislación europea en materia de derechos fundamentales (la protección de datos, la privacidad o la no discriminación, entre otros), protección de los consumidores y normas sobre la seguridad de los productos y responsabilidad civil»*. En esta línea, en el documento se comenta que parte de las directrices alumbradas el 19 de abril de 2019[29], por el grupo de expertos de alto nivel creado por la propia Comisión, estaban recogidas ya en la normativa. Que el Libro esté refrendado por la Comisión no es un hecho menor, tratándose de la institución a la que se le confía la guardia de los Tratados y se le encarga supervisar la aplicación del Derecho europeo bajo el control del Tribunal de Justicia de la Unión Europea.

Al margen del Derecho comunitario, pensando en cualquier Estado constitucional, como el español, toda actividad humana está condicionada *prima facie* por las reglas establecidas en la Constitución. El desarrollo de inteligencias artificiales por el sector privado se halla sujeto al principio de legalidad en su vertiente negativa, de modo que sus promotores gozan de un amplio margen de libertad en todo cuando no esté prohibido. Sin embargo, la idea de margen presupone correlativamente la de límites. Papel pautado y no folio en blanco.

29 En concreto, consistían en garantizar: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y medioambiental; y rendición de cuentas.

Es evidente que las inteligencias artificiales, lo diga o no explícitamente una norma, han de respetar de manera imperativa los derechos a la igualdad, al honor, a la intimidad personal y familia, al secreto de las comunicaciones, a la libertad ideológica y religiosa, a la propia imagen, a la propiedad privada... Y lo mismo sucede con el resto del ordenamiento jurídico, en la medida en que encaje con la casuística de la inteligencia artificial.

Así, los llamados «GPTS oscuros», disponibles en la *dark web*, serían fácilmente subsumibles en nociones como la de la letra a) del artículo 197 Ter del Código Penal, que se refiere a programas informáticos concebidos o adaptados para la comisión de ciertos delitos. A pesar de que dicho precepto fue introducido con motivo de la modificación del Código promovida por la Ley Orgánica 1/2015, de 30 de marzo, siete años antes de que Sam Altman se incorporará al «panteón» de celebridades de *Silicon Valley*, su contenido no precisa de ninguna modulación para su aplicación en el marco de la inteligencia artificial.

Menos cavilaba aún acerca de robots el legislador que tipificó el homicidio, delito de profundas raíces, que se remontan hasta el Código de Hammurabi. Sin embargo, si una persona se sirve de un autómata para matar a otra persona, cabría perseguir esta acción conforme a dicho tipo penal, sin necesidad de que se añadan a su enunciado palabras como «robot», «máquina» o «inteligencia artificial».

Cabe pensar también el ejercicio de ciertas actividades profesionales como la psicología o la abogacía que están sujetas a la disposición de un título habilitante. Si la inteligencia artificial se comporta como algo más que un repositorio de información y actúa de forma simulada como un profesional, podría tener que cumplir con las mismas obligaciones que un operador humano[30].

30 En EE. UU., encontramos algunas iniciativas que van en esa dirección. El 1 de agosto de 2025, el estado de Illinois concluyó el proceso legislativo de la *Wellness and Oversight for Psychological Resources Act* (HB 1806), donde se castiga la promoción de los chatbox de IA como una alternativa a la terapia tradicional de salud mental. Asimismo, se prohíbe el uso de esta tecnología para tomar decisiones o diagnosticar sobre el fondo, determinar el estado mental de la persona, o interactuar directamente con el paciente para proporcionarle consejo terapéutico. La norma orienta el recurso a estos sistemas sobre todo para labores administrativas o de apoyo para el profesional, como la transcripción de notas o analizar datos y tendencias, pero incluso en eso se exige supervisión última de una persona. El 25 de marzo de 2025, Utah enmendó su *Artificial Intelligencia Policy Act,* por medio de la Ley HB 452, para introducir una serie de obligaciones en salud mental. Sin llegar a la prohibición de Illinios, sí que sujeta el uso de estos chatbox a la superación de una serie de controles previos. Las autoridades del Estado serán las encargadas de elaborar los procedimientos técnicos que pongan a prueba de forma simulada la calidad del servicio. La AB 406 de Nevada, de 5 de junio de 2025, también censura de raíz la promoción de la inteligencia artificial generativa como sustitutiva de los cauces habituales de tratamiento en salud mental. Hay otros estados más como California, Pensilvania o Nueva Jersey tramitando normas semejantes. Asimismo, en el plano de la

Evidentemente, las transformaciones sociales y tecnológicas traen consigo nuevos retos al Derecho, que pueden exigir de soluciones no previstas inicialmente. En algunos casos, la interpretación basta para canalizar estas nuevas demandas, mientras que en otros la única vía será la actualización y reforma del ordenamiento jurídico. No obstante, es importante no perder de vista que ninguna de estas alteraciones anula la existencia del Derecho previo a ellas. Muchas veces no es tanto que no sea aplicable, sino que a raíz de estos cambios puede reevaluarse si quiere que lo sea de la misma forma. Los acontecimientos disruptivos actúan a menudo como catalizadores de nuevos juicios de oportunidad, es decir, como fuentes materiales del Derecho. Ahora bien, esa reconsideración solo surtirá efectos jurídicos a partir del momento en que se plasme en las fuentes formales de la Ley. Mientras tanto, persiste la letra de Ley, sin perjuicio del margen que ofrece el artículo 3 del Código Civil de 1889.

Esto es algo que se capta claramente al examinar el eventual reconocimiento de derechos de propiedad intelectual en relación a los *outputs* producidos de los modelos de inteligencia artificial generativa. De *lege lata*, son pocos los ordenamientos jurídicos que dan pie a que se atribuya la autoría de estos productos derivados a las personas que introducen los comandos necesarios para provocar su emisión. Menos aún, a los trabajos desarrollados de manera automática sin ningún tipo de intervención humana. La legislación en esa materia se asienta en la creatividad humana. El fundamento de protección puede ser diferente, dependiendo de la tradición jurídica de cada país, oscilando entre las teorías del trabajo, las personalistas, y las utilitaristas, pero al final todas ellas conectan con cualidades humanas que deben manifestarse en lo creado.

En EEUU, el 30 de enero de 2025, la *U.S. Copyright Office* admitió la solicitud presentada por Kent Keirsey para registrar la obra *«A single piece of american cheese»*. Esta resolución, que se ha interpretado por muchas voces como un cambio de doctrina respecto a denegaciones previas, como las que afectaron a *«Théâtre D'Opéra Spatial»* o *«Zarya of the dawn»*, en realidad no difiere de la pauta marcada por la *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, de 16 de marzo

ejecución legal, el fiscal general de Texas, Ken Paxton, abrió una investigación el 18 de agosto de 2025 por posible promoción engañosa de estas herramientas como soluciones de salud mental. A mayor abundamiento, y abandonando el campo sanitario, la Comisión Federal de Comercio de EEUU multó en 2025 con 193.000 dólares a la empresa detrás del *chatbox Do Not Pay* por anunciar que sus habilidades eran las mismas que las de un abogado humano. Merece la pena destacar una de las aseveraciones de la Comisión «*las empresas que prometen reemplazar a profesionales humanos deben tener la capacidad real para hacerlo, y deben respaldarlo con evidencia»*. Igualmente, en 2023 la empresa tuvo que abandonar su intentó de que su *chatbox* defendiera un caso en juicio, utilizando a una persona como «receptáculo de expresión» del mismo, y se enfrentó a una demanda colectiva, entendiendo que el *chatbox* ejercía la abogacía sin poseer un título de Derecho.

de 2023. La propia existencia de la guía refleja que la postura de la Oficina nunca ha sido de oposición absoluta a llevar a cabo registros de estas características, o, de lo contrario, no hubiera difundido un documento en el que se indica cómo deben prepararse los requerimientos de este tipo. Lo que siempre ha hecho la Oficina es sujetar la estimación de dichas peticiones al cumplimiento una serie de requisitos. Que hasta *«A single piece of american cheese»* considerara que ninguno de los trabajos examinados los acreditaba es una cuestión diferente.

Siguiendo el enfoque adoptado en el caso Urantia Found. v. Kristen Maaherra, de 1997, la Oficina entiende que para llevar a cabo una declaración favorable de *Copyright* es preciso que la persona solicitante haya realizado un proceso de selección y arreglos y que en él emerja su personalidad como creador humano. Exige control humano sobre la forma y el fondo de la obra. En el caso de *«Théâtre D'Opéra Spatial»*, el peticionario respaldó su reclamación en el hecho de que había introducido 642 instrucciones en el programa. Sin embargo, como apunta la Oficina desde la guía, por muchas instrucciones que se den, hoy por hoy, estas conllevan un componente de aleatoriedad, que impide equiparar la inteligencia artificial generativa con el manejo de otras herramientas donde es el usuario el que determina cada paso que se da con exactitud[31]. Prueba de ello es que con las mismas instrucciones cabe obtener resultados diferentes. Los modelos generativos, tal y como están configurados, no son deterministas.

Un ejemplo divertido es el de un periodista de La Vanguardia que buscaba recrear una escena realista en Babilonia y que, tras varios intentos e ir aproximándose a su objetivo, descubrió que todo era aceptable salvo un pequeño detalle: los habitantes llevaban un móvil. A pesar de todas las pautas que dio para que la imagen se ajustara al periodo histórico, la vertiente probabilística y estocástica le jugó una mala pasada. En vez de hacer un pronóstico acotado temporalmente, el modelo LLM lo hizo de la especie humana en abstracto. Como hoy en día hay más fotos de personas en la calle con móvil que sin él, «concluyó» que era algo que debía figurar necesariamente. Justo por casos como este ha surgido en Internet una corriente de memes basada en el contraste entre lo que se buscaba obtener con las instrucciones y lo que realmente se entrega, comparando a la inteligencia artificial con los genios de la lámpara, que en los cuentos les dan la vuelta a los deseos de sus amos con algún giro sorprendente.

En *«A single piece of american cheese»* se añadió a la imagen inicial, generada artificialmente, varios elementos, además de modificar o sustituir otros.

31 En la edición de imágenes, como se apuntaba en una nota anterior, algunos modelos están logrando mantener la consistencia con un grado de éxito muy superior a iteraciones pasadas. En ese sentido, es posible que jurídicamente aquí se pueda alcanzar una equiparación con los arreglos realizados por las personas que usan herramientas como Photoshop. En la generación *ex novo*, de momento el estado de situación es el descrito.

Estos cambios se ejecutaron en parte de forma manual, especificando áreas muy concretas de aplicación (a veces incluso de una línea) o llegando a dibujar directamente. Dicho de forma gráfica, Kent Keirsey tuvo que mover el cursor activamente y no limitarse solo a teclear palabras. El modelo que usó para ello, *Invoke IA*, contaba con recursos avanzados de edición de imagen propios de programas como Photoshop, Adobe, o Procreate. Posibilidades que ahora están más generalizadas, pero que no existían o estaban mucho más limitadas en 2022 o 2023. De forma acelerada, el vídeo que recoge el proceso dura diez minutos, lo que es ilustrativo de que no fue un caso ordinario y que tampoco se limitó a aplicar un par de ajustes menores.

Es discutible si esta intervención humana tenía entidad suficiente para que el resultado pasara a asimilarse legalmente al de las creaciones humanas convencionales, pero el fundamento en sí no es novedoso y no solo por el criterio manifestado en la guía del año 2023. En el arte, se ha visto muchas veces. Francis Bacon tomo el retrato del Papa Inocencio X de Velázquez para pintar su versión retorcida, del mismo modo que Pablo Picasso lo hizo con las Meninas en su reinterpretación del cuadro. Andy Warhol se hizo famoso —bastante más tiempo que quince minutos— gracias a su particular estilo *pop art*, en la que partiendo de fotografías o viñetas de cómics preexistentes llevaba alteraciones del color, así como otros cambios.

Distando de la calidad de estos trabajos aludidos —cuestión que en cualquier caso está sujeta a un grado enorme de subjetividad—, el precedente creado por Kent Keirsey iría por este camino. De hecho, en la defensa de su petición, parte de su argumentación se apoyaba en que se trataba de un collage formado por varios fragmentos y no de una imagen única. Si ha triunfado donde otros fracasaron es porque ha logrado demostrar a ojos de la Oficina que satisface, aunque sea en un grado mínimo, la exigencia legal de participación humana en el proyecto, en calidad, además, de creador.

Por consiguiente, la decisión favorable queda lejos de haber abierto la puerta de par en par de cara a que los *outputs* de inteligencia artificial generativa sean beneficiarios de derechos de propiedad intelectual en paridad con las obras creadas plenamente por seres humanos. Aparentemente, la Oficina continúa situando a la persona que introduce los *prompts* en el plano de un instigador o inductor del trabajo y no en el de los creativos. Si no trasciende de ese rol y ejerce un control más concreto y una selección amplia —elegir entre unas pocas posibilidades o una cifra similar no bastaría—, es poco probable que dé lugar a copyright.

Por otro lado, a mi forma de ver, resultaría capcioso pretender encajar toda imagen artificial con modificaciones en un collage. En este caso concreto lo hacía, pero en la mayoría sería forzar el concepto. Caso concreto donde, además, se ha documentado todo el proceso, punto por punto, de forma que la Oficina ha podido cotejar las incorporaciones propias del autor, deslindándolas de las puramente automatizadas, o resultado del *prompt* inicial. Es dudoso que todos los interesados potenciales vayan a ser tan minuciosos en

sus presentaciones. No en vano, Kent Keirsey es el CEO y fundador de *Invoke*, por lo que estaba interesado publicitaria y comercialmente en instaurar un *leading case*, pudiendo incluso haberse esforzado en que la ejecución se acomodara lo más posible al esquema legal; algo que de forma espontánea es menos probable que se dé.

Dentro de la esfera europea, son otras trabas las que se presentan, pero las consecuencias no se separan demasiado. La Sentencia del Tribunal de Justicia de la Unión Europea (TJUE), de 1 de diciembre de 2011, asunto C-145/10, Eva-Maria Painer v. Standard VerlagsGmbH y otros, interpreta la noción de originalidad exigida en la legislación de propiedad intelectual como subjetiva. Así se dice que las obras protegidas han de ser *«una creación intelectual del autor que refleje su personalidad y que se manifieste por las decisiones libres y creativas»*. La necesidad de dotar a las obras de un «toque personal», en palabras de Pablo Fernández Carballo-Calero, se ha mantenido también en otros pronunciamientos como los asuntos C-5/08, Infopaq, de 16 de julio de 2009, o C-604/10, Footbal-Dataco, de 1 de marzo de 2012.

En España, el Tribunal Supremo, en sentencias como la 1644/2017, de 26 de abril, ha sido más contundente al afirmar que los participantes en un proyecto solo pueden ser tutelados por la legislación de propiedad intelectual si su contribución presenta una cierta originalidad, que cumpla los requisitos de singularidad, individualidad y distinguibilidad. Entiende el Tribunal que *«actualmente prevalece el criterio de que la originalidad prevista por el art. 10.1 TRLPI exige un cierto grado de altura creativa. Esa concepción objetiva permite destacar el factor de recognoscibilidad o diferenciación de la obra respecto de las preexistentes, imprescindible para atribuir un derecho de exclusiva con aspectos morales y patrimoniales, lo que requiere que la originalidad tenga una relevancia mínima suficiente»*.

El uso del término «objetivo» puede inducir a cierta confusión, al parecer que remite a la asociación de originalidad con novedad, en la que bastaría la no preexistencia del trabajo. Sin embargo, desde una visión de conjunto de la argumentación, el resto del vocabulario usado conecta con la jurisprudencia europea y el principio de que los titulares de derechos de propiedad intelectual han de dejar su «huella personal» en lo creado. No basta con figurar, hay que hacer una aportación particular con impacto real. Particularmente, al señalarse que la originalidad demanda cierta altura creativa se está rechazando de plano que para ser original baste con la no repetición.

En España, como en EE. UU., la negativa a inscribir en el registro de propiedad intelectual se está apoyando, entre otros factores, en la pérdida de control del usuario una vez facilitadas las instrucciones. Si yo pido una imagen que represente una playa en verano, salvo que incurra la maquina en una alucinación, lo más probable es que obtenga lo solicitado, con mayor o menor calidad. Ahora bien, difícilmente se puede decir que el resultado nazca de un plan preconcebido. Aunque el *output* resulte satisfactorio, es poco pro-

bable que coincida con la imagen mental que se tuviera de antemano. Sin que yo lo haya pensado ni querido, la estampa puede incluir unos veraneantes que comen un helado o juegan al voleibol. No es un fallo, pero tampoco una elección. Es una inferencia estadística del modelo. De haberlo hecho por nosotros mismos, cada detalle respondería a una decisión consciente. Por muchos datos que se incluyan en las instrucciones, el dominio siempre será inferior[32], salvo que se vaya pixel por pixel determinando como debe ser, lo que sería sumamente ineficiente.

La clave está en artículos indefinidos y en los posesivos. El párrafo anterior podría haber sido escrito en un segundo empleando un comando de este estilo *«redacta un texto de unas once líneas con un lenguaje riguroso, pero accesible, en el que expongas el menor grado de control que tienen las personas en cuanto al resultado cuando piden a un modelo de inteligencia artificial que haga las cosas por ellos, en vez de hacerlas por sí mismos»*. A mí me ha costado varios minutos al reflexionar sobre qué decir y cómo hacerlo. De hecho, una parte se redactó en un día y otra al siguiente, al repasarlo. La cuestión es que si lo hubiera optado por delegar su redacción no sería MI párrafo, sino UN párrafo. No sabemos si lo generado sería mejor o peor, pero lo que es seguro que no estaría expresando mi criterio personal.

Aunque los modelos LLM poseen una adaptabilidad ampliamente superior a la de las cadenas de montaje clásicas, la filosofía subyacente es la misma: automatizar. Jurídicamente, esto es importante porque la automatización se compadece mal con la personalización que demanda la originalidad subjetiva requerida por la jurisprudencia del TJUE. Para el que quiere dibujar SU gato y no un GATO, o el que quiere escribir SU novela y no UNA novela, la IA generativa presenta importantes limitaciones. La IA responde mejor las demandas de aquellas personas que quieren alcanzar un resultado, pero les es relativamente indiferente la forma en la que se plasme.

Mientras revisaba las galeradas de mi capítulo, surgió una polémica en relación con la restauración de la Virgen de la Esperanza Macarena, en Sevilla. Al intentar rehabilitar la pintura de la imagen, el rostro se vio alterado en varios aspectos como la mirada, que pasó de dirigirse hacia el frente a proyectarse hacia abajo. A diferencia de otras recuperaciones fallidas de obras artísticas, como el famoso caso del *Ecce Homo* de Borja, el problema no radica en un deficiente trabajo, sino en que haya un cambio. La nueva cara de la Virgen puede ser estética, pero no es la que tenía antes y no es la que muchos miembros de la Hermandad quieren para ella.

32 El dibujante David López traza un interesante paralelismo entre la inteligencia artificial y el Papa Julio II. Para él, en manos de un usuario que no sabe, el rol que ocupa no es el de Miguel Angel, sino del que encarga los frescos de la Capilla Sixtina. Por muchas instrucciones que se den, el ejecutor nunca es el que solicita.

Si trasladamos este ejemplo humano a la inteligencia artificial generativa, tanto el rostro inicial previo a la restauración como el posterior serían opciones válidas para cumplir con unas instrucciones que demandarán una Virgen dolorida y llorosa de estilo barroco. Mientras solo quisiéramos conseguir un resultado que respete de forma general lo pedido, el número de lágrimas derramadas, la dirección de los ojos, la presencia o no de barniz en la madera, u otros tantos detalles serían irrelevantes para nosotros. Porque lo que querríamos es una Virgen. Pero si, como les ocurre a los sevillanos afectados, lo que se desea es una talla concreta, la que tenemos en nuestro pensamiento, afinar con los comandos no es sencillo.

Del mismo que tenemos dificultad para comunicarnos con otros seres humanos y trasmitirles nuestro mundo interior, con las máquinas también nos enfrentamos a este reto (al menos, mientras no se produzca una conexión cerebro-maquina plena, que fuera mucho más lejos de lo que tecnologías como Neuralink ofrecen en la actualidad). Nuestro azul, un azul inspirado en una puesta de sol que vimos, no lo va a conocer por un promedio estadístico plasmado píxel a píxel. Incluir «azul» en un *prompt* no va a trasladarlo. Si hay una referencia concreta, podremos indicarle que es «ese azul», pero si no, de contar con las habilidades precisas, normalmente será sencillo hacerlo uno mismo que pasar por un proceso de ensayo y error en las descripciones.

Esto, como afirmaba antes, pone cuesta arriba que en el grueso de los casos una imagen, una canción, un texto, o un vídeo producidos por inteligencia artificial expresen creatividad propia, por mucha información que se les proporcione. Porque al final es la máquina la que interpreta a su manera como ejecutarlos. Tanto en Europa como en EE. UU., los *«A single piece of american cheese»* van a ser la excepción y no la norma. O deberían serlo, conforme al tenor actual de la legislación y la jurisprudencia.

Naturalmente, esto se corresponde con el estado vigente de las cosas. Como reza el famoso aforismo jurídico, la Ley lo puede casi todo. De modo que una reforma legal podría otorgar a los instructores humanos la condición de autores, o una situación análoga que les permitiera reivindicar derechos morales y de explotación. Empero, como no dejo de insistir, el cambio, si llega, ha de materializarse de forma real para que surta efecto.

Cuando no gusta lo que dice el Derecho en vigor, con frecuencia, se opta por ignorarlo. Se abroga de facto y no de *iure*. Este proceder es contrario a las fuentes del Derecho puesto que la costumbre, por principio, solo rige en defecto de norma, nunca contra ella. No se trata de que sea mejor o peor. Dura o blanda, la ley es la ley, expresado de forma tautológica. Su acatamiento es la base sobre la que se construye la sociedad. Solo en situaciones extremas de dictaduras o violación de derechos humanos se puede abrir la cuestión del incumplimiento legítimo.

No me cabe duda de que en relación al supuesto de los *outputs* artificiales va a emerger la tentación de intentar forzar la legislación de propiedad inte-

lectual para que se homologuen a las manifestaciones ordinarias de creatividad humana. Desde mi punto de vista, las leyes de propiedad intelectual se han construido en torno a la humanidad y no puede incorporarse sin más la figura del creador robótico, autónomo o asistido. Si se quiere establecer un nuevo marco —y no todo el mundo lo desea—, lo idóneo sería modificar la norma previamente. Esta vía, sin embargo, resulta poco apetecible para algunos, porque exigiría entrar en un debate en el que cada parte tendría oportunidad de presentar argumentos. Reputarlo como un hecho consumado, algo que siempre estuvo ahí, evita pasar por el proceso y coloca en un plano de resignación a los que no estén satisfechos con la conclusión alcanzada: «lo siento, es que la Ley lo dice». De ahí que detecte cierto interés por trasmitir que lo que habría de ser, en todo caso, de *lege ferenda* es de *lege lata*.

Este proceder, a mi juicio, es poco honesto y riguroso jurídicamente. Del mismo modo que lo sería si se pretendiera dotar de protección al estilo, cuando siempre ha quedado fuera del ámbito tutelado por esta legislación de propiedad intelectual, siendo otros campos jurídicos los que han mostrado cierta preocupación por la réplica de la forma[33]. La capacidad de los modelos de inteligencia artificial generativa de mimetizar los manierismos propios de los autores y artistas puede llevar a reconsiderar esta exclusión, al menos respecto a las máquinas. Pero reclamar este giro sin renovar la normas internacionales y nacionales sería inadecuado.

Las leyes humanas, incluida la Constitución, prevén su reforma porque, a diferencia de los dictados divinos, no son absolutas. Hay margen de mejora y actualización, pero este ha de seguir los canales que el ordenamiento ha establecido para reformular la ponderación de los bienes jurídicos y de los juicios de oportunidad. No podemos ni debemos eludir estos cauces.

Quizá, suene paradójico cuando la mayor parte del capítulo se ha dedicado a mantener que el Derecho sin cambios puede regir situaciones novedosas. No obstante, esta antinomia solo es aparente. No es incompatible que haya partes del Derecho que puedan desplegar efectos sobre realidades inéditas y que otras queden desincronizadas, bien porque estas situaciones rebasen completamente lo regulado, o, más frecuentemente, porque, quedando dentro de sus contornos, la solución proporcionada se repute insatisfactoria en el nuevo contexto. O partes en las que pudiendo mantenerse tal como están, el cambio factico obligue a llevar a cabo un acto reafirmatorio en el que la sociedad exprese que no quiere variar el rumbo, explicitando lo que desde la eficacia general ya se desprendería.

Trasladado esto al objeto del tema, el ajuste al marco legal del sistema de entrenamiento de la inteligencia lo vamos a encontrar en el Derecho actual, sin perjuicio de las vías evolutivas ligadas a la interpretación. Si era posible hacer un *scraping* masivo de Internet, o utilizar millones de materiales

33 Véase, la legislación sobre propiedad industrial.

sujetos a propiedad intelectual sin autorización directa de sus titulares, lo determinará, dependiendo del país, el juego de excepciones de la Directiva (UE) 2019/790 del Parlamento Europeo y del Consejo, de 17 de abril de 2019 sobre los derechos de autor y derechos afines en el mercado único digital, o la doctrina del *Fair Use*. Nos convenga o no, será el Derecho el que tenga la última palabra. Al menos, mientras no se cambie esa palabra con su correspondiente modificación legal.

III. ¿Es demasiado pequeña la propiedad intelectual?

En el ámbito tecnológico, se dice que impera la filosofía de haz primero, pregunta después, porque esperar puede suponer quedar fuera de la carrera. Las batallas empresariales que se libran son muy encarnizadas. Es un campo donde es muy frecuente que el ganador se lo lleve todo, o casi todo, dejando espacio como mucho para otro competidor. Puede no haber plata ni bronce: solo oro. Basta pensar en la disputa entre el formato BETA y VHS o entre HD DVD y BLUE RAY. En el mercado de los sistemas operativos, el dominio de Windows es inexpugnable en lo que se refiere a los ordenadores de sobremesa, con una cuota de mercado del 70,14 % a junio de 2025. Posición que en móvil ostenta Android con un 74,18 %. En ambos terrenos, los productos de Apple, iOS y OS X, ocupan a gran distancia un segundo puesto, quedando ya segmentos muy pequeños para el resto[34]. En cuanto a navegadores, Chrome acapara entre un 65-70 % del mercado de los navegadores.

Mark Zuckenberg lo plasmó en su máxima *«muévete rápido y rompe cosas»*, que explicaba de la siguiente manera en una conferencia: *«muchas veces la gente es demasiado cuidadosa. Creo que es más útil hacer que las cosas sucedan, y luego disculparse más tarde*[35]*, en lugar de asegurarse de tener todos los puntos sobre las íes ahora, y luego no lograr hacer las cosas».* En 2021, Sam Altman ahondaría en la frase a través de Twitter, adaptándola a su gusto: *«muévete más rápido. La lentitud en cualquier parte justifica la lentitud en todas partes (...) 2021 en lugar de 2022. Esta semana en lugar de la próxima. Hoy en lugar de mañana. Moverse rápido compensa mucho más de lo que la gente cree»*.

Este enfoque implica que, a la hora de desarrollar aplicaciones, no sean las leyes humanas las que les preocupen sino las de la naturaleza. Interesa si es posible técnicamente, lo legal se deja para una fase posterior. En parte es comprensible, en la medida en que algunas dudas jurídicas se plantean conforme va tomando forma el producto y no antes. Sin embargo, sería deseable que la vertiente preventiva no se postergara tanto. Más aún que no se come-

34 Estadísticas extraídas de la página web statcounter.

35 Algo que ha tenido que hacer en un amplio número de ocasiones Zuckenberg, en sus comparecencias parlamentarias ante las instituciones de EE. UU. o la UE.

tieran actos de dudosa legalidad a sabiendas, esperando que no se descubran, o contando con poder convalidarlos posteriormente a través de una reforma legal *ad hoc.* Conductas en las que, por desgracia, parece que han incurrido varios desarrolladores destacados[36].

Junto al lema de Zuckenberg y sus derivados implícitamente hay otra idea troncal: *too big to fail*, demasiado grande para caer. Esta expresión popularizada a raíz de la crisis económica del año 2008, para describir la imposibilidad de no rescatar a la banca por el daño que hubiera causado sobre la estructura social, es una carta que tienden a jugar las grandes tecnológicas frente a los Estados.

A principios del 2025, Sam Altman planteó al gobierno de EE. UU. la necesidad de contemplar de manera expresa el entrenamiento de modelos de inteligencia artificial como uno de los supuestos de *fair use*, bajo los cuales está permitido utilizar material protegido por *copyright* sin consentimiento ni compensación a su titular. Para justificar su petición, defendió que, de no aplicarse esta dispensa, China se impondría en la disputa por la IA. Google y otras compañías del mismo sector han lanzado un mensaje similar, coincidiendo en que esta exoneración legal representa un asunto de seguridad nacional.

Este ideario parece haber sido asumido por el gobierno de EE. UU., a tenor del *AI Action Plan,* dado a conocer al 23 de julio de 2025, y del cese de Shira Pelmutter, directora de la Oficina de Copyright del país, después de que se difundiera en mayo de 2025 la versión en pre-publicación de la tercera parte del informe de la Oficina sobre inteligencia artificial y *copyright*, relativa al entrenamiento de los modelos de inteligencia artificial generativa. Informe que en varios puntos se alineaba más con las demandas de la comunidad artística y creativa crítica con el sistema actual de entrenamiento, que con las de la industria.

Si bien el *AI Action Plan* no se pronuncia de forma directa sobre la cuestión del Copyright, el Presidente de EEUU, Donald Trump, declaró en el marco de su presentación que *«no puedes esperar tener un programa de IA exitoso si se supone que tiene que pagar por cada artículo, libro o cualquier otra cosa que hayas leído o estudiado (...) Simplemente no puedes hacerlo, porque no es factible»*. En esa línea, añadió que *«cuando una persona lee un libro o un artículo, ha obtenido un gran conocimiento. Eso no significa que estés violando las leyes del copyright o que tengas que hacer acuerdos con cada proveedor de contenidos (...) Simplemente no puedes hacerlo. China no lo está haciendo...Tienes que jugar con el mismo set de reglas. Simplemente*

36 El grupo de Análisis de Internet de IMDEA Networks, dirigido por Narseo Vallina-Rodríguez, Gunes Acar y Tim Vlummens descubrió recientemente un posible abuso de la privacidad que involucraría a Meta y al gigante tecnológico ruso Yandex. Este escándalo se sumaría a otros incidentes como el de *Cambridge Analytica*. Habría muchos más ejemplos.

no funciona de esa manera. Por supuesto, no puedes copiar o plagiar un artículo, pero si has leído un artículo y aprendido de él, tenemos que permitir a la IA usar esa fuente de conocimiento sin pasar a través de unas complejas negociaciones contractuales, que se darían por miles cada vez que usáramos la IA»[37].

Estas afirmaciones son en esencia las que le trasmitieron las empresas, que ahora hace suyas. La única duda es si se plasmarán o no en forma de ley u orden ejecutiva, pero es evidente que la Administración Trump no pondrá trabas al «entrenamiento», salvo en lo que se refiere a la elección ideológica de las fuentes[38].

Irónicamente, Altman se queja de que DeepSeek se habría «entrenado» aprovechándose de ChatGP. Aseveración que David Sacks, asesor especial del ejecutivo estadounidense en materia de IA (el llamado «Zar de las criptomonedas y la IA»), respaldó: *«hay una técnica en la inteligencia artificial llamada destilación (...) cuando un modelo aprende de otro modelo y de alguna manera absorbe el conocimiento del modelo original. (...) Existen pruebas sustanciales de que lo que DeepSeek hizo aquí es extraer el conocimiento de los modelos de OpenIA»*. Pero como se ve de las palabras de Trump, no se discute sobre Derecho sino de geopolítica. Lo que presuntamente ha hecho DeepSeek es ilegal porque perjudica a los intereses estadounidenses, lo que hagan las compañías de EEUU será correcto en la medida que permita el establecimiento del patrón IA que desea su gobierno.

Esta vertiente del *too big to fail*, política y económica, se complementa con una segunda, que es el grado de implantación entre los usuarios. Una estrategia habitual de captación de clientes es ofrecer un producto a un precio inferior al coste o en mejores condiciones para crear un público fiel y luego, una vez se ha generado dependencia, encarecerlo o empeorar sus características.

Se ha visto estos años con Netflix, Disney +, Amazon Prime, Paramount +, o Max. El precio del servicio se ha incrementado, se ha limitado el uso de cuentas compartidas, o se han incluido anuncios en los planes de pago más básicos, cuando el punto de partida fue mucho más ventajoso. En relación a los modelos de inteligencia artificial generativa, en varios momentos sus usuarios han apreciado un descenso en la calidad de las respuestas pasados varios meses desde su lanzamiento. Más allá de picos de demanda, se sospecha que estas variaciones podrían ser fruto de la sustitución encubierta de la versión «completa» o «pura» de la herramienta por destilados de rendimiento inferior y menor coste para sus empresas. El lanzamiento de ChatGPT-5 ha estado envuelto en cierta polémica por un motivo parecido, por percibirse

37 Traducción propia.

38 Trump y su *AI Action Plan* indican que deberán evitarse lo que denominan como fuentes «woke».

que el enrutador que incorpora tiende a llevar al usuario a las opciones más baratas para OpenIA. Aun así, en las IAs generativas todavía predomina la seducción, siendo estas medidas un paliativo para compensar precisamente los costes de esta etapa de atracción.

Desde este punto de vista, no es ni mucho menos casual que ChatGPT y el resto de los grandes modelos LLM hayan contado con opciones gratuitas robustas, como forma de favorecer una penetración y consolidación acelerada en el mercado.

ChatGPT fue la plataforma de Internet que más rápido alcanzo los 100 millones de usuarios, hasta que Threads de Meta le desbancó seis meses después. Tardó menos de 60 días. Un récord que nunca se hubiera logrado de haber establecido un precio de utilización acorde a los gastos que supone sufragar el servicio. OpenIA, como sus rivales, asumen no obtener beneficios durante años y soportar cuantiosas pérdidas por las expectativas de lograr en el futuro ingentes ganancias y una posición de dominio en el sector, con la capacidad de control que lleva asociada. Pero además de eso, les interesa también que la necesidad creada lleve a que sea más difícil dar marcha atrás en cuanto al estándar que desean establecer.

En estos momentos, hay una honda preocupación por los *deep fake*, que si bien existían antes de que diera comienzo la era GPT, se han expandido drásticamente con la generalización de estas herramientas, que simplifican enormemente su realización. Dinamarca, por ejemplo, está redactando una ley que modificaría el régimen de propiedad intelectual para incluir en él la voz, el rostro, o el cuerpo de todas las personas consideradas individualmente. Ahora bien, de haber sido más gradual el acceso a los modelos de inteligencia artificial generativa tal vez se estarían debatiendo otras respuestas más restrictivas[39], porque la mayoría de la sociedad no disfrutaría de sus beneficios, pero sí que estaría expuesta a sus externalidades negativas.

Sam Altman admitió en enero de 2025 que la modalidad de ChatGPT Pro, cuyo precio es de 200 euros al mes, no cubría los costes asociados y generaba números negativos. En esta misma línea, en abril de 2023, se calculaba que OpenIA invertía 700.000 euros al día solo para mantener la aplicación en funcionamiento en su versión ChatGPT 3, a lo que debería sumarse el gasto pendiente de amortización por la creación en sí del programa y el pago de

39 Matizar que con esto no quiero decir que fueran necesariamente más adecuadas, solo que la sociedad estaría más abierta a medidas que siendo usuaria va a encontrar más inconvenientes. Generalizar un servicio es una forma de hacer corresponsable en cierta manera a más personas que los propios titulares del mismo. Un caso es el de los *riders*. No es lo mismo ser preguntado en abstracto sobre cuál es la fórmula laboral adecuada para la realización de esta actividad que cuando se está obteniendo un beneficio de ella. Aunque siempre va a haber personas dispuestas a sacrificar su interés propio en un asunto concreto, para ser fiel a sus ideales, otras se priorizaran, incluso si eso supone ir en contra de las ideas que normalmente defiendan de forma genérica.

su personal. De acuerdo con los cálculos facilitados por OpenIA, su previsión es empezar a generar beneficios en 2029, es decir, 7 años después del lanzamiento general de GPT-3, 11 tras GPT-1 y 14 desde su fundación. Unas estimaciones propias, en cualquier caso, que han de observarse dentro de su contexto, que es la captación de fondos. Por lo tanto, pudiendo llegar a cumplirse, no es improbable ni mucho menos que, incluso si mantiene una tendencia favorable, OpenIA requiera más tiempo para rentabilizar su gama de productos y que haya «endulzado» los números, quedándose con los pronósticos más optimistas. Sin valorar las repercusiones que puede tener la cada vez mayor competencia de sus rivales, que han salvado o incluso superado la ventaja técnica que mostró ChatGPT en noviembre de 2022.

Con estos datos, se capta que la prioridad de la organización ha sido formar una base de usuarios lo más amplia posible y no la rentabilidad a corto o medio plazo. Algo que ya se vio con Spotify, que arrastró pérdidas durante 18 años hasta 2024; año en el que consiguió un beneficio neto de 1.138 millones de euros. Lo más cerca que había estado antes de registrar un beneficio neto fue en 2021, cuando alcanzo un resultado de -34 millones de euros. Sin embargo, a pesar de estas pérdidas, Spotify no tardó en marcar las reglas del juego dentro del campo musical.

Al crear una comunidad que da por hecho lo que le ofreces, cualquier cambio va a enfrentar más resistencias. Porque aquello a lo que se ha acostumbrado es el *statu quo* y eso le dota de una inercia fuerte que no resulta sencillo alterar. Para cuando algunas autoridades en materia de protección de datos empezaron a suspender ChatGPT, la aplicación recibía 1.600 millones de visitas al mes. En ese marco, una paralización prolongada y estricta del servicio hubiera sido quimérica. Todo ello sin entrar a valorar el uso de VPN u otros mecanismos que permiten eludir las restricciones regionales.

Sin atribuir una actitud maquiavélica, no me parece disparatado afirmar que los principales proveedores de soluciones IA generativas, en pos de su negocio, han buscado la complicidad de la sociedad a través de su implementación masiva y explotando su vertiente más lúdica y aparentemente inofensiva. La moda de convertir las fotos personales a un supuesto estilo Ghibli o representarlas como muñecos de acción tiene una escasa rentabilidad económica, pero contribuye a dar forma a la percepción de la tecnología. Se trata, en definitiva, de que el ciudadano sea participe de la herramienta, como si formara parte del equipo creador, y que entienda que cualquier «ataque» a la misma va también contra él, como cliente[40]. Que sea impopular debatir sobre los condicionantes legales que perjudiquen a quienes proporcionan la

40 Que se busca esa implicación emocional más allá de lo utilitario se percibe en los discursos que plantean la idea de que con una IA se trabaja en equipo. Fuera de la literatura, rara vez se ha escuchado a nadie decir «Google y yo hemos hecho», «la calculadora y yo hemos hecho», o cualquier otra fórmula de ese tipo, en cambio con las IA sí que es fácil leer o escuchar a personas referirse de forma personificada a la IA.

aplicación. Que la llamada a la legalidad sea percibida como el vecino pesado al que le molesta el ruido en Nochevieja y quiere que pare la fiesta.

No obstante, nada de esto afecta de *lege lata*, como se ha indicado reiteradamente, a salvo de la previsión del artículo 3.1 del Código Civil, que llama a interpretar las normas de acuerdo con la realidad social del momento. El *too big to fail,* en sus dos facetas, es parte del contexto que ha de tenerse en cuenta (de hecho, en EE. UU., al analizar el *fair use*, la utilidad colectiva de un uso transformativo suele ser un argumento de peso para apreciar la aplicabilidad de la doctrina), sin embargo, el límite al final es la letra de la ley, que admite interpretaciones, pero no transmutaciones o desnaturalizaciones[41].

Refugiarse en el *too big to fail* sin someter esta argumentación al «careo» de la ciudadanía, no es correcto en un Estado democrático de Derecho. Debe darse oportunidad para cuestionar si no hay realmente margen para tropezar; o si de verdad un determinado enfoque regulatorio provocaría las caídas que se auguran; o si no hay otros valores que igualmente son demasiado grandes como para que podamos permitirnos perderlos en el camino. Y para esto, el proceso de elaboración de las normas, legal o reglamentario, representa el cauce idóneo.

Reino Unido, no sin polémica, está abordando una posible modificación legal de su régimen de propiedad intelectual. El 17 de diciembre de 2024 el gobierno del país lanzo una consulta pública que incluía cuatro opciones. La primera de ellas consistiría en mantener la regulación tal como está, sin introducir cambios. La segunda se orienta a reforzar la protección de los derechos de propiedad intelectual, exigiendo la solicitud de una licencia previa para el uso de contenido protegido. La tercera propone regular una excepción amplia de minería de textos y datos, que permita la extracción de datos sobre obras protegidas sin el permiso de los titulares de los derechos. La última constituye una variación de la anterior, configurando de forma más limitada la excepción, de tal manera que los titulares de derechos puedan hacer una reserva de los mismos que excluya el entrenamiento si no se alcanza un acuerdo. Dando libertad a la hora de seleccionar cualquiera de ellas, en la propia información facilitada de la consulta, el Gobierno de Keir Stramer mostró su preferencia por la tercera y la cuarta.

41 La publicación en redes sociales del Presidente Trump reaccionando a la decisión de un tribunal federal de apelaciones es muy representativo de la corriente que pretende dotar de valor legal al *too big to fail.* En su mensaje, Trump declaró que, si se permitía aplicar el fallo, el país quedaría literalmente destrozado, por lo que implícitamente solicitaba la colaboración del Tribunal Supremo para frenar la decisión de la Corte de Apelaciones y del resto de instancias que han estimado que debería haber sido en todo caso el Congreso de los EE. UU. quien aprobara los aranceles y no el Presidente mediante orden ejecutiva. Aun no versando sobre inteligencia artificial, la filosofía es la misma. Para ciertas personas, si una materia es lo suficientemente importante es trabajo de los jueces o de los operadores jurídicos buscar la manera en que se pueda aplicar como desean, lo permita o no el margen interpretativo de las normas.

Antes, entre el 29 de octubre de 2021 y el 7 de enero 2022, el ejecutivo de Boris Johnson ya había realizado una consulta pública en esta línea[42] y, a raíz de la misma, expresó la intención de proceder a una reforma legal que extendiera las excepciones vinculadas con la minería de datos. El plan se abandonó de facto, no obstante, en 2023 —con Rishi Sunak en Downing Street 10— ante las críticas recibidas desde las industrias culturales y la dificultad para alcanzar un acuerdo con las tecnológicas, que sirviera para el establecimiento de un código voluntario de derechos de autor en materia de IA.

Como se dejaba apuntando en el epígrafe anterior, el proceso legislativo lleva aparejados la celebración de debates, los cuales tienen derivadas que no siempre cabe prever. Mientras se tramitaba la *UK Data (Use and Access) Bill*, se planteó la posibilidad de incluir enmiendas que impusieran obligaciones de transparencia en cuanto al entrenamiento de IAs, de corte similar a las del artículo 53 del Reglamento (UE) 2024/1689, de 13 de junio, así como otras previsiones dirigidas a reforzar a los titulares de derechos de propiedad intelectual frente a la utilización no consentida. El objeto de la *UK Data (Use and Acces) Bill* no era disciplinar la inteligencia artificial desde la perspectiva de la propiedad intelectual —es más, el texto final no lo ha hecho, o como mínimo no de forma destacada; sería en todo caso una regulación más por omisión que por acción—, sin embargo, es tal el interés que hay por la cuestión, que la discusión saltó de lleno, aunque los promotores de la ley no tuvieran esa intención inicialmente.

En el proceso de deliberación, se han involucrado desde celebridades como Dua Lipa o Elton John pasando por figuras políticas y empresariales como Nick Clegg, antiguo líder de los Liberales Demócratas, ex–Vice Primer Ministro y Presidente de Asuntos Globales en Meta hasta 2025. Ha habido opiniones enfrentadas —y las volverá a haber cuando se aborde, esta vez sí, la hipotética actualización de la normativa inglesa de propiedad intelectual— pero todo eso es inherente al trabajo legislativo. Y además de ser intrínseco, es positivo, porque permite la contraposición de ideas. Que es justo lo que ha propiciado la tramitación de la *UK Data (Use and Access) Bill*.

Cabe remarcar que en las dos consultas públicas inglesas no se indicó que hubiera un vacío normativo. No aconsejándose por ninguno de los gobiernos que se han sucedido en Reino Unido desde el año 2021, la opción 0, mantener las cosas como están, es reconocida como una posibilidad real y tangible. Aunque las consultas, como es lógico, solo se refieren al ordenamiento inglés, la presencia de un escenario 0 es constatable de forma prácticamente general y, sin duda, se cumple en los Estados miembros de la Unión Europea.

42 Que se busca esa implicación emocional más allá de lo utilitario se percibe en los discursos que plantean la idea de que con una IA se trabaja en equipo. Fuera de la literatura, rara vez se ha escuchado a nadie decir «Google y yo hemos hecho», «la calculadora y yo hemos hecho», o cualquier otra fórmula de ese tipo, en cambio con las IA sí que es fácil leer o escuchar a personas referirse de forma personificada a la IA.

Tal como se ha dicho en el capítulo anterior, existe un Derecho totalmente aplicable y es a este al que debemos encomendarnos para resolver el interrogante que se formula en el título del capítulo: ¿cuál es el impacto del derecho de la propiedad intelectual en el entrenamiento de los modelos de inteligencia artificial generativa?

En el caso de Reino Unido, se trasluce de forma tácita que el marco vigente no da la respuesta que desearía el Gobierno en la voluntad de emprender una reforma. Siendo así, el camino que han escogido es el adecuado. El fondo de lo que pretenden podrá gustarnos más o menos, dependiendo de la opinión que tengamos[43], pero no resulta objetable en cuanto a la forma. Cuando el Derecho se reputa insatisfactorio por quienes representan a la sociedad en virtud de un proceso democrático, lo que debe hacerse es intentar modificarlo, no actuar como si ya dijera lo que les gustaría. Si se tienen las mayorías precisas, el cambio tendrá lugar. En caso contrario, no y se mantendrá lo vigente. Desde ese punto de vista, las autoridades inglesas estarían cogiendo el toro por los cuernos.

Lo mismo podría decirse de las españolas, cuando presentaron el cuestionado proyecto de Real Decreto por el que se regula la concesión de licencias colectivas ampliadas para la explotación masiva de obras y prestaciones protegidas por derechos de propiedad intelectual, para el desarrollo de modelos de inteligencia artificial de uso general.

Estos elogios, con todo, han de matizarse, puesto que no se termina de mirar de frente al elefante en la habitación. Los promotores de ambas normas, como poderes ejecutivos encargados de velar por la aplicación de las leyes, tendrían que estar adoptando algunas determinaciones mientras no se actualice el ordenamiento jurídico de la forma que pretenden[44]. Si ellos mismos estiman que ahora no hay una excepción que ampare ampliamente el sistema de «entrenamiento» masivo, utilizado en el desarrollo y soporte de los modelos de las IAs generativas y razonadoras, esto ha de acarrear necesariamente alguna consecuencia. De lo contrario, no serían precisas las reformas que impulsan o valoran impulsar. Sobre todo, en el caso inglés da la impresión de que al toro se lo agarre únicamente para legalizar a *posteriori* la desidia del presente.

Poniendo el foco en el Reglamento (UE) 2024/1689, de 13 de junio de 2024, sus considerandos 104, 105 y 106 dicen claramente que el resumen del con-

43 Personalmente, no es el sistema por el que abogó. Sin embargo, como parte del juego democrático, creo que es necesario acatar toda norma que se haya tramitado por los cauces legales preestablecidos y cuyo contenido no quiebre ni la Constitución ni el resto de disposiciones legales jerarquicamente superiores a ella.

44 Cabe criticar en ese sentido los rumores de que el Gobierno laborista de Stramer estaría valorando alcanzar un acuerdo con OpenIA para proporcionar a todos los ciudadanos la versión *Premium* de la Aplicación, sufragando su coste. Estaría actuando como un particular, privilegiando a un modelo frente a otro, y eludiendo las controversias legales de las que no pueden desligarse como poder ejecutivo.

tenido utilizado para el «entrenamiento» de los modelos de inteligencia artificial, previsto en la letra d) de su artículo 53.1, tiene como finalidad comprobar el respeto del Derecho de la Unión en materia de derechos de autor y derechos afines, y, en particular, garantizar que se ha respetado la reserva de derechos prevista en el artículo 4, apartado 3, de la Directiva (UE) 2019/790, de 17 de abril, del Parlamento Europeo y del Consejo, sobre los derechos de autor y derechos afines en el mercado único digital. Esta obligación enlaza, a su vez, con la de la letra anterior del precepto, que impone a los proveedores de modelos de IA de uso general la fijación de directrices para cumplir con este mismo entramado jurídico en materia de derechos de autor y afines.

El Reglamento en este aspecto nos sitúa en el escenario 0, tomando de base la nomenclatura de las consultas públicas de Reino Unido. No incorpora requisitos de fondo novedosos en el plano de los derechos de autor y afines, sino que remite a las normas preexistentes que, a diferencia del Reglamento europeo, se hallan plenamente en vigor. Lo que introduce son mandatos de tipo formal para facilitar la realización de controles en aplicación de esas leyes. Por eso, aunque hasta el 2 de agosto de 2026 no sean imperativas las secciones 1 y 2 del Capítulo V del Reglamento, poco importa al objeto de estas páginas, porque el canon de legalidad de fondo depende de normas externas al Reglamento en lo que se refiere a los derechos de autor, así como en otras áreas donde se apoya en disposiciones jurídicas previas, nacionales o europeas.

En el caso de la Directiva (UE), 2019/790, de 17 de abril, citada en el artículo 53, el plazo de transposición concluyó el 7 de junio de 2021, por lo que o bien su contenido se ha asimilado correctamente, o rige en parte mediante eficacia directa. España, en concreto, interiorizó su contenido tardíamente en el texto refundido de la Ley de Propiedad Intelectual, por medio del Real Decreto-Ley 24/2021, de 2 de noviembre.

Examinando el Código de Prácticas de Inteligencias Artificiales de Propósito General, de 10 de julio de 2025, que preparó un grupo de expertos independiente bajo los auspicios de la Comisión Europea y la Consejo Europeo de Inteligencia Artificial, se ratifica la impresión de que el Reglamento se centra en garantizar que se cumple lo que ya existe. De acuerdo con el considerando c) del capítulo sobre *Copyright* del Código, los firmantes reconocen que marco de la propiedad intelectual y otros derechos conexos emana esencialmente de las Directivas 2001/29/CE, (UE) 2019/790 y 2004/48/CE. Las medidas propuestas en esta materia, por su parte, se centran en garantizar aspectos concretos de las normas citadas, lo que contribuye a fortalecer todavía más la idea de que no se está obligando a nada nuevo de fondo (lo que supone que ya deberían haberlo cumplido, con o sin Código, con o sin Reglamento)[45].

45 La primera de las medidas insta a los suscriptores del Código a elaborar un documento

Una y otra vez nos encontramos ante la evidencia de que había un Derecho aplicable al «entrenamiento» antes de que ChatGP se diera a conocer masivamente, y que este Derecho no ha cambiado después, o apenas. Sin embargo, clarificado eso, lo que escasea es la concreción, la materialización de sus implicaciones prácticas. En los nuevos textos legales de *hard* y *soft law*, como el Reglamento o el Código voluntario de prácticas, se desprende como mucho cierta idea de lo que sería o hubiera sido para el legislador lo ideal, pero sabemos que no ha sido así, a veces porque las propias compañías lo han admitido. ¿Qué es lo que ocurre entonces?

¿El «entrenamiento» que incluye materiales protegidos por derechos de propiedad intelectual está amparado por la legislación y ha sido siempre válido, aún sin consentimiento de sus titulares? O, en cambio, ¿carece de cobertura y es una práctica fuera de la Ley? Sabemos que ha de ser una de las dos opciones, pero pocas autoridades se han decantado por una de ellas, colocándose más bien de perfil.

en el que se plasme su política en materia de derechos de autor, englobando este la implementación concreta del resto de medidas recogidas. La segunda se dedica, por un lado, a evitar que en el *scraping* se eludan las barreras técnicas que habilita el artículo 6 de la Directiva 2001/29/CE de 22 de mayo, y, por otro, de forma más general, a evitar el recurso a a webs identificadas como infractoras reiteradas en materia de derechos de propiedad intelectual y conexos (dicho de forma más coloquial, que no se recurra a las clásicas páginas con contenidos piratas). La tercera da algunos detalles sobre cómo realizar el rastreo y extracción de datos en Internet respetando el precitado artículo 4.3 de la Directiva (UE) 2019/790, de 17 de abril. La cuarta y la quinta posiblemente sean las más novedosas de todas, pero la base legal continúa siendo la misma en el plano sustantivito. La cuarta llama a introducir salvaguardias que técnicamente impidan o dificulten a los usuarios obtener resultados contrarios a los derechos de autor (aunque no lo menciona el texto del Código, un ejemplo sería limitar la capacidad de generar personajes o marcas preestablecidas). Asimismo, como submedida, dentro de la cuarta, también se plantea que los términos y condiciones de uso de los modelos prohíban expresamente cualquier acción que infrinja los derechos de autor y otros afines. Por último, la medida 1.5 configura el establecimiento de puntos de contacto para que los titulares de derechos puedan presentar sus reclamaciones a los firmantes, como una primera vía de resolución de conflicto pacifica, que no cierra la puerta a los canales administrativos o judiciales convencionales. Haciendo un análisis de conjunto se percibe que los expertos han considerado que las IAs potencialmente presentan riesgos mayores que otras tecnologías en relación al *Copyright*. No se entendería de lo contrario la figura de los puntos de contacto; si previeran que los autores apenas van a presentar objeciones a los firmantes, no sería necesario un instrumento así. Un punto de contacto solo tiene sentido cuando se presume que los choques van a ser frecuentes y que los titulares necesitarán saber a dónde acudir. En esta línea, hay muchas herramientas lícitas que pueden ser utilizadas también propósitos ilegales, pero no es habitual que se responsabilice de estos usos incorrectos al vendedor o al fabricante. De nuevo, los expertos parece que están anticipando que los riesgos se materializaran y no de forma excepcional (probablemente, porque es lo que ha ocurrido de estos años). Así que juzgan preciso asumir precauciones adicionales para que esos usos ilícitos no sé den, como las de la medida de mitigación 1.4.a.

Como no puede ser de otra forma, las Administraciones Públicas habrán de seguir la pauta que defina la jurisprudencia[46]. Ahora bien, no sabiendo cuando se resolverán los casos que se han y se van formulando, ni el impacto que finalmente desplegaran sobre la doctrina jurídica, sería conveniente que el sector público fijara entretanto un criterio propio, mientras aguarda el del poder judicial, en vez de tender a un acriticismo que no encaja con sus características, sino con las de un ciudadano ordinario sin deberes cualificados, cuya opinión es totalmente libre.

Bajo el paradigma constitucionalista europeo, el principio de legalidad vincula positivamente a los poderes públicos, de tal manera que solo pueden hacer aquello que de manera expresa les habilita el Derecho. Hay un plus de sometimiento frente a la Ley que exige extremar las precauciones en su actuación. Esto supone que, en contraste con los operadores privados, a los públicos no les puedan resultar indiferentes las controversias jurídicas que despiertan las IAs en relación con la propiedad intelectual, pero también respecto a otras áreas, como puede ser el derecho al honor de las personas o la defensa de las libertades fundamentales.

Haciendo un pequeño excurso, que retoma algunas de las ideas del epígrafe anterior del capítulo, piénsese en los casos en los que la respuesta de una inteligencia artificial generativa atribuye erróneamente hechos negativos o incluso delictivos. O en algunos incidentes que ha sufrido Grok, por citar un modelo concreto. Como que, en mayo de 2025, incluyera comentarios sobre un supuesto genocidio blanco en Sudáfrica, no habiendo sido consultada sobre el país africano o sobre cuestiones políticas. O que el 9 de julio de 2025, Grok se «autoidentificara» como «MechaHitler», tras una actualización fallida que buscaba alinear más sus respuestas con la visión personal de Elon Musk. ¿Cuál sería la reacción de los responsables públicos de consumo ante cualquier otro producto que registrara con relativa frecuencia problemas de este tipo, que afectan a sus funciones principales?

Sorprende la parsimonia con la que se acogen estos errores, a pesar de que su entidad no es escasa, ni social ni legalmente. Que una IA proponga a Hitler como un referente, o que exprese el «deseo» de que barrios específicos de una ciudad se consuman por las llamas, son declaraciones que podrían constituir un ilícito penal en boca de un ser humano.

Del mismo modo, aunque se advierta cómo descargo de legal que el programa puede sufrir alucinaciones, que una IA recomiende usar pegamento

46 Comienza a dictarse algún pronunciamiento aislado, como el de la sentencia del Tribunal Regional de Hamburgo, Sala Civil 10, de 27 de septiembre de 2024 —que se analizará más adelante—, pero no son tantos ni vienen de las máximas instancias judiciales de los países europeos, de modo que no puede hablarse todavía de una línea jurisprudencial definida y firme. Sentencias como la del Tribunal Regional de Hamburgo pueden quedar en anécdota o convertirse en la base sobre la que se edifique la futura exégesis en la materia. El tiempo lo dirá.

para favorecer la consistencia de la masa de una pizza (cómo le ocurrió a una primera versión de la función *AI Overview de Google*[47]), plantea posibles peligros para la salud pública cuando el defecto no sea tan obvio de constatar. Por ejemplo, si hablara de un fruto silvestre venenoso como comestible. Por lo tanto, la legislación sanitaria y alimentaria entrarían en juego, del mismo modo que se interviene ante personajes pintorescos como Josep Pamiés, por difundir la ingesta de lejía como tratamiento para varias enfermedades.

La protección de datos, una vez más, tendría mucho que decir ante el filtrado de 100.000 conversaciones de ChatGPT, que fueron indexadas para su consulta por los navegadores de Internet, tras utilizar una función experimental de la aplicación. Perplexity y Grok habrían experimentado un mismo problema similar, siendo en el último caso incluso de mayor magnitud, al estimarse que se habrían difundido públicamente alrededor de 370.000 chats.

Entre las prácticas prohibidas por el artículo 5.1 del Reglamento europeo 2024/1689, de 13 de junio de 2024, figura *«la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas»*. Se trata del primer supuesto del listado y en principio rige desde el 2 de febrero de 2025, de acuerdo con su artículo 113, aunque de forma general podría entenderse que estas prácticas ya se hallaban prohibidas como parte del escenario 0. Sin embargo, ninguna reacción ha habido por el momento a que el diseño de las herramientas de IA más populares aproveche los datos de sus usuarios para adaptarse al máximo a su carácter, favoreciendo la creación de lazos emocionales con las «máquinas»[48].

A título ejemplificativo, el gabinete de prensa de Microsoft explica Copilot de este modo en una comunicación de Internet: *«un compañero de IA es algo*

47 Las anécdotas son prácticamente infinitas. Lucie, un modelo desarrollado por el Grupo Linagora y el Consorcio OpenLLM France, con el respaldo del Gobierno francés, señalaba que los huevos de vaca eran un alimento nutritivo, además de ofrecer recetas para preparar metanfetaminas.

48 Se registran cada vez más casos de personas presas del efecto Eliza en sus interacciones con los modelos IA generativos o razonadores. No es extraño encontrar mensajes donde algunos consumidores se refieren a las IAs como un amigo o una pareja mejor que las humanas. Del mismo modo, ante el reseteo de la configuración de un modelo, debido a cambios de versión o por borrarse el historial de conversaciones previas, algunas personas han manifestado sentimientos de pérdida o duelo.

completamente personal; se construye en torno a las necesidades, valores y expectativas individuales. Por eso, en última instancia, habrá tantos Copilots diferentes como personas que los utilicen. Hoy, estamos embarcándonos en el viaje que hará que Copilot deje de ser "un" compañero de IA para convertirse en "tu" compañero de IA»[49]. Hay una narrativa que intenta trasmitir la idea de que las IAs son «colegas» o «amigos» y no simples herramientas, insertándose esta publicidad en esa línea.

Igual que en las redes sociales, los algoritmos en este caso están orientados a favorecer el uso recurrente mediante retroalimentación, pero aquí el perfil sobre el que trabajan se hace con mucho más detalle, encontrando y explotando patrones que incluso pueden ser desconocidos por la persona a la que se dirigen, al tratarse de elementos sutiles e inconscientes. Es algo que supone mucho más que ir ajustándose a sus necesidades o gustos. Si «capta» el programa IA que hay una serie de construcciones o términos a los que el usuario asocia connotaciones positivas, en sus respuestas los incluiría en un sofismo del lenguaje. Cuanto más se utilice la herramienta, con mayor precisión trazará la radiografía de la persona. Estas variaciones en la expresión formal pueden alterar por completo la percepción del contenido sustantivo de la respuesta y por lo tanto constituir una forma de manipulación no tan diferente a mostrar una imagen muy rápido sin que se pueda procesar racionalmente.

En España, la Ley 34/1988, de 11 de noviembre, General de Publicidad prohíbe la publicidad subliminal, construyéndose el concepto en torno a la idea clave de que pueda ser percibida. Quizá la definición del concepto que plantea su artículo 4 no se ajuste del todo a la situación que acabo de describir, al estar más pensada para lo visual que para lo escrito, pero tampoco sería imposible encontrarle acomodo. Como mínimo, con una norma así, lo lógico sería que los poderes públicos se hicieran algunas preguntas sobre al riesgo de una personalización que trasciende de la comodidad y que no queda tan lejos de la captación de la voluntad en los supuestos más extremos.

La novedad parcial de la tecnología genera desconcierto y esto, a su vez, parálisis. Pero esa no es la única razón de que la actitud de los Estados haya sido predominante contemplativa, con excepciones limitadas, como las que han protagonizado las autoridades de protección de datos. En la línea del *too big to fail*, se ha asumido la inevitabilidad de la IA, en singular. Solo cabría el *laissez faire*, *laissez passer*. Como mucho se podría matizarla, pero pretender alterarla de forma significativa a cómo es, o cómo se ha construido, se percibe cómo algo imposible. Ha calado la narrativa de que, igual que el mármol de Miguel Ángel, había una IA oculta que liberar, y el artesano solo contribuye a manifestarla, no a moldearla.

49 Se puede consultar en la siguiente dirección: https://news.microsoft.com/source/emea/features/tu-companero-de-ia/?lang=es.

Esto naturalmente no es verdad, ni tampoco lo era en el caso del famoso artista italiano. Era él y no el mármol quién decidía la forma que tomaban sus esculturas, del mismo modo que no hay una IA univoca, sino múltiples opciones que cabe explorar, tal y como apuntan los premios nobeles de economía Daron Acemoglu y Simon Johnson. Las IAs que disponemos en estos momentos son fruto de la suma de una serie de decisiones susceptibles de modificarse. Las IAs generativas que priorizan la verosimilitud a la exactitud podrían haberse configurado a la inversa. Hay una elección de diseño, con sus ventajas e inconvenientes.

El Reglamento (UE) 2024/1689, de 13 de junio, se abre declarando que como norma busca *«promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, proteger frente a los efectos perjudiciales de los sistemas de IA en la Unión, así como brindar apoyo a la innovación»*. Más allá del uso cuestionable del singular, que persiste en la equivocación de retratar a la IA como algo único y no polifacético, la idea clave que trasmite es que las IAs pueden ser moldeadas por las decisiones de la sociedad y, en concreto, de los legisladores. Por eso, hay que perder el miedo a ponderar, ni siquiera afirmar, que las IAs de moda pueden no ser intachables jurídicamente. Y una de esas eventuales tachas podría recaer en la gestión que han hecho, en cuanto al «entrenamiento», de la propiedad intelectual y de los datos personales de los usuarios disponibles en Internet, protegidos estos últimos a través de derechos como los del artículo 18 de la Constitución, en España.

Según dio a conocer la prensa en 2024, en una audiencia ante el comité selecto digital y de comunicaciones de la Cámara de los Lores de Reino Unido, OpenIA mantuvo que *«los derechos de autor cubren hoy prácticamente todo tipo de expresión humana —incluyendo publicaciones de blogs, fotografías, publicaciones en foros, fragmentos de código de software y documentos gubernamentales—»* y *«sería imposible entrenar los principales modelos de IA actuales sin utilizar materiales protegidos por derechos de autor»*. Concluía la compañía, asegurando que *«limitar los datos de entrenamiento a libros y dibujos de dominio público creados hace más de un siglo podría producir un experimento interesante, pero no proporcionaría sistemas de inteligencia artificial que satisfagan las necesidades de los ciudadanos de hoy»*. Junto a estos argumentos facticos, legalmente apeló a la doctrina del *fair dealing*, correlato de la del *fair use* de EEUU.

De estas aseveraciones, la única que posee relevancia directa en el plano legal es la última. El resto no son intrascendentes, porque no pueden entenderse las normas sin el contexto que las rodea, pero tienen mucho menos peso, ya que no es imposible recabar el consentimiento, solo inconveniente, laborioso, o más costoso para los fabricantes de IAs. Además, son los propios actos de las proveedoras de IAs los que desmienten esta presunta imposibilidad.

Google llegó a un acuerdo con Reddit, para utilizar las publicaciones y los comentarios de sus usuarios en el «entrenamiento» de sus modelos, pagando 60 millones de euros al año. El grupo editorial Axel Springer pactó una colaboración similar con OpenIA. Y como estos, ha habido otros muchos estos últimos años, no dejando de crecer su número. Por su parte, X en sus términos y condiciones prohíbe que otras compañías se beneficien de su repositorio de datos, los cuales se reserva en exclusiva para Grok y el resto de productos de la familia de X. Si estuvieran seguros de la viabilidad del paraguas legal que invocan, no estarían pagando para proveerse de contenidos (Claude se sirvió de Reddit sin negociar, por ejemplo, lo que ha llevado a que los propietarios del foro demanden a Anthorpic), o incluyendo cláusulas, como la de X, que sabrían carentes de valor. Por supuesto, tampoco habrían insistido tanto a los Gobiernos de EEUU y de UK acerca de la necesidad de reconocer como *fair use* o *fair dealing* el «entrenamiento» de IAs.

Esto no significa automáticamente una confirmación de la tesis de los críticos con el sistema de «alimentación» de las IAs. Los acuerdos podrían justificarse como una acción preventiva en el caso de que los tribunales no acepten sus planteamientos, o como una forma de sortear las «verjas» que se están levantando en el nuevo Internet ante el *scraping* o extracción de datos. Pero lo que es evidente es que si ese escenario lo vieran muy improbable no estarían abonando millones a estos grandes propietarios de contenidos. Las compañías saben que como mínimo hay brumas legales y que se lanzaron a la piscina sin comprobar si había agua, por una combinación de ahorro de costes y el azogue de no llegar más tarde que sus competidores.

En los primeros compases de la carrera de la IA generativa, Adobe Firefly presumía de ofrecer mayor seguridad jurídica, por haberse entrenado tomando solo datos licenciados. A cambio, sus resultados eran, a juicio de la mayoría de sus clientes, inferiores a los que lograban las herramientas de la competencia, que habían tomado referencias sin plantearse ningún límite en cuanto a de donde extraían la información. Si bien hay cierta polémica en cuanto a si realmente el diseño de la herramienta de Adobe Firefly se restringió a las fuentes que enumeran, como mínimo parece que habrían sido más propias que externas, ya que ante peticiones como «haz un fontanero con un mono azul y rojo, de origen italiano, con bigote y algo retaco» DALL-3 reproducía una versión apócrifa de Super Mario, mientras que el programa de Adobe creaba lo que haría una persona que jamás hubiera escuchado hablar del videojuego y el personaje de Nintendo.

Las conversaciones internas filtradas de varias compañías o los testimonios dados en algunos de los procesos judiciales que hay en curso muestran que la propiedad intelectual siempre fue para ellos un incómodo convidado de piedra. Del mismo modo, en el plano institucional encontramos textos como la primera consulta de Reino Unido o la resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial,

que, entre otras cuestiones, subraya: *«la necesidad de que la Comisión tenga como objetivo proporcionar una protección de la propiedad intelectual que sea equilibrada y se base en la innovación en beneficio de los desarrolladores europeos de IA, fortalezca la competitividad internacional de las empresas europeas, también ante posibles tácticas abusivas en los litigios, y garantice la máxima seguridad jurídica a los usuarios, especialmente en el marco de las negociaciones internacionales y, en particular, por lo que se refiere a los debates en curso sobre la IA y la revolución de los datos en el seno de la OMPI»*. Con estos antecedentes cabe plantearse por qué no se ha abierto antes el debate sobre los derechos de propiedad intelectual, si nunca hubo un vacío normativo y se era consciente de la presencia de ese Derecho susceptible de aplicarse. La respuesta es sencilla: principalmente porque hasta 2022 no existía riesgo de colusión comercial, o no lo percibían los potenciales perjudicados.

No dejan de aparecer informes que pronostican la inminente desaparición de profesiones, o la previsible reducción de la demanda de personal dentro de ellas. Pudiendo ser interesados estos presagios, como forma de mantener el entusiasmo o *hype* bursátil y seguir captando financiación, es incuestionable que las IAs en sus diferentes modalidades inciden sobre el empleo. Al margen del *marketing* desplegado, el diseño de los agentes de IA o las IAs generativas se dirige menos de lo que se publicita a la potenciación de las personas y más a su sustitución, como advierten Simon Johnson y Damon Acemoglu, en contraste con otros sistemas de IA que llevamos utilizando décadas, donde sí que se usan para obtener resultados que sería muy difícil que un ser humano lograría por sí solo. Y en este caso se da la particularidad, frente a otros procesos de automatización previos, que la base para lograr ese desplazamiento está en la propia obra de los desplazados, si no de forma individual sí de manera colectiva.

Dentro de la doctrina del *fair use*, uno de los factores objeto de valoración y que pueden llevar a su denegación es que se vea afectado el mercado del titular de los derechos de propiedad intelectual. En sentido parecido, el tercer paso que integra la regla de los tres pasos del Convenio de Berna promueve esta evaluación. Más allá de lo legal, sin esta incidencia, es sencillo que las personas que ostentan esos derechos no lleguen a percatarse de que se está produciendo una situación en la que quizá podrían reivindicar su posición. O que vean más costoso oponerse, si realmente no se están viendo perjudicados. O directamente que les resulte indiferente.

Cuando los primeros modelos LLM expedían imágenes de baja calidad, concebidas desde un literalismo exacerbado (basta examinar los resultados que daba la instrucción de «cerdo con armadura»[50]), ningún autor temía

50 A veces dotadas, sin embargo, de un surrealismo involuntario que las hacía hasta más interesantes que las actuales más depuradas, pero también más convencionales.

verse sustituido por la tecnología. Y lo mismo ocurría con otras vertientes. Un vídeo creado artificialmente del actor Will Smith comiendo espaguetis se convirtió en motivo de burla, por los fallos grotescos que mostraba, rozando la pesadilla[51]. Las IAs de ese tipo eran una simple curiosidad. Hoy no es así y la preocupación de dónde vienen sus datos es mayor. Mientras que la participación en la primera consulta de Reino Unido fue bastante limitada, en la segunda ha sido abrumadora[52].

El Derecho no tiene capacidad por sí mismo para imponerse. No son las leyes las que se defienden ni las que se aplican, siempre hay personas detrás, públicas o privadas. De hecho, esta es la razón de que junto a la tutela judicial declarativa exista la necesidad de contar además con una vertiente ejecutiva. En algunos casos, basta con que se constate el derecho de la persona, pero en otros, si quienes han de soportarlo se niegan, se ha de imponer, forzar a que la realidad se acomode a lo que el Derecho reconoce.

Durante años, los dos grandes bloques en el conflicto que existe a raíz del «entrenamiento» de las IAs con materiales sujetos a propiedad intelectual carecieron de incentivos para recabar el respaldo de los tribunales. Hubiera o no actos de reproducción, distribución, comunicación o transformación en el proceso de diseño y uso de las IAs, estos no interferían con los que se autorizaban al margen de la IA y que constituían, y todavía constituyen para muchos, la principal fuente de recursos económicos de los titulares de derechos de propiedad intelectual. Por su parte, materialmente, los desarrolladores no precisaban de la colaboración de estos últimos para acceder a sus contenidos, por estar disponibles en Internet de forma pública (lo que no es equiparable a estar libres de derechos o ser de dominio público), o en canales de fácil consulta.

En estos momentos, los dos lados del conflicto sí que requieren de ese auxilio legal o judicial expreso. Los titulares de derechos de propiedad intelectual, que intentan frenar la extracción no consentida, pero también los proveedores de IAs generativas y razonadoras, que son demandados por los primeros y se encuentran cada vez con más muros en su búsqueda de bases de información que utilizar[53].

---

51 Precisamente, por eso, conforme avanzó la tecnología, se usó este segmento como prueba de los avances. Todo ello sin el consentimiento del actor, que lo ha tolerado pasivamente.

52 Incluso en el marco de la primera consulta, se ve cómo el interés fue creciendo con el tiempo. La pauta que extrajo de ella el Gobierno británico sería objeto de mayores críticas en 2023 que el punto de partida en 2021, porque entre tanto también hubo una mayor concienciación de los agentes implicados.

53 Aunque la efectividad de estas medidas es cuestionable, se han dado casos de «envenenamiento» de los datos, con el fin de dificultar el procesamiento de la información. Asimismo, empresas como Cloudflare proporcionan servicios de bloqueo para las páginas web orientados a los bots de recopilación. Esta clase de prácticas podrían considerarse ilegales si se entiende que ya existe un derecho para el «entrenamiento» de IAs, en los

Todos estos factores hacen que lo que hubiera sido percibido en 2019, o incluso 2021, como una reflexión un tanto de laboratorio, hoy posea auténtica trascendencia práctica. Ahora bien, cómo se ha explicado reiteradamente, la base con la que se encara no es muy diferente a la que se habría usado entonces.

La relativa quietud del periodo en el que se «maceraban» las IAs generativas populares en la actualidad nos debe importar poco al enfocar la cuestión jurídicamente. La tolerancia pasiva no es equiparable a un consentimiento tácito. La falta de ejercicio en plazo puede acarrear consecuencias en cuanto a la eventual prescripción o la caducidad de las acciones de defensa o reivindicación, pero no anula la existencia de los derechos en el caso de que efectivamente se dispongan. No se puede perder de vista tampoco que, procesalmente, hasta el último día de plazo resulta viable acudir a los tribunales (cuestión distinta es que no sea conveniente apurar tanto). Basta comprobar cómo Disney y Universal han unido fuerzas para presentar una demanda conjunta contra Midjourney, en junio de 2025, por presunta infracción de derechos de propiedad intelectual.

Las dos empresas sostienen que la herramienta constituye una *«máquina expendedora virtual»* y un *«pozo sin fondo de plagio»* que alumbra *«copias interminables no autorizadas de las obras protegidas por derechos de autor de Disney y Universal»*. En apoyo de su postura, enumeran una serie de personajes de sus respectivos repositorios que son representados sin trabas por Midjourney ante las peticiones de sus usuarios (lo que nos retrotrae al ejemplo del fontanero italiano vs Super Mario, de cómo en un caso ofrecía la imagen del personaje y en otro no, porque no había sido «alimentada» con imágenes del famoso héroe del Reino Champiñón).

Prospere o no la reclamación, a la que se suma una posterior de Warner Bros, lo que reprochan en su escrito a Midjourney lo podrían haber esgrimido desde 2022. Que hayan actuado este año y no antes es fruto del cálculo de las empresas en relación con sus intereses, no de un cambio legal.

En definitiva, ni el *too big to fail*, ni la falta de reclamaciones durante años alteran de forma significativa el análisis legal que corresponde hacer conforme al Derecho vigente. Del mismo modo, contradiciendo las reflexiones del Presidente Trump, tal como se exponía en el primer apartado del capítulo, no podemos apelar a la inspiración humana para juzgar la legalidad del «entrenamiento» de las IAs, puesto que materialmente no nos hallamos ante

términos que plantean OpenIA o Google, y que estas medidas defensivas son un abuso de sus respectivos derechos como autores. Por el contrario, de rechazarse tal prerrogativa, serían plenamente legales, como parte de las facultades dominicales inherentes a la propiedad privada, del mismo modo que los dueños de una casa pueden instalar sistemas de seguridad para impedir o dificultar la entrada de intrusos.

procesos equiparables[54]. No hay ningún peso muerto que bloquee el pedal del acelerador jurídico.

Si el primer apartado se dirigía a separar seres humanos y máquinas, de forma muy modesta, y el segundo a reconocer la existencia de un Derecho vigente, este tercero se ha orientado a constatar que la aplicación de ese Derecho en vigor no es negociable, incluyendo la parte que afecta a los derechos de propiedad intelectual. Por consiguiente, el siguiente paso lógico es ver que dice concretamente el Derecho en materia de propiedad intelectual. La atención se centrará en el Derecho de la Unión Europea, pero también se tratará la jurisprudencia del *fair use*, en la medida en que puede servir de inspiración para los jueces de todo el mundo.

IV. La excepción de minería no basta para excavar esta veta

Como paso previo para determinar si los derechos de propiedad intelectual se han visto afectados de manera indebida por la pauta predominante de «entrenamiento» de los modelos de IA, lo primero que debe hacerse es concretar cuál es el contenido de estos derechos. Por ello, creo conveniente hacer un repaso fugaz de algunas nociones básicas sobre la propiedad intelectual.

Tal como se infiere del nombre, nos hallamos ante derechos de propiedad, igual que los que recaen sobre otros bienes, si bien matizados por su especial naturaleza. El artículo 348 del Código Civil de 1889 define la propiedad como *«el derecho de gozar y disponer de una cosa o de un animal, sin más limitaciones que las establecidas en las leyes»*. El 428 refiriéndose a la propiedad intelectual declara «*el autor de una obra literaria, científica o artística, tiene el derecho de explotarla y disponer de ella a su voluntad*», recogiendo los principales atributos que caracterizan a la propiedad en el 348. El artículo 2 del texto refundido de la Ley de Propiedad Intelectual, aprobado por el Real Decreto Legislativo 1/1996, de 12 de abril, reafirma esta idea al expresar que «*la propiedad intelectual está integrada por derechos de carácter personal y patrimonial, que atribuyen al autor la plena disposición y el derecho exclusivo a la explotación de la obra, sin más limitaciones que las establecidas en la Ley».*

54 Resulta curiosa la teoría que plantean al respecto Mason Kamb y Surya Ganguli en el artículo *An analytic theory of creativity in convolutional diffusion models*, publicado el 28 de diciembre de 2024, y revisado el 5 de junio de 2025. De acuerdo con su visión, la recombinación novedosa de la información utilizada en el «entrenamiento» sería más un «accidente» que algo dirigido. No siendo pacificas sus conclusiones, se ve que la falta de intencionalidad, derivada de la ausencia de verdadera inteligencia, hace que los procesos internos de las máquinas difieran de los nuestros.

La esencia de la propiedad es que el propietario, valga la redundancia, puede decidir con absoluta libertad dentro de los márgenes legales. Positivamente, significa que puede hacer lo que quiera con su propiedad, y, negativamente, que puede impedir a otros que hagan con ella lo que deseen. Los otros se relacionarán con esa propiedad en la medida de lo que él juzgue apropiado[55]. Esto se repite en el caso de la intelectual con importantes singularidades.

En comparación con la noción clásica y general de propiedad, la vertiente intelectual es de acuñación reciente. Durante muchos siglos, los autores poseían pocos o ningún derecho en relación a su trabajo, algo que se refleja en el conocido episodio del Quijote apócrifo de Avellaneda. Hasta llegar al *statu quo* vigente, la legislación pasó por varias etapas, como la de los privilegios reales con los que las revoluciones liberales rompieron, fijando las bases de la formulación moderna. Teniendo en cuenta esto, podemos decir que el Derecho de autor, tal y como se entiende hoy en día, tiene poco más de doscientos años. En el caso español, fue el Decreto CCLXV de las Cortes de Cádiz, de 10 de junio de 1813, titulado «Reglas para conservar a los escritores la propiedad de sus obras», el que plasmó el giro legal. Desde entonces, ha habido cambios, como es natural, pero el núcleo permanece y se aprecia el hilo conductor que une esta primera norma con las vigentes[56].

55 Por supuesto, esta formulación coloquial y simple habría que matizarla en muchos casos por las reglas del abuso de derecho, las servidumbres, las obligaciones especiales del propietario (piénsese en los bienes que cuenten con un nivel cualificado de protección natural o artístico, donde existe un deber activo de conservación), u otros factores que limitan el poder absoluto de la propiedad. A pesar de que el dominio en nuestro ordenamiento no es tan amplio como el literal descontextualizado del artículo 348 del Código Civil, sigue siendo el derecho real de contenido más amplio.

56 La regulación de 1813 establecía que: «*I. Siendo los escritos una propiedad de su autor, este solo, o quien tuviere su permiso, podrá imprimirlos durante la vida de aquellas cuantas veces le conviniera, y no otro, ni aun con pretexto de notas o adiciones. Muerto el autor, el derecho exclusivo de reimprimir la obra pasará a sus herederos por el espacio de diez años contados desde el fallecimiento de aquel. Pero sí al tiempo de la muerte del autor no, hubiese aun salido a luz su obra, los diez años, concedidos a los herederos se empezarán a contar desde la fecha de la primera edición que hicieren. II. Cuando el autor de una obra fuere un cuerpo colegiado, conservará la propiedad de ella por el término de cuarenta años contados desde la fecha, de la primera edición. III. Pasado el término de que hablan los dos artículos precedentes, quedaran los impresos en el concepto de propiedad común, y todos tendrán expedita la acción de reimprimirlos cuando les pareciera*». Aunque los plazos que plantea son mucho más breves en relación a los herederos, y los derechos del autor están mucho menos perfilados, centrándose en la parte económica de los mismos, sin valorar los de tipo moral, en lo fundamental podemos ver ya una conexión entre esta modesta regulación y la del texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia, aprobado por el Real Decreto Legislativo 1/1996, de 12 de abril. En ambos casos, los derechos nacen de la autoría, otorgan capacidad para excluir o autorizar, y pasado un tiempo tras la muerte, decaen en beneficio del dominio público.

Lógicamente, la normativa varía en función del país, no obstante, el Convenio de Berna para la Protección de las Obras Literarias y Artísticas ha garantizado un grado de homogeneidad elevado en los aspectos esenciales de la regulación mundial. La Asamblea de la Unión de Berna está compuesta por 179 Estados, lo que le dota de una enorme representatividad, a la altura de otros grandes pactos internacionales. El Tratado de la Organización Mundial de Propiedad Intelectual (OMPI) sobre Derecho de Autor, de 1996 complementa el Convenio de Berna, incorporando previsiones propias del contexto de difusión de aquel momento (mucho más limitadas a las que contamos en estos momentos, pero sin duda más cercanas a nuestra realidad que las que se tuvieron en consideración al preparar el texto primigenio de Berna en 1886). A pesar de que cuenta con menos signatarios, sigue gozando de un fuerte respaldo, al haberlo firmado 118 países entre los que figuran, al margen de los componentes de la Unión Europea, EEUU, China, Japón, Reino Unido, India, Rusia, Turquía, Singapur, Corea del Sur o Australia.

La Directiva (UE) 2019/790, del Parlamento Europeo y del Consejo, de 17 de abril de 2019 sobre los derechos de autor y derechos afines en el mercado único digital y el texto refundido de la Ley de Propiedad Intelectual, aprobado por el Real Decreto Legislativo 1/1996, de 12 de abril, se basan en este marco internacional. En consecuencia, muchas de las reflexiones que se hagan en estas páginas partiendo de estas normas, cabrá extenderlas, *mutatis mutandi*, a otros ordenamientos, con la salvedad importante del régimen de excepciones (quizá, uno de los puntos donde se aprecian más las diferencias de fondo entre legislaciones, al dejar también mayor libertad los textos internacionales para que los Estados las definan por su cuenta). Lo mismo sucedería al revés, si tomáramos de base otros Derechos basados en los principios de Berna y sus tratados asociados. Las conclusiones alcanzadas serían fácilmente replicables en las áreas de superposición jurídica y fáctica.

Metafóricamente, y usando al Derecho penal para construir el símil, cabría explicarlo de la siguiente manera: la legislación de los Estados que nos interesa al objeto del capítulo coincide sustantivamente en cuanto a los elementos objetivos y subjetivos del tipo (con oscilaciones a veces, en temas como la duración de los derechos en favor de los herederos tras la muerte del autor), pero varía en cuanto a las causas de justificación. Desde ese prisma, si se determina conforme a la Directiva o al texto refundido que se ha producido un acto de reproducción, distribución, comunicación pública o transformación sin consentimiento, previsiblemente, lo mismo sucederá en el resto de derechos ligados a Berna. Donde sí que haría falta un análisis separado es en la fase posterior. La falta de autorización individualizada del titular, en función de la normativa nacional de la que se trate, puede ser desde ilegal a irrelevante o subsanable. Aun así, no será raro que incluso aquí haya partes susceptibles de «reutilizarse» entre ordenamientos.

El artículo 30.4 de la Ley de Propiedad Intelectual japonesa, relativo a la minería de datos y textos, por ejemplo, está configurado en términos bas-

tante más generosos para el «minero» que la dupla formada por los artículos 3 y 4 de la Directiva 2019/790, de 17 de abril (si bien esto es matizable por el efecto restrictivo que teóricamente podría desplegar el requisito del «disfrute», presente en el precepto japonés y ausente en los comunitarios). La sección 244 de la norma de Copyright de Singapur, estando más cerca del enfoque europeo, es mucho más casuística en su redacción (y no solo por el hecho de que se trate de una Directiva, porque si miramos su traslación en varias normas nacionales, no se encuentra nada semejante a la letra a) del apartado d) de la sección 244.2, en la que se alude a la obtención de los datos saltándose los muros de pago o *paywalls*[57]). En el ámbito de EE. UU., ni siquiera hay una excepción de minería como tal, reconduciéndose su habilitación a la figura omnicomprensiva del *fair use*, que absorbe esta y otras situaciones. Habría que sumar, en los países donde estén reconocidas[58], el supuesto de las licencias colectivas ampliadas o las presunciones de representación, por las que es posible emplear material sujeto a propiedad intelectual sin la anuencia de su titular, cuando se haya alcanzado un acuerdo con alguna organización dotada de suficiente representatividad dentro del sector al que pertenezca el titular de los derechos. Precisamente, el Gobierno español pensó en ellas en el proyecto de Real Decreto que sometió a consulta pública en noviembre de 2024.

Como cabe observar, y es una muestra reducida, hay una importante diversidad que impide universalizar la respuesta jurídica al completo. Al mismo tiempo, se constata a través de estos ejemplos de que hasta en lo que se diferencian, no llegan a hacerlo del todo, habiendo varios solapamientos. Así, las excepciones de minería, sin ser idénticas, están construidas en torno a la misma noción extrajurídica y técnica que les da nombre, por lo que, si una actividad no es considerada minería, o no solo minería, a ojos de la Directiva o del texto refundido, lo ordinario, una vez más, será que tampoco goce de tal

57 El artículo 6.3 de la Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información, acuñó un concepto general que comprende esta clase de mecanismos: «*se entenderá por "medidas tecnológicas" toda técnica, dispositivo o componente que, en su funcionamiento normal, esté destinado a impedir o restringir actos referidos a obras o prestaciones protegidas que no cuenten con la autorización del titular de los derechos de autor o de los derechos afines a los derechos de autor establecidos por ley o el derecho sui generis previsto en el Capítulo III de la Directiva 96/9/CE. Las medidas tecnológicas se considerarán "eficaces" cuando el uso de la obra o prestación protegidas esté controlado por los titulares de los derechos mediante la aplicación de un control de acceso o un procedimiento de protección, por ejemplo, codificación, aleatorización u otra transformación de la obra o prestación o un mecanismo de control del copiado, que logre este objetivo de protección*». Para leer estas expresiones tan concretas hay que acudir a normas recientes de *soft law,* como el Código de Prácticas de Inteligencias Artificiales de Propósito General. La medida 1.2 del compromiso de *Copyright* menciona los modelos de suscripción o los muros de pago.

58 En Europa, Suecia, Finlandia y Dinamarca están reguladas, si bien con carácter previo al «boom» de la inteligencia artificial generativa.

consideración para los demás Derechos que contemplan una cláusula habilitante TDM (acrónimo de *Text data mining*).

En el caso de Japón, el artículo 30.4 de su Ley no utiliza la expresión *fair use*, pero uno de los límites, que no se cause un perjuicio irrazonable al dueño de los derechos afectados, recuerda a cuando en EE. UU. se examina el efecto que el presunto uso justo provoca o provocaría sobre el mercado o el valor de la obra procesada. Y ambas previsiones enlazan a su vez con la tercera prueba de la regla de los tres pasos del artículo 9.2 del Convenio de Berna. Por ello, aunque no resulte posible un traslado automático, es probable que algunas de las claves que surjan del estudio de las excepciones nacionales y europeas trasciendan su ámbito estricto de aplicación subjetiva.

Dentro del núcleo común, desde mi punto de vista, hay tres ideas que definen a la propiedad intelectual: una exclusividad matizada, un inmanente interés social, y la sociabilidad humana. Dada la pluralidad de bienes que existen en la realidad, existen muchas maneras de disfrutar de los mismos. La forma de aprovechar una moto, una consola de videojuegos, una casa, o un terreno de cultivo no es la misma. Cualquiera de estos cuatros bienes es susceptible de arrendarse, pero tres son más aptos para un disfrute directo por parte de su propietario, sobre todo si pensamos en un pequeño propietario no empresario. Este puede conducir la moto, vivir en la casa y jugar con la videoconsola. En relación a la tierra, si no sabe trabajarla o por su extensión supera sus capacidades propias, ahí sí que es más probable que los beneficios no vengan de su uso directo, sino de los frutos de otros al emplearla.

Ampliando el radio, se constata que la propiedad admite usos privativos y participativos, habiendo bienes más aptos para unos que otros, pero, en general, siendo posible cualquiera de los dos. Sin embargo, la propiedad intelectual, por su propia naturaleza, parece complicado que pueda disfrutarse en soledad.

Si vemos sobre que recae la propiedad intelectual, el artículo 1 del texto refundido español menciona tres tipos de obras: literarias, artísticas y científicas. Con más detalle, el artículo 2.1 del Convenio de Berna de 1886 las define así: *«"obras literarias y artísticas" comprenden todas las producciones en el campo literario, científico y artístico, cualquiera que sea el modo o forma de expresión, tales como los libros, folletos y otros escritos; las conferencias, alocuciones, sermones y otras obras de la misma naturaleza; las obras dramáticas o dramático-musicales; las obras coreográficas y las pantomimas; las composiciones musicales con o sin letra; las obras cinematográficas, a las cuales se asimilan las obras expresadas por procedimiento análogo a la cinematografía; las obras de dibujo, pintura, arquitectura, escultura, grabado, litografía; las obras fotográficas a las cuales se asimilan las expresadas por procedimiento análogo a la fotografía; las obras de artes aplicadas; las ilustraciones, mapas, planos, croquis y obras plásticas relativos a la geografía, a la topografía, a la arquitectura o a las ciencias*». A pesar de su minucio-

sidad, la continua aparición de nuevos medios y formatos hacen que algunas realidades como la del videojuego queden incluidas de forma analógica y no expresa. En cualquier caso, lo que se percibe es que las obras sujetas a propiedad intelectual nacen del deseo de comunicar de las personas.

Por supuesto, dentro del derecho soberano que corresponde a los autores, resulta legítimo, así como factible, que no deseen compartir lo que han creado y se lo reserven para sí. La mayor parte de la obra de Kafka fue conocida tras su muerte y solo porque su amigo Max Brod traicionó sus deseos y publicó sus abundantes escritos inéditos (de haber dependido de Kafka, su único público hubieran sido las llamas). Algo similar ocurrió con la poesía de Emily Dickinson, que dio a conocer su hermana póstumamente, al hallarla entre sus pertenencias personales. No obstante, estos son casos aislados y no la norma general del creador. Incluso si piensan en sí mismos al crear, están comunicando por medio de su arte o ciencia.

Un mensaje exige un remitente y un destinatario para que sea tal y lo mismo se puede decir de las obras sujetas a propiedad intelectual. El autor aspirará o no a la fama, a una audiencia pequeña o grande, pero siempre habla a alguien (puede que a sí mismo). Eso hace que, en oposición a otros derechos de propiedad, en el caso del intelectual casi se presuponga que habrá más personas que disfrutaran del bien en cuestión. No es un tesoro que dejar en una caja fuerte, o una casa que convertir en un castillo, cerrándola al resto del mundo. De ahí que líneas atrás aludiera a una exclusividad relativa y a la sociabilidad humana como elementos esenciales de la propiedad intelectual, entendiendo la creación como una manifestación del carácter político del animal humano en el sentido aristotélico. El titular del derecho rara vez va a dejar fuera a la totalidad de la humanidad; ahora bien, dentro de las facultades dominicales, está plenamente habilitado para excluir, para decidir con quién y cómo lo comparte.

Que la propiedad intelectual nació para compartirse es algo que está presente en el gran límite que se le impone: su duración. Mientras que cualquier otra propiedad es susceptible de legarse de generación a generación hasta el infinito, la intelectual está acotada temporalmente. Una vez transcurrido el plazo de aprovechamiento fijado legalmente pasan a integrar al dominio público con respeto a algunos derechos morales del autor original, como la acreditación de su autoría. E incluso en el periodo donde rigen los derechos de forma plena, tal y como se viene comentando, hay excepciones relevantes que permiten utilizar las obras por terceros sin encomendarse al autor, a sus herederos, o a quienes ostenten sus derechos de explotación, por habérseles cedido. Cada país dispone de su propio elenco, como indicaba antes, pero las que más se repiten son las ligadas a la crítica, el comentario, la enseñanza, la investigación, o la parodia. Excepciones que conectan, a su vez, con la filosofía implícita de que el objeto de la propiedad intelectual no descansa sobre realidades estancas, sino, por el contrario, interconectadas y predispuestas a la interacción.

No todas las constituciones del mundo contienen una referencia como la del artículo 33 de la Carta Magna española, en la que se sujeta la herencia y la propiedad a su función social; sin embargo, de facto, todos los ordenamientos lo hacen, en mayor o menor grado, siendo la propiedad intelectual unos de los exponentes más claros de este condicionamiento. Al restringir el goce y disfrute a un intervalo de tiempo, se está lanzando el mensaje de que interesa que todos puedan acceder a estas creaciones. Expresamente, el artículo 1.5 de la Ley 10/2007, de 22 de junio, de la lectura, del libro y de las bibliotecas proclama que *«se considera a la cultura, a todos los efectos, bien básico y de primera necesidad»*. Como el autor no puede vivir del aire, se le reconocen derechos de explotación exclusivos, que recompensen, o al menos compensen, su esfuerzo e ingenio, pero en un «mundo perfecto» los legisladores parecen creer que lo mejor sería un panóptico cultural y científico desprovisto de barreras. La plasmación realista de tal ambición es el diseño actual híbrido entre el dominio privado y el público, con predominancia privada hasta que se convierte totalmente en público[59].

Con esto en mente, la autoría otorga una serie de derechos que se dividen esencialmente en dos grupos: los de tipo moral y los de explotación. Habría una tercera categoría, a modo de cajón de sastre, que en España agrupa a los derechos de participación y compensación equitativa, no obstante, a mi juicio, suponen más bien emanaciones de algunos aspectos de los derechos morales y de los derechos de explotación. El de participación entronca con los de tipo moral, porque se sustenta en el principio de que es el artista el que aporta valor a su obra, primero creándola y luego por medio del resto de su carrera. Si un vendedor consigue incrementar el precio respecto al de la cesión original es porque el prestigio de la pieza o del autor respalda dicho aumento, fundamentalmente. De ahí que el Derecho juzgue apropiado que perciba un porcentaje de los beneficios cuando es la causa de los mismos (sin excluir la pericia que pueda tener el vendedor al negociar con el comprador), siendo otra forma de reconocimiento. La compensación equitativa, por otro lado, está ligada a las facultades de explotación del autor, en concreto a la renuncia de los autores a ejercer un control estricto de las de reproducción en favor de una mayor operatividad social.

El artículo 14 del texto refundido de la Ley de Propiedad Intelectual enumera lo derechos morales, abarcando los siguientes: *«1.º Decidir si su obra ha de ser divulgada y en qué forma. 2.º Determinar si tal divulgación ha de hacerse con su nombre, bajo seudónimo o signo, o anónimamente. 3.º Exigir el reconocimiento de su condición de autor de la obra. 4.º Exigir el res-*

59 Este diseño se complementa, asimismo, con la articulación de un sistema de apoyo estatal durante la fase de dominio privado que facilita el acceso de los ciudadanos, garantizando al mismo tiempo la percepción de una remuneración a los autores. Los préstamos a través de la red de bibliotecas, las subvenciones de consumo en bienes y servicios específicos, o la exhibición pública con cargo a fondos del Estado serían parte del mismo.

peto a la integridad de la obra e impedir cualquier deformación, modificación, alteración o atentado contra ella que suponga perjuicio a sus legítimos intereses o menoscabo a su reputación. 5.º Modificar la obra respetando los derechos adquiridos por terceros y las exigencias de protección de bienes de interés cultural. 6.º Retirar la obra del comercio, por cambio de sus convicciones intelectuales o morales, previa indemnización de daños y perjuicios a los titulares de derechos de explotación. Si, posteriormente, el autor decide reemprender la explotación de su obra deberá ofrecer preferentemente los correspondientes derechos al anterior titular de los mismos y en condiciones razonablemente similares a las originarias. 7.º Acceder al ejemplar único o raro de la obra, cuando se halle en poder de otro, a fin de ejercitar el derecho de divulgación o cualquier otro que le corresponda».

El primero de los apartados, aunque se caracterice como moral, no deja de encajar con la idea general de la propiedad. Resulta especialmente interesante el 6.º, por reflejar ese carácter «dialogado» de la propiedad intelectual. Los bienes que abarca sirven para comunicar, sin perjuicio de que cumplan además otras funciones, como puede ser entretener. Si el autor deja de sentirse representado por el mensaje que trasmitía a través de su creación, la legislación le otorga una facultad de retirada, que rebasa los contornos habituales del derecho a la propiedad. Ante una cesión de uso a un tercero formalizada, el régimen de propiedad por lo usual no admite una reversión unilateral, incluso mediando el pago de una indemnización por los daños y perjuicios que irrogue la decisión. El *pacta sunt servanda*, o su versión foral aragonesa, *stadum est chartae*, laten detrás de la regulación de los contratos. La opción de recuperación está pensando en la propiedad intelectual más allá de su utilidad práctica y bajo la consideración de que en ella se deja un poso personal superior que al fabricar o prestar otros bienes y servicios.

El 5.º también merece subrayarse, porque denota la sustantividad especial de la propiedad intelectual, tanto a la hora de permitir la modificación, como de limitarla en ciertos casos (aunque la creación fuera un acto libre, a partir de un momento pueden haber surgido intereses que trasciendan los deseos del autor y que lo constriñan hasta a él). Es algo que se percibe igualmente en los apartados 2.º y 3.º Si acudimos al Registro de la Propiedad, no cabe pensar en que un bien se inmatricule de forma anónima o con seudónimo (cuestión distinta es el uso en fraude de ley de personas interpuestas para ocultar el propietario real). En esta línea, al propietario de un bien mueble o inmueble solo le preocupa reivindicar la titularidad si alguien se la cuestiona, no teniendo especial interés, por lo normal, de que todo el mundo conozca ese coche o esa casa es suya, salvo si quiere presumir. Para los autores, en cambio, hay un componente habitualmente de orgullo en cuanto al fruto de su esfuerzo que lleva a querer que figure su nombre con claridad.

En relación a los derechos de explotación, aunque no se les adjetive como económicos, son los que están más orientados a proporcionar una remuneración para el autor, sin perjuicio de que puedan ejercerse de formas no

lucrativas. El artículo 17 del texto refundido indica que *«corresponde al autor el ejercicio exclusivo de los derechos de explotación de su obra en cualquier forma y, en especial, los derechos de reproducción, distribución, comunicación pública y transformación, que no podrán ser realizadas sin su autorización, salvo en los casos previstos en la presente Ley»*. A continuación, la ley los va definiendo uno por uno, teniendo todos ellos relevancia en el ámbito de las inteligencias artificiales generativas[60]. En puridad, los morales tampoco son ajenos a esta tecnología, pero los conflictos que han surgido y siguen apareciendo conectan más con esta segunda clase de derechos; razón por la cual me detendré más en las ramificaciones que presentan al entrar en contacto con las IAs.

Dependiendo de la fase del proceso técnico, la incidencia es desigual. Así, durante el «entrenamiento»[61] es complicado que se produzca un acto de distribución o de comunicación pública, pero, una vez puesta a disposición la herramienta al público, lo generado podría colisionar con estos derechos, especialmente en aquellos modelos que regurgitan sin alterar partes significativas de los datos usados durante su «aprendizaje»[62]. En este último

60 Hay que precisar esta afirmación porque existe cierto solapamiento entre el derecho de distribución y el de comunicación pública. Acudiendo a la dicción del artículo 19 del texto refundido de la Ley de Propiedad Intelectual *«se entiende por distribución la puesta a disposición del público del original o de las copias de la obra, en un soporte tangible, mediante su venta, alquiler, préstamo o de cualquier otra forma»*. Ese «cualquier otra forma» es el que podría llevar a entender que la regurgitación de datos de entrenamiento o de las fuentes consultadas por las IAs durante su funcionamiento es distribución. No obstante, a pesar de la apertura con la que se enuncia, no se puede descartar que haya quien no vea distribución por el soporte en el que se trasmite, o por la forma en que se hace. En ese caso, intervendría el derecho de comunicación pública subsidiariamente. Corolario de ello, si afecta a la distribución, en principio, no lo hace a la comunicación pública y viceversa.

61 Bajo este término y a efectos de simplificación se englobarán todas las etapas que se siguen en el diseño de un modelo y que requieren la utilización de datos. En puridad, habría que distinguir entre múltiples fases, cada una dotada de su propia sustantividad.

62 El estudio *Extracting memorized pieces of (copyrighted) books from open-weight language models* elaborado por A. Feder Cooper, Aaron Gokaslan, Ahmed Ahmed, Amy B. Cyphert, Christopher De Sa, Mark A. Lemley, Daniel E. Ho, Percy Liang, lo ha probado. Explican sus autores que *«mediante numerosos experimentos, demostramos que es posible extraer porciones sustanciales de al menos algunos libros de diferentes LLM. Esto demuestra que estos LLM han memorizado el texto extraído; este contenido memorizado se copia dentro de los parámetros del modelo. Sin embargo, los resultados son complejos: el grado de memorización varía tanto por modelo como por libro. Con nuestros experimentos específicos, observamos que los LLM más grandes no memorizan la mayoría de los libros, ni total ni parcialmente. No obstante, también observamos que Llama 3.1 70B memoriza algunos libros, como Harry Potter y la Piedra Filosofal y 1984, casi en su totalidad. De hecho, Harry Potter está tan memorizado que, utilizando una instrucción inicial que consiste solo en la primera línea del Capítulo 1, podemos generar determinísticamente el libro completo casi textualmente. Analizamos por qué nuestros resultados tienen implicaciones significativas para los casos de derechos de autor, aunque no favorecen inequívocamente a ninguna de las partes»*.

supuesto, se está valorando si no habría, también repercusiones respecto al derecho de reproducción, en la medida en que la réplica literal o casi exacta estaría reflejando una memorización de los datos, total o parcial. Con todo, incluso si es así, representaría un problema subsidiario al de las reproducciones iniciales, que se realizan para «enseñar» y «alimentar» con información a las máquinas, y que, de hecho, serían la base de esas eventuales reproducciones posteriores cuando, además de asimilar patrones, conservan el contenido concreto con el que se interiorizó tal conocimiento. A causa de ello, el derecho de reproducción donde sobresale es en el «entrenamiento».

Dejando un lado los datos licenciados de manera expresa para el desarrollo de modelos de inteligencia artificial, aquellos que son de dominio público y los que pertenecen a los propios desarrolladores de las IAs, que están exentos de controversias legales, las vías para proveerse de materiales han sido fundamentalmente: el *scraping* masivo de datos de Internet, el uso de bases de datos disponibles públicamente (lo que, una vez más, no es equivalente a que sean de libre disposición), la digitalización de soportes físicos, y la descarga pirata, usando protocolos de intercambio de archivos como BitTorrent u otros medios similares[63]. En todos estos casos, existe reproducción, dado que la información se aloja para su consulta en el «entrenamiento».

La reproducción, conforme al artículo 18 del texto refundido español, es «*la fijación directa o indirecta, provisional o permanente, por cualquier medio y en cualquier forma, de toda la obra o de parte de ella, que permita su comunicación o la obtención de copias*». Esta definición data del año 2006 y su única variación respecto a la original de 1996 es la incorporación de varias precisiones, no menores, eso sí. Se especifica que la fijación puede ser directa o indirecta y provisional o permanente —antes no distinguía—; asimismo, se enfatiza que puede ser de cualquier medio o cualquier forma. Tales adiciones traen causa de la Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información. Da la impresión de que el interés de estos añadidos venía sobre todo de adaptarse a la realidad técnica de las reproducciones provisionales, tan frecuentes y vitales en la transmisión de información por Internet. Sin embargo, la enorme amplitud de la redacción permite que englobe todo tipo de actividades novedosas.

63 El uso de copias ilegales ha sido constatado en la pionera sentencia el 23 de junio de 2025 dictada por el juez William Alsup, del distrito norte de California, que al mismo tiempo que apreció el «entrenamiento» como uso justo, consideró que Anthropic podía ser llevada a juicio de nuevo por haber incurrido en esta práctica, obteniendo bases de texto por canales no legales. Anthropic ha llegado a un acuerdo por 1.500 millones de dólares, más intereses, con los autores que iban a presentar una demanda colectiva con base a esta sentencia, con el fin de evitar la judicialización del conflicto. El montante final puede que varíe, pero al menos alcanzaría esta cuantía.

Probablemente, lo idóneo sería regular un nuevo derecho de explotación distinto al de la reproducción que atienda a estas nuevas vías de aprovechamiento de los trabajos sujetos a propiedad intelectual. Mientras tanto, no cabe duda de que hay reproducción, aunque esta sea instrumental para el rendimiento real que se pretende dar a la información (como ocurre en el «entrenamiento»).

Adelantando el estudio de las excepciones, algunos autores se han planteado si cabría invocar en favor del «entrenamiento» el supuesto de la reproducción provisional del artículo 31.1 del texto refundido español o el del artículo 5.1 de la Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001. Desde mi punto de vista, no se acomoda bien. La noción de «temporal» empleada en los enunciados de estos dos preceptos es un concepto jurídico indeterminado, pero si hacemos una lectura del contexto en el que se inserta y de la propia *ratio legis* de la excepción, se percibe que está pensando en duraciones más limitadas que la que requiere el «aprendizaje». La habilitación está concebida en torno al supuesto de la letra a) del artículo 5.1, el de la b) es uno residual; constituye una cláusula abierta o de evolución legislativa ante los imprevisibles desarrollos técnicos y/o legislativos, como los que habían dado pie a la letra a); un por si acaso en previsión de que haya un supuesto análogo a esta letra en el que no repararan en el momento de la redacción de la Directiva.

El recurso a la expresión «utilización lícita» tiene ecos de *«fair use»*, sin embargo, el Derecho de la Unión Europea no forma parte de esta tradición (la regla de los tres pasos del Convenio de Berna precisamente es una solución de compromiso entre el casuismo predominante en el Derecho continental y la abstracción del *fair use* y el *fair dealing*). El considerando 33 de la Directiva clarifica que *«la utilización se considerará lícita en caso de que la autorice el titular del derecho o de que no se encuentre restringida por la ley»*. Esta formulación queda muy lejos del *fair use* y es incluso contradictoria, porque si está autorizada la reproducción, no nos hallamos ante una excepción al derecho (algo que se hace más patente en la redacción del precepto español donde se viene a decir que no requerirá autorización el uso licito autorizado; un auténtico oxímoron).

En cualquier caso, la ausencia de significación económica se quiebra cuando se observan la cantidad de acuerdos de licencias alcanzados o iniciativas legislativas como las del proyecto de Real Decreto por el que se regulan las licencias colectivas ampliadas para la explotación masiva de obras y prestaciones protegidas por derechos de propiedad intelectual para el desarrollo de modelos de inteligencia artificial de uso general. Estas reproducciones no son insustanciales en el plano económico, sino todo lo contrario. Sumando los argumentos, se colige de forma clara que la excepción es inaplicable.

Hilando de nuevo con el factor económico, cabe descartar también la excepción de copia privada para la digitalización de material con el que ya se cuente o se adquiera *ex profeso*, por el ánimo de lucro presente en el impulso

de los modelos comerciales. Los dos artículos antes mencionados, 31 y 5, exigen enfáticamente que los fines no sean comerciales, ni directa ni indirectamente. A mayor abundamiento, quien realice la copia privada ha de ser una persona física, lo que deja fuera a las personas jurídicas.

El derecho de transformación, por su parte, intervendría antes y después de que los modelos de IA estén en funcionamiento de manera abierta. Durante su «aprendizaje», cabe entender que hay transformación si nos atenemos a la caracterización que se hace de ella en el artículo 21.1 del texto refundido. De acuerdo con este precepto, la traducción, la adaptación y los arreglos de una obra implican transformación y esto es lo que en definitiva se hace al etiquetar los datos. Del mismo modo que un ser humano no puede leer un texto en un idioma que no conoce y precisa que se lo proporcionen de manera que le resulte comprensible, en la creación de los modelos de IA es raro que se utilicen las fuentes de información en bruto. Una faceta crucial en el desarrollo de los modelos es la preparación de los datos de la forma en que se extraiga de ellos el máximo rendimiento. Empresas como Mercor, Surge, o Scale IA (cuya valoración era de 29.000 millones en junio de 2025) tienen como una de sus principales líneas de negocio el etiquetado profesional. En ese sentido, la conversión de los textos, imágenes, audios, y vídeos en algo «digerible» para la «máquina» acarrea una evidente transformación, bien como traducción, bien como adaptación o arreglos[64].

Si asumiéramos la lógica que equipara la inspiración de las personas con la inferencia de las máquinas, quizá se podrían salvar algunos de los interrogantes legales que suscitan otros derechos de explotación, sin embargo, el de la transformación persistiría. Porque igual que no puede una persona libremente traducir una obra a otro idioma sin autorización del titular para que otros la disfruten, lo mismo debería suceder con la parametrización de las máquinas. Caber recordar lo que se decía al principio del capítulo: el código binario es el lenguaje de los ordenadores y los lenguajes informáticos puentes para salvar la brecha comunicativa. Si se cae en una indebida antropomorfización de las inteligencias artificiales, ha de hacerse con todas las consecuencias; asumiendo que la conversión del lenguaje visual, sonoro y textual humano al de las máquinas sería traducción y/o adaptación. Personalmente, rechazó la equiparación entre los «cebreros maquina» y los humanos, pero creo que está bien hacer este apunte, *obiter dicta*. Fortalece la idea de que tanto concebidas en sus justos términos las IAs o presos del efecto Eliza va a haber actos de explotación, que solo encontrarán amparo legal si quedan comprendidos en alguna de las excepciones previstas legalmente.

64 En refuerzo de esta tesis, la sentencia del Tribunal de Justicia de la Unión Europea, de 22 de enero de 2015, relativa al asunto *Art & Allposters International BV v Stichting Pictoright* (Case C-419/13). En ella se comprueba que el TJUE mantiene un criterio estricto en cuanto al derecho de transformación. Como mínimo, la digitalización de obras a papel para el entrenamiento sería transformación, de acuerdo con su doctrina.

Una vez disponibles los modelos entre el público, aunque es la expresión formal específica de los autores lo que se protege por la legislación de propiedad intelectual, y no las ideas ni el estilo[65], se discute si no constituirían obras derivadas muchos de los *outputs* que toman de referencia trabajos previos. Directamente, la generación actual de IAs generativas permite al usuario seleccionar sonidos, fotografías o textos preexistentes para fusionarlos y combinarlos entre sí[66]. En puridad, hace tiempo que ya se podía, pero las sucesivas versiones lo han ido facilitando cada vez más, con interfaces tan básicas como las que se siguen para adjuntar un documento a un correo electrónico. Un ejemplo de estas mezclas sería el polémico tema *Heart on My Sleeve,* que engendró artificialmente un usuario llamado *Songwriter977* a partir de canciones reales de Drake y *The Weeknd*, no contando, naturalmente, con su permiso, ni del de las discográficas que gestionan sus derechos de explotación.

Sentado que el «entrenamiento» conlleva actos de explotación, si estos no son autorizados por las personas legitimadas para ello, únicamente habría un camino para no incurrir en una ilegalidad: el de las excepciones. Que los impulsores del modelo puedan acogerse válidamente a una. Varias han sido ya descartadas mientras se examinaba el contenido de los derechos de la propiedad intelectual y su relación en el campo de la IA, por lo que no tiene sentido incidir otra vez ellas. De las que quedan, la minería de datos es probablemente la que más se invoca en los discursos legitimistas de la extracción. En EEUU, el *fair use* ha devenido en una moderna apelación a «asilarse en sagrado», el bálsamo de Fierabrás que purga de todo mal jurídico; las cláusulas TDM lo son en la Unión Europea y en los demás estados que las han incorporado.

Con el título con el que he bautizado epígrafe, cabe imaginar cual será la conclusión respecto a la de minería de datos, no obstante, es necesario explicar cómo se ha llegado hasta ella. Lo primero que debemos tener presente es que la evolución tecnológica ha provocado que la expresión «minería» haya quedado algo anticuada.

---

65 Véase, entre otras referencias, la del artículo 2 del Tratado de la OMPI sobre Derecho de Autor de 1996.

66 A este respecto, téngase en cuenta la sentencia del TJUE, asunto C-476/17, de 29 de julio de 2019, en respuesta a una consulta prejudicial del Tribunal Supremo alemana. En ella se dice que *«el artículo 2, letra c), de la Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información, debe interpretarse, a la luz de la Carta de los Derechos Fundamentales de la Unión Europea, en el sentido de que el derecho exclusivo que confiere al productor de fonogramas, en virtud de dicha disposición, el derecho a reproducir y distribuir su fonograma le permite impedir que otra persona tome una muestra de sonido, aunque sea muy breve, de su fonograma para incluirla en otro fonograma, a menos que dicha muestra se incluya en el fonograma en una forma modificada irreconocible para el oído»*.

Tim W. Dornis y Sebastian Stobber hicieron el experimento de consultar a ChatGPT-4 si el «entrenamiento» de IA constituía minería de datos y si la excepción lo cubría. Las respuestas que obtuvieron incluían frases como *«el proceso requiere técnicas avanzadas de aprendizaje automático y redes neuronales profundas que van mucho más allá de los métodos de minería tradicionales. Además, el entrenamiento en IA generativa a menudo implica un procesamiento y modelado de datos más extensos para lograr resultados creativos y contextuales»* o *«debido a que el entrenamiento de modelos de IA generativa va más allá de la simple extracción y análisis de datos e incluye la creación de contenido nuevo, no necesariamente cae bajo esta barrera»*.

Aplicando las reglas del método científico, busqué replicar su experiencia, si bien alterando la instrucción que habían usado ellos, que, a mi juicio, podía predisponer el sentido del resultado[67], dada la configuración por defecto de los modelos GPT como agradadores de todos los Segismundos, al modo del personaje de Clarín en *La vida es sueño*. En mi caso, la pregunta que le formulé, el 22 de enero de 2025, fue *«¿ChatGPT ha sido desarrollado mediante minería de datos o se ha usado otra técnica?»*. A pesar de ser más neutro el enunciado, el texto proporcionado fue muy similar al de Dornis y Strobber, restando significación a la minería de datos y resaltando la significación del aprendizaje profundo. Asimismo, al delinear los conceptos dejaba claro que, aunque pudieran existir ciertas conexiones, se trataban de dos nociones diferentes.

En una última revisión del capítulo, el 5 de septiembre de 2025, con la misma pregunta, con una leve variación formal, obtuve estas contestaciones: *«¿ChatGPT ha sido desarrollado mediante minería de datos o a través de otra técnica?»*, exponía que: *«ChatGPT, como modelo de lenguaje desarrollado por OpenAI, no ha sido entrenado mediante minería de datos en el sentido tradicional, sino a través de un proceso conocido como entrenamiento de modelos de aprendizaje profundo, específicamente usando redes neuronales transformadoras (transformers)»*; *«🔷 ¿Se usó minería de datos? No directamente. Aunque el modelo se entrena sobre grandes volúmenes de texto, el objetivo no es extraer patrones de datos estructurados como en la minería de datos, sino aprender representaciones del lenguaje para generar texto coherente y relevante»*, y *«✅ Resumen Técnica utilizada: Aprendizaje profundo (deep learning), específicamente transformers. No es minería de datos, aunque se trabaja con grandes volúmenes de texto (big data). El objetivo es aprender cómo se estructura y se usa el lenguaje, no extraer patrones estadísticos de datos estructurados»*.

67 Las dos instrucciones eran *«¡Buenos días, ChatGPT! ¿Puedes explicarme en 5 frases por qué el entrenamiento de IA generativa no se clasifica como minería de datos y textos?»* y *«¡Muchas gracias! ¿Esto significa que la barrera de los derechos de autor para la minería de datos y textos no cubre la capacitación en IA para la IA generativa?»*.

Es bastante curioso que los propios modelos desautoricen a sus creadores en su pretensión de beneficiarse de ciertas exoneraciones legales. Como testigos serían su peor enemigo. No obstante, más allá de lo llamativo, no cabe tomar sus réplicas como argumento de autoridad. Conforme a lo explicado en el primer epígrafe de este capítulo, la arquitectura de los modelos busca predecir que palabra, píxel o sonido es más probable que siga al anterior conforme a los datos de entrenamiento y el contexto de la petición. No es la exactitud lo que buscan sino la verosimilitud, que es su producto natural frente a la veracidad, al margen de que luego se apliquen refuerzos para intentar potenciar el grado de acierto. Resulta necesario, por consiguiente, llevar a cabo comprobaciones adicionales que corroboren sus afirmaciones.

Anticipo que las respuestas de ChatGPT no se equivocan al minusvalorar la entidad de la minería de datos y a su apelación como paraguas legal. Parafraseando con menos gancho la expresión «nunca traigas un cuchillo a una pelea de pistolas», popularizada por el personaje de Sean Connery en *Los Intocable*s, nunca se debería traer una excepción TDM a un «entrenamiento» de IAs basado en aprendizaje profundo u otras técnicas que trascienden de la minería.

Conforme al artículo 2.2 de la Directiva 2019/790, de 17 de abril, la minería de textos y datos es *«toda técnica analítica automatizada destinada a analizar textos y datos en formato digital a fin de generar información que incluye, sin carácter exhaustivo, pautas, tendencias o correlaciones».* Para dar mayor contexto, el considerando 8 menciona que *«la minería de textos y datos posibilita el tratamiento de grandes cantidades de información con el fin de adquirir nuevos conocimientos y descubrir nuevas tendencias».* Aunque el artículo 2.2 configure el listado como *numerus apertus*, al señalar que no pretende ser exhaustivo, se ve una coherencia temática en cuanto a los ejemplos. También en el considerando se habla de conocimiento y no de nuevos textos o datos.

Lo que se produce, según la Directiva, es información y aunque en sentido extenso un vídeo, una foto, un audio, o un texto son información, como lo es todo cuanto captamos a través de nuestros sentidos, a mi parecer, el legislador europeo no estaba pensando por aquel entonces en esa clase de «información», ni en las posibilidades que se abrirían no mucho después desde el campo de la inteligencia artificial generativa. Solo en el considerando 18 podría insinuarse tímidamente cierta percepción al decir que la minería puede servir para el desarrollo de nuevas aplicaciones o tecnologías, con todo, es dudoso que esa alusión se hiciera teniendo en mente el «entrenamiento» de las IAs. Es verdad que la generalidad del Derecho podría llevar a que incluso sin haberlo valorado, luego encaje en la realidad, pero el engarce no es fuerte.

Leyendo la frase al completo se indica que *«en el contexto de la investigación científica, las técnicas de minería de textos y datos son utilizadas a*

gran escala por entidades tanto privadas como públicas para analizar grandes cantidades de datos en distintos ámbitos de la vida y con fines diversos, entre ellos para servicios gubernamentales, para la adopción de decisiones empresariales complejas y para el desarrollo de nuevas aplicaciones o tecnologías». A esto hay que añadir que en el considerando anterior se manifiesta que «*el posible perjuicio para los titulares de derechos que pueda derivarse de esta excepción sería mínimo. En consecuencia, los Estados miembros no deben prever una indemnización para los titulares de derechos por los usos al amparo de las excepciones aplicables a la minería de textos y datos establecidas en la presente Directiva»*.

Cuando la Unión Europea tomó mayor conciencia de la evolución de la IA generativa, uno de los riesgos que rápidamente señaló fue el de sus repercusiones sobre los titulares de derechos de propiedad intelectual. De haber sido plenamente consciente de este potencial en 2019, es difícil que luego hubiera infravalorado sus consecuencias económicas al redactar la Directiva. Resulta más probable que la referencia a las aplicaciones informáticas (de nuevo, nada concreto de IA) fuera ponderando el provecho del análisis de los datos en su mejora continua (saber que uso les dan sus clientes, las funciones más solicitadas, la forma en que interactúan, preferencias, etc), o en la detección de necesidades. Sea como fuere, esta gota, se interprete como se interprete, no altera la imagen global que se da de la minería en la Directiva y, sobre todo, su definición legal en el articulado, que es el contenido aplicable, frente a los considerandos que actúan solo como guías en la interpretación.

Siguiendo a Tim W. Dornis y Sebastian Stobber, la prospección de datos y textos tiene por objeto la información semántica de las obras, sin embargo, el «entrenamiento» de las IAs no se limita a esa vertiente, sino que alcanza también la información sintáctica, que es justo la que preocupa a la propiedad intelectual, en la medida en que las ideas no son objeto de su protección. Teleológicamente, como apuntan, la excepción de minería de datos se justificaba con facilidad porque recaía sobre lo sustantivo y no lo formal. Que el objeto de la propiedad intelectual haya dejado siempre fuera a lo semántico ha sido para no limitar el alumbramiento de nuevas ideas, que un conocimiento lleve a nuevos conocimientos. Sin embargo, emular lo sintáctico no genera por sí mismo nuevos conocimientos, porque es la «cascara» que envuelve el «fruto». Al alumbrar un modelo de IA directamente no se está alcanzando conocimiento, sino que se está preparando una herramienta con la que las personas adquieran conocimiento. Se puede decir que el modelo en sí es ese conocimiento, pero es estirar la redacción cuando resulta evidente que el legislador europeo tenía claro para qué estaba incorporando la excepción en 2019, no bastando la generalidad propia del Derecho para ensancharla. La letra de la Ley es límite de la generalidad y de las exegesis jurídicas.

A fortiori, no se puede decir que el «entrenamiento» sea automatizado. Hay múltiples enfoques desde los que abordarlo, a veces más manuales y otros más automatizados, donde los algoritmos se retroalimentan por medio

de simulaciones, pero al final la mano humana está presente, máxime en los primeros GPTs. Recuperando claves explicadas al tratar la transformación, si hay empresas encargadas del etiquetado es porque no basta echar los datos sin más a las máquinas y dejarles a ellas a operar. Tampoco el *Machine Learning* clásico, más vinculado a la noción estricta de minería de datos, era del todo automático, no obstante, la intervención humana era menor en todos los sentidos, porque no debía absorberse tanta información ni tan tamizada como la que requieren las IAs generativas o razonadoras para desenvolverse ante las solicitudes de los usuarios (no es lo mismo detectar cuantas veces se usa una palabra en un libro, que producir cualquier clase de texto, tanto en cuanto a temática como a registro literario). Las necesidades de un modelo como los LLM son mayores cuantitativa y cualitativamente que las asociadas a la prospección tradicional.

Se ha desaprovechado una oportunidad de actualizar el argot en el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, o de establecer nociones amplias que puedan adaptarse a la realidad cambiante. Seguir hablando solo de minería de datos a estas alturas es no captar el paisaje técnico del presente. Personalmente, me inclinaría más por una regulación que vaya al tipo de aprovechamiento que se hace y su incidencia en el mercado de derechos de propiedad intelectual y no tanto a la técnica que se utiliza. No se dejan de ensayar nuevas aproximaciones[68], por lo que el riesgo de obsolescencia es constante. Sin embargo, una descripción genérica del tipo *«cualquier proceso informático que a partir de la información semántica/sintáctica de los datos»* presumiblemente actuaría como una llave maestra capaz de prolongarse en el tiempo. Otro tema es el juicio de oportunidad.

Válidamente el legislador puede facilitar un tipo de extracción y limitar el otro. Y dentro de estas dos grandes tipologías (sintáctico-semántico) distinguir según el fin, igual que ya se hace la minería de datos en el Derecho de la Unión Europea, con una regulación diferenciada para la investigación científica en relación con los demás aprovechamientos potenciales. A título ejemplificativo, aunque tome información sintáctica y no solo semántica, que un modelo de IA médica aprenda de imágenes, audios, textos y vídeos sujetos a propiedad intelectual no relacionados con la medicina para que se familiarice con el cuerpo humano o interactúe con los pacientes, podría permitirse

68 De interés, el artículo *Hierarchical Reasoning Model*, de 4 agosto de 2025, que recoge las investigaciones de Guan Wang, Jin Li, Yuhao Sun, Xing Chen, Changling Liu, Yue Wu, Meng Lu, Sen Song, Yasin Abbasi Yadkori para superar desde un nuevo planeamiento las grandes limitaciones a las que se enfrentan los modelos GPT en cuanto a razonamiento. Igualmente, el 5 de septiembre de 2025, se difundió otro titulado *SpikingBrain Technical Report: Spiking Brain-inspired Large Models*, firmado por Yuqi Pan, Yupeng Feng, Jinghao Zhuang, Siyu Ding, Zehao Liu, Bohan Sun, Yuhong Chou, Han Xu, Xuerui Qiu, Anlin Deng, Anjie Hu, Peng Zho, Man Yao, Jibin Wu, Jian Yang, Guoliang Sun, Bo Xu, y Guoqi Li. No sabemos si prosperan estas vías, pero es una prueba del dinamismo del campo.

legalmente con pocas trabas, dada la función social que puede desempeñar una herramienta de este tipo y que no entra en colisión con el área de negocio de los titulares. En cambio, una IA generativa capaz de producir recursos creativos-artísticos que entren en competencia con sus fuentes de «aprendizaje» no debería contar, desde mi punto de vista, con las mismas ventajas, al hallarnos ante intereses prevalentemente privados; máxime cuando la producción humana ya colma las demandas existentes y supera nuestro margen de asimilación[69].

Recordando el artículo 33.3 de nuestra Constitución, en él se dice que nadie puede ser privado de su propiedad sin causa de utilidad pública o interés social. *Mutatis mutandi*, aunque no se traten de expropiaciones, las restricciones de la propiedad, incluida la intelectual, han de basarse en la función social de la misma. El progreso tecnológico es un valor de nuestra sociedad, pero no cabe convertirlo en una carta blanca que lo soporte todo, sobre todo cuando hay múltiples formas de avanzar en él, citando de nuevo a los premios nobeles de economía Daron Acemoglu y a Simon Johnson. Como ya se ha dicho, también trayendo a colación estos dos nombres, la IA en singular no existe, existen las IAs en plural. Para que se destinan y como se modelan es algo que decidimos los humanos —algunos más que otros— conforme a nuestras prioridades.

Leyendo el Reglamento 2024/1689, de 13 de junio de 2024, su principal preocupación en esta materia parece ser que se haya respetado la reserva de derechos a la que da opción el artículo 4.2 de la Directiva del año 2019. Con esto se trasluce que para el legislador europeo la minería es técnicamente la base de las IAs modernas, sin embargo, esto es cuestionable por lo que se acaba de exponer. Calificar solo de minería el proceso del «entrenamiento» con afección en la propiedad intelectual, en el mejor de los casos sería un ejercicio de analogía extensiva, lo que tratándose de una excepción no es lo más apropiado. Personalmente, siento que su afán por verificar la reserva de la minería es como preocuparse del código de circulación con un avión o un helicóptero. Con todo, como es muy posible que esta visión se imponga, al ser el único concepto al que alude el Reglamento con cierta conexión técnica con el «entrenamiento» y la propiedad intelectual, conviene comprobar las ramificaciones de esta perspectiva.

Al prestar atención al artículo 4.2 de la Directiva es posible que los redactores del Reglamento estuvieran pensando implícitamente en las nuevas formas de comprobar el ejercicio de este derecho, pero también que esa inquietud denote cierta desincronización por su parte con el estado de la técnica, sí

69 En 2019, se estrenaron en el mundo 4.600 series de ficción. En 2020, solo el catálogo de Netlflix sumaba 2,2 millones de minutos, lo que equivale a algo más de 4 años de visualización ininterrumpida. Leer todos los libros del mundo por esas fechas se calculaba que requeriría alrededor de 25.000 o 40.000 años. Y estas cifras no dejan de crecer, porque el ser humano sigue creando.

lo que pretende verificar con las obligaciones de transparencia es qué se han respetado los protocolos robot.txt.

A pesar de que la Directiva legalmente solo demanda que la exclusión se exprese de manera adecuada, el ejemplo que ponía de ello eran los medios de lectura mecánica, lo que nos conduce a este tipo de protocolos. Los robot.txt guían a los motores de búsqueda o a los bots dedicados al *scraping* para que «sepan» a que pueden o no acceder. Sin embargo, precisamente el avance de la inteligencia artificial en relación a la interpretación del lenguaje natural posibilita que los modernos bots cuenten con la capacidad de cribar en función de lo que de forma coloquial exprese un humano, sin necesidad de los robot.txt. Si un usuario en su cuenta de X, Instagram, Facebook, o BlueSky dice que no quiere que se usen sus obras para «entrenar» IAs, sería ya un medio adecuado en estos momentos. Para los desarrolladores sería más complejo de implementar, pero no imposible, como demuestran los denominados agentes de IA, que se relacionan con Internet de forma mucho más parecida a los humanos que los bots ordinarios, superando incluso los clásicos CAPTCHA (lo que está haciendo que se busquen nuevos medios cuando se quiere reservar el acceso exclusivamente a personas reales).

Desde el verano del año 2025, Universal Pictures ha incluido en los créditos de sus películas una prohibición de su uso para entrenar IAs. En los libros y cómics cada vez es más habitual ver previsiones de este tipo, hoy eficaces legal y técnicamente. Que «manera adecuada» sea un concepto jurídico indeterminado ayuda a interiorizar este salto.

En cualquier caso, lo tuvieran en cuenta o no los redactores del Reglamento, la realidad es que las nuevas formas de mostrar oposición a la minería colocan más piedras en el camino de los «entrenamientos», que de por sí son más que simple minería. Por eso se echa en falta mayor valentía en el Reglamento para atreverse a pronunciarse acerca de si quieren o no dar cobijo a la utilización de datos no licenciados para el «aprendizaje automático» sin fines de investigación científica. Políticamente, se entiende que no lo hicieran para lograr una solución de compromiso entre los Estados más reacios a plantear limitaciones y los más estrictos, pero esa no decisión al final es una decisión, solo que basada en el Derecho previo al Reglamento.

El posterior Código de Prácticas y, más específicamente, de la medida 1.3 del Capítulo de consagrado al *Copyright,* permite adentrarnos un poco más en la cosmovisión comunitaria, aunque no haya sido elaborado directamente por las instituciones que confeccionaron el Reglamento.

La letra a) de la medida dice que habrán de emplearse *«rastreadores web que lean y sigan instrucciones expresadas de acuerdo con el Protocolo de exclusión de robots (robots.txt), como se especifica en la Solicitud de comentarios n.º 9309 del Grupo de trabajo de ingeniería de Internet (IETF), y cualquier versión posterior de este Protocolo para la cual el IETF demuestre que es técnicamente factible e implementable por los proveedores de IA y*

los proveedores de contenido, incluidos los titulares de derechos». Con esta referencia se podría concluir que la UE se ha quedado congelada en el estándar, aunque se haga mención también a eventuales evaluaciones en el protocolo. Esta sensación muta, sin embargo, al leer el encabezamiento general de la medida y letra b). La introducción previa enuncia que para respetar las reservas de derechos se acudirá hasta a tecnologías de vanguardia (en el original en inglés, «including through state-of-the-art technologies»). Asimismo, la letra b) espolea a *«identificar y cumplir otros protocolos legibles por máquina adecuados para expresar reservas de derechos de conformidad con el artículo 4(3) de la Directiva (UE) 2019/790, por ejemplo a través de metadatos basados en activos o en la ubicación, que hayan sido adoptados por organizaciones de normalización internacionales o europeas, o que sean de última generación, incluidos los técnicamente implementables, y ampliamente adoptados por los titulares de derechos, considerando diferentes sectores culturales, y generalmente acordados a través de un proceso inclusivo basado en debates de buena fe que se facilitarán a nivel de la UE con la participación de los titulares de derechos, los proveedores de IA y otras partes interesadas pertinentes como una solución más inmediata»*.

Se aprecia cierta ambivalencia en Código de Prácticas en relación a esta cuestión, porque todavía mira mucho al robot.txt, pero al mismo tiempo se entrevé en la imagen de conjunto que este método no es ya el *state of the art* y que cabe esperar la llegada de otros nuevos, si es que no están ya ahí. Recuerda a esos pronunciamientos en los que se admite la corrección de algo, pero solo de forma provisional. En el presente. Un presente que casi se está difuminando. La sentencia del Tribunal Regional de Hamburgo de la Sala Civil 10, de 27 de septiembre de 2024, es una muestra de ello.

Cuando se conoció este pronunciamiento judicial alemán, en general fue percibido como un espaldarazo al «entrenamiento» de IAs, pero su efecto creo que es más limitado de lo que se ha reputado. Es más, se perciben grietas desfavorables cuando da pie a que se tengan en consideración los *opt-outs* expresados en lenguaje humano, dado el progreso de la técnica; lo que conecta con la tesis defendida de que hay que ir más allá de robot.txt al ponderar que es medio adecuado o no, a efectos del artículo 4.3 de la Directiva 2019/790, 17 de abril.

La decisión judicial, rechazando la reclamación de un fotógrafo contra Laoin e.V. *(Large-Scale-Artificial Open Network)*, se fundamentó en que la minería se habría producido en el ámbito de la investigación científica, donde, según el artículo 3 de la Directiva 2019/790, 17 de abril, no cabe que los titulares ejerzan este derecho de exclusión, que, por el contrario, sí reconoce el artículo 4 de la misma Directiva. Esto supuso que ese punto quedara sin juzgar y que no se entrara al fondo. Ahora bien, que se plantee es ya sintomático respecto al artículo 4, que es el que se tendría que aplicar, en su caso, para los modelos LLM comerciales. El tiempo en este caso juega a favor de que se consolide lo insinuado por el tribunal, porque la tecnología va a seguir perfeccionándose. Más dudoso es si otras partes de la sentencia prosperarán.

Valorando el acierto de la observación de la evolución técnica en cuanto al examen de la de los medios de lectura robotizada, merece, por el contrario, crítica la lectura superficial que se hace en la sentencia de lo que es la minería de datos. El Tribunal interpreta que ha de darse validez a la prospección que genere creaciones como las de los datos «excavados» y no solo información, y que la norma no cierra el paso a dicha posibilidad, sin embargo, la emulación exige mucho más de lo que ofrece la minería. El Tribunal, asimismo, está omitiendo las transformaciones que sufren los datos durante el «aprendizaje».

Dada la dificultad técnica de la materia y la persistencia del Reglamento por continuar con el vocabulario de la Directiva del año 2019, a pesar de todo lo que ha cambiado desde entonces, no sería extraño que la sentencia tenga réplica y que se continué equiparando el análisis semántico clásico con el que, además, asimila lo sintáctico. El Código de Prácticas parece decantarse por ese enfoque, según lo comentado. Eso sí, incluso si esto ocurre, de interiorizarse la visión extensiva de los medios adecuados en relación con la materialización de los *opt-outs*, la excepción perdería igualmente mucha virtualidad, lo que nos lleva al título del epígrafe: la excepción de minería no basta para excavar esta veta.

Bien por rechazar ontológicamente su aplicación, bien por haberse ampliado el margen de oposición, el futuro de la excepción de minería debería ser menos prometedor de lo que se augura, salvo que se caiga en el *too big to fail* y se acomode la interpretación de los medios adecuados de lectura mecánica no a lo factible técnicamente, sino a lo que en la práctica se esté haciendo.

Merece la pena dejar apuntado que la cláusula TDM japonesa, por lo general, mucho más amplia que la mayoría del resto del mundo y, en particular, de la de la Directiva comunitaria, valora el «disfrute». El artículo 30.4 de su Ley, según la traducción inglesa facilitada vía web por el *Copyright Research and Information Center* (CRIC) de Japon, declara que *«es lícito explotar una obra, en cualquier forma y en la medida que se considere necesaria, en cualquiera de los casos siguientes, o en cualquier otro caso en que no sea propósito de una persona disfrutar personalmente o hacer que otra persona disfrute de los pensamientos o sentimientos expresados en esa obra»*. Con esta regulación se quería distinguir entre los usos recreativos del material protegido, fuente de remuneración de los autores, y los que carecen de ese carácter, desposeyendo en estos últimos a los titulares de sus facultades habituales como propietarios. Esta fórmula, sin embargo, muestra otra cara cuando se trata de la IA generativa, donde los sentimientos y pensamientos del autor son reutilizados, a veces de forma clara y otras de forma más sutil. Normalmente, este contenido semántico no sería objeto de protección, pero aquí se quiere proteger la forma y el fondo a la vez.

Trayendo a colación la moda de convertir las imágenes a un presunto estudio Ghibli, a la que hacía referencia más atrás, creo que estaríamos ante la clase de disfrute que activa los resortes restrictivos de la cláusula TDM y al menos en 2024 esa parecía ser también la visión de la Agencia Japonesa

para Asuntos Culturales. Que esta excepción pueda ver rebasada hasta en el marco de la legislación más amplia del mundo dice mucho.

Como coda a este repaso de las excepciones de los derechos de explotación, y dejando atrás la minería, el concepto de pastiche ha ido cobrando un mayor protagonismo del que cabía esperar en un inicio. No siendo lo mismo —o sí, porque como veremos acotar hasta donde llegan los pastiches no resulta sencillo[70]—, la noción del collage fue clave en la validación legal de *«A single piece of american cheese»*, a los efectos de poder registrar su autoría. Conforme surjan más conflictos legales, es muy probable que haya quienes defiendan los *outputs* artificiales usando el pastiche como escudo jurídico en analogía con el criterio de la Oficina de Derechos de Autor de EE. UU. A este respecto, la lectura que hace Gema Rosés de la figura es digna de consideración, con el matiz crucial de que sus reflexiones se plantean en el marco del apropiacionismo, no entrando a valorar su hipotética interacción con las IAs generativas.

En la Directiva 2001/29/CE de 22 de mayo, que introdujo el término en el ámbito de la UE, no hallamos ninguna definición de pastiche. Tampoco la Directiva 2019/790, de 17 de abril, que convirtió en una obligación para los Estados miembros lo que había sido hasta entonces solo una posibilidad de regulación. Tanto el artículo 5.3 de la primera Directiva como el 17.7 de la posterior se limitan a emplear el término sin aportar ninguna explicación de él dentro del articulado[71]. Si bien es cierto que estos artículos hacen lo mismo con el resto de conceptos que utilizan (como cita o parodia), en el caso del pastiche resulta más problemático, porque esas otras nociones en líneas generales ya estaban presentes en los ordenamientos nacionales de una u otra forma. En cambio, la de pastiche proviene del Derecho francés y ni siquiera dentro de él están claros cuales sus contornos, habiendo dudas de si es un sinónimo de parodia y caricatura, o si posee una sustantividad diferenciada.

A pesar de esta falta de descripción del pastiche, y de la dificultad de su deslinde en el Derecho que sirvió de inspiración a la normativa comunitaria, caben extraer algunas notas de la Directiva 2019/790, de 17 de abril, partiendo del contexto en el que se inserta su artículo 17.7.b, así como del considerando número 70 en el que explican las razones detrás de la introducción de este precepto. Ahora bien, hay que tener claro que las características que se desprendan de la Directiva pueden no ser compartidas por la legislación de los Estados miembros, que ante las dudas que suscita el concepto, han

70 La armonización pretendida por la Directiva ha derivado en una pluralidad de enfoques, más allá de un núcleo común compartido.

71 Es verdad que lo mismo ocurre cuando citan la parodia o la caricatura como usos permitidos, pero estas nociones resultan probablemente mucho más intuitivas desde un sentido social, al margen de que pueda haber algunos casos ambiguos o conflictivos que terminen judicializándose y que exijan hilar más fino.

tendido a la dispersión más allá de cierto núcleo común. Un ejemplo de ello es que en Alemania la jurisprudencia llega a hablar de pastiches inversos.

En la Directiva de 2019, los rasgos esenciales de la excepción es que se constituye ligada al mundo *on line* (algo que también se capta en la Directiva 2001/29/CE de 22 de mayo) y que está vinculada de forma especial con la libertad de expresión. El considerando menciona que la regulación busca *«lograr un equilibrio entre los derechos fundamentales establecidos en la Carta de los Derechos Fundamentales de la Unión Europea (...), en particular la libertad de expresión y la libertad artística, y el derecho de propiedad, incluida la propiedad intelectual»*. No obstante, antes de llegar ahí, las primeras frases con las que se abre aluden solo a la libertad de expresión, sin mencionar la artística: *«las medidas tomadas por los prestadores de servicios para compartir contenidos en línea en cooperación con los titulares de derechos deben serlo sin perjuicio de la aplicación de las excepciones y limitaciones de los derechos de autor, incluyendo en particular aquellas que garantizan la libertad de expresión de los usuarios»*. Esto, desde mi punto de vista, supone que la preocupación principal detrás de esta limitación sea más expresiva que artística, al repetirse dos veces frente a una la apelación a la libertad de expresión.

Las explicaciones que da la Directiva de sí misma son confusas, pero si hacemos un análisis de conjunto creo que prevalecen los argumentos de que el fin del artículo 17.7 es proteger la libertad de expresión antes que la artística. Evidentemente, en la artística hay expresión también (el arte nunca es neutro), pero de manera intuitiva se capta la diferencia entre decir una opinión con un megáfono o un tuit a plasmarla en cualquiera de las múltiples manifestaciones artísticas concebidas por el ser humano. El enunciado del precepto no reconoce un derecho a la creación sino al alojamiento de contenidos ya generados. Cabría argüir, con sentido además que, si estos contenidos no fueran legales, la Directiva no ampararía que se compartieran, por lo que tan solo es un matiz de expresión el que el enunciado se construya desde la perspectiva de la difusión y no de la creación. No obstante, que se plantee así va en consonancia con la idea de que lo que quiere garantizarse es que haya una expresión espontánea y fluida.

La cita, la crítica, la reseña, la caricatura o la parodia no existen de forma autónoma, son necesariamente relacionales, partiendo de que el ser humano siempre está comunicando y, por lo tanto, las actuaciones puramente individualistas escasean más de lo que somos conscientes. Se critica, se caricaturiza, se cita, se reseña, o se parodia algo y al hacerlo se está adoptando una postura. Para que Cervantes parodiara las novelas de caballería, primero tuvieron que existir estas o no hubiera podido reconstruirlas desde su particular y quijotesca lente. En ese sentido, lo que proporciona el artículo 17.7 es un conjunto de herramientas jurídicas para que las personas se posicionen frente a ese algo, sin perjuicio de que a veces también desplieguen una faceta artística, dependiendo de la pericia y el gusto de su ejecutor.

El pastiche, en consecuencia, ha de considerarse como parte de esa caja de herramientas puestas a disposición de los usuarios, para que se manifiesten como sujetos políticos. Si la intención del legislador comunitario hubiera sido dotar de un marco general a la creación por medio de retazos, a lo Frankenstein, la excepción se encuadraría en un contexto más amplio y quizá se hablará de autores o artistas, en vez de usuarios[72].

Que el telón de fondo de las Directivas en la materia siempre haya sido lo digital puede justificar ciertos silencios, pero no todos, porque parte de ellos involucran a realidades propias del actual panorama tecnológico. En el 2016, se habían desarrollado mediante *machine learning* proyectos como *The Next Rembrandt,* en el que se creó un cuadro inédito imitando el estilo del artista neerlandés, recurriendo a 168.263 fragmentos de sus 326 obras. Corolario de ello, en el momento de redactar la Directiva del 2019, había una base para establecer una excepción de pastiche en la que se reconociera expresamente la creación de obras nuevas por combinación mecánica, incluyendo las que se generen utilizando trabajos no pertenecientes al dominio público.

Hay quién podría replicar que *The Next Rembrandt* es un exponente de la minería de datos, en la línea de la sentencia del Tribunal Regional de Hamburgo de la Sala Civil 10, de 27 de septiembre de 2024, y que, por ello, no hay ningún vacío respecto a estas técnicas. Sin embargo, de ser esto así, el resultado sería extraño, porque en unos casos la creación de pastiches mediante obras sujetas a derechos de propiedad intelectual sería libre (los hechos manualmente de forma física o digital) y en otros, los generativos, dependería del consentimiento de los autores o de que estos no hayan ejercitado sus derechos de exclusión. ¿Qué sentido tendría que ante resultados similares o idénticos el cauce jurídico fuera diferente? Sería tanto como decir que el plagio es válido si se usa un bolígrafo, pero no un lápiz o un ordenador. Atendiendo estos factores, no creo que la excepción de minería esté diseñada para suplir el encuadre limitado de la del pastiche.

Por otro lado, que la figura del pastiche se agrupe en la misma letra que la caricatura y la parodia resulta, a mi juicio, muy revelador del sentido y propósito que se quiere dar al precepto. La restricción en la Directiva del pastiche a los servicios de alojamiento de contenidos en línea es algo deliberado.

El legislador contaba con muchas posibilidades que gozaban de respaldo fáctico y, sin embargo, no las ejercitó. De haberlo deseado, hubiera podido dividir el apartado 7 del artículo 17 de la Directiva en tres letras, añadiendo una más, y separar el pastiche para remarcar su carácter propio frente a la parodia y la caricatura. Igual que nada le impedía refundir las dos letras en una única y convertirlas en una larga enumeración. Pero no lo hizo. En estas

72 El apropiacionismo no es un fenómeno restringido a lo virtual, es posible en el plano analógico, siendo precisamente este último de donde proviene, con Marcel Duchamp como uno de sus inspiradores.

decisiones de diseño hay, a mi juicio, intencionalidad. Separa lo que es diferente, agrupa lo parecido.

Los supuestos de las letras en los que está dividido el artículo 17.7 son coherentes temáticamente entre sí. En la letra a) se recogen las citas, reseñas y críticas, que, sin ser nociones idénticas, mantienen un hilo conductor común claro. Con la letra b), donde confluyen la parodia, la caricatura y el pastiche, sucede igual, al menos con la parodia y la caricatura (y, según mi criterio, también con el pastiche).

Si el pastiche de la Directiva se refiriera al collage artístico u otras técnicas asimiladas, desentonaría con sus compañeras de letra. La naturaleza de las creaciones de Bansky o Warhol difiere de las dos otras formas de adaptación pertenecientes a la letra b). Tanto en la parodia como en la caricatura hay presente un *animus iocandi* intrínseco que no tiene por qué aparecer, en cambio, en los collages y otras formas de agregación artística análogas (me remito de nuevo a los muchos exponentes del apropiacionismo que solo persiguen un efecto estético con la recombinación).

Las parodias y caricaturas implican, asimismo, la deformación del original en un grado elevado, mientras que los collages se realizan más como una combinación o selección, pudiendo no haber ni siquiera alteración. Lo que se toma de la obra u obras originales con frecuencia se conserva íntegramente, o se traslada de manera que pueda resultar fácilmente reconocible su conexión con la fuente de la que se ha extraído. Es algo que se percibe en el estudio de Francis Bacon del Papa Inocencio X, o en las numerosas reinterpretaciones del cuadro de las Meninas de Velázquez.

Debido a todo esto, me inclino a pensar que bajo el término de pastiche de la Directiva se ocultan dos supuestos: los memes (expresión de la que se rehuiría por resultar más coloquial y menos elegante, o bien por inercia de la nomenclatura de la Directiva del año 2001) y los *fan fictions*.

Qué es arte y qué no lo es representa uno de los grandes debates imposibles de resolver de la humanidad, por lo que sería atrevido negar la condición artística de los memes. Incluso si las personas que los hacen no se perciben como tales, según algunas teorías artísticas, lo serían, no constituyendo la autoidentificación un requisito para ostentar la condición de artista. A fin de cuentas, los primeros que hubo en Altamira o Leang Tedongnge es improbable que tuvieran conciencia de serlo. En cualquier caso, en un sentido más pedestre y social, el grueso de los memes no pretendería constituir arte sino algo distinto, de fronteras porosas, eso sí, como le ocurre al grafiti, que va desde el garabato (por ejemplo, «Tal estuvo aquí», «Viva mi partido político» o «Abajo mis rivales») a composiciones más complejas. La finalidad fundamental de los memes es la divertir y/o transmitir una idea, seria o cómica (aunque casi siempre revestida de cierta mordacidad), descuidando usualmente la forma. En definitiva, los memes cohonestan bastante bien con el pastiche, tanto por su intertextualidad (para construirlos, se toman toda

clase de referencias, sin atender a la titularidad del material base) como por su eminente valor expresivo, lo que les anuda con la libertad de expresión.

La otra vertiente del pastiche, de acuerdo con mi tren de razonamiento, serviría para dar cobertura a las obras derivadas de bajo o nulo impacto que crean sin autorización los fans y que es raro que den lugar a derechos de autor, precisamente por estar vinculadas a una obra que sí está protegida y que es ajena a ellos[73] (véase, los *fan fiction* o *fan arts*). Estos pastiches serían obras *a priori* no profesionalizadas (aunque a veces puedan ser ejecutadas con gran maestría), donde prevalecería no tanto la pretensión artística sino el actuar como muestras espontáneas de admiración o como un intento de dialogar con el original. El final alternativo que escribe el admirador de una historia previa no pretende exhibirse ante el mundo como original, ni sustituir a sus auténticos creadores, sino formar parte de algo que aprecia dentro de la dinámica relacional de la cultura; lo que conecta con la sociabilidad inmanente de la propiedad intelectual.

Antes de Internet, existía ya el fenómeno de los fanzines, en el que de forma paralela a los circuitos oficiales se publicaban una amplia variedad de contenidos. Por su carácter más *underground* a través de ellos a veces se daba circulación a versiones apócrifas de personajes o historias preexistentes. Un ejemplo de ello son las famosas Tortugas Ninjas que comenzaron como un homenaje autoeditado de los cómics de *Daredevil* y *Ronin* de Frank Miller. Empero, la verdadera explosión de la creación fan vino con Internet, al favorecer aún más la auto publicación y la comunicación entre personas con intereses similares. De ahí que tenga sentido que la Directiva del año 2019 tomara interés por dar respuesta jurídica a lo que ya era una realidad de facto.

El problema de lo expuesto es que, convincente o no, no lo encontramos en la norma de manera directa. España engarzó con los memes en la exposición de motivos del Real Decreto-ley 24/2021, de 2 de noviembre, que traspuso la Directiva *«se incluye por primera vez en el derecho español el límite de "pastiche", que refuerza la cobertura legal de las expresiones y construcciones multimedia que se replican y transmiten mediante Internet de persona a persona hasta alcanzar una amplia difusión —"los conocidos como «memes"—»*. Pero nada se dice de los productos fan; tema que, por el contrario, sí que abordó el legislador alemán al hacer la traslación junto a algún extra[74]. No

73 Se cuenta que la saga de las populares, aunque mediocres, novelas de 50 Sombras de Grey, en origen eran un *fan fiction* de las novelas de la saga Crepúsculo. En su concepción original como *fan fiction de* E.L. James, hubiera sido imposible registrarlas. Pero nada impidió que con unos cambios de nombre y ambientación se convirtieran en novelas autónomas.

74 Alemania parece haber hecho una lectura amplia de la figura, basada seguramente en la alusión puntual a la libertad artística que hacía el considerando 70 de la Directiva 2019/790, de 17 de abril de 2019. En el informe que acompañó al proyecto de Ley alemán se decía *«citar, imitar y tomar prestadas técnicas culturales es un elemento definitorio de la*

tenemos una idea unívoca, por lo que, según el país, habría margen para discutir que el collage o los *outputs* artificiales son exponentes del pastiche. Aclarar, eso sí, que el que exista espacio para el debate no es equivalente a que efectivamente se desprenda que son pastiches.

La regulación del artículo 70 del Real Decreto-Ley, por ejemplo, va más allá de lo que explica su exposición de motivos y del espíritu de la Directiva, empezando por el hecho de que se reconozca expresamente los pastiches fuera de lo digital; lo que desentona con la *ratio legis* dada en la exposición, que se centra únicamente en las características de Internet. Los pastiches son descritos como la *«transformación de una obra divulgada que consista en tomar determinados elementos característicos de la obra de un artista y combinarlos, de forma que den la impresión de ser una creación independiente, siempre que no implique riesgo de confusión con las obras o prestaciones originales ni se infiera un daño a la obra original o a su autor. Este límite será también aplicable a usos diferentes de los digitales».*

La definición está tomada directamente del Diccionario de la Lengua Española, solo que, prescindiendo de su encabezamiento, que vincula al pastiche con el plagio y la imitación con tintes peyorativos. Esta elección por parte del legislador fue un tanto arriesgada, porque suponía unir la excepción de la Directiva con algo que considera pernicioso la Real Academia Española. No obstante, si consideramos la otra parte de contenidos que cobijaría la noción de pastiche, tiene sentido, porque hablamos de una imitación no disimulada, eso sí, carente de la intención maliciosa del plagio, que supone usurpación. El inconveniente es que no se limita ahí, sin saber hasta donde se quiere llegar.

¿Los *outputs* artificiales serían pastiches de acuerdo con esta definición? Sí y no. Hasta «creación independiente» encajarían con el enunciado, pero no cumplirían las otras dos condiciones adicionales que ha incorporado el legislador español, en la línea de la regla de los tres pasos del Convenio de Berna. Puede haber confusión (de hecho, eso es lo que se persigue cuando se solicita a los modelos que hagan imágenes, vídeos, o textos con un determinado estilo o usando motivos preexistentes), además de daños. Una petición aislada no perjudica de forma perceptible, pero gota a gota genera un proceso erosivo. Sobran los informes que hablan de importantes caídas de ingresos de los autores a causa de la irrupción de la IA generativa.

En Alemania, que se mencionaba como muestra de las divergencias en la transposición, se ha puesto un mayor énfasis en el componente artístico, en

intertextualidad y de la creación y comunicación cultural contemporánea en la "red social". En particular, nos vienen a la mente prácticas como el remix, el meme, el GIF, el mashup, el fan art, el fan fiction, el cover o el Sampling». La temprana sentencia de 2 de noviembre de 2021 del Tribunal Regional de Berlín relativa a la obra *Schorched Heart* asimiló esta perspectiva jurisprudencialmente. Ahora bien, al incluir el *sampling*, se estarían desviando de la pauta de la sentencia del TJUE, asunto C-476/17, en la que se resolvía precisamente una cuestión prejudicial planteada por el Tribunal Supremo alemán.

paridad con el expresivo, pero tampoco sería fácil calificar los *outputs* artificiales como pastiches, precisamente por la necesidad que se impone de cierto juego artístico que en la producción industrial algorítmica solo se cumpliría de forma puntual en supuestos como el de «*A single piece of american cheese*».

La legislación de estos dos Estados denota que la noción de pastiche se ha ensanchado, por lo que incluso habiendo llegado a la conclusión de que los *outputs* artificiales no serían pastiches, las razones no coinciden con las que cabría dar solo con la Directiva en la mano. Hay un componente impredecible en lo que ocurriría, fallando aquí el objetivo de armonización pretendido.

De todas formas, de entrar en juego la limitación derivada de los pastiches, lo haría en la fase de comercialización de las herramientas de IA, no afectando a los derechos de reproducción y transformación relacionados con el «entrenamiento». La conversión en «lenguaje máquina» de los datos usados en el aprendizaje no resulta subsumible en la idea de pastiche acuñada desde la legislación española o alemana (y seguramente con ninguna de la de los Estados de la Unión Europea), ni tampoco con la de la Directiva de 2019.

De haber alguna repercusión, podría ser en cuanto a las obligaciones de mitigación del capítulo de *Copyright* del Código de Prácticas, o al contenido de las demandas frente a Midjourney, en la parte que se censura que los usuarios puedan pedir imágenes basadas en personajes u otros elementos sujetos a propiedad intelectual[75]. Lo que no es poco, pero no tanto como para permitirse relajarse en cuanto a la política de *Copyright,* bajo la esperanza de que el pastiche actuará como salvavidas.

Conforme al Derecho de la Unión Europea, mi conclusión es que las IAs generativas deberían haberse construido sobre un sistema de contenidos licenciados en aquello que no sea de dominio público, al no tener cabida el «entrenamiento» plenamente en las excepciones previstas por el marco legal. Por inconveniente que sea para las empresas, sus últimos movimientos demuestran que es factible llegar a acuerdos y que los datos tienen un enorme valor económico como para tomarlos unilateralmente. Si el legislador ve insuficiente esta vía, ha de actualizar el ordenamiento jurídico. Pretender excavar con el pico de la minería o el pastiche aboca a que el instrumento se rompa, sin penetrar en la roca. Igualmente, si no fuera viable obtener el consentimiento, revertir. Porque los beneficios son privados. La herramienta es diseñada conforme a sus preferencias y prioridades. La generalidad del Derecho tiene sus límites.

75 Planteó este escenario por no descartar interpretaciones factibles, pero atendiendo a la finalidad de las figuras como el pastiche o el collage, el acomodo jurídico resulta a mi parecer muy poco orgánico. De ahí que sea más probable que no se puedan beneficiar de la excepción del pastiche los *outputs*.

En favor de la seguridad jurídica, se debería dar un paso más allá del Reglamento (UE) 2024/1689, de 13 de junio, y responder realmente cual se desea que sea la política de propiedad intelectual para las próximas décadas. De *lege lata*, las licencias voluntarias y acordadas son y deberían haber sido la base de construcción de la mayor parte del entramado de inteligencias artificiales.

V. Un breve apunte sobre la doctrina del fair use y su jurisprudencia

El *fair use* y su doctrina constituyen una prueba clara de que no debe subestimarse la generalidad del Derecho. Los procesos judiciales que están surgiendo alrededor de la inteligencia artificial en EE. UU. son quizás algunos de los más importantes de su historia judicial y para resolverlos el terreno de juego lo determinan estas dos palabras, cuyas raíces se remontan a las primeras décadas del siglo XVIII en Inglaterra, casi a la par que la aprobación del vanguardista Estatuto de Ana de 1710.

En EE. UU., el *fair use* se acogió por primera vez en la sentencia Folsom v. Marsh de 1841. Planteada inicialmente por el juez Joseph Story como una vía de ampliación de las facultades asociadas a la propiedad intelectual, no tardaría mucho en derivar en lo contrario. Nos hallamos, por lo tanto, ante una figura de entre dos o tres siglos de antigüedad, dependiendo del hito que tomemos de referencia a la hora de hacer el cómputo, y de cómo valoremos las construcciones premodernas que han actuado como cimientos.

El artículo 107 de la *Copyright Act* de los Estados Unidos condensa el periplo judicial que ha atravesado el *fair use* desde su introducción en el país. Resulta remarcable, a este respecto, que, desde la aprobación de la ley en el año 1976, solo haya experimentado dos modificaciones, en 1990 y 1992, respectivamente. Modificaciones que, de todas formas, no alteraron su redacción sustancialmente. Por consiguiente, incluso si nos limitáramos a contabilizar el *fair use* a partir de la norma de 1976, aún hablaríamos de una herramienta jurídica con casi 50 años a sus espaldas (33 con el tenor actual, tras las dos enmiendas de los años 90). Prácticamente tan vieja como las promesas de Marvin Minsky o Herbert Simon, que recordaba en la apertura del capítulo. Y, sin embargo, con todos esos años a sus espaldas, el *fair use* se siente lozano de cara a analizar las implicaciones del proceso de «aprendizaje» de las IAs en la propiedad intelectual. Más que las cláusulas TDM de los Estados «mineros», que se han quedado congeladas en una terminología prematuramente anticuada pocos años después de haberla positivado[76].

76 En ciertos aspectos, la definición del artículo 2.2 de la Directiva (UE) 2019/790, de 17 de abril, guarda similitudes con la del artículo 107 de la *Copyright Act* de EEUU, al optar ambos preceptos por establecer un listado casuístico *numerus apertus* y no *numerus*

Esto no significa necesariamente que se equivocara con su enfoque el Derecho de la Unión Europea, o el de los demás países que optaron por regular también una excepción de minería. La relativa indefinición de la fórmula estadounidense representa una ventaja y un defecto a la vez, según los ojos con los que se observe. En los sistemas de propiedad intelectual sostenidos por excepciones acotadas individualmente, cada una refleja un juicio de oportunidad del legislador, en el que se manifiesta, a su vez, el criterio del cuerpo electoral por medio de sus representantes democráticos. En EE. UU., al margen de las peculiaridades de su carrera judicial, y de los pocos ejemplos del artículo 107, la valoración descansa sobre los jueces, con una mayor desconexión del estado de opinión social. De hecho, lo contrario sería peligroso, porque su tarea es interpretar adecuadamente las normas, no representar a la sociedad; rol que encarnan otros poderes del Estado. Sin perjuicio, lógicamente, de la interpretación de las normas atendiendo a la realidad social[77].

En el ámbito de la UE, aunque los jueces y su jurisprudencia tienen un papel crucial, nunca podrían constituir una excepción *ex novo* sin un mínimo engarce legal que al menos permita llevar a cabo una exégesis de tipo analógico. Con el *fair use,* sí es posible. Teóricamente, la excepción nunca deja de ser la misma, pero en la práctica sus concreciones actúan como supuestos inéditos. La ingeniería inversa, por ejemplo, cuando se reconoció en la jurisprudencia estadounidense marcó un precedente inédito respecto a su trayectoria previa, por más que en la *Copyright Act* haya una invocación genérica a la investigación.

Por supuesto, en EE. UU. el legislador podría fijar nuevas excepciones expresas, ampliar el listado ya desfasado del artículo 107 de la *Copyright Act*. Las peticiones del sector de la inteligencia artificial generativa iban en la línea de consagrar el «entrenamiento» como uno de los casos concretos de uso legítimo, presumiendo de base el cumplimiento de los cuatro requisitos

clausus. La razón de esta elección, además, es la misma: sentir que se hallan ante una materia en continua evolución, por lo que cualquier enumeración se puede ver superada por la realidad con el transcurso del tiempo (lo que ocurrió efectivamente con la legislación americana, cuyos supuestos ejemplificativos están anclados a un paradigma predigital). Sin embargo, hay una importante diferencia y es que en EEUU la idea de «uso legitima» es mucho más amplia que la de minería. La sentencia del Tribunal Regional de Hamburgo de la Sala Civil 10, de 27 de septiembre de 2024, ha intentado obrar de forma parecida a los tribunales de EE. UU., cuando subsumen las novedades que van surgiendo en la realidad en la noción de uso legítimo. Como concepto jurídico indeterminado, la legitimidad es mucho más vasta que algo como la minería, que solo puede ser lo que técnicamente encaja con esa técnica. Incluso desde una lectura expansiva, el margen de ensanche va a ser limitado y dejar abierto los ejemplos no basta para que la minería sea equiparable al fair use. Por eso, ante la evolución tecnológica, el descuelgue legislativo en Europa era más probable que en EE. UU.

77 Algo siempre controvertido, pero más en EE. UU. por las disputas entre las corrientes originalistas y las evolutivas.

que contiene el precepto. Pero la tendencia de EE. UU. no ha sido esa, aunque eventualmente pudiera cambiar en el futuro. Son los tribunales los que han aportado savia nueva, con más o menos acierto, al anquilosado muestrario del artículo 107, extrayendo de él supuestos de uso más allá de los explicitados que ni remotamente concebían los redactores de la *Copyright Act*.

Esto es un arma de doble filo, porque facilita la evolución dinámica del ordenamiento, pero también lo expone a un grado de mayor inseguridad jurídica que si se determina desde un inicio que se puede y que no (aunque lo cierto es que luego la realidad siempre rebasa hasta lo más detallado con giros imprevistos, convirtiendo en abierto lo que formalmente se supone cerrado). Entraña, asimismo, dejar en manos de la hipotética judicialización de los particulares la determinación de la legalidad de ciertas prácticas. No solo de particulares, sino de su capacidad de defensa (en las primeras sentencias sobre inteligencia artificial generativa de 2025, los jueces apuntan a que, con otros argumentos, quizá el sentido de sus sentencias hubiera sido diferente).

El casuismo por el que se ha optado en Europa o en otros Derechos permite conocer con mayor certeza la legalidad de ciertas prácticas[78], sin necesidad de que se produzca un conflicto judicial (estos en su caso ayudaran a terminar de perfilar los contornos exactos de la norma). A cambio, está más expuesto a sufrir una obsolescencia sobrevenida. De ahí que no se pueda decir categóricamente que un modelo sea superior o inferior al otro. Lo que sí es evidente es que, bueno o malo, en EE. UU. es más fácil saber a dónde acudir para juzgar la legalidad del «entrenamiento», tanto si lo avala como si no.

El artículo 107 de la *Copyright Act* es la referencia ineludible y nos proporciona todas los elementos de juicio que necesitamos: «*sin perjuicio de lo dispuesto en los artículos 106 y 106A, el uso legítimo de una obra protegida por derechos de autor, incluido su uso mediante reproducción en copias o fonogramas o por cualquier otro medio especificado en dicho artículo, con fines como la crítica, el comentario, la información periodística, la enseñanza (incluidas las copias múltiples para uso en el aula), la investigación o la beca, no constituye una infracción de los derechos de autor. Para determinar si el uso de una obra en un caso particular constituye un uso legítimo, los factores a considerar incluirán: (1) la finalidad y el carácter del uso, incluyendo si dicho uso es de naturaleza comercial o con fines educativos sin fines de lucro; (2) la naturaleza de la obra protegida por derechos de autor; (3) la cantidad y la sustancialidad de la parte utilizada en relación con la obra protegida por derechos de autor en su conjunto; y (4) el efecto del uso sobre el mercado potencial o el valor de la obra protegida por derechos de autor. El hecho de*

78 Naturalmente, esta afirmación ha de matizarse porque en las propias páginas del artículo se ha podido ver las dificultades que entrañaba el concepto de pastiche, o las dudas acerca de la limitación de la minería alcanza el contenido sintáctico y no solo semántico de la información.

que una obra no esté publicada no impedirá por sí mismo que se determine que existe un uso justo si dicha determinación se realiza considerando todos los factores anteriores».

Todos los pleitos relativos al uso legítimo giran a las cuatro exigencias del artículo 107 y los que están emergiendo con las IAs generativas no son diferentes.

De forma prematura, ha habido quienes han percibido las resoluciones dictadas por dos jueces de la Corte 4 del Distrito Norte de California en 2025 como un punto decisivo en la disputa entre los creadores críticos con las IAs generativas y las empresas que las fabrican, sin embargo, como sucedía con la sentencia de Hamburgo, hay espinas para los desarrolladores de IAs. Para empezar, hay divergencias entre ambas, a pesar de que forman parte del mismo distrito judicial.

La resolución del caso *RICHARD KADREY v. META PLATFORMS, Inc* comienza en su primera página indicando lo siguiente: *«dado que el rendimiento de un modelo de IA generativa depende de la cantidad y la calidad de los datos que absorbe como parte de su entrenamiento, las empresas no han podido resistir la tentación de incorporar materiales protegidos por derechos de autor a sus modelos sin obtener el permiso de los titulares de los derechos ni pagarles por el derecho a usar sus obras con este fin. Este caso plantea la cuestión de si dicha conducta es ilegal. Aunque la clave está en los detalles, en la mayoría de los casos la respuesta probablemente será sí. Lo que más preocupa a la legislación sobre derechos de autor es preservar el incentivo de los seres humanos para crear obras artísticas y científicas. Por lo tanto, generalmente es ilegal copiar obras protegidas sin permiso. Y la doctrina del "uso legítimo", que ofrece una defensa ante ciertas reclamaciones por infracción de derechos de autor, no suele aplicarse a las copias que reducen significativamente la capacidad de los titulares de los derechos de autor para obtener beneficios económicos de sus obras (lo que disminuye significativamente el incentivo para crear en el futuro)»*.

BARTZ v. ANTHROPIC, en cambio, resulta más favorable a las pretensiones de las empresas del sector tecnológico, pero, aun así, como ya se había adelantado en páginas anteriores, este pronunciamiento también constataba la aparente utilización de material pirateado[79]. Esta circunstancia constituye una infracción del copyright, lo que ha dado pie a un acuerdo de más de 1.500 millones, que abonará Anthropic para compensar a los autores.

En sentido contrario, el caso *ROSS INTELIGENCE v. THOMSON REUTERS*, visto superficialmente como un tanto a favor de los autores, es poco prometedor en realidad porque fue la falta de un mayor carácter transformativo (propio del primer requisito del *fair use*) uno de los factores clave para que

79 Algo que en otras resoluciones judiciales preocupo menos, como PERFECT 10 v. AMAZON, de 2007.

Reuters se impusiera. Y no se puede negar que las IAs generativas cumplen sobradamente el requisito de tratarse de un uso transformativo, de acuerdo con los parámetros seguidos en *AUTHORS GUILD, INC. v. GOOGLE INC*, así como otras sentencias conexas.

El panorama es que las sentencias que debían ser un espaldarazo para el modelo vigente de desarrollo de las IAs generativas han resultado ser collejas y que algunas de las que les perjudicaban son una flecha que pasa sin rozarles.

Aunque no se refiere al campo de las IAs, la sentencia del Tribunal Supremo de EEUU, de 18 de mayo de 2023, sobre *ANDY WARHOL FOUNDATION FOR THE VISUAL ARTS, INC. v. GOLDSMITH* a mi juicio resulta mucho más significativa que la de ROSS INTELIGENCE, tanto por la instancia judicial, por sus mayorías (siete contra dos, lo que supera la división entre magistrados en función de su designación), y por el criterio que adopta respecto a los derechos de explotación de los autores, especialmente protector. Dado que el Tribunal Supremo incluye los trabajos derivativos que alteren el original, su doctrina se puede trasladar a los *outputs* en fase de comercialización de las IAs y, por ende, a los juicios que afronta en estos momentos, a título de muestra, *Midjourney.*

Reviste de notoriedad, igualmente, el caso *HACHETTE v. INTERNET ARCHIVE*, confirmado en apelación en 2024, por el que se rechazó la aplicabilidad del *fair use* y el derecho de primera compra o *first use* (presente en la legislación europea de derechos de autor), como base para constituir una biblioteca *on line* que permitiera un acceso menos limitado a los libros digitales del que permiten los libros electrónicos.

Aunque ya se ha establecido que la regurgitación de datos íntegros no es habitual en las IAs generativas, si son capaces de resumir la esencia de los mismos, o hacer replicas muy próximas, y disuaden de la consulta de los originales, esto es algo que repercutiría en la apreciación del *fair use*. Un caso evidente de posible sustitución es la del periodismo. Incluso si los propietarios de los modelos abonan la suscripción de pago de los periódicos, si luego reproducen lo que se cuenta en ellos, estaríamos muy próximos a los presupuestos del caso de INTERNET ARCHIVE.

El repaso jurisprudencial podría ser más extenso, porque casos históricos no faltan, pero al final, yendo a lo esencial, el destino del «entrenamiento» de las IAs generativas se resolverá en el conflicto entre el primer y el cuarto factor del *fair use,* y en esto el peso del pasado es relativo, porque los jueces nunca se han encontrado con un caso de esta entidad. No en vano se dice que todo o prácticamente todo el conocimiento que había disponible en Internet ya habría sido empleado (cuestión distinta es que cada día se renueve y amplié con millones de expresiones humanas nuevas). Nunca había habido un caso de uso no autorizado de materiales protegidos por el *Copyright* de tal amplitud, cuantitativa y cualitativamente. Al mismo tiempo, aunque los fun-

damentos técnicos son en parte viejos (no todos, porque el paradigma Transformer, proviene del artículo de 2017 *Attention Is All You Need*), la manera de integrarlos no lo es en absoluto. Las IAs actuales son percibidas como un salto enorme en la interacción entre el ser humano y las maquinas.

Normalmente, al uso transformativo se le ha dado gran importancia en la jurisprudencia de EE. UU., hasta el punto de que podía anular la valoración del resto de criterios, en aras de promover el progreso técnico, con especial énfasis en los casos donde este revierte en aplicaciones destinadas a un uso general por parte de la ciudadanía (como es el caso de los modelos LLM de amplio espectro). Sin embargo, las proclamas apocalípticas o utopistas lanzadas por los promotores de las IAs generativas juegan en su contra. Si continuamente hablan de cómo van a convertirse en superfluos millones de empleos, difícilmente se me ocurre un escenario en el que cuarto factor brille con más fuerza.

A título ilustrativo, Getty Images y Shutterstock, competidores tradicionales, anunciaron su fusión a principios de 2025, ante el descenso en la demanda de los repositorios de imágenes, en favor de las generadas artificialmente (campo en el que también intentarán participar). De forma intuitiva, es algo que se percibe en muchas webs, donde cada vez es más raro encontrar una fotografía real para acompañar sus noticias o artículos.

Con independencia de si surgen nuevas ocupaciones, el punto de análisis ha de ser la situación en la que coloca a los autores originales y al mercado del que forman parte. Es posible que ellos no accedan a las nuevas oportunidades, o que no deseen participar de las mismas (pasar de ilustrador o escritor a corrector o verificador de imágenes o textos artificiales no suena muy estimulante).

A diferencia de precedentes como *AUTHORS GUILD v. GOOGLE*, donde el efecto de la herramienta era más marginal y no fácil de medir (¿cuántas personas dejaban de adquirir un libro por poder leer un fragmento con *Google Books*?), la implantación de determinadas funciones como *Google AI Overviews* o el *AI Mode* de Google han supuesto caídas objetivas en el tráfico de las webs de las que obtienen su información. Google defiende en público que los «clics» de calidad estarían aumentado, pero, más allá de esta subjetiva apreciación, la realidad es que, excelentes o mediocres, las consultas a las páginas webs se reducen, habiendo sido las visitas el principal modelo de monetización hasta ahora.

La Ley no tiene la obligación de preservar por tiempo indefinido la forma en que se desarrollan los mercados, por lo que los nuevos descubrimientos pueden cambiar las reglas de juego válidamente. Ahora bien, a diferencia de otros ejemplos que suelen ponerse, el tractor no nació del trabajo de los granjeros más que remotamente (en el sentido de identificar las labores que debía cumplir la herramienta), mientras que, en todos estos casos, sin la asistencia directa de los materiales protegidos por derechos de autor no sería

posible haber desarrollado ni mantener los grandes modelos generales. De modo que, si una profesión desaparece o se ve alterada sustancialmente, no bastaría con apelar a que ya ha sucedido en el pasado, porque para que suceda, esta vez el sustituto estaría demandando al sustituido ayuda para lograr sustituirlo.

La saturación del mercado con *outputs* artificiales es algo que ha llegado a inquietar hasta a Sam Altman, por el riesgo de que termine cumpliéndose la teoría del Internet Muerto, que advierte del riesgo de que el contenido no humano supere al creado por las personas. Considerando que con estas reflexiones se tira piedras contra su propio tejado, creo que debemos darle como mínimo cierto crédito[80].

No se pueden seguir, en definitiva, los criterios usuales de la doctrina del *fair use* tecnológico, cuando el presunto uso justo podría derivar en una depauperación general de las condiciones de trabajo de aquellos campos de donde se extrae la información (con un uso intenso además de los elementos esenciales de sus trabajos, obteniendo lo mejor de ellos para proyectarlo en el funcionamiento y en los resultados de las IAs, lo que es también objeto de interés a efectos del tercer factor). Hace falta fijar un nuevo criterio que valore la especificidad del asunto.

En el plano fáctico, los magistrados lo tienen bastante fácil, porque los cuatro factores se desprenden prácticamente solos de la naturaleza de la tecnología. Lo difícil es como resolver el conflicto entre ellos. Una decisión de Sophie para una jurisprudencia estadounidense más acostumbrada a validar que a negar las apelaciones al *fair use* en el ámbito tecnológico. Los jueces deben anteponer necesariamente un bien jurídico a otro, salvo que opten por alguna fórmula intermedia, que combine la legitimidad de cierto «entrenamiento» con la obligatoriedad de licenciar otra parte de él. Esta es una opción que se intuye en la Parte 3 del informe de Copyright e Inteligencia Artificial, elaborado por la Oficina de Copyright de EE. UU.[81].

La línea reafirmativa de la propiedad intelectual que se capta en los últimos años, con el gran respaldo del Tribunal Supremo de EE. UU., llevaría a

80 La parte 3 del informe de Copyright e Inteligencia Artificial de la Oficina de Copyright de EE. UU., prepublicada en mayo de 2025, comparte esta preocupación, lo que hace del escenario algo más que una quimera abstracta, fruto de un presunto miedo al progreso.

81 Texto que merece la pena leer con atención, por ser hasta el momento el análisis institucional más completo del «entrenamiento», además de estar bien razonado (las sentencias de 2025 al estar limitadas por el petitum de las partes, a veces no llegan tan lejos como el informe, aunque su valor jurídico sea mayor). Llamativamente no habla de minería para tratar los procesos técnicos que se producen, salvo cuando menciona el Derecho comparado, prefiriendo la expresión «*machine learning*», que englobaría tanto la minería como otros procesos. En el informe también se apunta que el *machine learning* de la IA trascendería del contenido semántico del material usado durante el «aprendizaje» y abarcaría la pura expresión, lo sintáctico.

pensar que los autores críticos con las IAs generativas cuentan con buenas cartas. Pero el *too big to fail* predispone a un nuevo *AUTHORS GUILD v. GOOGLE*. Ninguna opción está cerrada, la moneda esté en el aire y veremos qué factores se imponen, si el 1 o el 4 con el refuerzo del 3. En cuestión de probabilidades, de *lege lata* (o mejor dicho de *iurisprudentia lata*), veo que hay más asideros legales para sostener la legalidad del entrenamiento que en la UE.

Desde mi punto de vista, un *fair use* incondicionado sería desproporcionado y contrario al espíritu que late en la regulación legal de la propiedad intelectual. Estimo que hay suficientes argumentos jurisprudenciales y legales en EE. UU. para sostener esta posición. No obstante, esto no excluye que la munición en la que se apoya la perspectiva contraria sea sólida igualmente. Y tiene a su favor esa tendencia mencionada del pasado, por más que las novedosas circunstancias aconsejen no seguirla a la ligera, incluso si se optara por validar el «entrenamiento». Por eso, las probabilidades de éxito para las empresas detrás de los grandes IAs generativas o razonadoras son mayores en EE. UU. que en el ámbito de la UE o, en general, en los estados «mineros». Porque en EE. UU. las balas de las que se sirven en su defensa jurídica tienen pólvora, mientras que las de la UE no; aunque se quiera fingir a veces que sí están cargadas o que son de un calibre distinto, equiparando *machine learning* con minería.

VI. Conclusiones

A pesar de la extensión de este capítulo, es tal la magnitud que reviste la cuestión que podrían dedicarse diez veces más y seguirían quedando aspectos por detallar. Como he ido comentando en las páginas previas, con los argumentos expuestos mi aspiración ha sido rebatir la homologación errónea entre seres humanos y maquinas, la ausencia de un Derecho para la IA antes de que se aprobaran normas que aludieran, o los atajos legales en cuanto a validar el «entrenamiento» de las IAs. Sin ignorar la vertiente geopolítica de la carrera por la IA, debemos ser cautos a la hora de aceptar la legalidad de cualquier práctica, porque el respeto al ordenamiento jurídico o el valor de la creación humana son también muy grandes como para dejarlas caer.

De *lege lata*, el modelo de licencia es la opción más segura jurídicamente en EE. UU., en UE y en la mayor parte del mundo. Si hay contenidos que no son posibles obtener por ese canal acordado y los Gobiernos entienden que es imprescindible que dispongan de ellos, como mínimo antes de permitir un uso gratuito libre, habría de buscarse un sistema que permita revertir en favor de la sociedad los beneficios que se logran por medio de ese aprovechamiento. No puede perderse de vista que incluso si el progreso tecnológico nos beneficia a todos, este no es lineal y admitiría alternativas de desarrollo. Si un grupo de particulares no electos lo definen y además lo hacen con datos ajenos a él, sin autorización de sus titulares, hay que equilibrar

el reparto de ventajas, más allá de que los ciudadanos tengan la opción de emplear los servicios que desarrollen (entre otros motivos, porque estos servicios ofertados no son gratuitos, se pagan con dinero o con más datos). Hay que insistir en que, aunque útiles, lo que se puede hacer y lo que no a través de estos servicios viene determinado unilateralmente, al margen del impacto no menor de la competencia del mercado. Sería incongruente una colectivización del total del conocimiento humano alumbrado en nuestra historia, desconociendo el régimen de propiedad privada en lo que no integre todavía el dominio público, para que luego su provecho vuelva a ser privado una vez tamizado y empaquetado tecnológicamente.

Bibliografía

Acemoglu, D. y Johnson, S., *Poder y progreso. Nuestra lucha milenaria por la tecnología y la prosperidad*, 1.ª edición, Deusto, Barcelona, 2023.

Andres Aucejo, E. y Ramón, F., «Inteligencia artificial: "chat GPT" versus la ley y el Derecho. Jaque al derecho de propiedad intelectual», en *Revista de Educación y Derecho,* n.º 28, 2023.

Calleja Reina, M.R., «La inteligencia artificial y su derivada en los derechos de propiedad intelectual en la cultura: retos y amenazas», en *Periférica Internacional. Revista para el análisis de la Cultura y el Territorio*, n.º 24, 2023.

Díaz-Noci, J., «Todos los caminos llevan a Ro,a: propiedad intelectual, inteligencia artficiail y medios de comunicación», en *Anuario ThinkEPI,* 2024.

Dornis, T.W. y Stober, S., *Urheberrecht und Training generativer KI-Modelle – technologische and juristiche Grundlagen*, Nomos, 2024. Accesible a través de este enlace: https://adepi.net/wp-content/uploads/2024/09/ssrn-4946214.pdf

Fernández Carballo-Calero, P., *25 things you should know about Artificial Intelligence Art and Copyright*, Aranzadi, Pamplona, 2023.

Fernández Carballo-Calero, P., *La propiedad intelectual de las obras creadas por inteligencia artificial*, Aranzadi, Pamplona, 2021.

García Vidal, A., «Propiedad intelectual, minería de textos y datos y entrenamiento de la inteligencia artificial», en Gómez-Acebo & Pombo, 2024. Accesible a través de este enlace: https://ga-p.com/publicaciones/propiedad-intelectual-mineria-de-textos-y-datos-y-entrenamiento-de-la-inteligencia-artificial/

Ibañéz Kollman, C., «Origenes del fair use», en Revista Chilena de Derecho y Tecnología, Volumen 2, n.º 2, 2013.

Kaigeng, L., Hong.W y Yupeng, D., «Copyright protectiong during the training stage of generativa AI: industry-oriented U.S. Law, rights-oriented EU Law, and fair remumeration rights for generative AI training under the UN,s International Governance Regime for AI», en *Law & Security Review*, Volumen 55, 2024.

Lacruz Mantecón, M.L., Inteligencia artificial y derecho de autor, Reus, Madrid, 2021.

Lassalle, J.M, *Civilización Artificial*, 1.ª edición, Arpa, 2024.

Martins Fernandes, P., «AI training and copyright: should intellectual property law allow machines to learn», en Bioethica, Volumen 10, n.º 2, 2024.

Moreno, A., «Los derechos de autor en la era de la Inteligencia Artificial: el caso Silverman y OpenIA», en *Derecom,* n.º 35, 2023.

Muñóz Vela, J.M., «Inteligencia Artificial Generativa. Desafíos para la propiedad intelectual», en *Revista de Derecho UNED*, n.º 33, 2024.

Osorio Umaña, F., «Inteligencia artificial y derecho de autor: un estudio sobre la regulación británica», en *Revista Justicia & Derecho*, Volumen 5, n.º 1, 2022.

Perdices Huetos, A., «El Pastiche, ¿más acá o más allá de la parodia? A propósito del artículo 70 del Real Decreto-ley 24/2021, de 2 de noviembre y su subsunción en el artículo 39 de la Ley de Propiedad Intelectual», en García-Cruces Gónzalez (Coord), *De Iure Mercatus. Libro Homenaje al Prof. Dr. Dr.h.c. Alberto Bercovitz Rodríguez-Cano,* Tirant lo Blanch, 2023.

Rodríguez Maffioli, D., «Copyright in generativa AI training: balancing fair use through standardization and transparency», en SSRN, 2023. Accesible a través de este enlace: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4579322

Rogel Vide, C., Serrano Gómez, E. Lacrúz Mantecón, M.L., *Manual de Derecho de Autor*, 2.ª edición, Reus, Madrid, 2021.

Ríos Ruiz, W.R., «Los sistemas de inteligencia artificial y la propiedad intelectual de las obras creadas, producidas o generadas mediante ordenador», en *Revista la Propiedad Inmaterial*, n.º 3, 2001.

Rosés Sáiz, G., «Apropiacionismo y derecho de autor una comparativa de la situación jurídica del arte del apropiacionismo en Estados Unidos y en Europa», en *Revista General de Legislación y Jurisprudencia*, n.º 1, 2025.

Sanjun Rodríguez, N., «Inteligencia artificial y propiedad intelectual», en *Actualidad Jurídica Uria Menéndez*, n.º 52, 2019.

5.

DESENTRAÑANDO EL REGLAMENTO DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA: UNA GUÍA ESENCIAL

Lorenzo Cotino Hueso[1]

Presidente de la Agencia Española de Protección de Datos
Catedrático de Derecho Constitucional
Universidad de Valencia. ValgrAI

SUMARIO: I. A modo de presentación. II. La primera regulación global de la inteligencia artificial: un marco pionero y sus elementos clave. 2.1. La aceleración de la regulación mundial de la IA al tiempo del Reglamento de la Unión Europea. 2.2. Una visión general del Reglamento: fundamentos y disposiciones generales. III. Qué IA se considera inaceptable y prohibida en la Unión Europea. IV. El modelo de riesgos: sólo se quiere regular la ia más peligrosa, de «alto riesgo». 4.1. El modelo de riesgos para regular la inteligencia artificial. 4.2. Identificación de sistemas de IA de alto riesgo: ¿Cuáles son y por qué? V. El «corazón» del reglamento: ¿qué obligaciones deben cumplir los sistemas de alto riesgo? 5.1. Obligaciones esenciales que deben cumplir los sistemas de IA parar ser seguros. 5.2. Responsabilidades de proveedores, implementadores y otros actores clave. 5.3. La evaluación de conformidad que se obtiene si se cumplen estándares y normas y el registro en la base de datos de sistemas de alto riesgo. VI. ¿Qué debemos saber los humanos cuando interactuamos con determinados sistemas de IA? (Artículo 50). VII. Obligaciones respecto de los modelos de ia

1 cotino@uv.es. OdiseIA. El presente estudio, remitido en 2024, es resultado de investigación de los siguientes proyectos: MICINN Proyecto «Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas» 2023-2025 (PID2022-136439OB-I00) financiado por MCIN/AEI/10.13039/501100011033/; Proyecto «Algoritmic law» (Prometeo/2021/009, 2021-24 Generalitat Valenciana); «Algorithmic Decisions and the Law: Opening the Black Box» (TED2021-131472A-I00) y «Transición digital de las Administraciones públicas e inteligencia artificial» (TED2021-132191B-I00) del Plan de Recuperación, Transformación y Resiliencia. Estancia Generalitat Valenciana CIAEST/2022/1, Convenio de Derechos Digitales-SEDIA Ámbito 5 (2023/C046/00228673) y Ámbito 6. (2023/C046/00229475).

general como «ChatGPT». VIII. Innovación bajo control: sandboxes y otras medidas para fomentar la IA. IX. Gobernanza de la IA en Europa: ¿quiénes supervisan el cumplimiento del Reglamento en España y en la Unión? X. ¿Cómo se supervisan los sistemas de IA después de su comercialización? XI. Otros aspectos del reglamento: derechos, sanciones y aplicación gradual, códigos de conducta y delegaciones a la Comisión. 11.1. ¿Reconoce el Reglamento derechos a los afectados por la IA? 11.2. Códigos de conducta: Estándares voluntarios para sistemas no regulados. 11.3. Régimen sancionador: ¿cuáles son las consecuencias del incumplimiento? 11.4. Delegación de poderes a la Comisión para ir adaptando el Reglamento. 11.5. Disposiciones finales y la (lenta) aplicación del Reglamento. XII. Para acabar, también en 2024 un Convenio de Inteligencia Artificial del Consejo de Europa.

I. A modo de presentación

El Reglamento de Inteligencia Artificial de la Unión Europea (RIA)[2] es un documento pionero, extraordinariamente complejo y detallado, diseñado para establecer un marco legal general de la inteligencia artificial en la Unión Europea, con pretensiones de influencia mundial. Para dar una idea de su extensión, es comparable a cada tomo del El Señor de los Anillos, que cuales ronda las 100.000 palabras, como el RIA.

Ni más ni menos que se pretende en estas páginas hacer una síntesis del mismo que permita dar al menos una respuesta esencial a cuestiones como las que siguen: ¿Se está regulando la IA en el mundo? ¿Qué caracteriza al Reglamento de Inteligencia Artificial de la Unión Europea como pionero a nivel mundial? ¿Cuáles son los sistemas de IA que se consideran inaceptables y prohibidos bajo el RIA? ¿Qué obligaciones deben cumplir? ¿Qué obligaciones añade el RIA para que sepamos que interactuamos con algunos sistemas de IA? ¿Cómo se regulan modelos de IA de propósito general, como ChatGPT? ¿Qué papel juegan los sandboxes regulatorios en el fomento de la innovación en IA bajo el RIA? ¿Quiénes supervisarán y controlarán que se cumple el RIA en España y en la Unión Europea? ¿Cómo se controlará que un sistema siga siendo seguro después de su comercialización? ¿Hay sanciones por incumplimiento? ¿Qué papel jugará el Convenio de inteligencia artificial del Consejo de Europa junto al Reglamento?

En la actualidad, la inteligencia artificial (en adelante, IA) se ha consolidado como una de las tecnologías más disruptivas e influyentes de nuestro tiempo,

2 Su nombre finalmente es Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.° 300/2008, (UE) n.° 167/2013, (UE) n.° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), DOUE núm. 1689, de 12 de julio de 2024, DOUE-L-2024-81079. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

por su ingente capacidad de rápida transformación de sectores productivos, desde la medicina hasta la seguridad, y, también, entre muchas otras revoluciones, por su posible integración en objetos, productos o servicios digitales. Al menos en la Unión Europea (en adelante, UE) se percibe desde hace años la necesidad de un marco regulatorio general robusto y coherente que acompañe y encauce el desarrollo de esta tecnología. La Unión Europea, para bien o para mal, ha sido la pionera al establecer esta regulación general tanto para sí misma como para intentar influir en el resto del mundo con lo que se ha llamado el *efecto Bruselas,* algo que en cierta medida consiguió con la aprobación del Reglamento General de Protección de Datos (en adelante, RGPD).

Los pasos para abordar la IA en la UE se iniciaron hace más de siete años con intervención de Comisión, Parlamento y Consejo. Se busca una IA «made in Europe» caracterizada por estar diseñada éticamente para el respeto de los derechos y los principios democráticos[3]. El RIA ha seguido un largo y costoso procedimiento legislativo[4], del que cabe ahora destacar especialmente la propuesta de la Comisión Europea de 2021[5] y la posición del Consejo de la Unión de diciembre de 2022[6]. También son muy relevantes las enmiendas del Parlamento en junio de 2023[7]. El acuerdo político fue de 8 de diciembre de 2023 y desde entonces se ha ido concretando el texto y sus traducciones en las diversas fases de su adopción final hasta su publicación en el DOUE.

La propuesta de la Comisión Europea pretendía, de un lado, garantizar que los sistemas de IA utilizados en la UE e introducidos en el mercado europeo sean seguros y respeten los derechos de los ciudadanos; del otro, estimular la inversión y la innovación en el ámbito de la IA en Europa.

El texto definitivo del Reglamento de IA de la UE pretende armonizar las normas sobre inteligencia artificial, y lo hace con un enfoque basado en el riesgo, estableciendo un marco de obligaciones y requisitos distintos en función del nivel de riesgo de la tecnología de IA aplicable y su uso concreto. La técnica del enfoque armonizador exige, por ejemplo, unas obligaciones reforzadas para los sistemas de inteligencia artificial calificados como de

---

3 Un análisis exhaustivo de los pasos y políticas de la UE en la materia hasta 2019, Cotino Hueso, L., «Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho» en *Revista Catalana de Derecho Público,* n.º 58, 2019. http://revistes.eapc.gencat.cat/index.php/rcdp/issue/view/n58 http://dx.doi.org/10.2436/rcdp.i58.2019.3303

4 Puede seguirse el mismo en los siguientes enlaces de referencia. Parlamento Europeo: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en Y *Legislative Train Schedule*, también del Parlamento: https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence

5 https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A52021PC0206

6 https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf

7 https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html

alto riesgo; lo que deberá concretarse con la elaboración de estándares técnicos y el establecimiento de controles ad hoc para su aplicación.

Aunque el RIA entró en vigor en 2024, ciertamente cuenta con una aplicación y obligatoriedad muy escalonada, que llega incluso a los seis años. En cualquier caso, es posible prever un enorme impacto en el mercado y la sociedad. Pronto seremos testigos de las prohibiciones de tecnologías y usos concretos, de la implantación y actividad de las oficinas y autoridades europeas y nacionales de supervisión de la IA. Ya en febrero de 2024 se creó la nueva Oficina de la IA. Veremos la incoación y resolución de procedimientos sancionadores, de la aprobación de estándares técnicos y de la actividad de los organismos de certificación, de la proliferación de sistemas de gestión de riegos específicos, de la aparición de códigos de conducta voluntarios sectoriales, entre muchas otras cuestiones que serán reflejo de la norma que es objeto de estudio en el presente trabajo.

En modo alguno resulta sencillo hacer una síntesis de 106 mil palabras de una norma extraordinariamente compleja[8]. Valga este intento, eso sí, reenviando al lector interesado a los más de 35 estudios que analizan el RIA de modo sistemático en el *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, Aranzadi, 2024 que con Simó he coordinado y está publicado en abierto en español[9] y en inglés.

II. La primera regulación global de la inteligencia artificial: un marco pionero y sus elementos clave

2.1. La aceleración de la regulación mundial de la IA al tiempo del Reglamento de la Unión Europea

El RIA en su tramitación y aprobación final ha estimulado el interés global en la regulación de la IA[10]. Por ejemplo, el 16 de junio de 2022, Canadá presentó la *Digital Charter Implementation Act* de 2022[11] como parte de la *Arti-*

8 Una visión general del RIA, Presno Linera, M. Á., «La propuesta de "Ley de Inteligencia Artificial" europea», en *Revista de las Cortes Generales*, n.º 116, 2023, págs. 81-133. De este autor, de interés, Derechos fundamentales e inteligencia artificial, Marcial Pons, Madrid, 2023.

9 https://links.uv.es/uf7Pqmh

10 Una visión general de las propuestas de regulación mundial, muy reciente, Unesco, *Consultation paper on AI regulation. Emerging Approaches Across the World*, 16 de agosto de 2024, Unesco. https://unesdoc.unesco.org/ark:/48223/pf0000390979 De 2023, IALAB, Propuestas de regulación y recomendaciones de inteligencia artificial en el mundo, IALAB-Thomson Reuters, Buenos Aires, https://ialab.com.ar/wp-content/uploads/2023/08/Propuestas-de-regulacion-y-recomendaciones-de-IA-en-el-mundo-1.pdf

11 https://www.parl.ca/legisinfo/en/bill/44-1/c-27

ficial Intelligence and Data Act (AIDA)[12]. En diciembre de 2022, Brasil[13] inició la tramitación del proyecto de ley n.º 2338 en el Senado para regular la IA, mientras que el Reino Unido adoptó un enfoque más flexible en la regulación tras el *Brexit*[14].

El surgimiento de tecnologías como ChatGPT a finales de 2022 impulsó aún más estos desarrollos. Por ejemplo, China implementó las *«Medidas provisionales para la gestión de servicios de IA generativa»* el 13 de julio de 2023[15]. En septiembre de 2023, Canadá dio a conocer su *«Código de Conducta Voluntario para el Desarrollo y Gestión Responsable de Sistemas Generativos Avanzados de IA»*. Desde el G7, el *«Proceso de Hiroshima»* resultó en el *«Código Internacional de Conducta para el Desarrollo de Sistemas Avanzados de IA»* el 30 de octubre de 2023[16]. Ese mismo día, Estados Unidos adoptó la significativa *Orden ejecutiva sobre el desarrollo y la utilización seguros y fiables de la inteligencia artificial*, también el 30 de octubre de 2023[17]. De manera más específica, California aprobó el *Artificial Intelligence Accountability Act*[18] el 3 de enero de 2024, mientras que Illinois cuenta con su *Artificial Intelligence Video Interview Act (820 ILCS 42/)*[19]. El 17 de mayo de 2024 el gobernador de Colorado firmó la ley de protección del consumidor para las interacciones con la inteligencia artificial, que entrará en vigor del 1 de febrero de 2026, se trata de la primera legislación integral en los Esta-

12 El texto en https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading El estado de tramitación en https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act y https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

13 *Projeto de lei n.º 2338 do Senado, que regulamenta a IA*, texto inicial en https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1702407086098&disposition=inline&_gl=1*1ifop8*_ga*MTg0Njk3ODg3MS4xNzA1Mzk4MDU2*_ga_CW3ZH25XMK*MTcwNTM5ODA1Ni4xLjAuMTcwNTM5ODA1Ni4wLjAuMA Procedimiento en https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

14 Así, destaca Gobierno de Reino Unido: *A pro-innovation approach to AI regulation*, Department for Science, Innovation and Technology and Office for Artificial Intelligence, 29 de marzo 2023, actualización 3 de agosto. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

15 http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

16 https://www.mofa.go.jp/files/100573473.pdf

17 *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ Cabe seguir también la nota informativa https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ e

18 *California Senate Bill 896*, https://legiscan.com/CA/text/SB896/id/2868456

19 https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68#:~:text=An%20employer%20may%20not%20use,use%20of%20artificial%20intelligence%20analysis

dos Unidos dirigida a los «sistemas de inteligencia artificial de alto riesgo»[20]. En Iberoamérica, existen diversas iniciativas legislativas en curso[21]. Como se señala al final de este estudio, 2024 ha traído no sólo el RIA final, sino un Convenio del Consejo de Europa.

2.2. Una visión general del Reglamento: fundamentos y disposiciones generales

El RIA tiene como objetivo regular el uso de la inteligencia artificial dentro de la Unión Europea, con la finalidad de mitigar los riesgos inherentes a su aplicación. El ámbito de aplicación del RIA sigue la estela del Reglamento de protección de datos (RGPD) buscando no sólo el efecto Bruselas de influencia, sino la extraterritorialidad en su aplicación. Así, conforme al artículo 2, abarca tanto a los proveedores de sistemas de IA que se comercialicen o pongan en servicio en la UE, independientemente de su origen, como a aquellos cuya salida o resultados se utilicen en el territorio de la Unión. Asimismo, incluye a los implementadores, entendidos como los operadores de estos sistemas, excluyendo a los usuarios finales o afectados. El RIA también es aplicable a los proveedores de estos sistemas, a sus representantes autorizados, importadores y distribuidores, asegurando así un marco normativo amplio y comprensivo para la regulación de la inteligencia artificial en el contexto europeo.

El RIA excluye de su aplicación a las autoridades públicas de terceros países y a las organizaciones internacionales que empleen sistemas de IA en el marco de la cooperación policial o judicial con la UE o sus Estados miembros. De igual modo, no se aplica a sistemas destinados exclusivamente a fines militares, seguridad nacional o actividades de investigación y desarrollo científico.

El artículo 3 del RIA define un sistema de inteligencia artificial como aquel que, operando con distintos niveles de autonomía, puede adaptarse después de su despliegue. Este tipo de sistemas, con objetivos explícitos o implícitos, procesan la entrada de datos para generar resultados como predicciones, contenido, recomendaciones o decisiones que influyen en entornos físicos o virtuales. El RIA establece una jerarquía de riesgos basada en el uso de la IA y asigna obligaciones proporcionales a los riesgos identificados.

Gran parte del articulado del RIA se centra en la regulación de los sistemas de alto riesgo, que se clasifican en dos grandes categorías: finalidades de

20 *SB24-205, Consumer Protections for Artificial Intelligence. Concerning consumer protections in interactions with artificial intelligence systems*. https://leg.colorado.gov/bills/sb24-205

21 Existen propuestas normativas de legisladores particulares y otras vinculadas a estrategias nacionales en la materia como en Perú, Chile, Colombia, por mencionar algunos.

alto riesgo y productos de alto riesgo. En la primera categoría, el RIA enumera y describe un conjunto de sistemas cuya finalidad o uso se considera de alto riesgo, como los sistemas de identificación biométrica, protección de infraestructuras críticas, selección y promoción de personal, control fronterizo, o los empleados por las Fuerzas y Cuerpos de Seguridad del Estado o la Administración de Justicia. La Comisión Europea puede actualizar esta lista mediante actos delegados. Estos sistemas no serán considerados de alto riesgo en casos específicos, tales como cuando el sistema realice una tarea procedimental limitada, mejore el resultado de una actividad humana previamente completada, o detecte patrones de toma de decisiones sin reemplazar o influir en la evaluación humana sin una revisión adecuada.

Por otro lado, existen productos ya regulados por normativas armonizadas de la UE, sujetos a evaluaciones de conformidad. Entre ellos se encuentran dispositivos médicos, ascensores, juguetes o maquinaria. Un sistema de IA que forme parte de estos productos, o que constituya un componente de seguridad de ellos, estará sujeto a la normativa armonizada correspondiente y será considerado de alto riesgo bajo el RIA.

El RIA impone a los sistemas de alto riesgo una serie de requisitos obligatorios que deben demostrarse mediante pruebas de conformidad. Estos requisitos incluyen la gestión de riesgos, la gestión de datos, la documentación técnica, la transparencia hacia el implementador, la precisión, la robustez ante fallos, la ciberseguridad, la trazabilidad y la supervisión humana.

Además de los sistemas de alto riesgo, el RIA también regula los sistemas de inteligencia artificial de propósito general, que pueden ser utilizados para una variedad de finalidades, tanto directamente como integrados en otros sistemas de IA. El RIA establece obligaciones específicas para los proveedores de estos sistemas, siendo más estrictas cuando dichos sistemas se consideran de riesgo sistémico.

El RIA define diversas entidades de supervisión. Entre ellas, los organismos de evaluación de conformidad, encargados de realizar las pruebas de conformidad, deben estar habilitados por una autoridad notificante. Esta última notificará a la Comisión Europea sobre los organismos habilitados, documentando sus procedimientos. Las autoridades de supervisión de mercado serán responsables de monitorizar el correcto funcionamiento de los sistemas de IA de alto riesgo ya en el mercado, identificando riesgos, incidentes u otras situaciones que requieran medidas correctivas. Los productos de IA considerados de alto riesgo estarán supervisados por la autoridad designada en la normativa correspondiente.

En el caso de sistemas de identificación biométrica utilizados por las Fuerzas y Cuerpos de Seguridad del Estado, en el ámbito migratorio o en la Administración de Justicia, la supervisión recaerá sobre las autoridades de protección de datos competentes, como la Agencia Española de Protección de Datos en algunos casos.

Finalmente, para favorecer la gobernanza del RIA, se establecen varios órganos, incluyendo la Oficina Europea de la IA y el Consejo Europeo de Inteligencia Artificial, además de dos órganos consultivos: el grupo científico de expertos y el foro consultivo. Por último, el RIA consta de 113 artículos repartidos en trece capítulos, más trece anexos. En el Capítulo I («Disposiciones generales»), se establecen las bases del RIA, incluyendo el objeto, ámbito de aplicación, definiciones clave, y alfabetización en IA. En el Capítulo II («Prácticas prohibidas de IA»), se especifican las prácticas de inteligencia artificial que están completamente prohibidas dentro de la UE, asegurando la prohibición de usos peligrosos o éticamente inaceptables de la IA.

El Capítulo III («Sistemas de IA de alto riesgo») es uno de los más extensos, dividido en cinco secciones. Este capítulo abarca desde la clasificación de los sistemas de IA como de alto riesgo, los requisitos y obligaciones de los proveedores e implementadores, hasta las normas de evaluación de conformidad y la cooperación con las autoridades competentes. En el Capítulo IV («Obligaciones de transparencia para los proveedores e implementadores de determinados sistemas de IA»), se establecen las responsabilidades de transparencia para asegurar que la información relevante sobre los sistemas de IA esté disponible para los usuarios.

El Capítulo V («Modelos de IA de uso general») regula los modelos de IA de propósito general, abarcando su clasificación, obligaciones de los proveedores, y los códigos de buenas prácticas. El Capítulo VI («Medidas de apoyo a la innovación») fomenta la innovación en IA mediante la creación de sandboxes regulatorios y otras medidas que facilitan las pruebas en condiciones reales, especialmente para pymes y startups. El Capítulo VII («Gobernanza») se divide en dos secciones: la primera aborda la gobernanza a nivel de la Unión, incluyendo la creación de la Oficina de IA y del Comité Europeo de Inteligencia Artificial, mientras que la segunda trata sobre las autoridades nacionales competentes.

En el Capítulo VIII («Base de datos de la UE para sistemas de IA de alto riesgo»), se crea una base de datos centralizada para los sistemas de IA de alto riesgo, garantizando transparencia y trazabilidad. El Capítulo IX («Seguimiento poscomercialización, intercambio de información y vigilancia del mercado») detalla las obligaciones de seguimiento de los sistemas de IA tras su comercialización, el intercambio de información sobre incidentes graves, y la vigilancia del mercado.

En el Capítulo X («Códigos de conducta y directrices»), se fomenta la creación de códigos de conducta para la aplicación voluntaria de requisitos específicos, y se establecen directrices por parte de la Comisión. El Capítulo XI («Delegación de poderes y procedimiento de comité») regula la delegación de poderes a la Comisión y el procedimiento de comités para la toma de decisiones relacionadas con la IA.

El Capítulo XII («Sanciones») detalla las sanciones aplicables por incumplimiento, incluyendo multas administrativas para instituciones, órganos,

organismos de la Unión y proveedores de modelos de IA de uso general. Finalmente, en el Capítulo XIII («Disposiciones finales»), se incluyen modificaciones a otros reglamentos y directivas de la UE, además de disposiciones sobre la evaluación y revisión del reglamento.

Los trece anexos del RIA detallan la legislación de armonización aplicable, listan las infracciones penales relevantes y describen los requisitos técnicos y procedimientos necesarios para la conformidad de los sistemas de IA, incluyendo los de alto riesgo. Además, proporcionan pautas para la transparencia, registros y documentación técnica, y especifican los criterios para la designación de modelos de IA con riesgo sistémico[22].

Son más de cincuenta las definiciones que incluye el RIA, por destacar alguna, alineándose con la definición de la OCDE[23], el RIA define al sistema

22 En el Anexo I se presenta una lista de la legislación de armonización de la Unión, estableciendo el marco legal que interacciona con las disposiciones del Reglamento de IA. El Anexo II contiene una lista detallada de las infracciones penales a las que se refiere el artículo 5, apartado 1, párrafo primero, letra h), inciso iii), especificando las conductas delictivas que están prohibidas cuando se emplea inteligencia artificial. El Anexo III identifica los sistemas de IA de alto riesgo según lo establecido en el artículo 6, apartado 2, detallando aquellos sistemas que requieren un control y vigilancia más estrictos debido a los riesgos que implican. En el Anexo IV se describen los requisitos de la documentación técnica que deben cumplir los sistemas de IA de alto riesgo, de acuerdo con el artículo 11, apartado 1. Esta documentación es esencial para demostrar el cumplimiento de las normas de seguridad y calidad. El Anexo V proporciona el formato y contenido de la Declaración UE de conformidad, necesaria para que los sistemas de IA puedan ser comercializados dentro del mercado único europeo. El Anexo VI detalla el procedimiento de evaluación de la conformidad basado en el control interno, un proceso que los proveedores deben seguir para asegurar que sus sistemas cumplen con los requisitos establecidos. El Anexo VIII incluye información relacionada con la evaluación del sistema de gestión de la calidad y la documentación técnica. Este anexo se divide en varias secciones, cada una especificando la información que los proveedores y desplegadores de sistemas de IA de alto riesgo deben presentar al momento del registro, en conformidad con el artículo 49. El Anexo IX describe la información que debe presentarse en el registro de los sistemas de IA de alto riesgo enumerados en el anexo III, especialmente en lo que respecta a los ensayos en condiciones reales, de conformidad con el artículo 60. El Anexo X enumera los actos legislativos de la Unión relativos a los sistemas informáticos de gran magnitud en el espacio de libertad, seguridad y justicia, proporcionando el contexto normativo más amplio en el que se sitúan las regulaciones de IA. El Anexo XI se refiere a la documentación técnica exigida para los proveedores de modelos de IA de uso general, dividiéndose en dos secciones que especifican tanto la información básica que todos los proveedores deben proporcionar, como la información adicional requerida para aquellos modelos con riesgo sistémico. El Anexo XII detalla los requisitos de transparencia que deben cumplir los proveedores de modelos de IA de uso general, especialmente en lo que respecta a la documentación técnica que deben facilitar a los proveedores posteriores que integren el modelo en su sistema de IA. Finalmente, el Anexo XIII establece los criterios para la designación de modelos de IA de uso general con riesgo sistémico, como se describe en el artículo 51, proporcionando un marco para la clasificación y gestión de estos sistemas de mayor riesgo.

23 https://www.oecd-ilibrary.org/science-and-technology/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en

de inteligencia artificial como aquel que opera con diferentes niveles de autonomía, capaz de adaptarse tras su despliegue y de inferir, a partir de los datos que recibe, cómo generar resultados tales como predicciones, contenido, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales. La figura del «Proveedor» se refiere a la persona física o jurídica, o entidad pública, que desarrolla o para quien se desarrolla un sistema de IA, y lo pone en servicio o lo comercializa bajo su nombre o marca, ya sea con fines lucrativos o no. Por otro lado, se define al «Implementador» como la persona física o jurídica, pública o privada, bajo cuya autoridad se utiliza el sistema. Cabe destacar que el RIA excluye a los implementadores que sean personas físicas utilizando sistemas de IA en un contexto puramente personal y no profesional.

III. Qué IA se considera inaceptable y prohibida en la Unión Europea

El artículo 5 (Capítulo II) es uno de los más importantes. Ahí, el RIA prohíbe una serie de sistemas de inteligencia artificial que se consideran particularmente peligrosos debido a su potencial para causar daños físicos o psicológicos, o para afectar negativamente los derechos fundamentales de las personas. Entre estos se incluyen aquellos sistemas que utilizan técnicas subliminales con la intención de alterar el comportamiento de una persona, de manera que puedan provocarle daños a ella o a terceros.

Asimismo, se prohíben los sistemas que se aprovechan de las vulnerabilidades específicas de ciertos grupos, como aquellos definidos por su edad, discapacidad o situación social o económica, y que de esta forma podrían distorsionar su comportamiento, generando posibles daños. Los sistemas que categoricen a personas a partir de datos biométricos para inferir información sensible también están prohibidos, ya que estos procesos pueden conducir a la toma de decisiones discriminatorias.

Además, se consideran inaceptables los sistemas de inteligencia artificial diseñados para crear perfiles de personas basados en su comportamiento, si esto conduce a la elaboración de un «baremo social» que podría resultar en un trato desproporcionadamente desfavorable para ciertas personas o grupos, especialmente en contextos distintos a aquellos en los que se recopilaron los datos. Otro tipo de sistemas prohibidos son aquellos que pretenden evaluar o predecir la probabilidad de que una persona cometa un delito, basándose únicamente en perfiles o características personales. Esta prohibición admite una excepción: cuando estos sistemas se utilicen para apoyar la evaluación humana en la determinación de la implicación de una persona en actividades delictivas.

También están prohibidos los sistemas que buscan crear o expandir bases de datos de reconocimiento facial mediante la recopilación indiscriminada de imágenes faciales de internet o de grabaciones de sistemas de videovigilancia. Por último, se prohíbe la utilización de sistemas que infieran emocio-

nes de personas físicas en el entorno laboral o en instituciones educativas, salvo que dichos sistemas se empleen con fines médicos o de seguridad.

El uso de la identificación biométrica en tiempo real en lugares públicos por parte de fuerzas de seguridad o autoridades públicas ha sido desde el inicio, uno de los caballos de batalla, más discutidos política y socialmente que ha supuesto una regulación final muy compleja que, además, debe superponerse con el RGPD[24]. El artículo 5 establece una prohibición general, con algunas excepciones para casos específicos como la búsqueda de víctimas de delitos, la prevención de amenazas graves, la prevención de ataques terroristas, y la persecución de crímenes como el terrorismo o la trata de seres humanos. Para la autorización de estos sistemas se requerirá una evaluación previa de la escala y probabilidad del daño potencial, además de la obtención de una autorización judicial o administrativa, y se deberán imponer restricciones temporales, geográficas y personales. Fuera de estas excepciones, el uso de identificación biométrica en tiempo real sigue estando sujeto al Reglamento General de Protección de Datos (RGPD) 2016/679 y a la Directiva 2016/680 sobre el tratamiento de datos personales por parte de autoridades.

Debe subrayarse la grave dificultad que entraña la interpretación concreta de algunos de los sistemas prohibidos, y sobre todo hay que tener en cuenta que este artículo cinco de prohibiciones y sistemas inaceptables debe leerse conjuntamente con la descripción de los sistemas de alto riesgo, especialmente establecida en el anexo III. Y es que debe señalarse que muchos de los sistemas que no están prohibidos por excepcionarse en este artículo 5 pasan a ser de alto riesgo. Asimismo, en ocasiones no es fácil distinguir cuando el sistema es prohibido o es de alto riesgo, como puede suceder con los relacionados con el cumplimiento de la ley.

IV. El modelo de riesgos: solo se quiere regular la IA más peligrosa, de «alto riesgo»

4.1. El modelo de riesgos para regular la inteligencia artificial

Ante la IA, es fundamental adoptar técnicas de responsabilidad «proactiva», «demostrada» o «accountability» desde el diseño, siguiendo el princi-

24 Sobre el tema, mis estudios, «Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos», en Balaguer Callejón, F. y Cotino Hueso, L., en *Derecho público de la inteligencia artificial*, F. Jiménez Abad-Marcial Pons, Zaragoza, 2023. Acceso, más breve, «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», en *El Cronista del Estado Social*, IUSTEL, *monográfico Inteligencia artificial*, n.º 100, septiembre-octubre 2022, pp. 68-79. Acceso.

pio jurídico de «más vale prevenir que curar»[25]. Inspirándose en las leyes de la robótica de Asimov, esto implica la integración del cumplimiento ético y legal directamente en el «Código», conforme a los términos expuestos por Lessig. Este enfoque, conocido como *compliance* o cumplimiento normativo de origen anglosajón, comienza a ser reconocido en el Derecho continental europeo, aplicándose en áreas como el medio ambiente o los estudios de impacto de género en disposiciones generales[26], y extendiéndose para incluir la discriminación algorítmica pública. Este modelo se generaliza en la UE en diversos ámbitos digitales[27] y en la legislación europea sobre comercialización de productos potencialmente peligrosos.

El modelo de responsabilidad proactiva se ha proyectado claramente en la UE como el «Ethics & Rule of law by design X-by design», insignia de la «IA confiable» «Made in Europe»[28], que fundamenta las políticas y normativas de IA en la UE[29]. Además, el alto grupo de expertos de la UE adoptó un modelo de ética en el diseño con su exhaustiva lista de evaluación[30], su modelo ALTAI de evaluación o FUTURE-AI para salud[31]. El Libro Blanco de la IA de 2020 clarificó el modelo de riesgos, que se concretaría un año después en el RIA. Este enfoque supone un aumento de obligaciones y garantías en proporción al impacto o riesgo del sistema de IA. Es notable destacar las contribuciones

25 Sobre la aplicación del principio de responsabilidad proactiva y el diseño para el ámbito de la IA y el big data, Hernández Peña, J. C., *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*, Aranzadi, Cízur Menor, 2022, págs. 119-172; Martínez Martínez, R.: «Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo», Revista catalana de dret públic, n.º 58, 2019, págs. 64-81. También, mi estudio, «Nuevo paradigma en las garantías de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial», en Cotino Hueso, L. (editor), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizadas*, Thompson-Reuters Aranzadi, FIADI, Cizur Menor, 2022, págs. 69-105.

26 Impuestos a nivel estatal por artículos 22.2 para leyes y 24.1.b para reglamentos, Ley 50/1997, de 27 de noviembre, del Gobierno; artículo 19 Ley Orgánica 3/2007 para disposiciones de carácter general y los planes de especial relevancia económica y artículos 26. 3 f) Ley 40/2015. Existe abundante normativa autonómica.

27 Barrio Andrés, M., «El cumplimiento basado en el riesgo o risk-based compliance, pieza cardinal del nuevo Derecho digital europeo», en Análisis del Real Instituto Elcano (ARI), n.º 34, 2023.https://media.realinstitutoelcano.org/wp-content/uploads/2023/04/el-cumplimiento-basado-en-el-riesgo-o-risk-based-compliance-pieza-cardinal-del-nuevo-derecho-digital-europeo-real-instituto-elcano.pdf

28 Comisión Europea, *Directrices éticas para una IA fiable*, 2019, https://data.europa.eu/doi/10.2759/14078 versión 2018, págs. 19, 21, 29.

29 Un análisis exhaustivo en mi estudio «Ética en el diseño... cit.

30 Comisión Europea, *Directrices éticas... cit.*, en especial Capítulo III y listado, págs. 33-41.

31 Portal ALTAI https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal

de Mantelero[32] y Simón en España[33] en el ámbito de la protección de datos y posteriormente para los sistemas de IA.

El Capítulo III inicialmente regula la «Clasificación de sistemas IA de alto riesgo». Es de interés contextualizar para una mejor comprensión, que el RIA ha encajado la regulación de la IA en el ámbito de la seguridad y garantía de los productos, las normas de armonización y el modelo del llamado «nuevo marco legislativo». Se trata del marco por el que se establecen unas bases comunes para la comercialización, evaluación y vigilancia de productos en la Unión Europea[34]. Todo sea dicho, es un modelo con el que los juristas en general no estamos muy familiarizados. Ello influye también en la definición de lo que es un sistema de alto riesgo.

4.2. Identificación de sistemas de IA de alto riesgo: ¿Cuáles son y por qué?

Así, en primer término, estos sistemas incluyen aquellos que se implementan como productos o componentes de seguridad ya cubiertos por la legislación armonizada de la Unión Europea, lo que implica que están sujetos a evaluaciones de conformidad realizadas por terceros, como es el caso de dispositivos médicos, ascensores, aviones o maquinaria. Las normas armonizadas aplicables a estos sistemas están enumeradas en el Anexo II del RIA. Sin poder entrar ahora en muchos detalles que serían necesarios, hay que señalar que la clave está en delimitar qué productos peligrosos de la Unión Europea se considera que lo son tanto como para que esta evaluación de conformidad debe ser realizada por un tercero (un llamado «organismo, notificado»), pues por lo general, se suelen establecer auto evaluaciones por los propios proveedores.

---

32 Respecto del ámbito de los datos Mantelero, A., «Toward a New Approach to Data Protection in the Big Data Era», en Gasser U. y otros (dir.): *Internet Monitor 2014: Reflections on the Digital World.* Cambridge (MA): Berkman Center for Internet and Society at Harvard University, 2014, págs. 84 y ss. Para la IA, Mantelero A.: *Beyond Data. Human Rights, Ethical and Social Impact Assessment*, Information Technology and Law Series IT&LAW 36, Springer, Hague, 2022, https://link.springer.com/book/10.1007/978-94-6265-531-7

33 Simón Castellano, P.: *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, Cízur Menor, 2023. También, «Taxonomía de las garantías jurídicas en el empleo de los sistemas de inteligencia artificial», en *Revista de Derecho Político*, n.º 117, 2023, págs. 153-196.

34 Las tres textos legales que conforman el Nuevo Marco Legislativo son: el Reglamento (CE) n.º 765/2008 del Parlamento Europeo y del Consejo por el que se establecen los requisitos de acreditación y vigilancia del mercado de los productos; la Decisión n.º 768/2008/CE del Parlamento Europeo y del Consejo sobre un marco común para la comercialización de los productos y; el Reglamento (UE) 2019/1020 del Parlamento Europeo y del Consejo relativo a la vigilancia del mercado y la conformidad de los productos.

En segundo lugar, son de alto riesgo los sistemas IA utilizados para «influir sustancialmente en el resultado de la toma de decisiones» (art. 6.3 RIA) para hasta 25 finalidades enumeradas en el Anexo III. Así, el Anexo III del RIA especifica una serie de sistemas que se clasifican en diversas categorías de alto riesgo, tales como: sistemas de identificación biométrica que operan sin la participación activa del sujeto, infraestructuras críticas como el tráfico, la electricidad o el agua, así como la gestión del acceso a la educación o la planificación del desarrollo académico en el ámbito educativo y de formación profesional.

Asimismo, se incluyen sistemas relacionados con la selección de personal y la gestión de relaciones laborales, la gestión del acceso a servicios esenciales tanto públicos como privados, y actividades vinculadas a las fuerzas y cuerpos de seguridad, como la valoración de pruebas o sospechosos. También se consideran de alto riesgo los sistemas empleados en migración, asilo y control de fronteras, como polígrafos o la valoración de solicitudes, y aquellos utilizados en la administración de justicia y procesos democráticos.

Es importante destacar que la Comisión Europea posee una capacidad para añadir o retirar sistemas de estas categorías, adaptándose a las circunstancias cambiantes, así como para ir elaborando criterios y determinaciones que permitan concretar cuando un sistema es de alto riesgo.

No obstante, hay situaciones en las que estos sistemas de IA no se considerarán de alto riesgo. Por ejemplo, si el sistema realiza una tarea procedimental limitada, mejora el resultado de una actividad humana previamente completada o detecta patrones de toma de decisiones o desviaciones sin pretender reemplazar o influir en la evaluación humana sin una revisión adecuada. En tales casos, el proveedor debe documentar el proceso que llevó a considerar que su sistema no es de alto riesgo y debe proceder a su registro en una base de datos específica.

V. El «corazón» del reglamento: ¿qué obligaciones deben cumplir los sistemas de alto riesgo?

5.1. Obligaciones esenciales que deben cumplir los sistemas de IA parar ser seguros

El RIA implica esencialmente que los sistemas de IA de alto riesgo (artículo 6, anexo I y III) deben cumplir ciertos requisitos. «La clasificación de un sistema de IA como "de alto riesgo" debe limitarse a aquellos sistemas de IA que tengan un efecto perjudicial importante en la salud, la seguridad y los derechos fundamentales de las personas de la Unión» (Cons. 46). Como se señaló al inicio, el Considerando 48 menciona veintiún derechos fundamentales afectados por estos sistemas, además de derechos específicos de menores.

En caso de sistemas de alto riesgo, se requiere cumplir con un conjunto de obligaciones preventivas que, de verificarse, permiten una evaluación de conformidad. Ello permite que el sistema IA de alto riesgo se incluya en un registro y pueda ponerse en el mercado bajo el marcado «CE» de la UE. Así, en la Sección dos se regulan los requisitos para los sistemas de IA de alto riesgo. Los sistemas de inteligencia artificial clasificados como de alto riesgo deben cumplir con una serie de exigencias específicas que garantizan su seguridad y la protección de los derechos fundamentales. En primer lugar, se requiere la implementación de un sistema de gestión de riesgos que tenga en cuenta, especialmente, los riesgos que estos sistemas puedan presentar para la salud, la seguridad y los derechos fundamentales en función de su propósito.

Además, es imprescindible establecer una gobernanza adecuada y una gestión rigurosa de los datos utilizados para el entrenamiento y prueba de los sistemas. use datos de alta calidad (art. 10 y 16 RIA). Se parte de que los datos deben «estar libres de errores y ser completos en vista de la finalidad prevista del sistema» (Considerando 67). Si los datos de los que se alimenta la IA son basura, sus resultados necesariamente serán basura. El proveedor de un sistema de alto riesgo ha de generar documentación técnica que demuestra el cumplimiento de los requisitos y los elementos técnicos del sistema, para ponerla a disposición de las autoridades (art. 11). La documentación debe incluir un contenido mínimo especificado, que puede ser modificado por la Comisión cuando se considere necesario.

También se ha de diseñar la IA de alto riesgo para que genere registros y logs de funcionamiento (art. 12). Se trata de los sistemas de IA más peligrosos estén localizados y que se puedan vigilar, así como trazar exactamente cómo han funcionado, así como verificar que cumplen las obligaciones preventivas. Asimismo, al menos para los sistemas peligrosos, se hace preciso diseñar el sistema algorítmico de manera que pueda ser auditable, esto es, creando registros entendibles para que se pueda investigar si funcionó bien en todo momento. Y obviamente el sistema de IA ha de estar bien diseñado y comprobarse que funciona bien en todo momento. También hay obligaciones de «transparencia y comunicación de información a los responsables del despliegue» (art. 13), esto es, que los sistemas tengan buenos «manuales de instrucciones» y las empresas o administraciones —los implementadores— que los utilicen los conozcan y puedan también dar suficiente información a la ciudadanía, que son los afectados últimos por la IA. Así, se debe proporcionar información detallada a los implementadores sobre las capacidades del sistema, los requisitos de equipamiento, el ámbito de aplicación, el nivel de precisión, y las condiciones de uso que puedan implicar riesgos, así como los mecanismos de supervisión humana disponibles.

Las medidas proactivas pasan también en general por las obligaciones de supervisión humana (art. 14). Es fundamental que estos sistemas permitan la supervisión humana durante su uso, con el fin de minimizar cualquier riesgo

residual para la salud, la seguridad y los derechos fundamentales, incluso después de que se hayan implementado medidas de mitigación. Los implementadores deben poder monitorizar los sistemas y entender las decisiones o resultados que estos generan. En casos de identificación biométrica remota, se requiere que la salida del sistema sea verificada por al menos una persona física, y posiblemente dos, para garantizar su fiabilidad.

Finalmente, los sistemas de IA evitarán los sesgos y la discriminación si están diseñados con «solidez» técnica, cumpliendo obligaciones para prevenir o minimizar «errores, fallos, incoherencias, situaciones inesperadas» (Considerando 75) y la eliminación o reducción de resultados sesgados una vez comercializados (art. 15. 3.º). Estos sistemas deben ser diseñados con tolerancia a errores o inconsistencias que puedan surgir en su interacción con el entorno, especialmente cuando interactúan con personas u otros sistemas. Además, deben incorporar medidas de ciberseguridad —que no sean *hackeables*—. adecuadas y proporcionadas a las circunstancias específicas del sistema, especialmente para proteger contra la manipulación de los datos de entrenamiento.

Todas estas exigencias serán desarrolladas además por todo un armazón de normas técnicas y de armonización, que serán acompañadas de la normalización técnica de la IA que está actualmente en fase de desarrollo. Para los sistemas que no son de alto riesgo, todas estas obligaciones serán los referentes de las buenas prácticas a seguir voluntariamente en códigos, sellos o certificaciones a las que se sumen (Considerandos 165-166) los proveedores públicos o privados de esos sistemas de IA.

5.2. Responsabilidades de proveedores, implementadores y otros actores clave

En el marco del RIA, se establecen obligaciones claras para las diferentes partes que intervienen en la puesta en marcha, comercialización y funcionamiento de los sistemas de IA de alto riesgo. Los proveedores asumen la mayor parte de estas responsabilidades, debiendo cumplir estrictamente con los requisitos establecidos y superar las pruebas de conformidad. Además, están obligados a mantener un sistema de gestión de calidad bien documentado y actualizado, que incluya la preservación de todos los registros pertinentes relacionados con el sistema. Asimismo, deben colaborar activamente con las autoridades, lo cual implica registrar el sistema, demostrar su conformidad con los requisitos reglamentarios cuando sea necesario, y notificar cualquier incumplimiento o riesgo identificado, así como las medidas correctivas adoptadas en consecuencia.

Los implementadores fueron llamados «usuarios» en el periodo de elaboración, pero que nada tienen que ver con personas afectadas por una IA. Pues bien, los implementadores de sistemas IA de alto riesgo, especialmente si se

trata de organismos públicos o empresas que prestan servicios públicos, tienen la obligación adicional de realizar una evaluación de impacto sobre los derechos fundamentales relacionada con el sistema de IA. Por su parte, los importadores y distribuidores están encargados de conservar la documentación técnica y asegurarse de que los productos hayan pasado por la evaluación de conformidad requerida y cuenten con el marcado CE. Además, deben garantizar que exista un representante autorizado del proveedor dentro de la Unión Europea y colaborar con las autoridades cuando sea necesario.

Los implementadores también tienen responsabilidades clave en la supervisión humana de los sistemas, debiendo monitorizar su funcionamiento, guardar registros detallados y cooperar con las autoridades competentes. Cualquier modificación sustancial de un sistema de alto riesgo existente o de un sistema de inteligencia artificial de propósito general, que lo convierta en un sistema de alto riesgo, se considera como la creación de un nuevo sistema de alto riesgo, lo cual genera las mismas obligaciones mencionadas anteriormente.

5.3. La evaluación de conformidad que se obtiene si se cumplen estándares y normas y el registro en la base de datos de sistemas de alto riesgo

En la sección 5 de este capítulo III, se abordan las normas armonizadas, las evaluaciones de conformidad, los certificados y el registro de los sistemas de IA. Si un sistema de IA de alto riesgo o un sistema de IA de propósito general cumple con las normas técnicas armonizadas publicadas en el Diario Oficial de la Unión Europea, se presumirá que cumple con los requisitos del RIA. En ausencia de estas normas, la Comisión tiene la facultad de aprobar especificaciones comunes que actuarán como sustituto de las normas técnicas hasta que estas estén disponibles.

En la mayoría de los casos, la conformidad de los sistemas de alto riesgo y de propósito general se puede demostrar a través de un procedimiento de evaluación interna. No obstante, para sistemas de identificación biométrica regulados por el RIA, si no se han aplicado normas armonizadas o solo parcialmente, se requerirá la intervención de un organismo notificado[35].

Los sistemas que superen la evaluación de conformidad recibirán un certificado emitido por los organismos notificados, con una validez máxima de cuatro años para los sistemas de IA del Anexo II y de cinco años para los sistemas del Anexo III, pudiendo ser renovado si se superan las pruebas perti-

35 Sobre este tema, de especial complejidad, Palma Ortigosa, A.: «¿Quién es quién en el Reglamento Europeo de Inteligencia Artificial? Las autoridades notificantes y los organismos notificados», en *Actualidad Jurídica Iberoamericana*, n.º 21, monográfico, 2024, pp. 598-617 https://revista-aji.com/numero-21/

nentes. Estos sistemas también deberán llevar el marcado CE, y los incluidos en el Anexo III deberán registrarse en un registro europeo de sistemas de alto riesgo.

Finalmente, se prevé que una autoridad de supervisión de mercado puede autorizar, en casos excepcionales y justificados, y por un tiempo limitado, la operación de un sistema de alto riesgo que no haya completado las pruebas de conformidad, cuando esto sea necesario para proteger la vida de las personas, el medio ambiente o infraestructuras clave. Sin embargo, estas pruebas deberán realizarse lo antes posible, y si no se superan, el sistema deberá ser retirado.

Antes de que un sistema de alto riesgo sea puesto en servicio o comercializado, el proveedor está obligado a registrarlo en la base de datos de la UE. Asimismo, los implementadores del sector público deberán registrarse antes de usar el sistema. No obstante, no están obligados al registro los sistemas de alto riesgo destinados al orden público, control de fronteras, inmigración o asilo, ni sus implementadores.

El artículo 71, Capítulo VII del RIA se encarga de regular la creación y el mantenimiento de una base de datos a nivel de la UE, destinada a registrar los sistemas de inteligencia artificial de alto riesgo, tal como se definen en el anexo III. Esta base de datos será gestionada por la Comisión Europea en coordinación con los Estados miembros. Además, se requiere que los implementadores del sector público que utilicen estos sistemas de IA proporcionen información adicional específica durante el proceso de registro.

VI. ¿Qué debemos saber los humanos cuando interactuamos con determinados sistemas de IA? (artículo 50)

El artículo 50 supone un Capítulo, sobre Obligaciones de transparencia de ciertos sistemas de IA y genera una tipología especial de sistemas de inteligencia artificial con un régimen jurídico específico. Se establecen una serie de requisitos esenciales que los proveedores e implementadores de sistemas de inteligencia artificial deben cumplir para garantizar la transparencia en el uso de estos sistemas, especialmente cuando interactúan directamente con personas. Los proveedores tienen la responsabilidad de asegurarse de que las personas físicas sepan que están interactuando con un sistema de IA, salvo que dicha interacción sea evidentemente obvia, como podría ser el caso en ciertos contextos donde la intervención de IA es claramente identificable. Existen excepciones para este requisito, particularmente en casos relacionados con la persecución de delitos, donde revelar la presencia de un sistema de IA podría comprometer las investigaciones.

En cuanto a los sistemas de IA que generan contenido sintético, como audio o video, los proveedores deben garantizar que exista un mecanismo

técnico que permita corroborar que dicho contenido ha sido creado o manipulado artificialmente. Esta medida busca prevenir el engaño o la confusión en la percepción del contenido por parte del público.

Además, los implementadores de sistemas de categorización biométrica o de reconocimiento de emociones tienen la obligación de informar a las personas sobre el uso de dichos sistemas en su interacción. No obstante, se prevén excepciones en contextos de persecución del crimen, donde la divulgación de esta información podría interferir con la eficacia de las investigaciones.

Por último, en relación con los sistemas de IA que producen imágenes o sonidos que simulan ser reales, como en el caso de los llamados «deep fakes», los implementadores deben advertir sobre la naturaleza artificial de este contenido, con algunas excepciones, como en la persecución de delitos o en el caso de obras que sean claramente creativas, satíricas o ficticias.

VII. Obligaciones respecto de los modelos de IA general como «ChatGPT»

El RIA en su fase final y tras el surgimiento de tecnologías como ChatGPT y similares, introdujo regulaciones específicas para los sistemas de IA de propósito general, imponiéndoles una serie de obligaciones generales que se intensifican para aquellos considerados de riesgo sistémico. Estas regulaciones están bajo el control, dirección e impulso de la nueva Oficina de IA de la Unión Europea, que será responsable de establecer códigos de conducta para precisar estas obligaciones. Cabe apuntar que el modelo más o menos regulatorio e incisivo fue uno de los grandes caballos de batalla en la negociación final en la que especialmente Alemania Francia e Italia se oponían a otros países que pretendían una regulación más taxativa, llevándose a un modelo híbrido.

El Capítulo 1 establece «Normas de clasificación de los modelos de IA de propósito general». Así´, el RIA distingue entre modelos de IA de propósito general y aquellos clasificados como de propósito general con riesgo sistémico. Estos últimos son definidos como aquellos que, o bien poseen capacidades de alto impacto, o bien han sido designados como tales por la Comisión, o como resultado de una alerta emitida por el Grupo Científico de Expertos Independientes. El Capítulo 2 regula «Obligaciones de los proveedores de modelos de IA de propósito general». Todos los proveedores de sistemas de inteligencia artificial de propósito general están sujetos a una serie de obligaciones, entre las que se incluyen: la elaboración de documentación técnica adecuada, el diseño de una política de cumplimiento con la Directiva sobre derechos de autor, y la creación de un resumen suficientemente detallado sobre el contenido utilizado durante el entrenamiento del sistema de IA.

Además de las obligaciones generales aplicables a todos los sistemas de IA de propósito general, el Capítulo 3 dispone «Obligaciones de los proveedores de modelos de IA de propósito general con riesgo sistémico». Los proveedores de aquellos clasificados como de riesgo sistémico deberán llevar a cabo evaluaciones de los modelos con el fin de identificar posibles riesgos sistémicos, tomar medidas para mitigar dichos riesgos en su caso, documentar y notificar incidentes graves, y garantizar un nivel adecuado de protección de la ciberseguridad de estos sistemas.

VIII. Innovación bajo control: sandboxes y otras medidas para fomentar la IA

Este Capítulo de Medidas de apoyo a la innovación, esencialmente sobre *sandboxes*, establece las directrices para implementar medidas de apoyo a la innovación, centrándose en los *sandboxes* regulatorios. Según el RIA, las autoridades nacionales deben crear en dos años al menos un *sandbox* regulatorio destinado a desarrollar, entrenar, probar y validar sistemas de inteligencia artificial bajo su orientación, supervisión y apoyo. Cualquier proveedor, actual o potencial, que cumpla con los criterios establecidos y sea seleccionado, podrá participar en estos *sandboxes*.

La Comisión Europea, a través de actos de ejecución, definirá las características comunes que deberán incluir todos los *sandboxes* implementados por las autoridades nacionales. Estas características abarcarán aspectos como los procedimientos de planificación y ejecución, incluyendo la elegibilidad, el proceso de solicitud, selección, participación, seguimiento y salida.

La participación en un *sandbox* estará limitada temporalmente, dependiendo de la escala del proyecto. Basado en un plan específico, si los participantes cumplen con las directrices establecidas por las autoridades, estarán exentos de sanciones administrativas por infracciones legales relativas al sistema supervisado en el *sandbox*. Sin embargo, esta exención no limita los poderes de supervisión de las autoridades ni exime a los participantes de responsabilidad por cualquier daño causado durante su participación.

Dentro del *sandbox*, y bajo ciertas condiciones, se permitirá el tratamiento de datos personales recolectados para otros fines, siempre que sean esenciales para el desarrollo de sistemas que persigan un interés público sustancial, como en las áreas de salud, medio ambiente, sostenibilidad energética, movilidad, calidad del servicio público y seguridad de infraestructuras críticas.

Las autoridades nacionales deberán acordar con los participantes las condiciones para realizar pruebas en entornos reales, garantizando siempre la seguridad, la salud y el respeto a los derechos fundamentales. El RIA establece términos y limitaciones, que incluyen, entre otros, la aprobación previa y el registro del plan de pruebas del participante, el consentimiento de los afectados y la reversibilidad de las decisiones tomadas durante las pruebas.

Al finalizar su participación, los participantes recibirán una certificación de las actividades realizadas y un informe final, el cual podrá ser valorado por organismos notificados durante la evaluación de conformidad. Si los participantes lo consienten, estos informes estarán disponibles para la Comisión Europea y el Consejo Europeo de Inteligencia Artificial, y, en algunos casos, podrían publicarse en la plataforma de información única.

Además, se han establecido medidas específicas para PYMEs, incluidas las start-ups, tales como formación adaptada sobre el RIA, canales de comunicación específicos, apoyo en el cumplimiento del RIA y la publicación de información y mejores prácticas. Cabe destacar que las microempresas no estarán obligadas a mantener un sistema de gestión de calidad, aunque sí deberán gestionar los riesgos asociados.

IX. Gobernanza de la IA en Europa: ¿quiénes supervisan el cumplimiento del reglamento en España y en la Unión?

Es importante señalar que en el RIA se introduce la creación de diversos organismos clave para la gobernanza de la inteligencia artificial en la Unión Europea. Entre ellos, se encuentra la Oficina Europea de la IA[36], establecida el 24 de enero de 2024, cuya función principal es centralizar el desarrollo y fortalecimiento de las capacidades de la Unión en materia de IA. Asimismo, el RIA regula la conformación del Consejo Europeo de Inteligencia Artificial y la creación de dos órganos consultivos: un grupo científico de expertos y un foro consultivo.

El Capítulo VI del RIA se enfoca en la gobernanza, detallando las competencias asignadas a la Oficina Europea de la IA y al Consejo Europeo de Inteligencia Artificial. Este Consejo, compuesto por un representante de cada Estado miembro y con la participación del Supervisor Europeo de Protección de Datos como observador, tendrá como tareas principales la coordinación entre autoridades competentes, la recopilación de conocimientos técnicos y la promoción de mejores prácticas a nivel europeo. Además, contribuirá a la armonización de pruebas de conformidad y la supervisión de *sandboxes* y pruebas en condiciones reales, en colaboración con organismos expertos en diversas áreas, como servicios digitales, financieros, ciberseguridad, y criptomonedas.

El Consejo también asesorará a la Comisión Europea en la definición de directrices para la implementación del RIA, en la preparación de actos de

36 *Decisión de la Comisión de 24.1.2024 por la que se crea la Oficina Europea de Inteligencia Artificial*, Bruselas, 24.1.2024C(2024) 390 final, https://ec.europa.eu/newsroom/dae/redirection/document/101625

ejecución o delegados, y en asuntos internacionales relacionados con la IA. Asimismo, ofrecerá recomendaciones sobre especificaciones, estándares y guías, y apoyará a las autoridades nacionales en la supervisión del mercado de IA.

Por último, los órganos consultivos, como el foro consultivo y el grupo científico de expertos independientes, complementarán las actividades del Consejo proporcionando asesoramiento técnico y científico en la ejecución de las disposiciones del RIA.

Ya respecto de la gobernanza en cada Estado miembro, se designarán autoridades de supervisión del mercado (art. 59 del RIA) encargadas de monitorizar el desempeño de los sistemas de IA, incluidos aquellos considerados de alto riesgo, identificando riesgos emergentes, incidentes y situaciones que puedan requerir intervención. Además, los organismos de evaluación de conformidad, responsables de realizar las pruebas correspondientes, deberán estar habilitados por una autoridad notificante. Cada Estado miembro tiene la facultad de establecer nuevas instituciones o asignar estas competencias a autoridades existentes, como la Agencia Española de Protección de Datos o, para el uso de algoritmos en el ámbito judicial (Anexo III.8.º), el Consejo General del Poder Judicial[37]. España ha sido pionera en la UE al crear la Autoridad Española de Supervisión de la IA (AESIA)[38]. Estas autoridades deben actuar de manera independiente, si bien es cuestionable que en el caso español de la AESIA su regulación permita contar con este requisito[39].

En el mecanismo de gobernanza nacional hay que ubicar a las autoridades notificantes y organismos notificados, regulados en la Sección IV del Capítulo III. El RIA establece que cada Estado miembro debe designar una autoridad notificante. Esta autoridad es responsable de establecer y aplicar los

37 Para el caso de sistemas de identificación biométrica utilizados por Fuerzas y Cuerpos de Seguridad del Estado, en migración o en Administración de Justicia, las autoridades de supervisión serán, las autoridades de protección de datos competente en esos ámbitos, pudiendo ser en algunos supuestos la AEPD.

38 La creación legal y habilitación fue por la Ley 28/2022, de 21 de diciembre, en concreto, su D.A 7.ª reguló la creación. Asimismo, según Ley 28/2022, la AESIA está adscrita orgánica al Ministerio de Asuntos Económicos y Transformación Digital, a través de su Secretaria-o SEDIA.El Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto.

39 Se creó como un organismo público bajo el artículo 91 de la Ley 40/2015 y no como una agencia estatal independiente de acuerdo con los artículos 108 bis y siguientes de la Ley 40/2015. En sus estatutos regulados por el Real Decreto 729/2023, de 22 de agosto, el artículo 8 menciona como «Principios de actuación de la Agencia» la «Autonomía» y la «Independencia técnica» y señala que «la Agencia actuará con plena autonomía». No obstante, estas afirmaciones parecen insuficientes, especialmente dada una evidente falta de independencia orgánica que afecta a la independencia funcional. Incluso se establece explícitamente la «dependencia» de la Dirección de un órgano político como el Secretario de Estado, quien es el Presidente y desempeña algunas funciones materiales (art. 23.1.º).

procedimientos necesarios para evaluar, designar, notificar y supervisar a los organismos de evaluación de conformidad. La autoridad notificante tiene la opción de delegar la evaluación y supervisión en una organización de acreditación, lo que permite garantizar que los organismos notificados puedan verificar la conformidad de los sistemas de IA de alto riesgo con el RIA.

Los organismos de evaluación de conformidad que deseen ser «notificados», esto es, ser autorizados a poder realizar evaluaciones de conformidad) deben obtener una certificación que acredite el cumplimiento de los requisitos establecidos por un organismo nacional de acreditación. Luego, deben presentar su solicitud y la documentación correspondiente a la autoridad notificante, que a su vez notificará a la Comisión Europea una vez que se hayan cumplido los requisitos necesarios.

X. ¿Cómo se supervisan los sistemas de IA después de su comercialización

Resulta de interés el Capítulo VIII que regula la Monitorización post-comercialización, intercambio de información y vigilancia del mercado. Así, los proveedores de sistemas de alto riesgo deberán implementar un sistema de monitorización post-comercialización que sea adecuado a los riesgos identificados y a la tecnología de IA utilizada. Este sistema recogerá los datos necesarios para asegurar el cumplimiento de los requisitos establecidos para los sistemas de alto riesgo en el Capítulos III. La estructura del sistema deberá estar documentada en un plan que se ajuste al esquema propuesto por la Comisión.

Respecto del Intercambio de información sobre incidentes graves, los proveedores estarán obligados a notificar cualquier incidente a las autoridades de vigilancia del mercado del Estado miembro donde se haya producido. A su vez, estas autoridades informarán del incidente a las autoridades encargadas de la protección de los derechos fundamentales.

Asimismo, se dispone que las autoridades de supervisión del mercado reportarán a la Comisión las actividades realizadas en relación con el RIA. Para los sistemas de alto riesgo regulados por normativa armonizada, la autoridad supervisora del mercado será la indicada por dicha normativa, salvo que el Estado miembro justifique la designación de otra autoridad. En el caso de los sistemas de alto riesgo utilizados por instituciones financieras, el supervisor financiero será la autoridad competente, a menos que el Estado miembro designe otra autoridad justificada.

Para los sistemas de alto riesgo destinados a la identificación biométrica en contextos policiales, migratorios, de aplicación de la ley y justicia, las autoridades de vigilancia del mercado serán las correspondientes autoridades de protección de datos, como la Agencia Española de Protección de Datos en algunos casos.

Las instituciones de la Unión que estén bajo el ámbito del RIA serán supervisadas por el Supervisor Europeo de Protección de Datos. Los Estados miembros deberán facilitar la coordinación entre sus autoridades nacionales de supervisión del mercado, incluidas aquellas designadas en virtud del RIA y las que supervisen la aplicación de normativa armonizada. La Oficina Europea de IA prestará asistencia en estas tareas de coordinación.

Si es necesario, las autoridades de supervisión del mercado podrán acceder a los datos utilizados para el entrenamiento, validación y testeo de los sistemas a través de APIs, así como al código fuente, si fuera necesario, para evaluar la conformidad del sistema de alto riesgo con los requisitos del RIA o si otros métodos han sido insuficientes para verificar dicho cumplimiento.

Estas autoridades tendrán competencia y poderes para asegurar que las pruebas en condiciones reales se realicen de acuerdo con el RIA. Podrán otorgar o denegar permisos para realizar dichas pruebas, y, en caso de incidentes graves, exigir modificaciones al proveedor, suspender o finalizar las pruebas.

Las autoridades nacionales que supervisen o impongan obligaciones relacionadas con la protección de derechos fundamentales, cuando ejerzan sus competencias respecto al uso de sistemas de alto riesgo, podrán acceder a la documentación generada conforme a este RIA. Será necesario notificar a la Comisión cuáles son esas autoridades. Si la documentación resulta insuficiente, podrán solicitar a la autoridad de supervisión del mercado que organice pruebas específicas.

Si una autoridad de supervisión del mercado considera que un sistema presenta riesgos, podrá evaluarlo y, si descubre que no cumple con el RIA, requerir al proveedor que tome medidas correctivas. Si el incumplimiento es transnacional, la autoridad informará a la Comisión y a otros Estados miembros. Si el proveedor no toma las acciones correctivas necesarias, la autoridad podrá tomar medidas para prohibir o restringir el sistema de alto riesgo. El RIA detalla cómo se aplican estas medidas, especialmente en casos transnacionales, y establece procedimientos para resolver desacuerdos entre Estados miembros.

Las autoridades también podrán exigir medidas correctivas o incluso retirar del mercado sistemas de alto riesgo que, aunque cumplan con los requisitos, puedan presentar riesgos o no demuestren adecuadamente su conformidad.

Por cuanto a la confidencialidad, en virtud de este capítulo, toda persona física o jurídica involucrada en la aplicación del RIA deberá garantizar la confidencialidad de los datos obtenidos como resultado de dichas actividades. La información intercambiada entre las autoridades nacionales y la Comisión no se revelará sin previa consulta con la autoridad que la originó, especialmente cuando se trate de sistemas utilizados en actividades de orden público, control fronterizo, inmigración o asilo. El intercambio de datos operativos sobre

estas actividades no es obligatorio. Los proveedores tendrán la responsabilidad de custodiar la documentación de estos sistemas y de proporcionar copias a los agentes autorizados de las autoridades de supervisión.

XI. Otros aspectos del reglamento: derechos, sanciones y aplicación gradual, códigos de conducta y delegaciones a la comisión

11.1. ¿Reconoce el Reglamento derechos a los afectados por la IA?

En relación con el reconocimiento de nuevos derechos dentro del RIA, es fundamental destacar que este se enfoca principalmente en los fabricantes (proveedores) y en quienes implementan productos de IA. Sin embargo, ha recibido críticas por pasar por alto a las personas afectadas por estos sistemas de IA, quienes ni siquiera son mencionadas en el texto[40]. Respecto a la incorporación de derechos, las enmiendas 628 y 629 de junio de 2023 propusieron incluir ciertos derechos, culminando en el reconocimiento del derecho a presentar una reclamación ante una autoridad de vigilancia del mercado, así como el derecho a obtener explicaciones sobre decisiones individuales, como se indica en los artículos 85-87. Es relevante señalar que el texto final aprobado atenúa la propuesta inicial del Parlamento. Como menciona López-Tarruella, estas disposiciones se mantienen más en el ámbito de las buenas intenciones que en medidas concretas[41].

Así, el derecho a presentar una reclamación, a diferencia del derecho establecido en el artículo 77 del RGPD, se articula en el RIA como un derecho a presentar una solicitud ante la autoridad de vigilancia del mercado, la cual podría tomar en consideración para acciones o investigaciones futuras. No obstante, según el autor mencionado, estas solicitudes podrían verse entorpecidas por la compleja distribución de competencias entre las distintas autoridades de vigilancia del mercado, derivada del artículo 74 del RIA. En cuanto al derecho a recibir explicaciones sobre decisiones individuales, este se restringe a ciertos sistemas de alto riesgo, y en el caso de personas físicas, la obtención de dichas explicaciones deberá seguir los procedimientos establecidos en el artículo 22 del RGPD.

---

40 Hasta junio de 2023 no se incluyó entre las definiciones la de «persona afectada», junto con diversos derechos en las enmiendas del Parlamento (Enmienda 174, art. 3. 1.º, 8 bis). Las «personas afectadas» finalmente no se definen, pero sí que forman parte del «alcance» del artículo 2.1.g) RIA. «personas afectadas que estén establecidas en la Unión».

41 El primer estudio sobre el tema López-Tarruella Martínez, A.: «Vías de recurso para los particulares en el reglamento de inteligencia artificial», en Cotino Hueso, L. y Simón Castellanos, P. (coords.): *Tratado sobre el Reglamento ... cit.*

11.2. Códigos de conducta: Estándares voluntarios para sistemas no regulados

Cómo se ha señalado los sistemas que no están regulados por el RIA de inteligencia artificial (esto es los que no estén prohibidos, no sean de alto riesgo, no sean de inteligencia artificial general o que tengan las especiales obligaciones de transparencia del artículo 50, no contarán con ninguna obligación por el RIA. No obstante, claro está, que tendrán que cumplir cualquier obligación de cualquier otra normativa, como pueda ser el reglamento de protección de datos, por ejemplo.

El capítulo IX sobre Códigos de conducta regula que la Oficina de la IA y los Estados Miembros impulsarán la creación de códigos de conducta destinados a fomentar la adopción voluntaria de algunos requisitos del RIA para sistemas de alto riesgo en aquellos sistemas de IA que no se consideran de alto riesgo. Además, promoverán la aplicación voluntaria de otros requisitos, como los relacionados con la sostenibilidad medioambiental o la accesibilidad. Estos códigos de conducta podrán ser elaborados por empresas, organizaciones de la sociedad civil y otros agentes relevantes. Se apuesta así por un desarrollo del mercado de las certificaciones respecto de la inteligencia artificial que no sea de alto riesgo.

11.3. Régimen sancionador: ¿cuáles son las consecuencias del incumplimiento?

Por cuanto al régimen sancionador, cabe señalar que será preciso un desarrollo y concreción por los estados miembros para que reúna los mínimos exigibles por el principio de legalidad, sancionadora y tipicidad. En todo caso, el RIA sí que establece unos parámetros básicos. Respecto de las sanciones, se ajustarán en función de las circunstancias y considerarán el tamaño del proveedor.

El montante es importante, así, hasta 35 millones de euros, o el 7 % del volumen anual de negocio para empresas (3 % en el caso de PYMEs) por incumplir las prohibiciones del RIA. Hasta 15 millones de euros o el 3 % del volumen anual de negocio (2 % para PYMEs) por incumplir las obligaciones del Reglamento aplicables a proveedores, importadores, distribuidores e implementadores. Hasta 7.5 millones de euros o el 1 % del volumen anual de negocio (1 % para PYMEs) por suministrar información incorrecta a organismos notificados o autoridades nacionales competentes. Hasta 1.5 millones de euros para entidades de la Unión por incumplimiento de prohibiciones, o 750 mil euros por incumplimiento de obligaciones. Para los proveedores de modelos de IA de propósito general se contempla una sanción máxima de 15 millones de euros o el 3 % del volumen anual de negocio.

11.4. Delegación de poderes a la Comisión para ir adaptando el Reglamento

En los próximos meses y años, la Comisión Europea desempeñará un papel esencial en el desarrollo y concreción de los criterios y normas establecidas en el RIA, ya sea de manera directa o a través de la Oficina de Inteligencia Artificial.

Uno de los mecanismos para llevar a cabo estas funciones es la delegación de poderes, regulada en el Capítulo XI del RIA. En virtud de esta delegación, la Comisión tiene la facultad de modificar la lista de tipos de sistemas que se consideran de alto riesgo en el Anexo III. Asegurar que la documentación requerida sea suficiente para garantizar la conformidad de los sistemas con el RIA, teniendo en cuenta los avances técnicos. Revisar y actualizar los Anexos IV (documentación técnica), VI (evaluación de conformidad mediante control interno) y VII (evaluación de conformidad a través del sistema de control de calidad y documentación) para adecuarlos al progreso tecnológico. Exigir que ciertos sistemas de alto riesgo, para los cuales el RIA permite la demostración de conformidad mediante control interno, sean sometidos a pruebas de conformidad por parte de una entidad notificada y, finalmente, modificar el Anexo V (contenido de la declaración de conformidad) para adaptarlo a los avances técnicos.

11.5. Disposiciones finales y la (lenta) aplicación del Reglamento

Este Capítulo final se centra en la modificación de otras disposiciones para alinearlas con el nuevo Reglamento. Además, establece un cronograma para la implementación del RIA. Así, aunque el RIA ya está en vigor tras 20 días desde su publicación, entrará en plena aplicación dos años después de su publicación en el DOUE. Las prohibiciones específicas relativas a los sistemas de IA comenzarán a regir a los seis meses de dicha publicación. A los doce meses después de la publicación, entrarán en vigor las disposiciones relativas a las autoridades notificantes, las autoridades nacionales de supervisión, los modelos de IA de propósito general, el marco de gobernanza del RIA, así como las sanciones establecidas. Finalmente, a los treinta y seis meses, se incorporará la consideración de alto riesgo para los sistemas de IA regulados por legislación de armonización, tales como juguetes, ascensores o productos sanitarios. Debe notarse que las obligaciones sólo serán plenamente exigibles al sector público, ni más ni menos que después de seis años.

XII. Para acabar, también en 2024 un convenio de inteligencia artificial del consejo de Europa

Se expuso al inicio el contexto actual de aceleración regulatoria mundial de la IA, y particularmente bajo la influencia del RIA de la Unión Europea. Pues en este contexto se sitúa la adopción del Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa del Consejo de Europa[42], finalmente aprobado el 17 de mayo de 2024 durante la 133.ª Sesión del Comité de Ministros en Estrasburgo. Este Convenio, que será el tratado número 225 en la serie de tratados[43], se abrirá a la firma el 5 de septiembre de 2024 durante la Conferencia de Ministros de Justicia en Vilnius. El documento es amplio, con diecisiete considerandos y treinta y seis artículos distribuidos en ocho capítulos[44].

Este Convenio representa un esfuerzo considerable para armonizar y establecer un marco común mínimo en Europa, orientado hacia una visión global centrada en los derechos humanos, la democracia y el Estado de Derecho, en lugar de centrarse únicamente en la economía y el mercado. Europa aspira a posicionarse como un *faro luminoso* y una piedra angular en el ámbito normativo de la IA, no solo a nivel continental sino también global. Para los Estados que no son miembros de la UE, este Convenio de IA puede desempeñar un papel normativo y jurídico fundamental. En el caso de la UE y sus Estados miembros que ratifiquen el Convenio, este podría complementar, fortalecer y elevar la protección de los derechos. Es importante resaltar que, a diferencia del RIA, el Convenio de IA no se restringe únicamente a los sistemas de IA de alto riesgo. Además, para los Estados que ratifiquen el Convenio, este integra como «principios» en el ámbito jurídico y normativo aquello que hasta ahora eran solo principios éticos consensuados en torno a la IA. Estos principios jurídicos tienen un gran potencial en manos de los operadores legales para derivar normas y obligaciones concretas, siguiendo la tradición establecida en las últimas décadas con los principios de protección de datos. Además, el Convenio de IA otorga a las personas afectadas por la IA ciertos derechos y garantías, estableciendo mecanismos para la protección de estos derechos ante autoridades que deben ser independientes. Un aspecto especialmente relevante del Convenio de IA es su enfoque basado en el riesgo, particular-

42 Sobre el mismo, mi estudio The Council of Europe Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, forthcoming in CERIDAP Journal, n.º 3/2024, acceso en https://ceridap.eu/fascicoli/ También en español en SSRN y una versión inicial en *Revista Administración & Cidadanía*, EGAP, 2024.

43 https://rm.coe.int/1680afae3c

44 Un Preámbulo con diecisiete Considerandos y treinta y seis artículos estructurados en ocho capítulos: Disposiciones generales (I); Obligaciones generales (II); Principios (III); Recursos (IV); Evaluación y Mitigación de Riesgos e Impactos adversos (V) y los relativos a Aplicación (VI); Mecanismo de seguimiento y cooperación (VII) y Cláusulas finales (VIII).

mente en el artículo 16, que requiere una evaluación y mitigación continua de los riesgos e impactos adversos de cualquier tipo de sistema de IA.

Por lo tanto, el Convenio de IA tiene el potencial de complementar el RIA de la UE. Más allá de esta perspectiva, el valor del Convenio va más allá. Si se me permite, *el Convenio aporta una dimensión lírica a la prosa técnica que representa el RIA*. Mientras que el RIA establece las bases y estructuras para un ecosistema de IA seguro y confiable, el Convenio se centra en su impacto en las personas y en la sociedad democrática. El RIA es meticuloso, detallado y preciso, trazando un camino claro a través de la complejidad técnica y jurídica, estableciendo estándares firmes y obligaciones concretas para los proveedores y usuarios de sistemas de IA. En contraste, el Convenio introduce una perspectiva más elevada, integrando normativamente los valores fundamentales, los principios éticos y los derechos humanos que deben guiar la evolución de la IA. El Convenio no solo tiene un valor simbólico y metajurídico, sino que es un instrumento normativo con capacidad de integración casi constitucional en los ordenamientos jurídicos de los Estados parte, y posee un gran potencial interpretativo. Por ello, el Convenio de IA supera a muchos de los instrumentos declarativos y de *soft law* que ya resultaban redundantes, ineficaces e incluso tediosos. Naturalmente, será esencial seguir de cerca su proceso final de aprobación, ratificación y la implementación de sus efectos.

Para concluir. El RIA representa un hito fundamental en la regulación de tecnologías emergentes a nivel mundial. Su enfoque basado en la gestión de riesgos establece un marco normativo que no solo busca mitigar los posibles peligros asociados con la inteligencia artificial, sino también fomentar la innovación bajo estrictas garantías. A lo largo del análisis se ha expuesto cómo el RIA, identifica la IA más peligrosa para prohibirla o bien para imponer obligaciones específicas para los sistemas de alto riesgo. En los próximos años se comprobará la capacidad de influencia de esta regulación en el mundo, su «efecto Bruselas». Este fenómeno ya se aprecia en algunos casos. También veremos si el RIA supone un lastre al desarrollo de la IA en la UE o nos hace líderes en IA ética en el diseño. Igualmente podremos apreciar en los próximos años el desarrollo del Convenio de IA del Consejo de Europa en 2024, más fácil de asumir para muchos países.

Bibliografía

Barrio Andrés, M., «El cumplimiento basado en el riesgo o risk-based compliance, pieza cardinal del nuevo Derecho digital europeo», en Análisis del Real Instituto Elcano (ARI), n.º 34, 2023.

Cotino Hueso, L. y **Simón Castellano, P**. *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, Aranzadi, 2024, https://links.uv.es/uf7Pqmh

Cotino Hueso, L., «Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos», en Balaguer Callejón, F. y Cotino Hueso, L., Derecho público de la inteligencia artificial, F. Jiménez Abad-Marcial Pons, Zaragoza, 2023.

Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», en El Cronista del Estado Social, IUSTEL, monográfico Inteligencia artificial, n.º 100, septiembre-octubre 2022.

Cotino Hueso, L., «Nuevo paradigma en las garantías de los derechos fundamentales y una nueva protección de datos frente al impacto social y colectivo de la inteligencia artificial», en Cotino Hueso, L. (editor), *Derechos y garantías ante la inteligencia artificial y las decisiones automatizada*s, Thompson-Reuters Aranzadi, FIADI, Cizur Menor, 2022.

Cotino Hueso, L., «Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho» en *Revista Catalana de Derecho Público,* n.º 58, 2019.

Hernández Peña, J. C., *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*, Aranzadi, Cízur Menor, 2022.

Presno Linera, M. Á., «La propuesta de «Ley de Inteligencia Artificial» europea», en Revista de las Cortes Generales, n.º 116, 2023, págs. 81-133. De este autor, de interés, Derechos fundamentales e inteligencia artificial, Marcial Pons, Madrid, 2023.

López-Tarruella Martínez, A., «Vías de recurso para los particulares en el reglamento de inteligencia artificial», en Cotino Hueso, L. y Simón Castellanos, P. (coords.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, Aranzadi, 2024.

Mantelero, A., «Toward a New Approach to Data Protection in the Big Data Era», en Gasser U. y otros (dir.): Internet Monitor 2014: Reflections on the Digital World. Cambridge (MA): Berkman Center for Internet and Society at Harvard University, 2014.

Mantelero A, *Beyond Data. Human Rights, Ethical and Social Impact Assessment,* en Information Technology and Law Series IT&LAW 36, Sringer, Hague, 2022.

Martínez Martínez, R., «Inteligencia artificial desde el diseño. Retos y estrategias para el cumplimiento normativo», en Revista catalana de dret públic, n.º 58, 2019.

Palma Ortigosa, A.: «¿Quién es quién en el Reglamento Europeo de Inteligencia Artificial? Las autoridades notificantes y los organismos notificados», en Actualidad Jurídica Iberoamericana, n.º 21, monográfico, 2024.

Simón Castellano, P., *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi, Cízur Menor, 2023.

Simon Castellano, P., «Taxonomía de las garantías jurídicas en el empleo de los sistemas de inteligencia artificial», en Revista de Derecho Político, n.º 117, 2023.

6.

EL REGLAMENTO EUROPEO DE INTELIGENCIA ARTIFICIAL: EL SER HUMANO COMO MEDIDA DE TODAS LAS COSAS

Iban García del Blanco

Jurista. Eurodiputado 2019-24.
Miembro del equipo negociador de la Ley Europea de Inteligencia Artificial.

SUMARIO: I. El proceso regulador europeo: primeros pasos, la «soberanía digital europea». II. El reglamento europeo de inteligencia artificial, *iter* procedimental. III. La Ley de Inteligencia Artificial: contenido y estructura. 3.1. Usos Prohibidos. 3.2. Usos de alto riesgo. 3.3. La inteligencia artificial generativa. Obligaciones adicionales de transparencia para ciertos usos de IA. IV. Otros contenidos destacables de la *AI Act*. V. Gobernanza e implementación. VI. Régimen sancionador. VII. ¿Es Europa una isla de regulación en un mundo sin reglas? VIII. Punto y seguido: estamos ante una norma viva.

I. El proceso regulador europeo: primeros pasos, la «soberanía digital europea»

Al comienzo de la pasada legislatura europea, 2019-2024, se fijaron tres líneas de actuación prioritarias con el fin de situar a Europa a la cabeza de la transición digital sostenible, al tiempo que se construía el concepto y el objetivo de conseguir «recuperar la soberanía digital europea». Estas líneas eran:

El aumento de las inversiones desde una perspectiva estratégica europea, encuadrado este objetivo en la afectación de los fondos de programas ya existentes, y la creación de nuevas líneas, concretado inmediatamente en los propios planes de recuperación post COVID, que contribuyeron a financiar la doble transición ecológica y digital de la Unión. En ese sentido, y después de

tantos años, por fin en febrero de 2025 la Comisión anuncia una estrategia europea de IA que comportará 200.000 millones de euros de inversión, sin muchos más detalles por el momento.

La formación, basada en el incremento de contenidos digitales en la educación y planes para ofrecer capacidades y herramientas necesarias a la ciudadanía y cerrar la brecha digital.

Y establecer un marco regulatorio trasversal, holístico y completo para guiar la digitalización de acuerdo con los principios y valores europeos.

Con respecto a este último objetivo, la propia Ursula von der Leyen prometió en su discurso inaugural de 2019 avanzar desde el primer minuto y, con cierta ingenuidad, anunció que tendría una regulación en materia de inteligencia artificial en sus primeros cien días.

En esta dirección, no fue hasta enero de 2023 en que se realizó una «Declaración de principios y derechos digitales», firmada por las tres instituciones (Comisión Europea, Consejo Europeo y Parlamento Europeo), para proteger los derechos y la seguridad de todas las personas y las empresas cuando operan en el ámbito digital. Esta Declaración establecía los principios que deberían guiar el desarrollo de las nuevas tecnologías, y los derechos y deberes que tendrían que respetarse por parte de todos, administraciones públicas, partes interesadas, y ciudadanía, incluyendo nuestros derechos como usuarios, consumidores, y trabajadores y los derechos de los menores y las personas vulnerables.

No obstante, paralelamente, las instituciones han llevado un intenso trabajo regulatorio en esos cinco años para dotar a la Unión Europea de un marco jurídico único, justo y eficaz para regular la digitalización.

En Europa se han puesto en marcha un conjunto de Reglamentos, es decir, de normas de aplicación directa en todos los países de la UE, a los que se ha dado por llamar «leyes» *(«Acts»)* de manera harto simbólica, para mostrar el músculo legislador de la UE y trasladar que estamos ante un verdadero espacio único digital, con los mismos derechos y deberes, desde Riga hasta Lisboa.

Con este paquete legislativo se pretende regular el funcionamiento del mercado interior de la UE, en materia de mercados y servicios digitales, en datos, en privacidad, en capacidad tecnológica (chips)...

Pero Europa quería también, y quiere, ganar en soberanía, autonomía estratégica y capacidad de influencia. La regulación pretende establecer un *«golden standard»* mundial en el ámbito de la regulación digital, explotando lo que se ha dado en llamar el «efecto Bruselas», que comenzó con la aprobación del Reglamento de Protección de Datos Personales en 2016 (conocido como *«GDPR»,* por sus siglas en inglés). Es decir, un conjunto de reglas que esté centrado en nuestros propios principios y valores europeos, capaz de influir en el desarrollo normativo de la digitalización en otras partes del mundo.

Entre los reglamentos aprobados para configurar el marco normativo de la Europa digital, voy a mencionar aquellos que me parecen de mayor relevancia, antes de sumergirme en el análisis en profundidad del Reglamento de Inteligencia Artificial: leyes europeas de Mercados y Servicios Digitales *(«Digital Market Act»- «DMA» y «Digital Services Act»- «DSA»)*, aprobadas en 2022 y que entraron plenamente en vigor el pasado 17 de febrero de 2024. Con estas leyes se pone coto a la desregulación en internet, imponiendo obligaciones de control y transparencia a las plataformas digitales sobre los contenidos que se comparten, anuncian o venden en sus redes. También se han aprobado durante este mandato la Ley Europea de Datos y la Ley de Gobernanza de los Datos. Los datos son el carburante que necesitan las herramientas de inteligencia artificial para su desarrollo. Con estas leyes, Europa quiere explotar su potencial desarrollando un mercado de datos (no personales) que impulse a nuestra industria en la transformación digital (regulan también la nube).

II. El reglamento europeo de inteligencia artificial, *iter* procedimental

Sin duda, es la regulación de la Inteligencia Artificial la que ha estado en el centro del trabajo realizado durante estos años en las instituciones europeas y ha repercutido como un gran trueno en todo el globo, para alumbrar lo que sin duda es el colofón normativo del ecosistema digital europeo.

En diciembre de 2023, bajo la Presidencia Española del Consejo de la Unión Europea, y tras seis meses de negociaciones con los Estados Miembros, se alcanzó un acuerdo político entre los colegisladores europeos para la aprobación del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial. El acuerdo fue respaldado posteriormente por la Eurocámara con 523 votos a favor, 46 en contra y 49 abstenciones, en la sesión plenaria de marzo 2024. Tras su ratificación en el Consejo de la Unión Europea, su publicación se produjo el 2 de agosto de 2024, fecha a partir de la cual comienzan a contar los diferentes plazos en los que las diversas partes de la norma entran en vigor.

Fue noticia la duración maratoniana del último *«trílogo»* (diálogo tripartito, mejor expresado en español), pero, antes de ello, el Parlamento estuvo discutiendo sobre la propuesta de la Comisión Europea durante mucho tiempo, en un trabajo de acercamiento de posturas ideológicas e incluso filosóficas, que facilitó mucho el acuerdo posterior con el Consejo. Desde 2019, fueron varios los comités del mismo que aprobaron resoluciones y recomendaciones en diferentes ámbitos (seguridad, educación, cultura, propiedad intelectual, responsabilidad civil...), e incluso el Parlamento llegó a aprobar en marzo de 2020 la primera propuesta legislativa completa en esta materia en el mundo

—de la que tuve el honor de ser ponente— sobre «Aspectos Éticos de la Inteligencia Artificial, la Robótica y las Tecnologías conexas», que informó el texto que más tarde presentó la Comisión Europea y que acabó por convertirse en la Ley. Durante año y medio funcionó un Comité Especial sobre IA en el seno del Parlamento, en el que sus miembros tuvimos la ocasión de establecer un debate largo y sostenido en el tiempo con decenas de actores en este ámbito a escala mundial (sector tecnológico, civil, institucional, académico...), que acabó por culminar en un informe de recomendaciones no solo regulatorias, sino de desarrollo estratégico en la materia para Europa.

Como hemos dicho más arriba, la aprobación de la Ley Europea de IA es el colofón del sistema normativo en el ámbito digital europeo, fruto del intenso trabajo que en estos cinco años se ha venido realizando para dotar a la UE de un marco jurídico completo, justo y eficaz, para dirigir la digitalización y liderar la regulación de la inteligencia artificial en el mundo. Para la UE, dicha regulación debía hacerse respetando los principios y valores europeos, consagrados en los Tratados y en la Carta Europea de Derechos y Libertades Fundamentales de la Unión, para alumbrar una norma que situara en el centro a las personas. Es otro objetivo complementario a éste, que también la Ley dote de confianza a los usuarios, dé seguridad jurídica a las empresas y facilite la innovación y el desarrollo de IAs europeas.

III. La ley de inteligencia artificial: contenido y estructura

No sorprenderá con estos antecedentes que desvele que la primera gran piedra de toque en la discusión jurídico-regulatoria fuese la propia definición de la IA. El objetivo último fue que ésta fuese lo bastante amplia como para contener el suficiente espacio de realidad bajo el paraguas de la norma y, al mismo tiempo, conseguir una definición que estuviera alineada con los textos que se iban aprobando en el ámbito internacional. La definición que adopta la regulación, finalmente, tomó como referencia las que aparecían en los primeros textos de recomendaciones de organismos internacionales como la OCDE o la UNESCO.

El Artículo 3, apartado 1 de la Ley, define un sistema de IA como:

> «un software que, desarrollado con una o más de las técnicas y enfoques enumerados en el anexo I, puede, para un conjunto dado de objetivos definidos por humanos, generar resultados como contenido, predicciones, recomendaciones o decisiones que influyen en los entornos con los que interactúan»

Muy recientemente, a principios de febrero de 2025, la Oficina Europea de IA ha publicado una directriz (aunque no es vinculante y puede recoger posteriores precisiones) para acabar de perfilar qué se considera un sistema de

IA a efectos de la Ley. Concretamente, los rasgos que definen un sistema de IA, según la Oficina, son:

1. El aprendizaje automático o *machine learning*.
2. El aprendizaje supervisado, no supervisado y por refuerzo.
3. Los modelos generativos
4. Los sistemas basados en lógica y conocimiento experto.

Además, se establece un enfoque basado en el ciclo de vida del sistema, que divide la inteligencia artificial entre «fase de construcción» (cuando el sistema aprende y desarrolla modelos) y «fase de uso» (que trata cómo el sistema interactúa con el entorno para generar resultados).

Es interesante destacar que se excluyen expresamente los sistemas basados en reglas explícitas programadas por humanos, los *software* de procesamiento básico de datos (como las hojas de cálculo o bases de datos), los modelos de optimización matemática sin capacidad de inferencia avanzada, y los algoritmos clásicos que no aprenden o ajustan su comportamiento tras su implementación. En España, es común confundir el sistema «VIOGEN», utilizado para establecer criterios de riesgo en materia de violencia de género, con inteligencia artificial, pero no lo es (aunque parece que hay proyectos de hacer el sistema más sofisticado, incluyendo autoaprendizaje, por ejemplo).

La Ley Europea es aplicable a todas las tecnologías que incorporan IA y cuenta con una vocación geográfica global en cuanto a sus efectos. La virtualidad extraterritorial de la Ley garantiza su aplicación a todas las aplicaciones de IA que se producen o se usan en Europa, incluso cuando los modelos se hayan desarrollado o entrenado en terceros países, siempre que sus resultados —decisiones, recomendaciones, sugerencias o contenidos producidos por la IA— tengan un impacto en la Unión Europea. Es recomendable echarle un vistazo al artículo 2 de esta Ley, sobre el ámbito de aplicación, para tener una mayor concreción sobre todos estos ámbitos.

Se debe aclarar que los usos militares o de seguridad nacional están expresamente fuera del ámbito de aplicación de esta ley. La razón no es otra que la de que nos hallamos ante competencias exclusivas de los Estados Miembros y, por ello, la Unión no tiene capacidad para legislar. ¿Quiere esto decir que esos ámbitos quedan sin regular? No, pero le corresponde a cada Estado miembro hacerlo y sería deseable que en un futuro no lejano haya un tratado internacional que ponga en común algunas limitaciones y reglas.

Una apostilla muy importante: sí será de aplicación lo contenido en la Ley siempre que nos hallemos ante tecnologías o herramientas susceptibles de ser usadas híbridamente o para alguna otra tarea adicional (pensemos en aquellas utilizadas para el control de fronteras).

Uno de los grandes retos cuando se aborda la regulación de una materia tan compleja y en permanente evolución es no evacuar un texto que pueda

quedar obsoleto en muy poco tiempo. Por eso la técnica jurídica utilizada se centra en el uso y no en la tecnología en sí. Se trata de una Ley basada en el riesgo. Las obligaciones giran alrededor de la necesidad de garantizar los derechos fundamentales, asegurar que no se produce ningún daño a las personas ni una quiebra en nuestras sociedades democráticas. Gracias a los esfuerzos del Parlamento Europeo, también se incluyeron la protección de la democracia, el Estado de derecho y el medioambiente como intereses generales a proteger.

De este modo, de acuerdo con el uso que se vaya a hacer de la IA, se establece una clasificación basada en el riesgo. Se identifican prácticas prohibidas dentro de la UE —aquellas que presentan riesgos inasumibles para nuestras sociedades democráticas—, unos sistemas que se califican de alto riesgo, y unos usos que requieren obligaciones extraordinarias de transparencia. Los demás usos, que un estudio de la Comisión Europea estima en más del 80 % de los sistemas (recordemos ahora a esas voces que hablan de «sobrerregulación»), no estarían sujetos a ninguna obligación derivada de esta Ley, aunque se estimula la adopción de códigos de buenas prácticas.

3.1. Usos prohibidos

Comencemos nuestro análisis de la estructura de riesgos con las prácticas prohibidas. Es de resaltar que, gracias al trabajo del Parlamento Europeo, se ampliaron considerablemente los usos prohibidos en el proyecto de la Comisión Europea. Los usos prohibidos en la UE, definidos en el en el Artículo 5 de la Ley son:

- Los marcadores sociales que conllevan un perjuicio social. Pensemos en el famoso carné por puntos ciudadano que se dijo que China quería implementar, con castigos y premios derivados del buen o mal comportamiento como ciudadano, decidido por un Gran Hermano estatal capaz de establecer un seguimiento permanente de los más de mil millones de habitantes chinos y procesar toda esa información.
- Las IAs que realizan una manipulación subliminal o explotan vulnerabilidades de personas o grupos de personas.
- Los sistemas de reconocimiento de emociones empleados en el ámbito del trabajo y la educación.
- El uso de IA en tareas de policía predictiva (¿nos acordamos de *Minority Report*? Otra vez la ciencia ficción contándonos lo que podría suceder en un futuro no tan lejano).
- El rastreo de imágenes faciales indiscriminado desde la red.
- La prohibición de uso de sistemas de categorización biométrica que puedan deducir datos sensibles, como son las opiniones políticas, religiosas o filosóficas, pertenencia a sindicatos, en la vida u orientación sexual o la etnia.

- La vigilancia biométrica indiscriminada. En este caso, para poder utilizar una IA de identificación biométrica en tiempo real en lugares públicos, por parte de las fuerzas y cuerpos de seguridad, se exceptúan exclusivamente tres casos incluidos en la Ley:
  - » Encontrar a las víctimas de delitos de secuestro, trata o abusos sexuales.
  - » Cuando exista peligro o riesgo inminente de ataque terrorista.
  - » Para buscar a los culpables de delitos muy graves, delitos que se encuentran incluidos en un listado en el anexo a la ley. En cualquier caso, antes de su posible uso deberán contar con autorización judicial, un análisis de impacto en derechos fundamentales, y someterse a enormes garantías y salvaguardas, legales y judiciales.

Me detendré en este último punto, porque fue el último en acordarse en la mesa de negociación y escenifica el juego de equilibrios que hay que manejar en la tarea legislativa. Teníamos a un Parlamento que, tras un arduo debate interno, proponía la prohibición absoluta del uso de la IA para este propósito, frente a un Consejo que pretendía su utilización sin ninguna restricción. No se trataba de un juego de buenos y malos, ambas instituciones aducían poderosas razones de defensa de derechos fundamentales y bienes jurídicos esenciales que, como tantas veces en el Derecho, colisionan y hay que calibrar cuál tiene más valor en cada caso. Estimo que finalmente se llegó a una solución bastante equilibrada.

En febrero de 2025, curiosamente tras la puesta en vigor de esta parte de la Ley (2 de febrero, a los 6 meses de la publicación de la norma), la Oficina Europea de IA ha publicado unas guías de aplicación de las disposiciones sobre usos prohibidos, ¡que supera las cien páginas de comentarios!

3.2. Usos de alto riesgo

Se consideran de alto riesgo aquellos usos de la IA que pueden tener un impacto significativo en la seguridad, la salud o los derechos fundamentales de las personas. Esto puede producirse, bien porque se integra la IA en productos altamente regulados —como dispositivos médicos o juguetes— o bien porque se vaya a utilizar en áreas consideradas sensibles. Para un análisis de qué usos y en qué condiciones se consideran de alto riesgo, hay que acudir al artículo 6 en combinación con el Anexo III de la Ley —en el que se incluyen los usos y áreas que se califican como sensibles y susceptibles de provocar ese impacto en nuestra salud, seguridad o derechos fundamentales el listado de los productos ya regulados y sometidos a estrictos controles de calidad y seguridad—.

De este modo, se consideran de alto riesgo:

- Las herramientas de IA biométricas, usos de la IA relacionados con el empleo o en el lugar de trabajo.

– La calificación o el acceso a educación y formación.
– El uso de IA en el control de fronteras.
– El acceso a la justicia o a servicios públicos esenciales —incluidos los créditos financieros o seguros (curiosamente la introducción de estos dos últimos ámbitos en la categoría de alto riesgo, fueron uno de los aspectos más problemáticos de la negociación)—.
– También se considera de alto riesgo la IA que puede provocar un impacto en los resultados de procesos electorales.

En todos estos casos, los desarrolladores deberán identificar y declarar el alto riesgo de sus productos y asumir una serie de obligaciones con carácter previo a su comercialización y despliegue: realizar una evaluación de los riesgos y poner en marcha planes de prevención, garantizar la buena gobernanza de los datos para evitar sesgos y discriminación, asegurar información y transparencia en el funcionamiento de la IA, una continua supervisión humana, requisitos de ciberseguridad y controles técnicos exhaustivos para conseguir una IA robusta, y la obligación de inscribir la IA de alto riesgo en un Registro Europeo.

Por otra parte, cuando la IA se use en el marco de la prestación de un servicio público se deberá realizar una evaluación de impacto de derechos fundamentales. Para ello, las Administraciones o empresas que despliegan los sistemas de IA de alto riesgo deberán identificar a las personas o grupos de personas que pueden verse afectadas por sesgos o discriminación, evaluar los riesgos potenciales y poner en marcha planes concretos de prevención que eviten la quiebra en los derechos fundamentales.

3.3. La inteligencia artificial generativa. Obligaciones adicionales de transparencia para ciertos usos de IA

La irrupción y el impacto de las IA generativas puso de manifiesto la necesidad de una regulación específica para los modelos de IA de uso general. En cierta medida, es la única excepción metodológica en una norma que se centra en el uso y no en el concreto desarrollo tecnológico. Para su regulación, la Ley establece un enfoque en dos niveles:

Un primer nivel establece obligaciones de transparencia, incluyendo el respeto de los derechos de autor y la elaboración de un índice con las bases de datos utilizadas en el entrenamiento de la IA. Con estas medidas se pretende poner límite al uso indiscriminado de obras protegidas por derechos de autor en el entrenamiento de los modelos generales como Chat GPT, Dall-e o tantos otros, que han hecho saltar las alarmas en el sector cultural y creativo.

En un segundo nivel, cuyas obligaciones y características se detallan en el artículo 52 de la Ley, se integran aquellas que por su alta capacidad pre-

sentan un «riesgo sistémico» (sistemas tan potentes como los de Open AI), es decir, que tienen un impacto en el mercado de la Unión capaz de provocar repercusiones considerables debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor. Para estos modelos se prevén obligaciones adicionales de prevención y reducción de riesgos, seguridad y ciberseguridad. Y también una obligación de informar sobre el consumo energético, ya que se trata de modelos altamente *«energívoros»*. Finalmente, es obligatoria su inscripción en un Registro Europeo.

Para las obligaciones especiales de transparencia hay que acudir al Capítulo IV de la Ley y más específicamente al artículo 50:

> Las IAs diseñadas para interactuar con personas naturales deberán siempre informar a los usuarios, como un requisito adicional de transparencia.

Además, se exige una transparencia total para la producción y difusión de imágenes, voces, videos, pero también de textos, que sean creados o manipulados por IA y que pueden pasar por auténticos, los conocidos como *«deepfakes»* o «ultrafalsificaciones» en español. Los desarrolladores de aplicaciones que permitan estas falsificaciones y las plataformas que las difundan, deberán marcar las creaciones e informar que se trata de contenidos generados por IA. El objetivo es limitar la difusión de desinformación y otro tipo de abusos en la difusión de estos contenidos manipulados por IA. Sólo existe una excepción, en lo referido a textos, cuando estemos ante aquellos utilizados en el ámbito periodístico, amparados por la responsabilidad editorial.

IV. Otros contenidos destacables de la *AI act*

El texto legal también incluye nuevos derechos para las personas, como el derecho a ser informado y el derecho a recibir una explicación razonable cuando se vean afectadas por el uso de IA; si bien su ejercicio está todavía definido de forma imprecisa.

Del mismo modo, se ha velado por los derechos sociales y laborales de los trabajadores, debiéndose consultar a sus representantes antes de la implantación de la IA en el lugar de trabajo. En el ámbito laboral, la Ley prevé una excepcionalidad en cuanto a que las prevenciones contenidas en la misma pueden ser ampliadas como consecuencia de la negociación colectiva o mediante una ley estatal. Este es el único capítulo en el que una Ley pensada para ser lo más y lo menos en todos los países miembros, permite cierta flexibilidad regulatoria adicional (en favor de los derechos de los trabajadores).

Para favorecer el desarrollo de esta IA europea, se incluye en la Ley un sistema de incentivos, a través de los «espacios de pruebas controlados» *(«san-*

dboxes»). La idea es crear al menos un espacio de pruebas en cada uno de los países miembros, pero también se prevé la posibilidad de espacios compartidos. España fue el primer país en anunciar el lanzamiento de esta herramienta, incluso antes de la aprobación de la propia Ley, pero no fue hasta enero de este mismo 2025 en el que puso en marcha de forma efectiva la convocatoria para que los posibles usuarios hiciesen propuestas.

Finalmente, una tecnología tan disruptiva debe acompañarse de garantías para que la ciudadanía pueda ejercer un control democrático. La Ley en su artículo 4, incluye una obligación sobre alfabetización en materia de IA, para que las empresas garanticen que su personal tiene los conocimientos necesarios para evaluar los riesgos que pueda provocar el uso de la IA, cumplir con las obligaciones establecidas en la ley e informar a los posibles usuarios de sus servicios. Comienza a discutirse si la omisión de esta obligación por parte de desarrolladores o comercializadores de una aplicación de inteligencia artificial, o su deficiente puesta en marcha, puede ser objeto de sanción con arreglo a esta propia norma. La opinión mayoritaria se inclina hacia la respuesta positiva —por mi parte no tengo duda—, pero es un debate sin acabar de cerrar. La Oficina Europea de IA ha publicado, en febrero de 2025, las guías prácticas para el cumplimiento de esta obligación, si bien no se trata de una lista exhaustiva, sino que se irá ampliando. Éstas, aunque no son vinculantes, sí pueden arrojar luz sobre posibles interpretaciones y facilitar soluciones para su cumplimiento.

Durante la negociación de la Ley fui especialmente insistente en este punto referido a la alfabetización digital, porque, a mi juicio, la próxima brecha social se planteará entre quienes se adapten a este cambio y quienes no lo hagan, e incluso acerca del ritmo de adopción de esta tecnología. No se trata sólo de la muy necesaria adaptación del mercado laboral y el establecimiento de planes de transición socioeconómicos —este es el gran elefante en la habitación de los gobiernos de todo el mundo y, de no empezar pronto a mirar para él, corremos el riesgo de ser aplastados bajo sus patas—, o de la necesaria formación masiva de perfiles técnicos, sino que el propio ejercicio de ciudadanía será afectado —ya lo es— dramáticamente por la capacidad de comprender la transformación: entendimiento del cambio civilizatorio que vivimos, conciencia de nuestros derechos, desarrollo del sentido crítico... Por las previamente aludidas y aducidas razones competenciales, no fue posible en el marco de la negociación establecer un mandato específico a las Administraciones europeas para que desarrollen campañas masivas de formación y alfabetización digital, si bien sí que se recomienda en la parte dispositiva, no vinculante.

V. Gobernanza e implementación

Para asegurar el cumplimiento de estas obligaciones, se ha previsto en la Ley el desarrollo de códigos de buenas prácticas con metodologías concretas. A su vez, para su elaboración, se ha creado una Oficina Europea de IA

en el seno de la Comisión Europea (aunque con cierta autonomía funcional, como la disposición de presupuesto y plantillas propias). También la Oficina de IA es la encargada para la supervisión y el control de los modelos de uso general. La Oficina Europea, asimismo, prestará apoyo técnico a las autoridades nacionales de los países miembros. Son estos, los países miembros, los que tendrán que vigilar ordinariamente el cumplimiento de la Ley, aunque contarán con la supervisión de esta *«AI Office»*, que también ayudará a una implementación armonizada de la Ley dentro de todo el territorio europeo. En España, ya se ha señalado a la Agencia Española de Inteligencia Artificial (AESIA), con sede en Coruña, como autoridad supervisora nacional.

Igualmente, la Oficina Europea es la encargada de desarrollar códigos de buenas prácticas de asunción voluntaria por el mercado, y de mantener un diálogo permanente con la sociedad (expertos, empresas tecnológicas, representantes sociales, instituciones...) sobre cómo se va aplicando la norma y cómo va evolucionando la tecnología. Y es que cuando se estima que la IA no presenta riesgos se incentiva el establecimiento de códigos de conducta voluntarios. En cualquier caso, la IA europea deberá guiarse por unos principios éticos y se recomienda desarrollar códigos de conducta para favorecer una IA europea que sea más justa, segura, transparente, robusta, democrática, sostenible y socialmente responsable.

En relación a los plazos, la Ley entró en vigor el 1 de agosto de 2024, 20 días después su publicación, y tiene diferentes hitos temporales para su puesta en marcha dependiendo de la materia. La Ley será de plena aplicación veinticuatro meses después de su entrada en vigor, con la excepción de:

Las prohibiciones de prácticas y usos, que se aplicarán seis meses después de la fecha de entrada en vigor. Éstas han comenzado a desplegar efectos este mismo 2 de febrero de 2025, e inmediatamente después se ha publicado una interesante guía práctica de aplicación para usuarios, por parte de la Oficina Europea de IA. Ésta, si bien no es vinculante en sus indicaciones, sí despeja algunas dudas y simplifica el método de cumplimiento.

Los códigos de buenas prácticas deberán estar aprobados por la Oficina de IA nueve meses después de la publicación de la norma. Después de un largo proceso de diálogo, en el que se han consultado a más de 1.000 personas y colectivos interesados, se publicaron en febrero de 2025.

Las normas sobre la IA de uso general, incluida la gobernanza, entrarán en vigor a los doce meses después de la publicación.

Las obligaciones que se aplicarán a los sistemas de IA de alto riesgo que no estén específicamente enumerados en el Anexo III del AI Act, pero que estén destinados a ser utilizados como componentes de seguridad de un producto, o cuando la IA sea en sí misma un producto, y este deba someterse a una evaluación de conformidad por terceros según la legislación específica vigente de la UE, comenzarán a los treinta y seis meses tras la fecha de publicación de la Ley.

VI. Régimen sancionador

Entiendo interesante aprovechar este capítulo, unos de los que suelen generar más interés en la opinión pública, para subrayar algo que suelo recordar cada vez que se me pregunta sobre esta cuestión: la *IA Act* es una capa adicional de garantía y precaución, añadida al normal ejercicio del Derecho. Esto significa, simplemente, que el resto de la regulación no deja de aplicarse por el hecho de que se esté utilizando inteligencia artificial. Se pueden cometer delitos y tener responsabilidad personal en diferentes órdenes, independientemente de que se esté usando una piedra o la última tecnología.

Dicho esto, el Reglamento de Inteligencia Artificial establece un régimen sancionador específico que busca garantizar el cumplimiento de las normas en torno al desarrollo, comercialización y uso de sistemas de inteligencia artificial. Este régimen sancionador se detalla en el Capítulo XII. Las sanciones más severas que contempla están orientadas a garantizar el respeto por los derechos fundamentales: privacidad, no discriminación...

El Capítulo XII comienza por señalar que las infracciones administrativas serán proporcionales, disuasorias y establecidas según la gravedad de la infracción:

Multas de hasta 30 millones de euros o el 6 % del volumen de negocios anual global (la cantidad que sea mayor, pensado para las sanciones sobre grandes compañías multinacionales), por incumplir las prohibiciones absolutas establecidas en el Artículo 5 o los requisitos de los sistemas de alto riesgo.

Multas de hasta 20 millones de euros o el 4 % del volumen de negocios anual global (la cantidad que sea mayor), por incumplir las obligaciones de los distribuidores, importadores o usuarios de IA.

Multas de hasta 10 millones de euros o el 2 % del volumen de negocios anual global (la cantidad que sea mayor), por no cooperar con las autoridades competentes o por proporcionar información falsa o incompleta.

Son los Estados miembros, a través de las autoridades nacionales responsables de supervisar el cumplimiento del reglamento (lo que incluye la potestad sancionadora), los que decidirán si las sanciones administrativas se aplican además de otras sanciones civiles o penales. Cada Estado debe establecer sus propios procedimientos administrativos.

La Ley pretende contemplar toda la cadena de valor del producto en todas sus prescripciones y el régimen sancionador no es una excepción. No sólo se dirige a los desarrolladores de sistemas de IA, sino también a importadores, distribuidores y usuarios que incumplan sus obligaciones (aunque en este último caso, también en relación al principio de proporcionalidad, esté más restringida su virtualidad, considerando cuestiones como la mala fe, etc.).

VII. ¿Es Europa una isla de regulación en un mundo sin reglas?

En noviembre de 2021, una pequeña delegación de europarlamentarios, miembros del Comité de Asuntos Jurídicos del Parlamento Europeo, nos desplazamos a Washington, Estados Unidos. Acabábamos de aprobar en el Pleno del Parlamento, en el mes de octubre, la primera propuesta legislativa del mundo sobre ética aplicada a la Inteligencia Artificial; junto con otros dos informes no legislativos, que contenían interesantes reflexiones en el ámbito de las relaciones entre la inteligencia artificial, la responsabilidad civil y la propiedad intelectual. Como ponente de la propuesta sobre «Ética aplicada a la inteligencia artificial, la robótica u otras tecnologías relacionadas», tras haber trabajado en el texto y su negociación en pleno confinamiento, con una alta dosis de voluntarismo, recuerdo mi excitación por conocer los puntos de vista sobre la IA y su regulación al otro lado del Atlántico.

Era justo el momento en el que se había producido el cambio de la Administración Trump a la de Biden, y éste había creado una oficina de estudio sobre inteligencia artificial en la Casa Blanca. Tras nuestro periplo, no pude evitar sentirme decepcionado, porque aparte de algunos «*think tanks*» y miembros de la academia, mi sensación fue que el mundo de las instituciones ni había comenzado la reflexión sobre la cuestión en sí. Hasta la oficina de la Casa Blanca no reprimía su gesto de extrañeza ante nuestra pretensión de regular la IA. Es decir, los responsables de la mayor potencia del mundo, repositorio de las compañías que estaban protagonizando la revolución de la IA, ni siquiera parecían ser conscientes entonces de los riesgos de un desarrollo tecnológico y un uso sin control de la misma; la pretensión de regular era, para ellos, poco menos que una nueva extravagancia de «estos europeos que no saben hacer otra cosa que regular». Durante el tiempo que siguió, he tenido la ocasión de mantenerme en permanente contacto con instituciones y academia norteamericanas y de ver cómo progresivamente ha ido cambiando esta fotografía.

Hoy, en EEUU, existen cientos de iniciativas en materia de regulación sectorial o más general de los efectos y usos de la IA, desde el nivel local al federal (varios Estados están desarrollando leyes propias de amplio contenido); y, muy recientemente, a mediados de diciembre del pasado año, el Congreso de los EEUU aprobó el «*House Bipartisan Task Force on Artificial Intelligence Delivers Report*» («Informe del Grupo de Trabajo Bipartito sobre Inteligencia Artificial»), un extenso trabajo de análisis sobre oportunidades y riesgos de la IA en el país de Lincoln, y lo que puede ser la antesala de una regulación federal completa que supere ampliamente lo contenido en la Orden Ejecutiva que el presidente Biden aprobó a finales de 2023, y que, en su primer día de segundo mandato, el nuevo presidente Trump suprimió, junto con otros tantos decretos del Gobierno anterior.

Pero no sólo es EE. UU., el mundo también se está moviendo en el terreno regulatorio, cada vez más consciente de los retos de carácter civilizatorio y no sólo económico que implica esta revolución tecnológica. En Asia, China fue la primera en publicar un protocolo para la regulación y uso de la IA generativa, aún antes de que los legisladores europeos avanzásemos con nuestra propia propuesta y, posteriormente, lo ha hecho sobre otros aspectos. Lo más reciente es la propuesta de Ley integral realizada por un grupo de expertos. Corea del Sur o Japón, pretendidamente tierras de promisión para el desarrollo tecnológico sin cortapisas, cada vez se acercan más a un modelo tuitivo, también en relación con la protección de la propiedad intelectual. La propia Hong Kong, en el ejercicio de su autonomía, está trabajando una actualización de su normativa sobre derechos de autor para enfrentar el reto del uso masivo de datos para la ingesta en motores de IA.

En América del Sur, son varios los países inmersos en la preparación de una Ley específica para la IA. Brasil, después de un amplio debate con la sociedad organizada, acaba de aprobar en el Senado (enero 2025) su Proyecto de Ley sobre Inteligencia Artificial, con declarada aproximación humano-céntrica y estructurada en cuanto al riesgo, desplegando una melodía que nos sonará mucho a los europeos. El Congreso de Colombia trabaja en al menos tres textos sectoriales; Chile avanza con su propio texto integral; Ecuador, México o Costa Rica han anunciado que trabajarán inminentemente en una Ley de IA. Mención aparte merece la denominada «Ley Turing» argentina, una de las primeras iniciativas en el mundo y que ha sido paralizada en su tramitación tras las últimas elecciones presidenciales.

Australia ha abierto un período de presentación de propuestas legislativas para apoyar el uso seguro y responsable de la IA, y cuenta con una Estrategia Nacional sobre IA, que sitúa también a la ética como el centro de las políticas de desarrollo y protección. La información que obtenga en el curso de este proceso será utilizada por el país para reflexionar sobre las respuestas normativas y políticas adecuadas. Canadá lleva trabajando desde 2022 en su Ley de Inteligencia Artificial y Datos («AIDA», por su acrónimo) y muchas voces en el Parlamento urgen a terminar este mismo año el texto, también para la necesaria protección del ser humano.

La Unión Africana, recientemente (mediados de 2024), ha publicado unas recomendaciones sobre regulación y ha aconsejado unificar en lo posible la misma a escala continental.

Sería deseable avanzar también en acuerdos internacionales en la materia, que implicaran también al sector tecnológico privado más prominente, habida cuenta de que esta tecnología se halla mayoritariamente en sus manos y hay que contar necesariamente con su concurso para diseñar el futuro. Como señalaba Henry Kissinger en el que fue prácticamente su último artículo antes de su muerte, en el *Financial Times*, la diferencia con la energía nuclear y la consecución de tratados internacionales sobre ella, es que ésta se encontraba mayoritariamente en manos públicas.

En este terreno, el movimiento más serio ha sido la firma del «Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho», el pasado 17 de mayo de 2024. Es el primer tratado internacional legalmente vinculante, destinado a garantizar que el uso de los sistemas de inteligencia artificial *«es totalmente coherente con los derechos humanos, la democracia y el Estado de derecho»*. El convenio marco fue acordado, además de por los 46 miembros del Consejo de Europa —recordemos que Rusia está expulsada—, por la Unión Europea y por 11 Estados no miembros (Argentina, Australia, Canadá, Costa Rica, Estados Unidos de América, Israel, Japón, México, Perú, la Santa Sede y Uruguay). Es una declaración de principios éticos completa y útil, que está sujeta ahora a la ratificación y a la adopción de regulaciones coherentes con el mismo por parte de los signatarios (el convenio se halla abierto para que eventualmente se puedan sumar otros países).

Más recientemente, en el marco del Pacto para el Futuro, aprobado a finales de septiembre de 2024, las Naciones Unidas han propuesto el Pacto Mundial Digital. En éste se esbozan compromisos para garantizar que las tecnologías digitales contribuyan al desarrollo sostenible y a los derechos humanos, al tiempo que se señalan riesgos como las brechas digitales, la ciberseguridad y el uso indebido de la tecnología, al tiempo que se proponen medidas de corrección. El Pacto *«pretende reducir la brecha digital y garantizar que las tecnologías de IA se utilicen de forma responsable, fomentando la cooperación mundial tanto en las capacidades de IA como en las amenazas a la seguridad»*. El texto propone formar un Panel Científico mundial sobre IA en el seno de las Naciones Unidas con participación del sector privado, capaz de emitir análisis y recomendaciones a los Estados. Se aspira a *«iniciar una conversación internacional sobre la gobernanza de la IA en el seno de la ONU»*.

Retomo entonces la pregunta con la que titulaba este capítulo: ¿es Europa una isla regulatoria en un mar desregulado? Este lugar común, como acabamos de ver en este repaso, se compadece mal con la realidad. Son cada vez más los espacios que se incorporan a tratar de «civilizar» un desarrollo vertiginoso que plantea retos en materia de derechos fundamentales, en el propio ejercicio de ciudadanía (e incluso volitivo), y, por supuesto, también en materia de soberanía, desde la perspectiva del ejercicio del poder. Obviamente, existen matices y categorías propias en cada uno de estos ámbitos, pero existe un núcleo centrado en el abordamiento del riesgo y la defensa del ser humano que parece compartido. Y una gran necesidad de avanzar en armonizaciones regionales (en esto la Unión Europea va por delante) y globales.

VIII. Punto y seguido: estamos ante una norma viva

El acento humano-céntrico y los principios éticos han guiado la Ley Europea de IA. Su objetivo es velar por los intereses generales sin impedir el desa-

rrollo tecnológico. Puede que se requiera de mucho estudio y práctica para la implementación efectiva de algunos de los grandes objetivos recogidos en la norma: cumplimiento de obligaciones en formación e ilustración digital, sumarios de transparencia en materia de propiedad intelectual, ejercicio de algunos derechos como el de información... En muchos casos la Oficina Europea de IA (y los supervisores nacionales) está trabajando con cientos de partes interesadas en guías de buenas prácticas y modelos prácticos de aplicación de la Ley; en otros, serán los desarrollos a nivel local o incluso la práctica jurisprudencial los que irán definiendo límites.

Frente a modelos tecnocrático-economicistas o a los de control de Estado, en Europa hemos querido fomentar la confianza en el uso de las tecnologías, garantizando los derechos fundamentales y los valores democráticos como señas de identidad de la Unión Europea, marcando de este modo una dirección clara para el desarrollo de la Inteligencia Artificial en el mundo, de acuerdo con nuestros principios. Hay quien dice que la regulación se queda corta y alerta de «lagunas» y «omisiones burocráticas» en la defensa de los derechos de los/as ciudadanos/as. También hay quien afirma que se ha «sobre regulado» y ello lastrará nuestra competitividad. Incluso una desafortunada parte del, por otro lado, muy interesante «Informe Draghi», imputa a una inexistente regulación digital como causa de nuestro diferencial déficit de productividad con EE. UU. o China... ¡desde los años 90 del siglo pasado!

No parece, no obstante, esas voces y las presiones de las grandes compañías estadounidenses, que Europa se vaya a transformar de una isla regulatoria en un mar de desregulación. Como ya hemos visto, cada vez más países y regiones del mundo anuncian el desarrollo de una regulación propia y ya son algunos los tratados internacionales propuestos en esta materia. De hecho, el principal problema que puede afrontar EE. UU. en adelante, de no prosperar los diversos intentos de asentar una regulación clara en la materia —que también conllevan un riesgo adicional de fragmentación regulatoria interna—, es sustituir el imperio de la ley por el del estado de ánimo de su presidente (como hemos visto en el caso TIKTOK).

Con independencia de esto, resulta indudable que tendrán también efectos en todo el mundo las resoluciones judiciales pendientes en tribunales norteamericanos acerca de cuestiones de privacidad y propiedad intelectual, que, como primer lugar en el mundo en suscitarse, pueden influir en la interpretación que se da en otros lugares a muchos de los dilemas del uso de la IA, aun con diferencias regulatorias.

Por otra parte, hemos visto cómo muchos de aquellos que gritaban por desregular Europa, desde dentro y desde fuera, se aferraban a la regulación frente al terremoto que supuso la irrupción de *DeepSeek*, el modelo de inteligencia artificial generativa chino. Unos, los de dentro, acudiendo al Reglamento General de Protección de Datos, tan demonizado por los amigos de la «simplificación regulatoria»; otros, los de fuera, señalando también los ries-

gos para los datos personales de los usuarios, y acusando de robo industrial a los desarrolladores chinos, después de que esos mismos que acusan hoy, se hayan dedicado a amasar datos protegidos sin autorización de sus legítimos tenedores, para el entrenamiento de sus propios modelos, y encastillándose luego para no tener ni siquiera que compensar a sus propietarios.

Pero, mientras todo esto ocurre, es a Europa a quien todos los ojos del mundo miran y mirarán, buscando la piedra filosofal del equilibrio entre la protección del ser humano y el desarrollo tecnológico.

7.

LA GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL EN ESTADOS UNIDOS Y CHINA, MODELOS REGULATORIOS ALTERNATIVOS AL EUROPEO

Ana Gascón Marcén[1]

Profesora Contratada Doctora de Derecho Internacional Público
Universidad de Zaragoza

SUMARIO: I. Introducción. II. La regulación de la inteligencia artificial a nivel internacional. III. La regulación de la inteligencia artificial en Estados Unidos. IV. La regulación de la inteligencia artificial en China. V. Conclusiones.

I. Introducción

La Inteligencia Artificial (IA) es una tecnología con una fuerte presencia en nuestras vidas altamente digitalizadas que tiene además un tremendo potencial. Sus posibles aplicaciones son cuantiosas en casi todos los sectores con la capacidad de mejorar la producción, eficiencia, toma de decisiones, etc., pero también con tremendos retos para el ejercicio de nuestros derechos fundamentales[2]. Por lo tanto, su regulación es perentoria tanto para aprovechar sus ventajas como para mitigar sus riesgos.

1 Miembro del equipo de investigación del proyecto «Hacia una transición digital centrada en la persona en la Unión Europea». Esta publicación es parte del proyecto TED2021-129307A-I00, financiado por MICIU/AEI/10.13039/501100011033. También se enmarca en el proyecto «Análisis jurídico-internacional de las potenciales soluciones de la digitalización a los retos demográficos de Europa y Japón en el marco de la Sociedad 5.0 y de la Agenda digital europea» (UZ2023-CSJ-04) y el proyecto «Privacidad y carpeta de identidad digital europea» (EUDIWAPRY- PID2023-150123OB-I00).

2 Gascón Marcén, A., «Derechos humanos e inteligencia artificial», en *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*. Centro de Estudios

La Unión Europea (UE) se ha mostrado pionera en este ámbito y adoptó en 2024 el Reglamento de Inteligencia Artificial (RIA, conocido como *AI Act* en inglés)[3]. La primera norma horizontal que regula de una manera avanzada la IA para armonizar las reglas del mercado único europeo y proteger los derechos fundamentales. El RIA utiliza un enfoque basado en el riesgo, prohibiendo algunos usos concretos de la IA por su impacto en los derechos fundamentales y creando una serie de requisitos para la puesta en el mercado y desarrollo de otros sistemas considerados de alto riesgo. Esta norma ha sido ampliamente estudiada[4] y criticada[5] (por lo que no será objeto del presente estudio), pero resulta innegable que es un hito en la materia y un punto de referencia para todo el mundo.

Además, tendrá efecto extraterritorial, como muchas otras normas destinadas al ámbito digital de la UE[6], porque se aplicará no sólo dentro de la misma, sino también a los proveedores que introduzcan en el mercado o pongan en servicio sistemas de IA o que introduzcan en el mercado modelos de IA de uso general en la UE, con independencia de si dichos proveedores están establecidos o ubicados en la UE o en un tercer país; e incluso a los proveedores y responsables del despliegue de sistemas de IA que estén establecidos o ubicados en un tercer país, cuando los resultados de salida generados por el sistema de IA se utilicen en la UE (art. 2 RIA).

Políticos y Constitucionales, 2020, pp. 335-350; Presno Linera, M. Á., *Derechos fundamentales e inteligencia artificial*. Marcial Pons, Madrid, 2023.

3 Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828, DOUE L, 2024/1689, 12.7.2024.

4 Algunas de las obras más completas sobre la materia son: *El Reglamento Europeo de Inteligencia Artificial*, Barrio Andrés, M. (dir.), Tirant lo Blanch, Valencia, 2024; *Comentarios al Reglamento Europeo de Inteligencia Artificia*l, Barrio Andrés, M. (dir.), La Ley, Madrid, 2024; *Tratado sobre el Reglamento de inteligencia artificial de la Unión Europea*, Cotino Hueso, L. y Simón Castellano, P. (dir.) Aranzadi, Madrid, 2024; y Schwartmann, R., Keber, T. O., y Zenner, K., *AI Act: A practical guide,* C.F. Müller GmbH, 2024.

5 Veale, M. y Zuiderveen Borgesius, F., «Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22, n.º 4, 2021, pp. 97-112; Gascón Marcén, A., «La propuesta de Reglamento de la Comisión Europea por el que se establecen normas armonizadas en materia de inteligencia artificial», en *Retos de la sociedad digital. Regulación y fiscalidad en un contexto internacional*, Reus, Madrid, 2022, pp. 15-40; y Edwards, L., *Regulating AI in Europe: four problems and four solutions*. Ada Lovelace Institute, 2022.

6 Gascón Marcén, A., «El Reglamento General de Protección de Datos como modelo de las recientes propuestas de legislación digital europea», *Cuadernos de Derecho Transnacional,* vol. 13, no 2, 2021, pp. 209-232.

Se ha teorizado sobre un posible «efecto Bruselas» del RIA, un término acuñado por BRADFORD[7], según la cual, las empresas de todo el mundo terminan adoptando las normas europeas por el atractivo de su mercado y su poder normativo y aplicándolas fuera de ella[8]. El ejemplo paradigmático en este sentido es el Reglamento General de Protección de Datos (RGPD)[9] de la UE. ARNAL y JORGE argumentan que la UE podría volver a beneficiarse de este fenómeno en el ámbito de la IA, pero avisan de que enfoques regulatorios divergentes sobre la IA podrían dar lugar a efectos desglobalizadores, llevar a una fragmentación de la gobernanza tecnológica y a reducir el impacto tradicional del «efecto Bruselas»[10].

SMUHA defiende que en un contexto de competencia global para aprovechar las oportunidades que promete la IA, muchos países y regiones participan explícitamente en una «carrera hacia la IA». Sin embargo, la mayor visibilidad de los riesgos de la tecnología ha llevado a peticiones cada vez más fuertes para que los reguladores miren más allá de los beneficios y también aseguren una regulación adecuada para garantizar que la IA sea «digna de confianza», es decir, legal, ética y sólida. Además de minimizar los riesgos, dicha regulación podría facilitar la adopción de la IA, impulsar la seguridad jurídica y, por lo tanto, también contribuir a mejorar la posición de los países en la carrera. En consecuencia, esta autora sostiene que la «carrera hacia la IA» también genera una «carrera hacia la regulación de la IA»[11].

BRADFORD en su reciente libro titulado *Digital Empires: The Global Battle to Regulate Technology* examina los orígenes ideológicos, las implicaciones sociales y la influencia global relativa de los tres principales enfoques regulatorios de la economía digital (también aplicable a la IA) que son muy distintos: el modelo estadounidense impulsado por el mercado, el chino impulsado

7 BRADFORD, A., *The Brussels Effect: How the European Union Rules the World*. Oxford University Press, New York, 2020.

8 Aunque éste es el sentido específico de «efecto Bruselas», tal como lo definió BRADFORD, GREENLEAF ha defendido que en una interpretación más amplia pueden distinguirse cuatro tipos de «efectos Bruselas» que tienen que considerarse para ver la influencia en terceros países del RIA y que serían: su aplicación extraterritorial; la adopción corporativa de facto, ya descrita; la emulación legislativa por parte de terceros países; y la adopción en acuerdos y estándares internacionales (como el del Consejo de Europa que se explicará *infra*). GREENLEAF, G., «EU AI Act: Brussels Effect(s) or a Race to the Bottom?», *Privacy Laws & Business International Report*, vol. 190, 2024, pp. 3-6.

9 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE, DOUE L 119 de 4.5.2016, pp. 1-88.

10 ARNAL, J. y JORGE RICART, R., *Inteligencia artificial: el «efecto Bruselas», en juego*, ARI 88/2023, Real Instituto Elcano 3/10/2023.

11 SMUHA, N. A., «From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence», *Law, Innovation and Technology*, vol. 13, n.º 1, 2021, pp. 57-84.

por el Estado y el europeo impulsado por los derechos. Estos tres modelos reflejan diferentes teorías sobre la relación entre los mercados, el Estado y los derechos individuales y colectivos, y colisionan con frecuencia en el ámbito internacional. Bradford identifica el conflicto entre libertad y control, la batalla por la supremacía tecnológica entre China y Estados Unidos y la pugna entre los modelos estadounidense y europeo que le dan un peso diferente a los derechos, el mercado y la seguridad. Además, para esta autora la influencia global del tecnoliberalismo americano se está debilitando, China está exportando autoritarismo digital a través de infraestructuras y Europa busca globalizar su modelo de derechos a través de su poder regulatorio.

El libro analiza cómo los gobiernos y las empresas tecnológicas abordan estos conflictos inevitables que surgen cuando estos enfoques regulatorios contrastantes chocan en el ámbito internacional, y finalmente se pregunta qué imperio digital prevalecerá en su competencia mutua por la influencia global. Estos conflictos regulatorios tienen lugar en un momento en el que las sociedades digitales están en un punto de inflexión. La cascada de regulación digital que ahora se está elaborando en todo el mundo será crucial para dar forma a la economía y la sociedad digitales durante los próximos años y décadas.

Bradford no es la única que ha fijado estos tres ejes, sino que es una categorización que utilizan otros autores como Gao que al examinar las disposiciones de los acuerdos comerciales que abordan cuestiones de soberanía de los datos, como las restricciones al flujo de datos (filtrado y censura de Internet) y los requisitos de localización de datos, incluidos los requisitos para utilizar determinadas tecnologías, reconoce un enfrentamiento entre tres reinos tecnológicos que son Estados Unidos, China y la UE, que optan por defender la soberanía de la empresa, el Estado y el individuo, respectivamente[12].

Según el *AI Index Annual Report* de Standford, estas tres áreas son las principales fuentes de modelos de IA de primer nivel del mundo, especialmente Estados Unidos[13], si bien, por ejemplo, China lidera en número de patentes a nivel internacional. Además, son las que más invierten en IA, destacando especialmente la inversión privada estadounidense.

Lee-Makiyama *et al*. consideran que solo la UE ha optado por legislar mediante leyes vinculantes debido a sus deficiencias estructurales únicas[14]. Explican que la UE carece de una constitución supranacional que salvaguarde los derechos humanos y proteja a los ciudadanos contra la vigilancia o el con-

12 Gao, H. S., «Data sovereignty and trade agreements: Three digital kingdoms», en *Data sovereignty: From the digital Silk Road to the return of the state*, Oxford University Press, New York, 2024, pp. 213-239.

13 En 2023, 61 modelos de IA notables surgieron de instituciones con sede en Estados Unidos, lo que supera con creces a los 21 de la UE y los 15 de China.

14 Lee-Makiyama, H. *et al., AI and India's National Interest,* ECIPE, 2024.

trol policial basados en la IA por parte de sus Estados miembros y por eso debe promulgar normas vinculantes para adelantarse a las leyes de IA de los gobiernos nacionales que, de lo contrario, fragmentarían su mercado único. Estos autores, estudiando concretamente el caso de la India (otro país que puede ser relevante en el sector), argumentan que ésta se ha inspirado anteriormente en las leyes de la UE o de EE. UU., pero que en esta materia debe seguir su propio camino y perseguir sus intereses nacionales basándose en su perfil industrial impulsado por los servicios y la competencia bidireccional con China y los EE. UU., que han optado por no imponer ningún impedimento regulatorio al uso o desarrollo de la IA. Otros países probablemente no contarán con esa oportunidad.

Idealmente, una cuestión como la IA debería regularse a nivel internacional[15], por eso este trabajo comienza mostrando los principales avances en este plano (apartado 2). No obstante, actualmente no hay un tratado internacional global sobre IA, por lo tanto, es relevante conocer más allá del RIA, sobre el que tanto se ha escrito, lo que están haciendo Estados Unidos (apartado 3) y China (apartado 4), tanto porque muchos de los modelos y sistemas de IA se están desarrollando en su territorio como por la influencia que sus modelos regulatorios pueden tener más allá de sus fronteras.

Antes de presentar las diferentes iniciativas regulatorias hay que destacar que la gobernanza de la IA es mucho más amplia y, aunque aquí se analicen cuestiones relacionadas sobre todo con acuerdos internacionales y normas nacionales (aunque con una probable aplicación extraterritorial), Veale *et al.* explican que hay otras modalidades de gobernanza con un impacto relevante como códigos éticos, gobernanza de la propia industria, contratos y licencias y estándares técnicos y, dado que el panorama global del desarrollo de la IA se caracteriza por niveles muy altos de control e influencia de la industria, la cuestión del poder privado es crucial[16].

15 Dennis *et al*. argumentan que, a medida que avanza la IA, los Estados reconocen cada vez más la necesidad de una gobernanza internacional para abordar los beneficios y desafíos compartidos. Sin embargo, la cooperación internacional es compleja y costosa, y no todas las cuestiones de IA requieren cooperación a este nivel. Estos autores analizaron nueve áreas de políticas críticas en la gobernanza de datos, computación y modelos utilizando cuatro factores que incentivan ampliamente a los Estados a internacionalizar los esfuerzos de gobernanza: externalidades transfronterizas, arbitraje regulatorio, capacidad de gobernanza desigual e interoperabilidad. Encontraron fuertes beneficios de la internacionalización en la supervisión de los proveedores de computación, la procedencia del contenido, las evaluaciones de modelos, el monitoreo de incidentes y los protocolos de gestión de riesgos. En contraste, en su opinión, los beneficios de la internacionalización son menores o mixtos en la privacidad y procedencia de los datos, la distribución de chips y la mitigación de sesgos. Dennis, C. *et al., What Should Be Internationalised in AI Governance?*, Oxford Martin AI Governance Initiative, 2024.

16 Veale, M. *et al.*, «AI and global governance: Modalities, rationales, tensions», *Annual Review of Law and Social Science*, vol. 19, n.º 1, 2023, pp. 255-275.

II. La regulación de la inteligencia artificial a nivel internacional

Son muchísimas las organizaciones internacionales que han buscado facilitar la gobernanza de la IA y posicionarse como líderes en la materia. El Consejo de la OCDE adoptó en 2019 una Recomendación con sus Principios sobre Inteligencia Artificial que se actualizó en mayo de 2024. Se trata de un texto que no es jurídicamente vinculante pero varios Estados y otras organizaciones internacionales (entre las que se encuentra la UE con el RIA o el Consejo de Europa que veremos *infra*) han convergido hacia el uso de la definición de la OCDE de un sistema de IA y su ciclo de vida en sus marcos legislativos, estrategias y orientaciones. En concreto define un «sistema de IA», como aquel basado en una máquina que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales. A esto añade que diferentes sistemas de IA pueden funcionar con distintos niveles de autonomía y capacidad de adaptación tras el despliegue. Los principios de la OCDE fueron replicados en los principios del G20 para la IA y la OCDE también ha tenido un papel muy importante en la respuesta del G7 a estas cuestiones a través del Proceso de Hiroshima[17].

Por otra parte, la UNESCO adoptó en 2021 la Recomendación sobre la ética de la inteligencia artificial que fue el primer marco normativo universal sobre ética de la IA y su valor estriba en que muestra el consenso de la comunidad internacional en la materia al ser adoptado por sus 193 Estados miembros.

No obstante, cabe destacar la labor del Consejo de Europa, que en 2024 adoptó el Convenio Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho[18]. Lo que hace especialmente relevante este acuerdo es que se trata del primer tratado internacional sobre esa materia. Hasta ahora las iniciativas en este sector habían sido de *soft-law*, como las ya explicadas. Sin embargo, este Convenio será jurídicamente vinculante para los Estados que lo ratifiquen. Cotino ha alabado la adopción del Convenio explicando que introduce la lírica de la protección de los derechos y los principios en la prosa del RIA más técnico y detallado y elogia el potencial normativo e interpretativo del Convenio junto con su valor simbólico y metajurídico[19].

---

17 Gascón Marcén, A., «La influencia de Japón en el G-20 y el G-7 como instrumento de su política exterior y liderazgo en el ámbito digital», en *Las contribuciones de Japón desde la perspectiva del siglo XXI: aspectos jurídicos y económicos en el plano internacional*, Aranzadi, Madrid, 2024, pp. 209-232.

18 Gascón Marcén, A., «Análisis del Convenio Marco sobre Inteligencia Artificial del Consejo de Europa», en *Desafíos actuales de la Inteligencia Artificial*, UOC, 2024, pp. 141-152.

19 Cotino Hueso, L., «The Council of Europe's Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law», *Rivista Interdisciplinare Sul Diritto Delle Amministrazioni Pubbliche*. n. 2/2024.

Su vocación es crear un marco regulador no sólo para sus 46 Estados miembros, sino ir más allá de las fronteras europeas. Han participado en su negociación otros once Estados: Argentina, Australia, Canadá, Costa Rica, Estados Unidos, la Santa Sede, Israel, Japón, México, Perú y Uruguay. Estos Estados pueden ratificar el tratado, pero además también podrá hacerlo cualquier otro Estado del mundo que reciba una invitación de la organización. De hecho, Estados Unidos e Israel ya lo han firmado[20].

ROBERTS *et al.* han definido el Convenio como más prometedor que iniciativas voluntarias, como las del G20 o la UNESCO, por ser jurídicamente vinculante, ya que se espera que los Estados que lo ratifiquen lo traduzcan a su legislación nacional. Sin embargo, también avisan de que la ratificación de las convenciones del Consejo de Europa ha sido históricamente lenta, lo que resulta problemático dado el rápido ritmo del desarrollo de la IA y que los Estados que no participan en la redacción de las convenciones de la organización también se han negado anteriormente a ratificarlas debido a una percepción de falta de legitimidad, y en lugar de ello han presionado por un proceso más representativo en la ONU[21].

El Supervisor Europeo de Protección de Datos también mostró su preocupación por el altísimo nivel de generalidad de las disposiciones jurídicas del Convenio, denunciando que su naturaleza en gran medida declarativa conduciría a una aplicación divergente del mismo, socavando así la seguridad jurídica y su valor añadido[22]. Esto en parte se debe a que los enfoques muy alejados de los diferentes negociadores han llevado a que el texto final sea poco ambicioso al terminar prevaleciendo en muchos casos el mínimo común denominador.

Una de las principales controversias durante la negociación fue el ámbito de aplicación del Convenio y si debía aplicarse sólo al sector público o también al privado. Básicamente, mientras que la Comisión Europea abogaba por obligaciones también para el sector privado, replicando el modelo del RIA, Estados Unidos (apoyado por Israel, Japón, Reino Unido y Canadá) era partidario de dejar al sector privado fuera. Esto se puede ver como un intento

20 Andorra, Georgia, Islandia, Israel, Noruega, Moldavia, San Marino, Reino Unido, Estados Unidos de América y la UE lo firmaron el 5 de septiembre de 2024, que es el día que se abrió a la firma. No obstante, con la nueva Administración Trump es poco probable que Estados Unidos lo ratifique, en todo caso, la mera firma supone que Estados Unidos no puede hacer nada que frustre la consecución del objeto y el fin del tratado (art. 18 de la Convención de Viena sobre el Derecho de los Tratados).

21 ROBERTS, H. *et al*. «Global AI governance: barriers and pathways forward». *International Affairs,* vol. 100, n. 3, 2024, pp. 1275-1286.

22 SUPERVISOR EUROPEO DE PROTECCIÓN DE DATOS, *Statement in view of the 10th and last Plenary Meeting of the Committee on Artificial Intelligence (CAI) of the Council of Europe drafting the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, 2024.

de continuar con la tradición liberal estadounidense de una escasa regulación de las cuestiones digitales muy protectora de sus propias empresas y también de alinear el Convenio con su propia legislación.

Finalmente, el Convenio se aplica a las actividades dentro del ciclo de vida de los sistemas de IA realizadas por autoridades públicas (o actores privados que actúen en su nombre) y cada Parte abordará los riesgos e impactos que surjan de las actividades por parte del resto de actores privados de manera conforme con el objeto y propósito del Convenio. Cada Parte especificará en una declaración al ratificar el Convenio cómo se propone implementar esta obligación, ya sea aplicando los principios y obligaciones establecidas en el Convenio a actividades de actores privados o adoptando otras medidas apropiadas.

Esto hace que se pierda gran parte de la utilidad del Convenio al crear un régimen fragmentado y mucho más débil para el sector privado. Se trata de un compromiso/cesión porque al ser muchas de las grandes compañías del sector estadounidenses era una prioridad no desincentivar la ratificación por parte de Estados Unidos (y el resto de Estados que le apoyaban).

Almada y Radu argumentan que el RIA tendrá un «efecto Bruselas» secundario que socavará la ambición de la UE de extender sus valores en la gobernanza de la IA a nivel global, porque el RIA sigue la legislación de la UE de seguridad de los productos y sus disposiciones ofrecen una protección limitada a algunos de los valores que la política de la UE pretende proteger, como los derechos fundamentales. En relación con la negociación del Convenio, consideran que la UE ha decidió resolver los posibles conflictos con el enfoque del RIA empujando al Consejo de Europa hacia converger con él[23].

En el marco de las Naciones Unidas, sus relatores de Derechos Humanos llevan mucho tiempo denunciando los riesgos de la IA[24] y numerosos organismos y agencias están colaborando en el contexto de la *AI for Good Global Summit* para lograr una IA que fomente el desarrollo sostenible liderada por la Unión Internacional de Telecomunicaciones. Sobre su gobernanza, recientemente, el órgano consultivo de alto nivel sobre IA de Naciones Unidas resaltó en su informe final que, en términos de representación, partes enteras del mundo han quedado fuera de las conversaciones internacionales sobre gobernanza de ésta, siendo que pueden ser especialmente vulnerables a algunos de los riesgos que puede provocar[25]. De siete iniciativas de IA des-

23 Almada, M. y Radu, A., «The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy», *German Law Journal*, 25, 2023, pp. 646-663.

24 Por ejemplo, aunque no sólo: Kaye, D., *Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión,* A/73/348, 2018: y Alston, P., *Informe del Relator Especial sobre la extrema pobreza y los derechos humanos*, A/74/493, 2019.

25 United Nations High-Level Advisory Body on Artificial Intelligence, *Governing AI for humanity*, 2024.

tacadas que estudiaron, solo siete Estados miembros de la ONU eran parte de todas, mientras que 118 Estados no eran parte de ninguna (principalmente del Sur Global).

Adan *et al.* denuncian que el acceso a la IA y la participación en su desarrollo y gobernanza se concentran en unos pocos países con capacidades avanzadas en materia de IA, mientras que la «mayoría mundial de la IA» (definida como la población de países que abarcan principalmente África, América Latina, el sur y el sudeste de Asia y partes de Europa del Este) está en gran medida excluida. Estas regiones, si bien son diversas, comparten desafíos comunes en el acceso a tecnologías avanzadas de IA e influencia sobre ellas[26].

Según el *AI Index Annual Report 2024* de Standford, hasta su publicación se habían adoptado 75 estrategias nacionales de IA. En 2023, se agregaron ocho nuevas estrategias de países de Medio Oriente, África y el Caribe, lo que demuestra la expansión mundial de formulación de políticas de IA. La lista de nuevos países que estaban desarrollando estrategias nacionales de IA en ese momento incluía: Antigua y Barbuda, Barbados, Bielorrusia, Costa Rica, Jamaica, Pakistán o Senegal. Pero una mirada al mapa mundial representativo mostraba prácticamente un vacío en África muy por detrás del resto de continentes[27].

Robles ya había criticado que el estudio de la práctica muestra que, en el marco internacional, la mayoría de las iniciativas están localizadas en organizaciones o foros regionales o interregionales participados por países tecnológicamente desarrollados; y el ordenamiento jurídico internacional ha de activar mecanismos de respuesta de alcance universal frente a esa situación y al desafío global que implica la gobernanza de la IA[28]. En ese sentido, hay que destacar que la Unión Africana aprobó una Estrategia Continental de IA en 2024. Esta estrategia subraya el compromiso de África con un enfoque de la IA centrado en su continente y orientado al desarrollo, que promueva prácticas éticas, responsables y equitativas. También es loable la labor de la Coalición del Foro de la Gobernanza de Internet sobre gobernanza de datos e inteligencia artificial[29].

En opinión del órgano consultivo de alto nivel, las Naciones Unidas pueden ser el vehículo para un nuevo contrato social para la IA que garantice la

---

26 Adan, S. N. *et al*. *Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance*, Oxford Martin School AI Governance White Paper, 2024.

27 Standford, *AI Index Annual Report 2024*, 2024, p. 26.

28 Carrillo Robles, M., «La gobernanza de la inteligencia artificial: contexto y parámetros generales», *Revista Electrónica de Estudios Internacionales*, n.º 39, 2020.

29 *AI from the Global Majority. Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition*. FGV Direito Rio, Rio de Janeiro, 2024.

aceptación global de un régimen de gobernanza que nos proteja y empodere a todos[30]. No obstante, esto parece difícil en la práctica y no hay avances tangibles en esa dirección. Por eso es tan importante estudiar la regulación de algunos de los países más importantes en esta materia.

III. La regulación de la inteligencia artificial en Estados Unidos

En 2024, las cinco empresas con mayor capitalización de mercado del mundo eran estadounidenses y estaban relacionadas de múltiples maneras con el entorno digital y la IA[31]. Se trataba de Apple, NVIDIA, Microsoft, Alphabet (Google) y Amazon, con Meta (Facebook) séptima, que desarrollan sus actividades a nivel mundial. Es evidente que cómo regule Estados Unidos la IA tiene un elevadísimo impacto no sólo en su territorio, sino fuera de él.

Se considera que en el mundo hay más de 160 leyes nacionales de protección de privacidad de datos[32]. No obstante, Estados Unidos a pesar de ser uno de los países más relevantes por albergar a muchas de las compañías más importantes en la materia carece de una ley general federal de protección de datos personales y sólo cuenta con normativa sectorial, por ejemplo, para proteger a los menores o leyes de Estados concretos como California[33]. Esta anomalía y la exoneración de responsabilidad de los prestadores de servicios de la información en la sección 230 de la *Communications Decency Act* estadounidense[34], entre otras, son elementos claros del mencionado régimen de tecnoliberalismo estadounidense.

Chander ha sostenido que las decisiones de los legisladores estadounidenses a finales de 1990 fueron cruciales para el éxito de Silicon Valley, ya que redujeron los riesgos legales a los que se enfrentaban las nuevas empresas de Internet, al tiempo que se abstuvieron en gran medida de regular los nuevos riesgos que introducían[35]. Según este autor, los jueces y legisladores estadounidenses modificaron la ley a principios del milenio para promover el desarrollo de las empresas de Internet. Europa y Asia, en cambio, impusieron

30 United Nations High-level Advisory Body on Artificial Intelligence, *op. cit*.

31 A 9 de diciembre de 2024.

32 Greenleaf, G., «Global Data Privacy Laws 2023: 162 National Laws and 20 Bills», *Privacy Laws and Business International Report,* vol. 181, 2023.

33 Barrio Andrés, M., «La regulación del derecho a la protección de datos en los Estados Unidos», *Cuadernos de Derecho Transnacional*, 2022, vol. 14, no 2, pp. 186-193.

34 Castelló Pastor, J. J., «Exoneración de responsabilidad de los prestadores de servicios de la información en la sección 230 de la *Communications Decency Act* estadounidense», *Revista Aranzadi de derecho y nuevas tecnologías*, n.º 39, 2015, pp. 175-194.

35 Chander, A., «How law made Silicon Valley», *Emory Law Journal,* 2013, vol. 63, pp. 639-694.

regímenes de responsabilidad de los intermediarios, normas sobre propiedad intelectual y salvaguardias para la privacidad, lo que impidió el florecimiento de empresas locales de Internet más exitosas. Las empresas estadounidenses aprovecharon su base local liberal para convertirse en líderes globales en el ciberespacio.

Como se ve, es un enfoque totalmente diferente al del RIA. De hecho, la Representante para el Comercio de los Estados Unidos, en su Informe sobre Barreras al Comercio Exterior de 2024, que proporciona una revisión integral de las barreras extranjeras significativas a las exportaciones estadounidenses de bienes y servicios argumentó que el proyecto del RIA propuesto por la Comisión Europea en 2021 no brindaba esquemas de protección sólidos, consistentes y transparentes para la información innovadora y exclusiva incorporada en el código fuente cuando se requiere que este sea revelado a las autoridades[36].

Por su parte, Bradford ha cuestionado la opinión común de que una regulación más estricta de la economía digital inevitablemente compromete la innovación y socava el progreso tecnológico[37]. Según ella, esta opinión, defendida vigorosamente por la industria tecnológica, ha dado forma al discurso público en los Estados Unidos, donde la próspera economía tecnológica del país suele asociarse con un firme compromiso con los mercados libres. Los legisladores estadounidenses también han adoptado tradicionalmente esta perspectiva, lo que explica su vacilación a la hora de regular la industria tecnológica hasta la fecha. Según los críticos de la UE, su regulación tecnológica de gran alcance ha tenido como consecuencia la falta de innovación, lo que explica la incapacidad de la UE para fomentar las empresas tecnológicas y competir con Estados Unidos y China en la carrera tecnológica. Sin embargo, Bradford sostiene que la relación entre la regulación digital y el progreso tecnológico es considerablemente más compleja que lo que la conversación pública, los legisladores estadounidenses, las empresas tecnológicas y varios académicos han sugerido hasta la fecha. Por esta razón, la brecha tecnológica existente entre Estados Unidos y la UE no debe atribuirse a la laxitud de las leyes estadounidenses y la rigurosidad de la regulación digital europea[38].

---

36 United States Trade Representative, *2024 National Trade Estimate Report on Foreign Trade Barriers*, 2024.

37 Bradford, A., «The False Choice Between Digital Regulation and Innovation», *Northwestern University Law Review*, 2024, vol. 118, n.º 2, pp. 377-453.

38 En particular, Bradford expone como causas las diferentes culturas y los idiomas de Europa que suponen que no se pueden escalar operaciones tan fácilmente como en Estados Unidos, que a su vez tiene mejores mercados de capital para invertir en innovación (que no están fragmentados a diferencia del europeo), las normas concursales europeas que disuaden de la toma de riesgos y las estrictas leyes de inmigración de Europa que perjudican su acceso al talento.

En todo caso, STREINZ muestra como el «consenso de Silicon Valley» que es el conjunto de prescripciones de políticas para el desarrollo económico basado sobre la creencia de que la mejor manera de desarrollar las economías digitales es preservando los flujos libres de datos sin restringir los negocios digitales a través de la privacidad u otras formas de regulación, no sólo es prevalente en Estados Unidos, sino que este ha llegado a incluirlo en varios tratados comerciales multilaterales[39].

Por todo esto no es de extrañar que el primer acercamiento de la Administración Biden a la materia fuera reunirse con los representantes de algunas de las principales empresas de IA (Google, Microsoft, Meta, Amazon, OpenAI, Anthropic e Inflection) y obtener una serie de compromisos voluntarios para promover el desarrollo de una IA segura y confiable. Estos compromisos implicaban realizar evaluaciones de seguridad internas y externas de los sistemas de IA antes del lanzamiento, compartir información sobre los riesgos identificados, permitir la denuncia pública de problemas y revelar cuándo el contenido es generado por IA[40].

No obstante, se consideró que eso no era realmente suficiente y el presidente Biden adoptó la *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* el 30 de septiembre de 2023, que establecía nuevos parámetros de referencia para la seguridad de la IA, el avance de la equidad y los derechos civiles, y el fomento de la competencia y la innovación. A diferencia del RIA, no es una norma horizontal, sino sectorial, si bien tocaba muchos ámbitos, desde los derechos de los trabajadores, a la educación o la seguridad nacional. Ordenaba a un elevado número de agencias que crearan estándares y reglas para el desarrollo, uso y supervisión de la IA y también se encargaron informes, evaluaciones de impacto, desarrollo de pautas y mejores prácticas. Buscaba crear un mercado de IA abierto y competitivo que enfatice la innovación estadounidense y aprovechar el poder que tienen las agencias como mayores contratistas del país a través de las compras y contratos públicos. Alentaba a la Comisión Federal de Comunicaciones (FTC) a evaluar el impacto de la IA en las telecomunicaciones y el Instituto Nacional de Estándares y Tecnología (NIST) tenía instrucciones de formular pautas y mejores prácticas para respaldar el consenso de la industria sobre el desarrollo y la implementación de una IA segura, confia-

39 STREINZ, T., «Digital megaregulation uncontested? TPP's model for the global digital economy» en KINGSBURY, B. *et al* (eds), *Megaregulation Contested: Global Economic Ordering After TPP,* Oxford University Press, New York, 2019, pp. 312-342.

40 WHITE HOUSE, «FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI», 21/07/2023. [en línea], (2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [Consulta: 17/12/2024].

ble y ética. En el marco de este último, se creó el Instituto para la Seguridad de la Inteligencia Artificial de Estados Unidos (AISI).

La Orden Ejecutiva ordenó a las agencias federales abordar varias áreas fundamentales: gestionar modelos de IA de doble uso, implementar protocolos de prueba rigurosos para sistemas de IA de alto riesgo, hacer cumplir las medidas de rendición de cuentas, salvaguardar los derechos civiles y promover la transparencia en todo el ciclo de vida de la IA. Estas iniciativas estaban diseñadas para mitigar los posibles riesgos de seguridad y defender los valores democráticos, al tiempo que fomentaban la confianza pública en el campo de la IA.

En su balance después de un año de la entrada en vigor de la orden, la Administración Biden se felicitó al constatar que las agencias federales habían completado según lo previsto cada acción que la Orden Ejecutiva encomendado para ese año, más de cien en total[41]. Entre estos se puede mencionar la adopción del *National Security Memorandum on Artificial Intelligence* o el *Advancing the Responsible Acquisition of Artificial Intelligence in Government Memorandum*.

En lo referente al Congreso estadounidense existe un gran interés por la materia. En sólo un año el número de proyectos de ley relacionados con la IA aumentó más del doble, pasando de 88 en 2022 a 181 en 2023, no obstante, lo normal en el Congreso americano es que se propongan muchas leyes, pero casi todas fracasen, como fue aquí el caso pues finalmente sólo una fue adoptada[42]. El grupo de trabajo bipartidista sobre Inteligencia Artificial de la Cámara de Representantes de Estados Unidos adoptó en diciembre de 2024 un Informe que reflexionaba sobre el mejor enfoque regulatorio para la IA que equilibrara la necesidad de promover la innovación y liderazgo americano y al mismo tiempo protegiera a los estadounidenses de posibles daños y que recomendó una serie de principios para las políticas y sobre todo una aproximación sectorial e incremental[43].

En cualquier caso, con la nueva Administración Trump existen muchas dudas sobre cómo avanzará la cuestión, pero, teniendo en cuenta que Trump ha sido apoyado por varios empresarios tecnológicos, entre los que destaca

41 White House, «Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order», 30/11/2024 [en línea], (2024), https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/30/fact-sheet-key-ai-accomplishments-in-the-year-since-the-biden-harris-administrations-landmark-executive-order/ y también KLEIN, A. et al., «1 year later, how has the White House AI Executive Order delivered on its promises?» Brookings, 4/11/2024 [en línea], (2024), https://www.brookings.edu/articles/one-year-later-how-has-the-white-house-ai-executive-order-delivered-on-its-promises/ [Consulta ambos: 17/12/2024].

42 Standford, *AI Index Annual Report 2024*, 2024, p. 17.

43 Bipartisan Artificial Intelligence Task Force, *Report on Artificial Intelligence*, US House of Representatives, 2024.

Elon Musk, puede predecirse que es muy probable que impere la desregulación en el sector digital, incluida la IA. Trump derogó la Orden ejecutiva de Biden pocas horas después de jurar su cargo, aunque dado que gran parte del trabajo se había realizado ya por y dentro de las agencias federales, algunos de sus principios pueden sobrevivir a su derogación. También es probable que la nueva administración relaje la regulación de las agencias, se concentre más en la competencia con China (con mayores limitaciones a la exportación) y reduzca la aplicación de leyes antimonopolio relacionadas con la IA, entre otras posibilidades[44]. Esto se ha visto confirmado con el anuncio del nombramiento de Andrew Ferguson para dirigir la FTC en sustitución de Lina Khan y de la creación de un puesto *ad hoc* de «zar de la inteligencia artificial y las criptomonedas» para David Sacks. Ambos tienen como mandato «fomentar el liderazgo y la innovación en estas áreas y luchar contra la censura».

No obstante, igual que ocurre con la regulación de la protección de datos, es a nivel de los estados (o incluso local) donde se está compensando la falta de normas a nivel federal, al menos en algunos casos. Entre 2016 y 2023, California fue el Estado que más leyes promulgó relacionadas con la IA, con un total de 13, seguida de Maryland (10) y Washington (7), mientras que otros muchos Estados no promulgaron ninguna[45], por lo que el panorama estadounidense está bastante fragmentado en la materia. En 2024, el congreso de California aprobó la *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act* (SB 1047) que buscaba mejorar la seguridad de algunos de los grandes sistemas de IA frente a riesgos existenciales, pero Gavin Newsom, Gobernador de California, bloqueó su adopción por considerar que podía obstaculizar la innovación, lo cual representa un buen ejemplo de lo controvertido de la acción legislativa en este país[46].

También existen algunas normas relevantes a nivel local, por ejemplo, la Ley 144 de 2021 de Nueva York sobre herramientas automatizadas de toma de decisiones laborales prohíbe a los empleadores y agencias de empleo utilizar una herramienta automatizada de toma de decisiones laborales a menos que la herramienta haya sido objeto de una auditoría de sesgo dentro del año posterior a su puesta en funcionamiento, la información sobre la auditoría esté disponible públicamente y se hayan proporcionado ciertos avisos a los empleados o candidatos a un puesto de trabajo[47].

---

44 Villasenor, J. y Turner, J., «AI policy directions in the new Trump administration», Brookings, 14/11/2024. [en línea], (2024), https://www.brookings.edu/articles/ai-policy-directions-in-the-new-trump-administration/ [Consulta: 17/12/2024].

45 Standford, *AI Index Annual Report 2024*, 2024, p. 18.

46 Lee-Makiyama, H. y Lozano, C., *Empires of Exceptionalism: Lessons from the EU AI Act and Attempts at AI Legislation in California*, ECIPE, 2024.

47 Si bien, según GROVES, esta ley fracasó rotundamente en la creación de un sólido ecosistema de auditoría de terceros que mejorara los resultados de equidad en la contratación. Algunos de los desafíos estaban relacionados con la ley en sí, incluida

No obstante, la nueva administración de Trump podría introducir una legislación federal que bloquee el desarrollo de leyes estatales más estrictas con la excusa de salvaguardar la competitividad de la IA estadounidense[48]. En el citado Informe del grupo de trabajo bipartidista sobre Inteligencia Artificial de la Cámara de Representantes se consideraba una ley federal de esta naturaleza como una herramienta que el Congreso podría utilizar para lograr diversos objetivos, sin embargo, se reconocía que esto plantearía cuestiones jurídicas y de política complejas que deberían tenerse en cuenta.

Según Tréhu y Jorge, la UE y los Estados Unidos deberían establecer una colaboración internacional concreta y pragmática en materia de IA. En su opinión, esto se puede lograr a pesar de las diferencias entre sus respectivos enfoques y las incertidumbres políticas en torno a la nueva administración de los Estados Unidos y el realineamiento y, además, es necesario porque, sin esos esfuerzos, los socios de terceros países pueden verse incentivados a adoptar normas o implementar la IA de maneras que socaven el impulso global a favor de un modelo de gobernanza que respete los derechos y, en especial, si quieren competir con China en las asociaciones con terceros países[49].

IV. La regulación de la inteligencia artificial en China

China es, después de los Estados Unidos, otro de los países que alberga a algunos de los mayores gigantes tecnológicos. A diferencia de los Estados Unidos y de lo que mucha gente piensa, China cuenta con una Ley de protección de la información personal, que curiosamente tiene algunos elementos comunes con el RGPD de la UE[50]. Sin embargo, en su sistema de manera general se privilegian los derechos e intereses colectivos como la seguridad y el orden sobre las libertades individuales, lo que los ha llevado a crear un sistema de crédito social alimentado por una ingente cantidad de datos de

una «teoría del cambio» defectuosa que no requería que las empresas dejaran de usar herramientas sesgadas, sino solo que pusieran los resultados de la auditoría a disposición de los candidatos. La ley también adolece de una definición estrecha y manipulable de qué sistemas estaban dentro del alcance y una falta de mecanismos significativos para su aplicación por parte de los reguladores. Otras razones para el aparente fracaso de la ley en algunas áreas clave se debían a la compleja dinámica en torno a la auditoría de algoritmos. Estos incluían desafíos culturales y prácticos que enfrentaban los auditores para obtener los datos necesarios para realizar auditorías, y una falta de estándares claros sobre qué roles y prácticas deberían adoptar los auditores. GROVES, L., *Code & conduct. How to create third-party auditing regimes for AI systems*. Ada Lovelace Institute, 2024.

48 Villasenor, J. y Turner, J., *op. cit.*

49 Tréhu, J. y Jorge Ricart, R., *Global AI Governance: Key Steps for Transatlantic Cooperation*, German Marshall Fund of the United States, 2024.

50 Ke, X. *et al*., «Analyzing China's PIPL and how it compares to the EU's GDPR», *IAPP,* 24/08/2021. [en línea], (2021), https://iapp.org/news/a/analyzing-chinas-pipl-and-how-it-compares-to-the-eus-gdpr [Consulta: 17/12/2024].

sus ciudadanos y una red de cámaras y sensores por todo el país que hacen que prácticamente sea imposible escapar a los mismos. Según GREENLEAF, su Sistema de Crédito Social ha emergido como el sistema de vigilancia más generalizado y potencialmente totalitario del mundo y entraña grandes riesgos para los derechos humanos[51]. Esto también se ha visto reflejado en un férreo control de Internet para controlar la información y a la población con un sistema de censura general.

De hecho, en la UE el RIA prohíbe los sistemas de IA para evaluar o clasificar a personas físicas atendiendo a su comportamiento social o a características personales o de su personalidad conocidas, inferidas o predichas de forma que la puntuación ciudadana resultante provoque un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente, o un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos de personas que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este, por su riesgo para los derechos fundamentales[52].

Según ROLF, la regulación digital de China responde a un doble imperativo: un esfuerzo tecnonacionalista que busca fomentar la innovación y que las empresas chinas avancen hacia la vanguardia de las tecnologías digitales, al tiempo que se preserva la estabilidad social durante un período de agitación sociotécnica[53]. ZENG explica que las prácticas del gobierno chino son particularmente interesantes dada la enorme inversión del país en IA combinada con sus objetivos autoritarios[54]. Su necesidad primordial de mantener (si no fortalecer) su régimen autoritario ha proporcionado un modo chino único de gobernanza de la IA. A falta de controles y contrapesos por parte de un poder legislativo fuerte, el enorme presupuesto de seguridad interna de China permite a su aparato estatal invertir en costosas tecnologías de vanguardia para desarrollar sus fuerzas de seguridad. En su intento de mantener su régimen autoritario, el Estado tiene fuertes incentivos para fortalecer su control sobre la sociedad, y la tecnología digital es una de sus herramientas.

SHEEHAN ha destacado que, en Occidente, no se suele prestar atención a la regulación china de la IA por ser considerada irrelevante o vista pura-

51 GREENLEAF, G., «Asia's Data Privacy Dilemmas 2014-19: National Divergences, Cross Border Gridlock». *Revista Uruguaya de Protección de Datos Personales*, vol. 4, 2019, pp. 49-73.

52 GARCÍA MAJADO, P., «Los sistemas de puntuación social en el Reglamento Europeo de Inteligencia Artificial», *RETINA-DER*, 28/11/2024 [en línea], (2024), < https://retina-der.uniovi.es/index.php/2024/11/28/los-sistemas-de-puntuacion-social-en-el-reglamento-europeo-de-inteligencia-artificial-2/ [Consulta: 17/12/2024].

53 ROLF, S., *China's regulations on algorithms, Context, impact, and comparisons with the EU,* Friedrich-Ebert-Stiftung, 2023.

54 ZENG, J., «Artificial intelligence and China's authoritarian governance», *International Affairs,* vol. 96, n.º 6, 2020, pp. 1441-1459.

mente a través de la lente de una competencia geopolítica para escribir las reglas de la IA[55]. Es cierto que la creciente huella digital de China y su apoyo a las empresas tecnológicas nacionales plantean numerosos desafíos económicos y de seguridad para la UE y los EE. UU., como potencias digitales tradicionalmente dominantes en múltiples sectores (no sólo el de la IA)[56], no obstante, SHEEHAN defiende que estas regulaciones merecen un estudio cuidadoso para ver cómo afectarán la trayectoria de la IA en China y qué pueden presentar lecciones interesantes para los responsables de las políticas de todo el mundo sobre la regulación de la tecnología. Incluso si los países discrepan fundamentalmente sobre el contenido específico de una regulación, aún pueden aprender unos de otros en lo que respecta a las estructuras subyacentes y la viabilidad técnica de diferentes enfoques regulatorios. Incluso si no se comparten muchos de sus planteamientos, no se puede negar el liderazgo chino tanto en la parte técnica como en el desarrollo de sistemas de IA, como en la reguladora, siendo los pioneros en varios campos.

Por ahora China carece de una ley horizontal de IA, aunque esto podría cambiar pronto, porque parece haber interés por una norma más amplia[57]. No obstante, hasta ahora ha seguido un acercamiento sectorial regulando aquellos aspectos que le parecían más urgentes.

GONG considera que China ha seguido tres etapas en lo referente a la gobernanza de la IA que serían la primera de planificación estratégica y autodisciplina de la industria, la segunda de estándares voluntarios y comienzo de la supervisión regulatoria y la última en la que se encontraría ahora de supervisión directa de las tecnologías de IA[58]. En esta última, se han materializado las tres normas más relevantes y específicas sobre algoritmos e IA que son su regulación de 2021 sobre algoritmos de recomendación, las reglas de 2022 para síntesis profunda (contenido generado sintéticamente) y las reglas de 2023 sobre IA generativa. Como se puede ver, todas adoptadas antes de que la UE consiguiera llegar a un acuerdo sobre el RIA.

---

55 SHEEHAN, M., *China's AI Regulations and How They Get Made*, Working Paper, Carnegie Endowment for International Peace. 2023.

56 BROWN, S. A. W., «Beyond the great firewall: EU and US responses to the China challenge in the global digital economy», *Journal of European Integration*, vol. 46, n.º 7, 2024, pp. 1089-1110.

57 Véase *Artificial Intelligence Law of the People's Republic of China (Draft for Suggestions from Scholars),* Center for Security and Emerging Technology 2024 y Plan de trabajo legislativo del Consejo de Estado para 2023 (国务院办公厅关于印发 国务院2023年度立法工作计划的通知), [en línea], (2023), https://www.gov.cn/zhengce/content/202306/content_6884925.htm[Consulta: 17/12/2024].

58 GONG, J. *et al*, «AI Governance in China: Strategies, Initiatives, and Key Considerations», *Bird & Bird*, 14/03/2024 [en línea], (2024), https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations [Consulta: 17/12/2024].

Esta respuesta tan rápida a las novedades tecnológicas se debe a que China utiliza un modelo de gobernanza adaptativa en su política digital[59], que se ha caracterizado por la implementación de sucesivas directivas o directrices de carácter administrativo, emitidas por diferentes reguladores, lo que ha llevado posteriormente a la consolidación de ciertas normas mediante las primeras medidas legislativas integrales en materia digital, como la Ley de Comercio Electrónico y la Ley de Ciberseguridad, mientras que otras áreas aún se rigen por un «mosaico» de directrices administrativas, como es el caso de la IA[60]. Esta técnica da forma a lo que se ha denominado un modelo «iterativo» de gobernanza[61], en el que las directrices administrativas se suceden, se reemplazan y se superponen entre sí para responder a los problemas a medida que surgen. Este enfoque muestra al Estado chino no como una entidad monolítica, sino más bien como un sistema complejo de reguladores que compiten entre sí para mantener y expandir sus prerrogativas y control sobre aspectos de la formulación de políticas[62]. Esto permite una respuesta rápida y tener una especie de laboratorio de pruebas legislativo para ver lo que funciona y lo que no, si bien también plantea retos para las empresas tecnológicas, además de cuestiones de legitimidad democrática. La actitud de las grandes empresas tecnológicas chinas hacia las políticas y prioridades gubernamentales se ha descrito como, por un lado, complaciente, pero, por el otro, activa en la búsqueda de margen de maniobra cuando lo permiten estas reglas superpuestas[63].

Centrándonos en el contenido de las normas específicas sobre IA, en diciembre de 2021, la Administración de Ciberseguridad de China (conocida como CAC), que es el organismo de control de Internet de China, aprobó el Reglamento de Gestión de Algoritmos de Recomendación de Servicios de Información de Internet.

El Reglamento incluye normas interesantes para proteger a las personas mayores y a los niños, así como los derechos de los trabajadores cuando los algoritmos sean usados para la asignación de pedidos en una plataforma, la remuneración y el pago, las horas de trabajo, las recompensas y las sanciones, y garantizar los derechos e intereses legales de los trabajadores, como la remuneración, el tiempo de descanso y las vacaciones. También se prohíbe imponer un trato diferenciado a los consumidores en la venta de productos y servicios, como, por ejemplo, mediante la discriminación de precios basada

59 Zhang, A. H., «Agility Over Stability: China's Great Reversal in Regulating the Platform Economy», *Harvard International Law Journal,* vol. 63, 2022, pp. 457-514.

60 Migliorini, S., «China's Interim Measures on generative AI: Origin, content and significance», *Computer Law & Security Review*, vol. 53, 2024.

61 Sheehan, M., *op. cit.*

62 Migliorini, S., *op. cit.*

63 Zhang, A. H., *op. cit.*

en las preferencias y hábitos de transacción de los consumidores. Además, obliga a los proveedores a ofrecer a los usuarios opciones que no se basen en sus perfiles u ofrecerles una opción conveniente para desactivar los servicios de recomendación de algoritmos. Todo esto responde a lógicas parecidas a las perseguidas en muchos casos también por la normativa europea.

No obstante, la normativa china se diferencia sustancialmente en que los requisitos de moderación de contenidos estipulados en la norma responsabilizan a las plataformas de cualquier contenido ilegal o indeseable que se recomiende a los usuarios a través de algoritmos. Se trata de una extensión de estipulaciones similares de la Ley de Ciberseguridad de China que responsabilizan a las plataformas por alojar contenido ilegal o indeseable. También se exige a los operadores que actualicen y diseñen sus algoritmos para que se ajusten a los «valores dominantes» e incluso que promuevan «contenido positivo». Estas estipulaciones amplían el alcance de las responsabilidades de los operadores de algoritmos, que van desde la moderación y la prevención de la aparición de contenido ilegal en sus plataformas hasta la promoción activa de contenido «positivo» que siga la línea del Partido. Esto consiste principalmente en contenido patriótico, apto para toda la familia y centrado en historias positivas en línea con los «valores socialistas fundamentales» del Partido, mientras que se abstienen de contenido que promueva comportamientos indeseables (extravagancia y consumo excesivo, comportamiento violento o antisocial, promiscuidad sexual, adoración excesiva de ídolos famosos y otras figuras públicas y activismo político, por nombrar algunos)[64]. Las regulaciones también indican un cambio marcado en la forma en que el gobierno ve el papel de las grandes empresas y plataformas tecnológicas en la sociedad. Ya no será suficiente simplemente cumplir con las regulaciones y prevenir el comportamiento delictivo; las empresas también deben hacer un esfuerzo concertado para garantizar un entorno en línea seguro e ideológicamente saludable para todos los usuarios[65].

En 2022, la CAC, el Ministerio de Industria y Tecnología de la Información y el Ministerio de Seguridad Pública emitieron conjuntamente las Disposiciones sobre la administración de la síntesis profunda de los servicios de información basados en Internet (las Disposiciones de síntesis profunda) que entraron en vigor en 2023 y obligaban a añadir etiquetas a la información generada mediante el uso de tecnologías de síntesis profunda, como la simulación de voz, la conversación inteligente o la escritura que simula el estilo de una persona real, la síntesis de imágenes faciales o la manipulación

64 Huld, A., «China's Sweeping Recommendation Algorithm Regulations in Effect from March 1», *China Briefing*, 6/01/2022. [en línea], (2022), https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022/ [Consulta: 17/12/2024].

65 *Ibidem.*

de rostros[66]. Además, requerían la creación de un mecanismo para disipar las noticias falsas, de modo que cuando se utilicen servicios de síntesis profunda para producir, copiar, publicar y difundir información falsa, los proveedores de servicios de síntesis profunda estén obligados a tomar medidas para eliminar dichas noticias, mantener registros y denunciarlas a las autoridades pertinentes[67].

El julio de 2023, la CAC, junto con otras seis agencias gubernamentales de China, emitieron conjuntamente las Medidas provisionales para la administración de servicios de inteligencia artificial generativa, que entraron en vigor en agosto de ese mismo año[68] y se fundamentan en las medidas anteriores. Para Migliorini, estas Medidas contienen una serie de disposiciones dignas de mención y deberían ser consideradas por los legisladores y las partes interesadas a nivel mundial como un punto de referencia comparativo interesante para las reflexiones sobre la regulación de la IA generativa dentro de sus propios sistemas legales[69].

Los proveedores de servicios de inteligencia artificial generativa deben garantizar que todo el contenido creado y los servicios prestados cumplan con los requisitos establecidos en los siguientes principios: defender los valores socialistas fundamentales[70] y no generar contenido que pueda perjudicar la seguridad nacional o que esté prohibido por leyes o reglamentos, adoptar medidas efectivas para evitar que el contenido sea discriminatorio por cualquier motivo, respetar los derechos de propiedad intelectual y la ética comercial; respetar los derechos legítimos de los demás y no infringir su bienestar e intereses; y adoptar medidas estrictas y efectivas para aumentar

66 Esto es algo similar a lo que hace el RIA, que exige a los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación que hagan público que estos contenidos o imágenes han sido generados o manipulados de manera artificial.

67 Interesse, G., «China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023», *China Briefing,* 20/12/2022. [en línea], (2022), https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/ [Consulta: 17/12/2024].

68 A la UE la explosión de la IA generativa le pilló en plenas negociaciones del RIA, por lo que hubo que crear un apartado nuevo que no estaba en la propuesta de la Comisión y que no sigue exactamente la lógica del resto de la norma.

69 Migliorini, S., *op. cit.*

70 Esto puede resultar chocante desde la perspectiva europea, pero la UE también ha hecho del respeto de sus valores un requisito del RIA que busca la adopción de una IA centrada en el ser humano y fiable, garantizando un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, y proteger frente a los efectos perjudiciales de los sistemas de IA en la UE. Así podemos ver que tanto China como la UE enfatizan en el desarrollo de la IA sus valores, si bien estos son bastante diferentes.

la transparencia, precisión y confiabilidad de sus servicios[71], para lo que se requiere emplear medidas efectivas para mejorar la calidad de los datos de entrenamiento, su veracidad, precisión, objetividad y diversidad. Esto último se modificó respecto al borrador de las medidas publicado para consulta pública que exigía asegurar la veracidad de los datos, lo que se consideró prácticamente imposible.

Los proveedores de servicios no deben recopilar información personal innecesaria, almacenar información de entrada o registros de uso de una manera que pueda identificar los detalles personales de los usuarios, ni proporcionar dicha información o registros a terceros. Los proveedores de servicios deben establecer un mecanismo para recibir y gestionar las quejas de los usuarios. Además, también deben responder a cualquier solicitud de los usuarios para revisar, copiar, corregir, complementar y eliminar su información personal de manera oportuna[72].

Si algún servicio posee atributos de opinión pública o tiene capacidad de «movilización social», el proveedor de servicios debe realizar una evaluación de seguridad y cumplir con los requisitos relacionados con la presentación, modificación o cancelación establecidos. La evaluación de seguridad puede ser realizada por el proveedor de servicios o a través de un tercero, y debe presentarse ante la CAC y la Oficina de Seguridad Pública de China.

Gong *et al.* argumentan que el marco de gobernanza de la IA de China intenta abordar múltiples cuestiones jurídicas interconectadas pero distintas que son la moderación de los contenidos y la información generados y difundidos en línea, la protección y seguridad de los datos personales y el uso de algoritmos para tomar decisiones sobre las personas[73].

Sheehan considera que el control de la información es un objetivo central de las tres normas, pero también contienen muchas otras disposiciones notables. Las tres regulaciones exigen que los desarrolladores presenten una solicitud al registro de algoritmos de China, un repositorio gubernamental que se creó a partir de las disposiciones sobre algoritmos de recomendación y se extendió posteriormente al resto de normas y reúne información sobre cómo se entrenan los algoritmos, y también exigen que pasen una evaluación de seguridad[74]. Curiosamente en el RIA también se establece un registro con información básica de los sistemas de IA, pero mientras en China está pensado para las empresas, en la UE está pensado fundamentalmente para los poderes públicos, lo cual refleja una idiosincrasia diferente.

---

71 Cole, J., Sheng, M. y Leung H. T., «New Generative AI Measures in China», *Ashurst,* 26/08/2023 [en línea], (2023), https://www.ashurst.com/en/insights/new-generative-ai-measures-in-china/ [Consulta: 17/12/2024].

72 *Ibidem.*

73 Gong et al, *op. cit.*

74 Sheehan, M., *op. cit.,* p. 4.

Sheehan considera que, estructuralmente, las regulaciones contienen lecciones para los responsables de las políticas en el extranjero. Al implementar una serie de regulaciones de IA más específicas, los reguladores chinos están construyendo de manera constante su conocimiento burocrático y su capacidad regulatoria. Las herramientas regulatorias reutilizables como el registro de algoritmos pueden actuar como un andamiaje regulatorio que puede facilitar la construcción de cada regulación sucesiva, un paso particularmente útil mientras China se prepara para redactar una ley nacional de IA en los próximos años[75].

Erie y Streinz han teorizado sobre un posible «efecto Beijing»[76]. China moldea la gobernanza transnacional de datos al suministrar infraestructura digital a los mercados emergentes. La explicación predominante para este fenómeno es el «autoritarismo digital», por el cual China exporta no sólo su tecnología sino también sus valores y sistema de gobernanza a los Estados anfitriones. Contrariamente a la tesis del autoritarismo digital de talla única, estos autores teorizan sobre un «efecto Beijing» con una combinación de factores de «expulsión» y «atracción» que explica la creciente influencia de China en la gobernanza de datos más allá de sus fronteras y que también podría extenderse a la regulación de la IA. Los gobiernos de las economías emergentes demandan infraestructuras digitales construidas por China y emulan el enfoque de China en materia de gobernanza de datos en pos de la «soberanía de datos» y el desarrollo digital. La «Ruta de la Seda Digital» de China, un esfuerzo masivo para construir los componentes físicos de la infraestructura digital (por ejemplo, cables de fibra óptica, antenas y centros de datos) para mejorar la interoperabilidad de los ecosistemas digitales en esos Estados en desarrollo, materializa el efecto Beijing. Sus principales impulsores son las empresas tecnológicas chinas que cada vez brindan más servicios de telecomunicaciones y comercio electrónico en todo el mundo. El efecto Beijing, en su opinión, contrastaría con el «efecto Bruselas», en el que las operaciones globales de las empresas gravitan hacia las regulaciones de la UE. También se desvía de los esfuerzos de Estados Unidos por dar forma a la gobernanza global de datos mediante instrumentos de derecho económico internacional. Estos autores sostienen que la «soberanía de los datos» es ilusoria, ya que el partido-estado chino conserva diversos grados de control sobre las empresas chinas que suministran infraestructura digital e instan al desarrollo de infraestructuras legales acordes con las estrategias de desarrollo digital.

Además, China lleva varios años implicándose cada vez más es los organismos de estandarización técnica y lanzó la estrategia «Estándares de China

75 *Ibidem.*

76 Erie, M. S. y Streinz, T., «The Beijing effect: China's Digital Silk Road as transnational data governance», *New York University Journal of International Law and Politics*, vol. 54, 2021.

2035» en 2018, con el objetivo de crear un modelo para que su gobierno y sus principales empresas tecnológicas establezcan estándares globales para tecnologías emergentes, como 5G (y 6 G), Internet de las cosas e IA[77].

V. Conclusiones

Es necesario regular a nivel internacional la IA, porque los principios no son suficientes cuando hablamos de cuestiones que afectan a derechos humanos, responsabilidad, etc. Además, actualmente en la gobernanza de la IA, un puñado de compañías están teniendo mucha más importancia que la mayoría de los países. Pero cuando se quiere ir más allá de lo principios en los que casi todos los Estados sí que coinciden a un nivel muy general y elaborar una legislación vinculante granular y concreta no hay consenso internacional, ni siquiera entre aquellos que deberían ser aliados naturales por sus valores como Estados Unidos y la UE. Hasta ahora los avances están siendo decepcionantes. Incluso en el marco del Consejo de Europa donde sí se ha conseguido el hito de lograr un texto jurídicamente vinculante por negociarse entre Estados relativamente homogéneos el resultado ha sido un texto que fija un mínimo común denominador superficial y deja sectores muy relevantes fuera.

Parece que los modelos de regulación van a ser creados por unos pocos países o regiones y extenderse, por lo que debe estudiarse cuáles son los pasos que se están dando en aquellos que parecen marcar las principales tendencias. La UE ha logrado posicionarse como la primera gran potencia en tener una norma horizontal y completa que regula la IA, no obstante, se pone en duda que vaya a tener la misma capacidad para influir más allá de sus fronteras que tuvo el RGPD. Si la UE quiere exportar su modelo va a tener que invertir en ello de manera activa y a través de la cooperación internacional.

Estados Unidos podría ser un aliado, además como cuna de muchas de las empresas tecnológicas más relevantes en este sector podría tener una influencia extrema, no obstante, parece que va a optar (y más con la Administración Trump) por evitar la regulación a nivel federal como palanca para la innovación, siguiendo una estela anterior tecnoliberal, aunque sí pueden verse algunas novedades regulatorias interesantes a nivel estatal e incluso local que no deben ignorarse.

Por su parte, China va a tener una gran influencia, por su propio número de ciudadanos, pero también por la relevancia de sus empresas tecnológicas en las que está invirtiendo ingentes cantidades de dinero y por su capacidad para exportar su modelo normativo. Este país se ha erigido en pionero de las normas relativas a los algoritmos de recomendación y de la IA generativa regulando (a nivel administrativo) mucho antes incluso que la UE estos sec-

77 Zúñiga, N. *et al.*, «The geopolitics of technology standards: historical context for US, EU and Chinese approaches», *International Affairs,* Vol. 100, n.º 4, 2024, pp. 1635-1652.

tores. Su normativa presenta algunos objetivos comunes a los de la UE como proteger a los menores, a los consumidores o a los trabajadores o luchar contra la discriminación, e incluso algunas medidas que podrían mejorar la transparencia como el registro de algoritmos. Esta regulación puede resultar atractiva para Estados del Sur Global que pueden ver en China un ejemplo alternativo al europeo o estadounidense de soberanía digital. Además, China está invirtiendo en mejorar la conectividad en muchos de esos países a través de la nueva ruta de la seda o Iniciativa de la Franja y la Ruta, exportando no sólo infraestructura sino también estándares. La propuesta de una ley más general en China requerirá una gran atención y deberá ser tenida en cuenta por todos los demás actores. A esto se suma el posicionamiento estratégico de China en los organismos de estandarización técnica. No obstante, no se puede olvidar que, aunque su modelo legislativo pueda tener elementos útiles, debe rechazarse su modelo de control social de la población que reforzado por sistemas de IA tiene una inmensa capacidad para vulnerar sistemáticamente derechos fundamentales básicos.

En ese sentido, la UE puede llenar un vacío ofreciendo su apoyo y experiencia a los Estados que se encuentren actualmente ante el desafío de regular la IA, pero no con el objetivo de que dupliquen o copien su normativa, sino que puedan adaptar aquello que les resulte útil y también tener la capacidad y flexibilidad para innovar legislativamente con una base de derechos humanos. Por ejemplo, la UE puede apoyar los esfuerzos de legislación que se están produciendo actualmente en América Latina con los que compartimos una serie de valores fundamentales[78], pero sin imponerles nuestras propias normas para que ellos puedan adoptar normas que respondan a su situación y cultura diferentes de los que han dado como resultado el RIA. Así, la UE también debería mostrarse activa en los foros internacionales para apoyar que no se excluya a estos países y al resto de regiones del mundo que deberían poder participar en los debates sobre gobernanza de la IA, donde hasta ahora sólo han estado presentes unos pocos países.

Bibliografía

Adan, S. N., Trager, R., Blomquist, K., Dennis, C., Edom, G., Velasco, L., Abungu, C., Garfinkel, B., Jacobs, J., Okolo, C. T., Wu, B. y Vipra, J., *Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance,* Oxford Martin School AI Governance White Paper, 2024.

78 Por ejemplo, merecen atención los desarrollos de la legislación brasileña con el proyecto de ley (2.338/2023) que regula el desarrollo y uso de la inteligencia artificial aprobado a finales de 2024 por el Senado y pendiente de su adopción por la Cámara de Diputados (que contiene algunos elementos del RIA) y el proyecto de ley que regula los sistemas de inteligencia artificial de Chile presentado en mayo de 2024.

AI from the Global Majority. Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition. FGV Direito Rio, Rio de Janeiro, 2024.

Almada, M. y Radu, A., «The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy», *German Law Journal*, 25, 2023, pp. 646-663.

Alston, P., *Informe del Relator Especial sobre la extrema pobreza y los derechos humanos*, A/74/493, 2019.

Arnal, J. y Jorge Ricart, R., *Inteligencia artificial: el «efecto Bruselas», en juego*, ARI 88/2023, Real Instituto Elcano 3/10/2023.

Barrio Andrés, M., «La regulación del derecho a la protección de datos en los Estados Unidos», *Cuadernos de Derecho Transnacional*, 2022, vol. 14, no 2, pp. 186-193.

Bipartisan Artificial Intelligence Task Force, *Report on Artificial Intelligence*, US House of Representatives, 2024.

Bradford, A., «The False Choice Between Digital Regulation and Innovation», *Northwestern University Law Review*, 2024, vol. 118, n.º 2, pp. 377-453.

Bradford, A., *Digital empires: The global battle to regulate technology*. Oxford University Press, New York, 2023.

Bradford, A., *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, New York, 2020.

Brown, S. A. W., «Beyond the great firewall: EU and US responses to the China challenge in the global digital economy», *Journal of European Integration*, vol. 46, no 7, 2024, pp. 1089-1110.

Carrillo Robles, M., «La gobernanza de la inteligencia artificial: contexto y parámetros generales», *Revista Electrónica de Estudios Internacionales,* n.º 39, 2020.

Castelló Pastor, J. J., «Exoneración de responsabilidad de los prestadores de servicios de la información en la sección 230 de la Communications Decency Act estadounidense», *Revista Aranzadi de derecho y nuevas tecnologías*, N.º 39, 2015, pp. 175-194.

Chander, A., «How law made Silicon Valley», *Emory Law Journal,* 2013, vol. 63, pp. 639-694.

Cole, J., Sheng, M. y Leung H. T., «New Generative AI Measures in China», *Ashurst,* 26/08/2023 [en línea], (2023), https://www.ashurst.com/en/insights/new-generative-ai-measures-in-china/ [Consulta: 17/12/2024].

Comentarios al Reglamento Europeo de Inteligencia Artificial, Barrio Andrés, M. (dir.), La Ley, Madrid, 2024.

Cotino Hueso, L., «The Council of Europe's Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law», *Rivista Interdisciplinare Sul Diritto Delle Amministrazioni Pubbliche.* n. 2/2024.

Dennis, C. *et al.*, *What Should Be Internationalised in AI Governance?,* Oxford Martin AI Governance Initiative, 2024.

Edwards, L., *Regulating AI in Europe: four problems and four solutions.* Ada Lovelace Institute, 2022.

El Reglamento Europeo de Inteligencia Artificial, Barrio Andrés, M. (dir.), Tirant lo Blanch, Valencia, 2024.

Erie, M. S. y Streinz, T., «The Beijing effect: China's Digital Silk Road as transnational data governance», *New York University Journal of International Law and Politics*, 2021, vol. 54.

Gao, H. S., «Data sovereignty and trade agreements: Three digital kingdoms», en *Data sovereignty: From the digital Silk Road to the return of the state*, Oxford University Press, New York, 2024, pp. 213-239.

García Majado, P., «Los sistemas de puntuación social en el Reglamento Europeo de Inteligencia Artificial», RETINA-DER, 28/11/2024 [en línea], (2024), < https://retina-der.uniovi.es/index.php/2024/11/28/los-sistemas-de-puntuacion-social-en-el-reglamento-europeo-de-inteligencia-artificial-2/ [Consulta: 17/12/2024].

Gascón Marcén, A., «Análisis del Convenio Marco sobre Inteligencia Artificial del Consejo de Europa», en *Desafíos actuales de la Inteligencia Artificial*, UOC, 2024, pp. 141-152.

Gascón Marcén, A., «Derechos humanos e inteligencia artificial», en *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos y Constitucionales, 2020, pp. 335-350.

Gascón Marcén, A., «El Reglamento General de Protección de Datos como modelo de las recientes propuestas de legislación digital europea», *Cuadernos de Derecho Transnacional*, vol. 13, no 2, 2021, pp. 209-232.

Gascón Marcén, A., «La influencia de Japón en el G-20 y el G-7 como instrumento de su política exterior y liderazgo en el ámbito digital», en *Las contribuciones de Japón desde la perspectiva del siglo XXI: aspectos jurídicos y económicos en el plano internacional*, Aranzadi, Madrid, 2024, pp. 209-232.

Gascón Marcén, A., «La propuesta de Reglamento de la Comisión Europea por el que se establecen normas armonizadas en materia de inteligencia artificial», en *Retos de la sociedad digital. Regulación y fiscalidad en un contexto internacional*, Reus, Madrid, 2022. pp. 15-40.

Gong, J., Qu, H. y Dorwart, H., «AI Governance in China: Strategies, Initiatives, and Key Considerations», *Bird & Bird*, 14/03/2024 [en línea], (2024), https://www.twobirds.com/en/insights/2024/china/ai-governance-in-china-strategies-initiatives-and-key-considerations [Consulta: 17/12/2024].

Greenleaf, G., «Asia's Data Privacy Dilemmas 2014-19: National Divergences, Cross Border Gridlock», *Revista Uruguaya de Protección de Datos Personales*, vol. 4, 2019, pp. 49-73.

Greenleaf, G., «EU AI Act: Brussels Effect(s) or a Race to the Bottom?», *Privacy Laws & Business International Report*, vol. 190, 2024, pp. 3-6.

Greenleaf, G., «Global Data Privacy Laws 2023: 162 National Laws and 20 Bills*», Privacy Laws and Business International Report*, vol. 181, 2023.

Groves, L., *Code & conduct. How to create third-party auditing regimes for AI systems.* Ada Lovelace Institute, 2024.

Huld, A., «China's Sweeping Recommendation Algorithm Regulations in Effect from March 1», *China Briefing*, 6/01/2022. [en línea], (2022), https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022/ [Consulta: 17/12/2024].

Interesse, G., «China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023», *China Briefing*, 20/12/2022. [en línea], (2022), https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/ [Consulta: 17/12/2024].

Kaye, D., *Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión*, A/73/348, 2018.

Ke, X., Liu, V., Luo, Y. y Yu, Z., «Analyzing China's PIPL and how it compares to the EU's GDPR», *IAPP*, 24/08/2021. [en línea], (2021), https://iapp.org/news/a/analyzing-chinas-pipl-and-how-it-compares-to-the-eus-gdpr [Consulta: 17/12/2024].

Klein, A., Kerry, C. F., Radsch, C. C., Maccarthy, M., Friedler, S. y Turner Lee, N., «1 year later, how has the White House AI Executive Order delivered on its promises?» *Brookings*, 4/11/2024 [en línea], (2024), https://www.brookings.edu/articles/one-year-later-how-has-the-white-house-ai-executive-order-delivered-on-its-promises/ [Consulta: 17/12/2024].

Lee-Makiyama, H. y Lozano, C., *Empires of Exceptionalism: Lessons from the EU AI Act and Attempts at AI Legislation in California*, ECIPE, 2024.

Lee-Makiyama, H., Narayanan Gopalakrishnan, B., Lozano Rodriguez, C. y Pandya, D., *AI and India's National Interest,* ECIPE, 2024.

Migliorini, S., «China's Interim Measures on generative AI: Origin, content and significance», *Computer Law & Security Review*, vol. 53, 2024.

Presno Linera, M. Á., *Derechos fundamentales e inteligencia artificial,* Marcial Pons, Madrid, 2023.

Roberts, H., Hine, E., Taddeo, M., y Floridi, L., «Global AI governance: barriers and pathways forward», *International Affairs*, vol. 100, n. 3, 2024, pp. 1275-1286.

Rolf, S., *China's regulations on algorithms, Context, impact, and comparisons with the EU*, Friedrich-Ebert-Stiftung, 2023.

Schwartmann, R., Keber, T. O., y Zenner, K., *AI Act: A practical guide*, C.F. Müller GmbH, 2024.

Sheehan, M., *China's AI Regulations and How They Get Made*, Working Paper, Carnegie Endowment for International Peace. 2023.

Smuha, N. A., «From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence», *Law, Innovation and Technology*, vol. 13, n.º 1, 2021, pp. 57-84.

Standford, *AI Index Annual Report 2024*, 2024.

Streinz, T., «Digital megaregulation uncontested? TPP's model for the global digital economy» en Kingsbury, B. et al (eds), *Megaregulation Contested: Global Economic Ordering After TPP*, Oxford University Press, New York, 2019, pp. 312-342.

Supervisor Europeo de Protección de Datos, *Statement in view of the 10th and last Plenary Meeting of the Committee on Artificial Intelligence (CAI) of the Council of Europe drafting the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, 2024.

Tratado sobre el Reglamento de inteligencia artificial de la Unión Europea, Cotino Hueso, L. y Simón Castellano, P. (dir.), Aranzadi, Madrid, 2024.

Tréhu, J. y Jorge Ricart, R., *Global AI Governance: Key Steps for Transatlantic Cooperation*, German Marshall Fund of the United States, 2024.

United Nations High-Level Advisory Body On Artificial Intelligence, *Governing AI for humanity*, 2024.

United States Trade Representative, *2024 National Trade Estimate Report on Foreign Trade Barriers*, 2024.

Veale, M. y Zuiderveen Borgesius, F., «Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach», *Computer Law Review International*, vol. 22, n.º 4, 2021, pp. 97-112.

Villasenor, J. y Turner, J., «AI policy directions in the new Trump administration», *Brookings*, 14/11/2024 [en línea], (2024), https://www.brookings.edu/articles/ai-policy-directions-in-the-new-trump-administration/ [Consulta: 17/12/2024].

White House, «FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI», 21/07/2023 [en línea], (2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [Consulta: 17/12/2024].

White House, «Fact Sheet: Key AI Accomplishments in the Year Since the Biden-Harris Administration's Landmark Executive Order», 30/11/2024 [en línea], (2024), https://www.whitehouse.gov/briefing-room/statements-releases/2024/10/30/fact-sheet-key-ai-accomplishments-in-the-year-since-the-biden-harris-administrations-landmark-executive-order/ [Consulta: 17/12/2024].

Zeng, J., «Artificial intelligence and China's authoritarian governance», *International Affairs,* vol. 96, n.º 6, 2020, pp. 1441-1459.

Zhang, A. H., «Agility Over Stability: China's Great Reversal in Regulating the Platform Economy», *Harvard International Law Journal*, vol. 63, 2022, pp. 457-514.

Zúñiga, N., Datta Burton, S., Blancato, F. y Carr, M., «The geopolitics of technology standards: historical context for US, EU and Chinese approaches», *International Affairs*, vol. 100, n.º 4, 2024, pp. 1635-1652.

8.

INTELIGENCIA ARTIFICIAL: MUCHO MÁS QUE CIENCIA Y TECNOLOGÍA PARA EL DERECHO Y LA DEMOCRACIA

M.ª José González Ordovás
Catedrática de Filosofía del Derecho
Universidad de Zaragoza

SUMARIO: I. Lo mejor y lo peor. II. A propósito de la dimensión mitológica del derecho. III. Sobre la función civilizatoria del derecho. 3.1. ¿Sigue siendo posible el pensamiento complejo? Algunas consideraciones sobre la «tecnociencia». IV. Personas o máquinas ¿quién y por qué debe hacer justicia? 4.1. Inteligencia artificial: una vuelta de tuerca al dilema formalismo-antiformalismo jurídico. V. Algoritmos, valores y sesgos: algunos problemas.

I. Lo mejor y lo peor

En 2016, el astrofísico Stephen Hawking decía en su discurso de inauguración del «Centro Leverhulme para el futuro de la Inteligencia' (C.F.I.) de la Universidad de Cambridge: «la Inteligencia Artificial (I.A.) puede ser lo mejor o lo peor que nos ha sucedido a la humanidad, todavía no lo sabemos»[1].

Algunos años después y vistas su extensión e intensificación, pero sobre todo su creación y uso fuera de control tal vez convenga preguntarse si no sería conveniente cambiar la disyuntiva «o» —«o lo mejor que nos ha sucedido»—, por la copulativa «y», lo mejor y lo peor que nos ha sucedido. Que sea una cosa u otra va a depender de las decisiones adoptadas al respecto, decisiones que necesariamente han de provenir del Derecho y la ética.

1 El discurso fue pronunciado el 20 de octubre de 2016 y puede consultarse en la siguiente dirección electrónica: https://www.rtve.es/noticias/20161020/stephen-hawking-inteligencia-artificial-sera-mejor-peor-pase-humanidad/1428740.shtml

Uno de los mayores obstáculos y retos al respecto radica en la dificultad de crear una cultura internacional común que sea capaz de establecer los límites éticos y jurídicos que sienten las bases del desarrollo y aplicación de una I.A. que no suponga un peligro para la libertad humana tanto individual como colectiva. A este respecto trataremos de mostrar cómo el otorgar un cheque en blanco a la I.A., lo que prácticamente sucede hoy en algunas materias, aunque con diferencias según se trate del caso europeo, norteamericano o chino, por citar tres modelos distintos, supone un grave riesgo para la continuidad de la democracia liberal.

Conviene tener presentes, desde el comienzo de esta reflexión, las atinadas palabras de Paul Virilio cuando nos recuerda que «inventar el barco es inventar el naufragio; inventar el avión es inventar el accidente aéreo; inventar la electricidad es inventar la electrocución... Cada tecnología lleva consigo su propia negatividad que aparece al mismo tiempo que el progreso técnico»[2]. En realidad, es algo que no debiera sorprendernos porque, en general, «no hay ganancia sin pérdida» y, por tanto, «no hay adquisición tecnológica sin pérdida en el nivel del ser vivo, de lo vital»[3]. Dicho de otro modo, tanto como tiene de positivo y avance la I.A., tiene de riesgo y peligro. En última instancia, como ya dijera Hölderlin: «Pero donde hay peligro, crece lo que nos salva»[4].

Que la balanza se incline de uno u otro lado depende de las decisiones éticas, jurídicas y políticas adoptadas al respecto. Nada está dicho, no es el destino sino la acción humana quien tiene la palabra. Dicho lo cual no debe caerse en la complacencia ni el optimismo dada la escasa o nula toma en consideración de la apremiante necesidad de fijar condiciones a la I.A. máxime si, como sostenemos, la I.A. es mucho más que tecnología.

Un posible contraargumento a la acción prescriptiva del Derecho es la conocida sentencia de que «no pueden ponerse puertas al campo» destilando la idea de que resulta imposible frenar o «confinar» determinados hechos o fenómenos. Sin embargo, siguiendo con el símil naturalista la cuestión puede verse desde otra óptica: hay diques en el mar que contienen su potencial fuerza destructiva sin impedir su vital y enriquecedora presencia.

Desde una postura manifiestamente proclive a la «intervención» de la I.A. desde parámetros humanistas observamos cómo, a causa de la complejidad que reviste la I.A., pero también debido a la presión política y económica del reducidísimo número de oligopolios que controlan el mercado de la I.A. «los

2 Virilio, P., *El cibermundo, la política de lo peor. Entrevista con Philippe Petit*, 3.ª ed., Cátedra, Madrid, 2005, pág. 88

3 *Ibídem*, pág. 56.

4 Hölderlin, F.,«Patmos» en *Poesía completa (Edición bilingüe),* 5.ª ed., Ediciones 29, Barcelona, 1995, pág. 395.

gobiernos democráticos han renunciado a legislar sobre la revolución digital y las tensiones de equidad que provoca»[5].

Decíamos antes que la I.A. es mucho más que tecnología caso de que, erróneamente, se dé a este término un sentido únicamente instrumental[6]. La I.A. y con ella toda la transformación digital va mucho más allá de ello porque lejos de ser un accesorio, con la llegada del *homo digitalis* lo digital ha dejado de estar en la periferia de la vida pública y privada para pasar a ser el centro y universo donde ambas, difuminadas, se desenvuelven. Lo digital adquiere hoy la categoría de esencial, esencial es hoy para la información, la investigación, la comunicación, la producción y el consumo. La pregunta que a continuación hemos de formularnos es: ¿es el *homo digitalis* un *homo inhumanus*? ¿la cualidad de *digitalis* le hace perder al hombre sus rasgos distintivos o los erosiona de tal manera que, por contradictorio que pueda parecer, pasa a ser *homo inhumanus*? Ese es el *quid* de la cuestión.

Al decir de Sloterdijk,

> «Con las últimas revoluciones de las redes informáticas, en las sociedades actuales la coexistencia humana se ha instaurado sobre fundamentos nuevos. Estos son (...) decididamente post-literarios, post-epistolográficos, y, en consecuencia, post-humanísticos. Quien tenga por demasiado dramático el prefijo ''post-''de estas formulaciones, podría sustituirlo por el adverbio ''marginalmente'', de tal modo que nuestra tesis sería la siguiente: las sociedades modernas sólo ya marginalmente pueden producir síntesis políticas y culturales sobre la base de instrumentos literarios, epistolares y humanísticos (...) La síntesis social no es ya -ni siquiera aparentemente- cuestión ante todo de libros y cartas (...) La era del humanismo moderno ha pasado (...) Forma parte del credo del humanismo el convencimiento de que los hombres son ''animales sometidos a influencia'', y que es por ello indispensable hacerles llegar el tipo correcto de influjos (...) El humanismo, tanto en el fondo como en la forma, tiene siempre un ''contra qué'', pues supone el compromiso de rescatar a los hombres de la barbarie»[7].

---

5 Lasalle, J. M.ª, *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, Arpa, Barcelona, 2019, pág. 114.

6 Sabemos, a instancias de Heidegger, que «la técnica no es solamente un medio: es un modo de ''desvelamiento'' (...) el ''desvelamiento'' que rige la técnica moderna tiene el carácter de una interpelación en el sentido de una provocación (...) La esencia (ambigua) de la técnica reside en el apresamiento (...) La esencia de la técnica no es nada técnica (...) No es una obra diabólica. Al contrario, cuando nos abrimos limpiamente a la esencia de la técnica, nos hallamos tomados, de forma inesperada, por una llamada liberadora (...) La técnica exige de nosotros que nos pensemos en otra acepción», Heidegger, M., *Essais et Conférences*, Gallimard, París, 1992, págs. 18, 19, 22, 34 y 47.

7 Sloterdijk, P., *Normas para el parque humano. Una respuesta a la Carta sobre el Humanismo de Heidegger*, tr. T. Rocha, Siruela, Madrid, 2000, págs. 28, 31 y 32.

La concurrencia de, al menos, seis factores estaría sentando las bases de un radical cambio ontológico[8]:

1. Sujeto y objeto de un incesante acopio de datos, el ser humano está cada vez más cerca de convertirse en un algoritmo isomorfico.
2. Por influencia de la corriente científica y filosófica del transhumanismo, el ser humano trataría de superar y trascender las limitaciones propias de su condición material.
3. Se resignifica el valor tradicional del trabajo, directamente emparentado con la idea de dignidad.
4. La I.A. es considerada «superior» al hombre pues carece de las debilidades de éste.
5. Algoritmos biométricos lograrán interpretar las emociones humanas y sustituirán la empatía entre personas por la conectividad entre máquinas.
6. La revolución digital arrastra consigo la ciberpolítica y el capitalismo cognitivo. La primera «desestabiliza el conjunto de la arquitectura teórica de la Ilustración» y «pone en duda todos sus fundamentos políticos, culturales, económicos y sociales». Y, en virtud del segundo, las decisiones económicas y financieras del individuo se ven desplazadas por las que adopta una I.A. que jibariza cada vez más al individuo.

Desoyendo la recomendación de Jaspers[9] nos hemos convertido en vasallos del futuro, espoleados por la interesada y derrotista convicción de que pronto seremos superados por nuestra propia creación, damos por sentado que la mejor forma de gestionar la complejidad de nuestro mundo será en breve a través del algoritmo. Al conceder a la I.A. la soberanía de la interpretación y la decisión le dotamos de una peligrosa autoridad. No obstante, y aunque el contexto sea otro, no debería caer en saco roto la seria advertencia que en su día hiciese Hannah Arendt al respecto: «la autoridad, cualquiera que sea su forma, siempre significa una restricción o una limitación de la libertad, pero nunca su abolición. La denominación totalitaria, empero, se orienta a la abolición de la libertad, incluso a la eliminación de la espontaneidad humana en general, y en forma alguna a una restricción de la libertad, por tiránica que sea»[10].

---

8 Nos orientamos en este punto por la opinión del profesor Lasalle si bien hacemos una lectura libre de ella. Al respecto *cfr.* LASALLE, J. M.ª, *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, *op. cit.,* págs. 65, 66 y 69.

9 «Pecamos con nuestra misión cuando nos perdemos en el pasado o en el futuro. Sólo a través de la realidad actual es accesible lo intemporal», JASPERS, K.*, La filosofía desde el punto de vista de la existencia*, tr. J. Gaos, F.C.E Buenos Aires, 1973, pág. 119.

10 ARENDT, H.*, Los orígenes del totalitarismo*, tr. G. Solana, Taurus Madrid, 1998, pág 327.

El panorama actual es radicalmente nuevo respecto a épocas pasadas, la búsqueda del saber, encarnada históricamente en la filosofía y las ciencias, se ve hoy «hipotecada por la reivindicación de una ciencia totalizadora, siempre a un paso de expandirse en forma de efectos totalitarios»[11]. Ese es el contexto en que se incardina la I.A., en modo alguno indistinto para la Democracia, por ende, para su Derecho y, por supuesto, para la Filosofía, de la que «no hay manera de escapar»[12].

Democracia y ciencia, ciencia y democracia son las referencias absolutas de nuestra sociedad, una vez que la Religión dejó de serlo. Su condición de referentes absolutos hace de ambas la fuente originaria de las interpretaciones jurídicas y políticas lo cual comporta indudables consecuencias institucionales[13]. No obstante, sería erróneo colocarlas en un plano de igualdad, la acción y efecto de la aplicación de la I.A. en el ámbito político y jurídico acarrea, a su vez, novedosas implicaciones, en todos los terrenos y a todos los niveles. Si bien es verdad que desde el punto de vista estrictamente formal el concepto de soberanía se estaría viendo afectado por el tema de la globalización, pero no por la «tecnociencia», ello no deja de ser una pequeña falacia. Lo cierto es que, paulatinamente, estamos procediendo a generar una «algocracia». Sugiero este término porque creo que refleja bien la idea de que algo de gobierno está siendo cedido a los algoritmos, una cesión que no deja de crecer cuantitativa y cualitativamente.

Si a ese respecto nos preguntamos a propósito del porqué mantenemos esa afirmación y del cómo se está produciendo ese resultado, observamos que los poderes conferidos a los algoritmos no responden a decisiones que pasen por filtros democráticos expresos y específicos, sino que no hay más filtros que los que se hallen en las veladas motivaciones de las autoridades competentes. El problema proviene de que los algoritmos, y la I.A. que de ellos se deriva, administran y gobiernan los datos de un modo desconocido e inexplicable, a la manera de una especie de «caja negra» (como se mostrará más adelante). En la práctica, el algoritmo «ejerce una soberanía de facto sobre la circulación que circula por internet». El algoritmo actúa como si fuese una norma desde el momento en que «las instrucciones que da a los ordenadores para clasificar, tratar, agregar y representar la información que procesan son incuestionables para las máquinas y los hombres que interactúan bajo su poder»[14]. Las limitaciones técnicas son, hasta ahora, las más importantes, poco puede o hace el Derecho para acotar tal «soberanía». Sin

11 Legendre, P., *El Tajo. Discurso a jóvenes estudiantes sobre la ciencia y la ignorancia*, tr. I. Agoff, Amorrortu, Buenos Aires, 2008, pág. 11.

12 Jaspers, K., *La filosofía desde el punto de vista de la existencia, op. cit.*, pág. 11.

13 *Ibídem*, pág. 91.

14 Lasalle, J. M.ª, *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, *op. cit.*, pág. 91.

necesidad de recurrir a la violencia, su dominio es global. Las afecciones psicológicas, antropológicas[15], y culturales en general, son tan profundas que no parece exagerado asegurar que estamos siendo formateados con una estructura algorítmica: de la forma de entender la libertad y la ciudadanía al modelo de consumo todo queda a disposición del «tecnopoder».

Y aunque las voces, lo mismo de especialistas que de ciudadanos, reclamando una ordenación y regulación de su desarrollo van en aumento, las de las principales compañías lo hacen más rápidamente y con un poder de influencia incomparablemente mayor. Nadie discute que son muchas las ventajas que la I.A. puede aportar, lo que resulta mucho menos evidente es que deba aplicarse por igual en cualquier ámbito humano, esquivando o anulando las sustantivas diferencias que separan a unos de otros. Como si medicina y derecho, por ejemplo, fueran semejantes, como si la garantía, restricción, negación o vulneración de derechos y libertades fuese comparable a cualquier otro ejercicio profesional. La extensión global del «tecnopoder», desbordando la geografía, apenas y débilmente cercado por algunas fronteras territoriales (la Unión europea sería la que más preocupación muestra al respecto) nos conduce a un horizonte en que la «tecnociencia» se perfila como patria «operativa».

Pero ¿acaso se trata de un proceso irreversible frente al que solo cabe resignación en el sentido de una aceptación acrítica? No es esa la idea que defiende Edgar Morin. El filósofo francés es de la opinión de que «la desposesión del saber, muy mal compensado por la vulgarización mediática, plantea el problema histórico clave de la democracia cognitiva. La continuación del proceso tecnocientífico actual, proceso por lo demás ciego si escapa a la conciencia y voluntad de los mismos científicos, conduce a una regresión frente a la democracia. Para ello no existe una política inmediata a poner en funcionamiento. Existe la necesidad de una toma de conciencia política de la urgencia de funcionar por una democracia cognitiva»[16].

No caben soluciones fáciles para lo complejo. Solo a través de lo complejo puede tratarse lo complejo y de poca ayuda sirve la I.A. cuando, al sustituirnos, nos aleja de la necesidad de formular pensamientos complejos. El recurso a la complejidad técnica que encarna la I.A. produce en la práctica la paradoja de apartar al humano de las decisiones complejas y, por tanto, del pensamiento necesariamente complejo que conduce a ellas. «El problema crucial de nuestro tiempo (apostilla Morin) es el de la necesidad de un pensamiento apto para aceptar el desafío de la complejidad de lo real, es decir captar los enlaces, interacciones e implicaciones mutuas, los fenómenos

15 «Administrados *on line*, los hombres están siendo sometidos a un proceso de alienación técnica de su conciencia. Un cambio de paradigmas que afecta a las raíces antropológicas de la elección (...) El problema de este cambio es la feudalización de los usuarios por parte de las plataformas», LASALLE, J. M.ª, *Ibídem,* pág. 96.

16 MORIN, E., *El Método, 6, Ética*, 2.ª ed., tr. A. Sánchez, Cátedra, Madrid, 2009, pág. 170.

multidimensionales, las realidades que son a la vez solidarias y conflictivas, como la democracia misma, que es el sistema que se nutre de antagonismos al tiempo que los regula»[17].

II. A propósito de la dimensión mitológica del derecho

El Derecho puede ser definido como un sistema de normas dictadas por una autoridad destinadas a fijar la conducta individual y colectiva, que prevé anticipadamente consecuencias jurídicas en caso de incumplimiento y que, en última instancia, es sostenido por la fuerza. Desde ese punto de vista, el Derecho es, indefectiblemente un producto del poder. Siendo esto cierto, resulta incompleto por varias razones. En primer lugar, porque buena parte de la creación del Derecho (privado) recae sobre los ciudadanos a resultas del ejercicio de su libre albedrío y de la libertad contractual. Pero además de ello porque el Derecho es también un producto «cultural». El Derecho no nace *ex novo* ni al margen de todo y todos. El Derecho es pensamiento complejo en el que se hibridan poder, libertad, tradición, innovación, pasado y futuro. El Derecho es cultura y ¿acaso podemos liberarnos de nuestra cultura?

Son varias las causas que en los últimos tiempos nos llevan a despreciar o prescindir de todo aquello que no sea cuantificable y, por extensión, registrable y manejable a través de estadísticas. Si el Derecho tan sólo fuese un elevado número de prohibiciones, permisos u obligaciones podría también quedar «atrapado» en el dato. Pero no es así, el Derecho es mucho más que todo eso. El Derecho es un montaje en el que realidad y ficción intervienen entrelazadas y en el que permean, desde su momento fundacional moral, religión, ideología, mitología y teología.

A través de la conjunción de su formulación positiva escrita y de su dimensión abstracta de filosofía política, menos evidente, el Derecho tiene como misión instituir la vida *(vitam instituere)* y el tiempo. Pero para el Derecho existen dos clases de tiempos: el tiempo histórico y el tiempo mítico. Éste, «el de lo inaugural mítico es necesario para la fabricación de los fundamentos dogmáticos del poder de los que se alimenta la reproducción de toda sociedad»[18]. No caben dogmas sin mitos y no hay Derecho sin dogmas. El anclaje de toda la arquitectura institucional y jurídica depende de los dogmas, esto es, creencias tenidas como verdades indiscutibles. El Derecho no está compuesto por verdades sino por dogmas. Una norma no puede ser ni verdad ni mentira. De lo único que puede predicarse la verdad o falsedad es acerca del concreto contenido de una de ellas. Dicho de otro modo, no puede decirse que la norma 1 sea verdad o mentira. De la norma 1 sólo podrá afir-

17 *Ibídem*, pág. 171.

18 Legendre, P., *Le desir politique de Dieu. Étude sur les montages de l'État du Droit, Leçons* VII, 9.ª ed., Fayard, París, 2005, pág. 135.

marse que es verdad (o no) que dicha norma dice x. Así, el mundo del Derecho antes que de normas está formado por dogmas, a éstos deben las normas su existencia. Pero, para que algo sea considero verdad indiscutible sin que quepa demostración o prueba en contrario (dogma) se requiere que venga «avalado» por un mito.

La mitología en Derecho, como toda mitología es el discurso que, a través de representaciones, símbolos y ficciones, revela los principios primeros de división sobre los que toda sociedad se edifica: lo que se acepta y lo que se rechaza, lo que se encumbra y lo que se prohíbe, lo sagrado (patria, bandera, soberanía, pueblo...) y lo tabú. Ese discurso inaugural se encarna en una Referencia o más que establecen el origen de todo y lo que debe ser dicho al comienzo para fijar las certidumbres (Dios, revolución, nación, pueblo, Estado, Constitución...) certidumbres que sostendrán y perpetuarán la sociedad a través de las instituciones. Unas referencias toman el relevo de otras, a veces a través de lucha y otras, pacíficamente, con continuidad. En todo caso las referencias son el germen de las instituciones que estructuran la sociedad. La liturgia, representación y puesta en escena de las instituciones resulta clave para su perpetuación y eficacia. Así se construye el poder, sobre toda una arquitectura de referencias, imágenes, representaciones, instituciones, conceptos, categorías y, por último, normas. Las normas y las decisiones, resoluciones y sentencias en ellas basadas, son el último peldaño de la construcción.

Sin embargo, a partir de un determinado momento de la historia, aunque el entramado del poder y el Derecho continúe siendo en esencia así deja de ser percibido de ese modo. Su dimensión mitológica desparece, deja de ser algo cualitativo para pasar a ser mecánico y cuantitativo. Con Descartes primero y con el auge del positivismo y el empirismo después son las normas las que se sitúan en el centro del andamiaje. La mitología, no desaparece, no puede hacerlo, aunque ahora aparezca velada en construcciones dogmáticas. De hecho, «sin un culto de fundamentos intocables, (...) sin un fanatismo político que roza lo inmutable, no hay normatividad jurídica pensable»[19]. La sociedad se convierte en texto, en la genealogía de los textos que la sostienen. De entre ellos hoy en Occidente el jurídico es el más importante, pero, en todo caso, cada sociedad abraza sus textos y lucha por ellos (Biblia, Derecho romano, Derecho canónico, Declaración Universal de derechos, Tratado de la Unión europea, Constitución, Torá, Corán...). El texto, la suma de lo explícito y lo implícito, de su letra y su espíritu, se convierte en raíz y horizonte. El texto, todo él, no solo sus normas se convierten en un arma, un imperio posible[20]. Existirían por tanto varios textos o «tejidos», entre ellos habría uno llamado Occidente. Para comprender y apreciar la cultura jurídica de Occidente es necesario conocer su texto, y si se ignora o minusvalora su matriz teológi-

19 *Ibídem*, pág. 301.

20 *Ibídem*, págs. 374 y 375.

ca-jurídica: el encuentro entre el discurso religioso cristiano y el vivero de conceptos institucionales técnicos de la juridicidad romana, no se estará en condiciones de advertir «su capacidad (aparentemente paradójica) de integrar en su sistema normativo el individualismo antinormativo y su eficacia para laminar las otras culturas utilizando el arma institucional»[21]. Cada texto o, lo que es lo mismo, cada sociedad sobrevive si perdura su memoria dogmática[22], tesoro, pero al tiempo prisión, que perpetúa la cadena genealógica.

Históricamente cada uno de esos textos, de una u otra forma, ha aspirado a trascender sus fronteras y conquistar el mundo. Por ello, cada vez con más premura es preciso establecer nuevos equilibrios, a veces negociados, otras veces, batallados.

La novedad es que hoy la exigencia de un nuevo equilibrio no llega de otro texto sino de otro método, el de la I.A. Al hilo de ello, uno de los interrogantes que forzosamente nos sale al encuentro es, si so pretexto de mejorar la gestión y aplicación de ese Derecho ¿se procede a través de la tecnología a una forma de traición? En lo que sigue trataremos de reflexionar y dar respuesta a la conjetura de si con la aplicación de la I.A. se burla o no la complejidad del Derecho y si se trata de un grillete autoimpuesto, una herramienta auxiliar o la solución a todos sus problemas.

El Derecho debe su naturaleza a lo complejo y heterogéneo de su origen, composición, funciones, interpretación, aplicación... a todo ello habrá de atenderse para valorar adecuadamente el impacto de la I.A. en la esfera jurídica y no solo, como por otra parte viene siendo habitual, a una parte de lo que el Derecho es. Resulta falaz extender y elevar al conjunto del Derecho las consecuencias que para el mismo supone la aplicación de la I.A. si en el análisis únicamente se ha tenido en cuenta uno de sus aspectos o partes puesto que el Derecho es un sistema y, en consecuencia, repercute en el conjunto lo que a cada uno de sus elementos atañe.

Apuntalado por la mitología, los dogmas e incluso, los ritos, el Derecho instituye al individuo y a la sociedad a través de un entramado, un tejido, un texto en el que se plasma una determinada visión del mundo. Pues bien, las maneras en que se observa y describe el mundo afectan a la construcción del mismo. El Derecho, junto con otras formas de cultura como la filosofía, el arte o las ciencias participa de forma muy especial y activa en esa construcción del mundo dado lo privativo de su componente coercitivo. La forma en que el Derecho lo viene haciendo en Occidente desde el Derecho Romano es por completo diferente de la manera en que lo hace y puede hacerlo la I.A., entendiendo por diferente en este punto no reemplazable. En concreto, «el monumento de escritos al que llamamos Estado y Derecho es la construcción

21 Legendre, P., *Lo que Occidente no ve de Occidente. Conferencias en Japón*, tr. I. Agoff, Amorrortu, Buenos Aires, 2013, pág. 23

22 Al respecto *cfr. Ibidem*, págs. 46 y ss.

política más sobresaliente de la humanidad occidental exportada además a todo el planeta a través del industrialismo nacido en Europa»[23]. Un acopio de conocimiento destilado de la experiencia y los valores a través de los siglos.

Para mejor comprender el alcance del impacto de la I.A. sobre el Derecho conviene contextualizar el indudable logro tecnológico que la I.A. supone. La I.A. es una de las manifestaciones, probablemente la de mayor trascendencia social, de la «tecnicización del pensamiento» a la que venimos asistiendo en los últimos tiempos. «Enmascarada por una "interdisciplina" productora de algo por lo cual se desvive la investigación actual las "síntesis informacionales". Síntesis que se disgregan en cuanto al fondo no bien aparece la pregunta inevitable: ¿qué valor tienen nuestras categorías y modos de razonar?»[24] Ese pensamiento «tecnicizado» basado en lo informacional «pone en tela de juicio el modo de objetivación occidentalista de la organización humana, lo que llamamos sociedad, cultura, civilización»[25]. La ciencia, y de su mano la técnica, se han convertido en las «vencedoras cognitivas». Cuanto no pueda contabilizarse, medirse, informatizarse... se considera un vestigio de interés, pero incapaz de adquirir y proporcionar la categoría de verdad, nivel «mítico» atribuido ahora a la ciencia en exclusividad.

Como dice Legendre, «hubo un Dios, hoy es la Ciencia la que le dice al hombre (...) lo que hay que pensar del pensamiento. Así, la Ciencia es la heredera de los dogmas de Occidente, extiende su poder al dominio del desamparo y explica al hombre lo que él vive»[26]. A la sombra de la ciencia, la tecnología se ha convertido en lo más parecido a una patria global y la I.A. en su emblema. Ese dominio del mundo extendido a todo tipo de saberes y disciplinas se ha hecho posible por la síntesis de dos principios: el de eficiencia y el de fragmentación. El concurso del primero, en tanto que sublimación económica de la eficacia, y del segundo, método del análisis empírico científico-técnico con su correspondiente eco en el trabajo social se ha convertido en un imperativo totalizador para toda aquella sociedad que aspire a seguir existiendo en los actuales términos competitivos. Sin embargo, ese gobierno de la técnica, encabezado hoy por la I.A., tiene consecuencias sociales de primer orden. Por un lado, estarían las posibles injusticias y desigualdades derivadas de prácticas opacas, como tratamos de evidenciar[27]. Por otro, en lo que

23 *Ibídem*, pág. 37.

24 *Ibídem*, págs. 18 y 19

25 *Ibídem*, pág. 19.

26 Legendre, P., *La fabrique de l'homme occidental*, Fayard, París, 1997, pág. 26.

27 Ferrer, I., «Países Bajos veta un algoritmo acusado de estigmatizar a los más desfavorecidos», *El País,* 13 de febrero de 2020. «El Gobierno usaba el programa SyRI para determinar qué ciudadanos son supuestamente más proclives a defraudar al Estado. Un tribunal neerlandés ha decidido que SyRI (acrónimo de System Risk Indication), un sistema de análisis para rastrear posibles fraudes al Estado utilizado por el Gobierno, "no respeta la privacidad del ciudadano y vulnera el artículo 8 de la Convención Europea de Derechos

concierne al núcleo más profunda y radicalmente humano nos estaríamos aproximando a un horizonte en que ambas, ciencia y técnica, laminarían la idea de límite consustancial a la noción de libertad. Sin la fijación y asunción de ciertos límites nos situamos en otro escenario pues sin ellos, sin la presencia de límites no cabe un ejercicio colectivo de la libertad. Lo imposible y/o lo prohibido se desdibujan por acción y efecto de las posibilidades inagotables de la tecnología en todos los ámbitos de la vida.

El caso del uso de la I.A. en el ámbito del Derecho es solo un ejemplo, pero quizás uno de los que más claramente muestre los riesgos de generar un individuo «des-instituido», al apartarlo de los referentes que originaron su cultura jurídica. Es característico del pensamiento humano preguntarse sobre sus límites, tender hacia ellos y desafiarlos, de hecho, es así como crece. El pensamiento, para serlo, ha de estar siempre en movimiento y es precisamente la noción de límite la que actúa como estímulo y acicate, si ella desaparece del consciente y el subconsciente nos hallaremos ante un ser «des-instituido», esto es, carente de referencias y filiación y por tanto individuos cuyos lazos sociales tienden a difuminarse y desaparecer.

Con su tratamiento fragmentado y fragmentario del fenómeno jurídico y tras su coartada de eficiencia, la I.A. convierte la normatividad en una mera herramienta de ingeniería social. Sin embargo, tanto la «insociable sociabilidad del hombre»[28] como su inconsciente escapan al imperativo de la eficiencia. En ese contexto, y sin ambages, Legendre denunciaría la irracionalidad a que conduce la racionalidad científico-técnica dejada a su suerte. Un leviatán digital que prescinde de la mitología del «Texto», de sus referentes fundacionales y, por tanto, del anclaje societario que la filiación jurídico-institucional ofrece al individuo, esto es, un «ciberleviatán»[29] sin más cortapisas que

Humanos relativo al derecho a la vida privada". Poco tiempo después y a consecuencia de las investigaciones realizadas, "El Gobierno holandés dimite en bloque por un escándalo de ayudas sociales"», Rachidi, Imane, *El Mundo*, 15 de enero de 2021. El Estado de Derecho neerlandés falló debido a que se habían incluidos datos (segunda nacionalidad de los solicitantes) en un programa de I.A. que produjeron un trato institucionalmente sesgado. «Un comité parlamentario establecido para investigar este escándalo, y cuyo informe fue la razón detrás de la caída del gobierno neerlandés, concluyó que las autoridades fiscales tuvieron un "trato institucionalmente sesgado" hacia las víctimas porque gestionaron la Ley de Cuidado Infantil —por la que se otorgan estas ayudas— como un "proceso masivo" y no como casos de familias en condiciones específicas e independientes. Su actitud hizo que muchas familias se hayan roto a causa de lo ocurrido, y algunos padres aún tienen que convivir con problemas psicológicos y financieros. Esta comisión alertó de que ministros, funcionarios, el parlamento y los jueces habían cometido errores».

28 Kant, I., «Idea de una historia universal en sentido cosmopolita» *en Filosofía de la historia*, F.C.E, México D.F, 1978, págs. 46 y 47.

29 Lasalle, J. M.ª, *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, *op. cit.* Arpa Barcelona, 2019.

las tecnológicas puede acabar tratando al ser humano como si fuese ganado, reinventando la noción de sacrificio[30].

Los distintos programas y sistemas de I.A. han demostrado estar muy preparados para llevar a cabo labores de búsqueda y predicción en el ámbito de la información jurídica: búsqueda y tratamiento de normas, decisiones, resoluciones, sentencias, precedentes... Pero ¿se distingue en la esfera de la I.A. entre información, conocimiento y sabiduría? ¿Es capaz la I.A. de generar conocimiento jurídico o solo de reproducir el ya creado? ¿Se puede, en ausencia de conciencia, generar conocimiento? ¿Y sabiduría? ¿Puede convertirse el uso «descontrolado» de la I.A. en el ámbito jurídico en el primer paso de la renuncia a ella? Como se observa, cada vez son más los puntos ciegos a que nos conduce un uso de la I.A. no testado ni evaluado desde diferentes ópticas como la teoría del conocimiento o la teoría de la justicia. El riesgo de que los agujeros de legitimidad de los Estados de Derecho aumenten y se profundicen en caso de que las aplicaciones de la I.A. se lleven a la práctica sin el cortafuegos de los derechos humanos es demasiado alto.

III. Sobre la función civilizatoria del derecho

La función civilizatoria del Derecho va estrecha y directamente vinculada a su naturaleza antropológica, de ella deriva el cometido humanizador implícito en toda civilización. En ese sentido ha de afirmarse que el Derecho es cultura, la nuestra, por ejemplo, la cultura jurídica occidental, viene desarrollando a lo largo de la historia una labor de filiación y humanización fundamental, la de vincular cada generación con las anteriores y establecer compromisos con las futuras. Ninguna generación lo es todo en Derecho, todas tienen una deuda con las pasadas y todas han de garantizar el traspaso a las siguientes[31]. A través del Derecho se sustancia el relevo generacional y el acopio del legado del pasado.

La naturaleza antropológica y la función civilizatoria del Derecho se lleva a cabo a través de las instituciones. Como afirma el profesor Martínez García, «el Derecho está hecho antes de instituciones que de normas: no se entiende bien si se considera sólo como un conjunto de reglas y principios. Es un conjunto de instituciones. La clave está en la institución y desde ahí podemos entender las normas»[32].

---

30 Legendre, P., *Leçons VI. Les enfants du texte*, p. 358. Al respecto y *cfr.* la excelente tesis doctoral *La antropología dogmática de Pierre Legendre,* de J.A. Bellido Añón, UAM, 2022, pp. 165-170.

31 *Cfr*. Legendre, P., *Ibidem*, pág. 395.

32 Martínez García, J. I., «Instituciones inteligentes», en Solar Cayon, J. I. (Ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho. Cuadernos Democracia y Derechos Humanos*, Alcalá de Henares, Universidad de Alcalá, 2020, pp.

La complejidad del mundo institucional proviene de la heterogeneidad de las mismas, y es que, como distingue Hauriou, las instituciones pueden ser de dos tipos: «instituciones-personas», las organizaciones, o «instituciones-cosas», los contratos[33]. Las instituciones construyen la realidad, hacen de ella un universo de significaciones compartidas hasta el punto de que puede decirse que la sociedad, más que un conjunto de individuos es una red de instituciones. La sociedad es, al tiempo, instituida e instituyente, desarrollándose en un permanente proceso de «autocreación»[34]. Dicho de otro modo, las instituciones no son tanto, inteligencia aplicada cuanto creadoras de inteligencia. A través de las instituciones los individuos comparten una forma de concebir y transformar la sociedad, las instituciones «estructuran» nuestra forma de ver la realidad y condicionan nuestra acción sobre ella. Es decir, las instituciones generan una forma de conocimiento que aporta un sustrato o fundamento común a quienes participan de ellas.

Tomemos como ejemplo los derechos humanos, sostenemos con el profesor Martínez García que «los derechos humanos no son solo normas: son también mecanismos cognitivos. Son modos de sentir y pensar que forman parte de nuestros circuitos mentales. Suscitan conexiones neuronales que modifican nuestra actividad cerebral. Impregnan toda la mentalidad contemporánea, incluso la de sus adversarios»[35]. Algo parecido puede decirse de la figura del contrato. El contrato es una «institución-cosa» (por utilizar la terminología de Hauriou) y a la vez, y por ello, un esquema a través del cual procesamos y razonamos cuanto sucede a nuestro alrededor. Una figura a través de la cual diseñamos y evaluamos la realidad, una figura que trasciende lo jurídico para pasar a ser un elemento básico de la vida individual y colectiva. Una invención de la familia jurídica occidental a tal punto importante que hubieron de adoptarla quienes vivían ajenos a ella para no verse «expulsados» de las relaciones internacionales[36]. En suma, uno de los más poderosos motores jurídicos de civilización.

---

26 y 27. El origen de esa teoría, profundamente asentada en el pensamiento jurídico occidental, se halla por lo que respecta al Derecho privado en las obras de Savigny, F. C. von, *Sistema del Derecho Romano actual*, Góngora, Madrid, 1878-1879, Tomo I, Cap. II, pág. 86 y *Metodología jurídica*, Depalma Buenos Aires, 1979, p. 14 y Ihering, R. Von, *El espíritu del Derecho Romano en las diversas fases de su desarrollo*, Comares, Granada, 1998. Y, en lo concerniente al Derecho Púbico obra de Hauriou, M., «La teoría de la institución y de la fundación (Ensayo de vitalismo social)» en Obra Escogida, Instituto de Estudios Administrativos, Madrid, 1976.

33 Hauriou, M., *op. cit*, p. 266.

34 Al respecto, *cfr* Martínez García, J. I., «Instituciones inteligentes», *op, cit*., p. 31.

35 *Ibídem*, pp. 34 y 35.

36 «Acceder a la cultura del contrato devino la condición de acceso a la Modernidad y al concierto de las naciones. Eso fue lo que ocurrió en el Japón de la era Meijí que, para escapar al yugo de los tratados de los "tratados desiguales", hubo de dotarse de un Derecho de contratos cuya naturaleza le era fundamentalmente extraña», Supiot, A., *Homo juridicus. Essai sur la fonction anthropologique du Droit*, Seuil, París, 2005, p. 137.

¿Que cuál es la relación de todo ello con la I.A.? Tratamos de mostrar que el hombre se explica por su Derecho y eso incluye, por supuesto, sus normas, pero también y, sobre todo, sus instituciones. Y que el Derecho es antropología pues lo que normas e instituciones hacen es primero crear para después mostrar y hacer patente la estructura profunda de la sociedad. En un proceso autorreferencial la sociedad crea el Derecho y éste a aquella pues uno y otro son la misma cosa. Una sociedad es su Derecho, su Derecho derogado, su Derecho modificado, su Derecho vigente, su Derecho eficaz, su Derecho vivo... A la dogmática jurídica le interesan las leyes positivas y a la antropología jurídica las leyes inconscientes, las subyacentes que provocan la creación de las positivas. Retando a su naturaleza sistémica y orgánica lo que la I.A. hace con el Derecho es dividirlo, trocearlo, fragmentarlo para dominarlo. En un afán de convertirlo en dato lo simplifica, le despoja de una complejidad que le es consustancial.

Así, en su audaz ambición de capturar lo que el Derecho es y dominar su aplicación, la I.A. a través, primero, de la reducción y conversión en dato de normas, disposiciones, resoluciones y sentencias y de su matematización algorítmica, después, «traiciona» al Derecho porque:

1.º Lo simplifica: como toda simplificación, la «algoritmización» del Derecho tiene como función la eliminación de las paradojas. El pensamiento jurídico es necesariamente paradójico pues ha de sintetizar elementos contrarios, opuestos e incluso absurdos pues siendo la política una de las fuentes del Derecho la idea del legislador racional no pasa de ser un mito.

La. I.A. falsea el Derecho cuando al fragmentarlo para contraerlo (normas, sentencias, hechos, circunstancias, informaciones, perfiles...) a la condición de dato acaba menguando una de sus partes y, una vez reducida, trata a esa parte como si fuese un todo. Pensar que todo el Derecho está en las normas, en las sentencias o en los libros es privarle de su mayor posibilidad de innovación, pues sólo a través de la práctica reflexiva, y esa es exclusivamente humana, se pueden explorar nuevos horizontes[37]. La lógica de la I.A. solo es oportuna con lo mensurable y cuantificable pero nunca puede sustituir al pensamiento y, por tanto, a la inteligencia, pues sabemos, como mínimo desde Pascal, que «todas las cosas son causadas y causantes, ayudadas y ayudantes, mediatas e inmediatas, y manteniéndose todas por un nexo natural e insensible que liga las más alejadas y las más diferentes, tengo por imposible conocer las partes sin conocer el todo, así como conocer el todo sin conocer particularmente más partes»[38].

2.º Lo opaca: de nada vale decir que los productos y programas de la I.A. recopilan y tratan datos previamente confeccionados por hombres, siendo en

37 Al respecto *cfr.* Rodríguez-Arana Muñoz, J., «La reflexión en la enseñanza del Derecho» en Rodríguez-Arana Muñoz, J. y Palomino Lozano, R., (Dirs.), *Enseñar Derecho en el siglo XXI. Una guía práctica sobre el Grado en Derecho*, Aranzadi, Cizur Menor, 2009, págs. 205-207.

38 Pascal, B., *Pensamientos*, 10.ª ed., Espasa Calpe, Madrid, 1995, pág. 93.

consecuencia resultantes de las muy diversas acciones humanas. Con arreglo al modo actual de proceder de la I.A., a la manera de un «caja negra», no es posible conocer ni el cómo ni el porqué de las decisiones que adopta por sí misma[39]. Pero, incluso en el caso de que sí fuese posible, tales sistemas, por sofisticados que sean, carecen de rasgos constitutivos de la naturaleza humana como son el libre albedrio, la conciencia, la libertad, la creatividad, la empatía, la imaginación, la inteligencia... ¿es conveniente, lícito o ético aceptar que la visión del mundo se construya al margen de los atributos específicamente humanos.

3.º Lo fosiliza: los sistemas predictivos y de resolución automática se basan en la búsqueda y aplicación de decisiones anteriores (precedentes judiciales) cercenándose la vivacidad propia del Derecho.

4.º Lo debilita: a la manera de una prótesis técnica, resulta indubitable que un uso indiscriminado de la I.A., acabará por generar una dependencia que debilita la capacidad de decisión del hombre y, como consecuencia, al Derecho mismo pues éste no es un dato sino una acción. En esa acción que el Derecho es convergen norma e intérprete. Suprimida por la I.A. la capacidad creativa e innovadora del interprete, se ve sustituida o por la predicción repetitiva de una máquina o por la desconocida conexión neuronal que, en el denominado «pensamiento profundo», desarrolle la máquina, desprovista en todo caso del factor causal, reemplazado por los de la correlación y conectividad.

Uno de los argumentos esgrimidos en favor del empleo de la I.A. en el ámbito jurídico es que fundamentalmente, hasta la fecha, se utiliza como apoyo auxiliar y no definitivo en la elaboración de sentencias. Herramienta en la que el juez puede apoyarse pero que en última instancia sigue siendo libre para hacerlo o no y, en ese sentido, sería inane. Se trata, no obstante, de una argumentación teórica algo débil por varias razones. En primer lugar, porque la aplicación de la I.A. en el ámbito jurídico desborda con mucho al estrictamente procedimental, de hecho, se han desarrollado herramientas consis-

39 «Las arquitecturas de redes neuronales profundas son capaces de identificar "patrones ocultos" y crear correlaciones inesperadas (e indetectables para una inteligencia humana) entre *inputs y outputs* simplemente a partir del análisis de datos masivos, generando automáticamente los modelos predictivos o las reglas de decisión que posteriormente aplican a nuevos conjuntos de datos. (...) Los modelos predictivos y las reglas de decisión son constantemente modificados y refinados a medida que el sistema va acumulando más información (...) En este tipo de sistemas, ni siquiera el acceso abierto al conjunto de datos utilizados para su entrenamiento y a su código fuente permitirá desvelar cómo los algoritmos realizan sus predicciones, en cuanto éstas son el resultado de la compleja interacción de una red de millones de neuronas artificiales distribuidas en múltiples capas o niveles de análisis que ponderan simultáneamente el valor predictivo de cientos o miles de factores potenciales y de sus múltiples posibilidades combinatorias, lo que imposibilita identificar el peso de cada factor en los resultados (...) Las herramientas de aprendizaje profundo son inescrutables, completamente opacas: ni el sistema ni sus diseñadores pueden proporcionar una explicación», Solar Cayón, J. I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riegos», en Solar Cayón, J.I. (Ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho, op. cit*., pp. 148 y 149

tentes en identificar los patrones de comportamiento de jueces y tribunales como estrategia que permite conocer cuál será la concreta actuación del magistrado que conozca del caso. Su inquietante extensión ha llevado a países como Francia a limitar su uso[40].

En segundo lugar, porque al agente jurídico que se le facilita y permite el uso de dicha herramienta le resulta más fácil y cómodo recurrir a ella que «enfrentarse» a ella. Es obvio que plegándose a su dictado no solo «ahorra tiempo» sino que evita posibles problemas de responsabilidad pues es tenido como casi infalible y en todo caso mucho más «inapelable», lo que determine un algoritmo, basado en cientos de miles de datos, frente a que lo que determine una sola persona cuya probabilidad de «error» en mucho mayor. La máquina evita la equivocación humana.

3.1. ¿Sigue siendo posible el pensamiento complejo? Algunas consideraciones sobre la «tecnociencia»

Lo que empuja a la vida es el pensamiento, todo tipo de pensamiento: el fabulador, el filosófico, el científico, el jurídico... cada uno de ellos posee códigos propios (analítico, abstracto, inductivo, deductivo...) con sus consiguientes racionalidades o irracionalidades, lógicas e incluso ilógicas, pero eso resulta de todo punto secundario pues lo primordial es que el pensamiento es, en cualquier caso, el punto de partida de toda creación. Llevados por la dinámica de la imparable división del trabajo social, de la que de forma precursora ya se ocupara Émile Durkheim en 1893[41], la progresiva especialización del pensamiento ha ido forjando una fragmentación del mismo hasta devenir hoy en verdadera atomización. Ello, que por un lado facilita indudables e innumerables avances tecnológicos, tiene como necesaria contrapartida la derrota del

40 *Cfr*. el artículo 33 de la LOI n.º 2019-222, de 23 de marzo de 2019 de programación y de reforma de la Justicia. En dicho artículo se sanciona penalmente la utilización de los datos de identidad de los jueces «con el propósito de evaluar, analizar, comparar o predecir sus prácticas profesionales reales o presuntas». Ello no obsta para que en ese país se haya implementado un proyecto de reforma de la Justicia (2018-2022) con el propósito de automatizar ciertos pasos del procedimiento judicial como, por ejemplo, la resolución de conflictos de menor entidad por medio de algoritmos de conflictos cuando el valor de la disputa sea entre 4.000 y 6.000 euros.

41 Es el volumen y densidad de las sociedades la que propicia la división del trabajo, es decir, su especialización que, a su vez, entronca directamente con los cambios civilizatorios. En ese sentido, a juicio de Durkheim, la civilización no es un fin conscientemente establecido sino sencillamente el resultado del número de individuos convivientes que ejercen una creciente presión entre sí. Eso que él denomina «intensidad de la vida», propia de la solidaridad orgánica, y que es motor del aumento de la especialización, no tiene por qué implicar un aumento ni de la felicidad ni de la moralidad. «Al determinar la causa principal de los progresos de la división del trabajo, hemos determinado al mismo tiempo el factor esencial de lo que se llama civilización», Durkheim, É., *De la division du travail social*, 11.ª ed., P.U.F., París, 1986, págs. 327 y 328.

todo, de la unidad. Pues con un pensamiento que solo puede crecer en añicos, el desconocimiento del casi-todo es inevitable dada la inmanejable cantidad de investigaciones cada vez más hiperespecializadas. Como señala Morin, «la hiperespecialización contribuye poderosamente a la pérdida de visión o concepción de conjunto pues las mentes encerradas en su disciplina no pueden aprehender las solidaridades que unen entre sí los conocimientos (...) La clausura disciplinar, junto a la inserción de la investigación científica en los cuadros tecnoburocráticos de la sociedad, produce irresponsabilidad para todo aquello que es exterior al dominio especializado»[42].

Así las cosas y habida cuenta de que, como nos recordara Pascal, no es posible conocer el todo sin las partes, ni conocer éstas ignorando aquél, navegamos desorientados en un archipiélago inconexo de datos e informaciones en un contexto en el que los medios no dejan de crecer sin que, ni lejanamente, ocurra lo mismo con los fines. Es el «inhumano» manejo de toda la «información» generado por el progresivo número de especialidades el que, por un lado, «invita» a confiar en la I.A. el uso eficiente del mismo pero que, por otro, complica la continuidad de un pensamiento «inteligente» en el sentido de transversal, integrador, profundo y clarividente. La fertilidad aportada por un pensamiento no dividido[43], tal como se produjera en etapas tan «fértiles» de la Historia como la Grecia Clásica o el Renacimiento cuando física y metafísica (ciencia y filosofía) agitaban conjuntamente el entendimiento, no es probable que vuelva a darse.

Afirmamos, sin embargo, que el pensamiento, para serlo de verdad y en cualquier época, ha de ser necesariamente complejo. Pero la específica complejidad del pensamiento contemporáneo no proviene del «encuentro» de disciplinas heterogéneas sino de su «desencuentro». La generación y «trasvase» de conocimientos favorecido por la sinergia de doctrinas, teorías y materias ya no parece posible. En los últimos años la referida especializa-

42 Morin, E., *El Método, 6, Ética*, *op. cit.*, p. 81.

43 Hacemos abstracción en este punto del sentido que da Heidegger al verbo pensar. Para él «la ciencia no piensa porque su enfoque y sus medios auxiliares son tales que no puede pensar —queremos decir pensar a la manera de los pensadores—. No hay que ver ningún defecto en el hecho de que la ciencia no pueda pensar, sino una ventaja. Esa ventaja es la que asegura a la ciencia un acceso posible a dominios de materias correspondiendo a sus modos de búsqueda (...) La proposición "la ciencia no piensa" choca con nuestra concepción habitual de la ciencia. Dejemos el carácter chocante de la afirmación pues hay aún otra que le sigue (...) la ciencia no puede nada sin el pensamiento. La relación de la ciencia con el pensamiento sólo es auténtica y fecunda cuando el abismo que les separa se hace visible y se hace patente que ningún puente puede salvar ese abismo. No hay puente que nos lleve de las ciencias al pensamiento, lo que hay es un salto», Heidegger, M., «Qué quiere decir pensar» en *Essais et Conférences, op. cit.,* pág. 157. En la base de tal disquisición estaría la distinción que el mismo autor realiza entre lo exacto y lo verdadero: «lo que simplemente es exacto aún no es lo verdadero», «La cuestión de la técnica», en *Essais et Conférences, op. cit.,* pág 11. De lo primero se ocuparían la ciencia y la técnica, de lo segundo el pensamiento filosófico.

ción ha venido acompañada de su matematización a través de los algoritmos. ¿A dónde nos lleva eso? Si Pascal afirmaba que solo conociendo el todo se conoce a las partes y, a la inversa, que sin tomar en cuenta a las partes conocer el todo es ilusorio[44], Lichtenberg da un paso más allá. A finales del XVIII este físico y pensador alemán, no sin cierta ironía sostenía en uno de sus aforismos que: «quien sólo entienda de química, tampoco la entenderá bien»[45]. En otras palabras, el que sólo sabe de Derecho tampoco sabe de Derecho o el que sólo sabe de física tampoco sabe de física, y así sucesivamente. Si Lichtenberg estaba en lo cierto debemos afirmar que, en la actualidad, a fuerza de la antedicha atomización apenas se sabe nada. La cantidad de datos acumulados es colosal y la información de que disponemos ingente, pero ello no debe confundirse con el conocimiento o el saber.

Nos hemos atenido hasta ahora a la distinción clásica que organiza «el saber» en tres compartimentos: ciencia, técnica y todo lo demás. Esto es, todo aquello que al no ser ni ciencia ni técnica iría a una especie de «cajón de sastre» y, reuniría a artes (todas), humanidades y ciencias sociales. Disponiendo de un lenguaje propio, el matemático, las primeras, ciencia y tecnología pueden avanzar porque, a base de acumular demostraciones, hacen acrecer su acervo en una progresión sin fin. Sin embargo, tengo el convencimiento de que tal clasificación, acertada en su tiempo no atina con la realidad contemporánea porque no la representa. Comparto con Edgar Morin la idea de que «la relación ciencia/técnica se ha vuelto tan indisociable» que se impone la expresión «tecnociencia». «Se ha puesto en marcha un proceso en el que la manipulación técnica está cada vez más al servicio de la ciencia, pero en el que, también, la ciencia está cada vez más al servicio de la manipulación técnica. El desarrollo del conocimiento por el conocimiento que es propiamente científico en adelante es inseparable del desarrollo del dominio que es propiamente técnico»[46].

La «mixtura» entre ciencia y técnica a la que se refiere Morin no es, en modo alguno, cuestión baladí. La aparición de ese *tertium genus* trastoca y trasciende el enfoque tradicional pues da origen a una nueva y turbadora realidad ya que, a diferencia de lo que ocurriera cuando la ciencia, pudiera ser calificada de aventura pura, en el sentido de apolítica, la «tecnociencia», convertida en motor, produce y acumula un poder gigantesco. Y, allí donde existe o se acapara poder se requiere de la presencia de criterios y, por tanto, valores para ordenar, ajustar y regular dicho poder. Pero si la «tecnociencia» se

44 *Cfr*. nota 41.

45 Lichtenberg, G. Ch., «Rousseau, creo, dijo una vez: "Un niño que sólo conozca a sus padres, no podrá conocerlos bien". Esta idea puede aplicarse a muchos otros conocimientos, yo diría que a todos los que no sean de naturaleza totalmente pura: quien sólo entienda de química, tampoco la entenderá bien», *Aforismos*, tr. J. del Solar, Edhasa Barcelona, 1990, aforismo 860, pág. 231.

46 Morin, E., *El Método, 6, Ética*, *op. cit*., pág. 78.

ha convertido en la palanca omnipresente que todo lo puede y todo lo mueve es porque son sus desarrollos los que impulsan la economía siendo la alianza de ambas la causa primera de la sustancial transformación que individuo y sociedad conocen. De la poderosa mano de la economía, «el vínculo ciencia/técnica/sociedad/política es evidente»[47], atrás quedó la época en que ciencia y tecnología caminaban por separado y podían caminar separadas de la reflexión ética. Pues bien, precisamente en ese contexto es en el que se halla la I.A. y respecto al cual, probablemente más que nunca, sean imperativos y urgentes los compromisos ético y jurídico.

A decir verdad, la preocupación por el papel asignado a la ciencia, hoy tecnociencia, viene de algo atrás. En 1935 en su conferencia «La crisis de las ciencias como expresión de la crisis vital radical de la humanidad europea» Husserl advertía de que la asignación disyuntiva en virtud de la cual la filosofía y las artes eran el exclusivo ámbito de la subjetividad mientras que a la ciencia quedaba reservado el de la objetividad, acababa generando un punto ciego. La ciencia, capaz de desvelar los secretos de la materia gracias a las más sofisticadas técnicas, es incapaz de tener conciencia de sí misma, de su marcha y, por supuesto, de la subjetividad humana. Siendo, precisamente, esa carencia de subjetividad la que le imposibilita pensarse y pensar acerca de su responsabilidad. Pero eso no siempre fue así:

> «''no siempre estuvieron las cuestiones y problemas específicos de la humanidad desterrados del dominio de las ciencias, ni quedaron siempre fuera de consideración las relaciones internas de dichos problemas y cuestiones con todas las ciencias, incluidas aquellas en las que el hombre no es el tema (...) Mientras las cosas no fueron aun así la ciencia pudo reivindicar una significación (...) Las ciencias en plural, todas las que aún habían de fundarse y todas las ya establecidas, no eran sino ramas no independientes de una filosofía una'' Pero la ciencia ''se volvió contra su modo precedente de existencia''»[48].

El punto de partida de la transformación hay que situarlo a finales del siglo XIX.

> «Una transformación relativa no solo al juicio sobre su cientificidad, sino a la percepción de lo que ciencia en general ha significado y puede significar para la existencia humana. La exclusividad con la que en la segunda mitad del siglo XIX se aceptó determinar la visión entera del mundo moderno por las ciencias positivas y se dejó deslumbrar por la *prosperity* hecha posible por ellas, significó paralelamente un desvío indiferente respecto de las cuestiones realmente decisivas para la humanidad auténtica (...) Las cuestiones que excluye por principio son precisamente las más candentes para unos seres sometidos, en esta época desventurada, a mutaciones decisivas: las cuestiones relativas al

47 *Ibídem*, pág. 79.

48 Husserl, E., *La crisis de las ciencias europeas y la fenomenología trascendental,* trs. J. Muñoz y S. Más, Crítica, Barcelona, 1991, págs. 7 y 8.

> sentido o sinsentido de esta entera existencia humana (...) ¿Qué tiene la ciencia que decirnos sobre la razón y sinrazón, qué sobre nosotros, los seres humanos en cuanto sujetos de esta libertad? (...) Nada tiene, evidentemente, que decirnos puesto que ha hecho abstracción de todo lo subjetivo»[49].

A partir de entonces y avalado por los éxitos del positivismo, el maniqueo e irreductible reparto de roles según el cual «la ciencia sigue siendo intrínsecamente buena, benefactora y moral, la técnica ambivalente (...) y la política, sí mala» de tal modo que «las utilizaciones malas de la ciencia se deben a la política» pervive aún en muchos foros. «Las mentes formadas por un modo de conocimiento que repudia la complejidad, por tanto, la ambivalencia, no saben concebir la ambivalencia inherente a la actividad científica, en la que conocimiento y manipulación son las dos caras del mismo proceso. Más generalmente, la mentalidad formada en un modo de pensamiento binario, que excluye la ambigüedad, no puede concebir que la ciencia sea a la vez "buena" y "mala", benefactora y perversa, útil y nefasta»[50].

La condición de conocimiento exclusivamente «objetivo» de la ciencia, al que se refería Husserl —hoy de la tecnociencia— genera dos efectos «emparentados». Por un lado, una «ceguera ética» como consecuencia de la deliberada desatención de lo subjetivo, requisito y condición de su meteórico ascenso y, por otro, el «aprovechamiento» de tal «autorrestricción» por parte de la política y la economía. «La ciencia, aventura desinteresada, es captada por los intereses económicos; la ciencia, aventura apolítica, es captada por las fuerzas políticas (...) Los científicos están totalmente desprovistos de los poderes que, sin embargo, emanan de sus laboratorios; estos poderes están reconcentrados entre las manos de sus empresas y de los poderes estatales»[51]. A día de hoy, no cabe duda alguna, «la ciencia es omnipresente, en interacciones-retroacciones innumerables en todos los dominios, creadora de poderes gigantescos, y totalmente impotentes para controlar sus poderes»[52].

En la actualidad, en nuestras sociedades, gracias al triunfo de lo operativo, la tecnociencia ha pasado a ser objeto de devoción a tal punto que «no es

---

49 *Ibídem*, pág. 6. La extensión de las citas de esta obra de Husserl se justifica por la importancia para el tema que aquí nos ocupa pues el autor apunta al origen de las causas del divorcio entre filosofía y ciencia: si solo es tenido por verdadero lo constatable empíricamente cuestiones como la autonomía y la libertad con la consiguiente carga de responsabilidad quedan por completo fuera del interés de la ciencia, se externalizan y al hacerlo o bien se vuelve la mirada a la Filosofía y la ética o bien se pierde el control del futuro.

50 Morin, E., *El Método, 6, Ética*, *op. cit*., pág. 80.

51 *Ibídem*, págs. 78 y 79.

52 *Ibídem*.

posible disociar las conquistas científicas de la ley del vivir»[53]. De esa indisociabilidad contemporánea el exponente más significativo es la I.A., que, con sus claroscuros, encarna a la perfección las fortalezas y debilidades que todo progreso lleva inherente. Pero ¿cuánto hay de punto ciego en la I.A.? ¿Conviene resignarse ante esa ceguera? ¿Cómo opera la I.A. y su ceguera sobre la cuestión de nuestra identidad? La identidad es el establecimiento de la no confusión. Afirma Legendre que «la identidad del humano resulta de una composición, de un ensamble». La tecnociencia, con la I.A. a la cabeza, es una de las piezas centrales de esa composición. Si la operatividad es la coartada para que sea la I.A. la que, en los más variados ámbitos, decida por nosotros, ¿en qué lugar queda nuestra identidad?, ¿será subalterna a ella?, ¿podrá seguir siendo autónoma si se confunde con ella? Al asomarnos a un mundo en el que la subjetividad será cada vez menos diferenciada, la posibilidad de generar pensamiento original, novedoso y por supuesto heterodoxo, será cada vez menor. Elevar la «tecnociencia» a la categoría de juez y parte de nuestro futuro no parece ni lo más justo ni lo más inteligente.

IV. Personas o máquinas ¿quién y por qué debe hacer justicia?

Al referirse a los efectos generados por la I.A. en el Derecho y en la Administración de justicia se ha convertido en un lugar común hablar de cambio de paradigma. Conocidos y muy reseñados son los riesgos que comporta especialmente en lo que a defensa y protección de los derechos humanos concierne. En este ámbito serían dos los principales problemas que presenta. Uno, el inconsciente o indeseado, pero no por ello menos grave, de los sesgos que, o bien por desconocimiento o bien por negligencia, implica un trato discriminatorio o desigualitario de las personas (racismo, xenofobia, machismo, clasismo...) al no haber reflejado correcta y justamente la diversidad humana en los datos que sirven de base al algoritmo. Y dos, porque existen y cada vez son más y más extendidos los programas de I.A. que desarrollan lo que se denomina aprendizaje continuo de manera autónoma. Lo característico de ellos es que «el rendimiento del sistema de I.A. evoluciona con el tiempo sin intervención humana a medida que se dispone de nuevos datos. Esto contrasta con los sistemas de I.A. cuyos algoritmos de producción son actualizados periódicamente»[54].

53 Legendre, P., *El Tajo. Discurso a jóvenes estudiantes sobre la ciencia y la ignorancia*, *op. cit*., pág. 60.

54 Benjamins, R., «Hacia una I.A. sostenible: una perspectiva 360 incluyendo negocio, sociedad, ética y cambio climático» en Peralta, A., Torres, L., y Herrera, F. (Coords.), *El derecho y la Inteligencia Artificial*, Granada, Editorial Universidad de Granda, 2022, pág. 123.

Conscientes de los riesgos que ello puede suponer para el marco de convivencia del Estado de Derecho son varias las iniciativas que, procedentes de distintas instituciones, tratan de evitar las consecuencias negativas que la implementación de la I.A. pueda suponer[55]. Sin embargo, como afirma Muñoz Vela, lo cierto es que «el derecho vigente no contempla marcos específicos de seguridad de la I.A.»[56]. O, lo que es lo mismo, no existe un derecho vinculante *erga omnes* en lo que a la I.A. se refiere. No en el plano internacional y apenas en el nacional como más adelante se expondrá.

La irreversible llegada de la I.A. a la esfera jurídica nos obliga a reformular los derechos y a plantearnos algunos interrogantes. Nuestra premisa es que la finalidad más importante del Derecho es la de garantizar los derechos humanos y fundamentales pero la cuestión es ¿cómo hacerlo ahora habida cuenta de las ventajas y desventajas que conlleva la I.A.? Si, como creemos, no hay otra perspectiva posible que la de los derechos humanos, el Derecho para «auto-protegerse» de los peligros inherentes a un uso «descontrolado» de la I.A. ha de ir de la mano de la ética. ¿Significa eso que hemos de desarrollar una ética exclusiva y privativa para la I.A.? No, creemos, con Adela Cortina que no se trata de crear «una ética de los sistemas inteligentes, sino de cómo orientar el uso humano de estos sistemas de forma ética»[57]. ¿Y puede esperase ética por parte de las máquinas? Es evidente que el componente ético ha de provenir de la acción humana y ha de hacerlo en las diversas fases en que interviene la I.A. Desde el momento en que la minería de datos *(data mining)*, basándose en la predictibilidad de las acciones humanas, realiza la búsqueda para establecer un patrón[58] hasta el «juez artifcial»[59].

Si, como se viene observando, no dejamos de plantearnos dudas y preguntas acerca de la idoneidad del uso de la I.A. en el ámbito jurídico es, en primer lugar, porque fuera de él, en las esferas propiamente científicas (medicina, ciencias de la tierra, ecología...) su aportación es incuestionable y, en segundo, porque que sea o no una valiosa herramienta en la esfera jurídica va a depender de que se consideren o no una serie de limitaciones y cautelas de carácter ético y jurídico. Nos serviremos de un ejemplo para mostrar cuáles son los principales peligros a que se somete un sistema de derechos y libertades cuando se opta por confiar a programas de I.A. determinadas decisiones en el proceso judicial.

---

55 A la cabeza de dichas instituciones se hallaría la Unión Europea pero también son destacables las acciones lideradas por la OCDE o la UNESCO en el horizonte internacional o el OdiseIA (Observatorio del impacto social y ético de la I.A.) en el nacional.

56 Muñoz Vela, J. M., *Retos, riesgos, responsabilidad y regulación de la I.A. Un enfoque de seguridad física, lógica, moral y jurídica*, Aranzadi, Cizur Menor, 2022, pág. 256.

57 Cortina, A., «Ética de la Inteligencia Artificial», *Anales de la Real Academia de Ciencias Morales y Políticas*, n.º 96, 2019, pág. 386.

58 Al respecto *cfr.* Latorre, J. I., *Ética para máquinas*, Ariel, Barcelona, 2019, págs. 91-95.

59 Martínez Zorrilla, D., «El juez artificial: ¿próxima parada?», *Oikonomics*, n.º 12, 2019, pp. 1-12.

Con el fin de reducir los gastos de la jurisdicción penal estadounidense no ha dejado de crecer la aplicación de instrumentos predictivos en distintas áreas y fases del procedimiento. En ese contexto, el programa más extendido tanto en la jurisdicción estatal como las locales es el *COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)*. Se trata de un programa de evaluación de riesgos propiedad de una compañía privada[60] que, al considerar que su algoritmo es un secreto industrial, no desvela de qué modo se calcula la puntuación adjudicada a cada acusado. Dicha puntuación resulta de todo punto determinante pues establece el riesgo de reincidencia del acusado, dato (totalmente opacado) que será tomado en consideración por el juez para adoptar su decisión al respecto. El Estado de Wisconsin es uno de los Estados que en USA recurre a dicho programa, de los casos en que fue empleado el de *State versus Loomis*[61] es, hasta hoy, el que más controversia ha despertado a causa de las reclamaciones judiciales interpuestas, las reacciones sociales suscitadas[62] y, sobre todo, porque sentó un importante precedente en la Justicia de aquel país. Sobre la base del alto riesgo de incidencia fijado por Compas, así consta en la motivación del juez, Eric L. Loomis fue condenado a seis años de prisión más cinco de libertad condicional.

El abogado de Loomis impugnó tal decisión ante el Tribunal del Estado de Wisconsin alegando tres motivos:

> 1.º su derecho a la información del modo en que funciona el algoritmo para preparar y ejercer su defensa (el secreto empresarial impide conocer cómo funciona COMPAS)
>
> 2.º su derecho a la imparcialidad, COMPAS incluye datos como el sexo o la raza para establecer sus predicciones
>
> 3.º su derecho a una sentencia individualizada[63]

60 El nombre actual de la compañía es Northpointe Inc. y el programa COMPAS comenzó a usarse en USA en 1998. Su algoritmo revisa 137 parámetros del sujeto para predecir sus posibilidades de reincidir en el futuro y otorga una puntuación del 1 al 10: riesgo bajo (1-4): riesgo medio (5-7); alto riesgo (8-10). Para una información completa sobre cómo se confecciona el COMPAS *cfr.* Roa Avella, Marcela del P., Sanabria Moyano, J. E. y Dinas-Hurtado, K., «Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos», *Revista Brasileira de Direito Processual Penal*, Vol. 8, n.º 1, 2022, especialmente págs. 275-293.

61 *Statte v. Loomis*, 881 N.W. 2d 749 (Wis. 2016).

62 La organización independiente y sin ánimo de lucro *ProPublica* (https://www.propublica.org/) realizó un estudio sobre el algoritmo COMPAS en el que advirtió «la probabilidad de una discriminación algorítmica hacía los individuos afrodescendientes, lo que conllevaría a un posible riesgo de vulneraciones de los derechos humanos», Roa Avella, Marcela del P., Sanabria Moyano, J. E. y Dinas-Hurtado, K., «Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos», *op. cit.*, pág. 291.

63 Solar Cayón, J. I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de la evaluación de riesgos», *op. cit.*, pág. 138 y ss.

Lo que se reclamaba, en definitiva, era el acceso al código fuente del algoritmo dado que, de otro modo, la carga probatoria que recae sobre el acusado es injustificada y desproporcionada. Sin embargo, la impugnación fue desestimada por el alto Tribunal al anteponer el secreto empresarial al derecho a la transparencia[64]. Como suele ocurrir lo complejo en el mundo jurídico se da cuando, como en este supuesto, hay conflicto de derechos: el de defensa frente al secreto empresarial.

Esa decisión no acaba de casar bien con la adoptada por la Corte Suprema de los EE. UU. en 1963 en el caso *Brady versus Maryland*[65] en el que se establecía que el fiscal ha de entregar tanto posibles pruebas exculpatorias como todas aquellas que puedan resultar favorables al reo cuando éste las haya requerido (petición conocida hoy como «solicitud Brady»). La denegación al código fuente puede ser interpretada como una denegación de prueba favorable ya que imposibilita el derecho a la legítima defensa y al debido proceso.

«Compas» es solo uno de los muchos programas que utilizan el aprendizaje automático con el objetivo de predecir delitos antes de que se cometan[66]. Se llaman *Crime Predicting Software* y se ocupan de recopilar y analizar datos sobre crímenes del pasado, para predecir los futuros. La de los *software* predictivos es una industria en rápido crecimiento. De hecho, como dice el experto Solar Cayón, «la automatización está contribuyendo a intensificar la privatización de la justicia penal estadounidense en todas sus etapas»[67].

Frente a los problemas jurídicos que el uso de esa I.A. muestra pueden defenderse dos tipos de contraargumentos. Uno, que la solución pasa por la correcta y completa recolección de los datos para que el algoritmo no resulte discriminatorio. Y dos: que, al menos hasta hoy, la predicción del algoritmo no es vinculante para el juez. Respecto al primero, para evitar sesgos, de cualquier clase, sería necesario que los datos fuesen recabados por un equipo interdisciplinar integrado no solo por matemáticos, físicos e ingenieros sino también por juristas, sociólogos, filósofos y éticos. No obstante, si bien es verdad que, de ese modo, podría solucionarse el grave problema de los sesgos tampoco de ese modo se resolvería el «delicado» asunto de la opacidad del funcionamiento del algoritmo resultante. Por lo que concierne al segundo de los argumentos, no parece descabellada la opinión del profesor Surden, según la cual «si un juez recibe un informe que indica que un acusado tiene un

64 El caso, a través del recurso de la sentencia del Estado de Wisconsin, llegó al Tribunal Supremo de los EE.UU. pero ello no supuso ninguna variación pues dicho Tribunal inadmitió el recurso presentado por el abogado de Loomis el 26 de junio de 2017.

65 *Brady v. Maryland*, 373 U.S. 83 (1963)

66 Otros algoritmos similares a COMPAS son el PSA, el Lsi-R, HART, CAS, NDAS, Roa Avella, M. del P, Sanabria Moyano, J.. E. y Dinas-Hurtado, K., «Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos», *op. cit*., pág. 303.

67 Solar Cayón, J. I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de la evaluación de riesgos», *op. cit*, pág. 144.

80'2 % de probabilidades de volver a delinquir según el modelo de aprendizaje automático, esa predicción tiene un aura de neutralidad e infalibilidad mecánica»[68] de la que resulta muy difícil sustraerse y motivar su desistimiento. Si eso es lo que ocurre durante el proceso, ¿cómo confiar a los algoritmos la decisión última? Ante ello una de las respuestas viene a decir: «corríjanse las deficiencias del algoritmo y resuelto el problema» pero ¿Cómo se construye la decisión? ¿Qué pasa con la opacidad, la individualidad y la motivación? ¿Quién y cómo prevé las consecuencias de la decisión?

Por lo que se refiere a la construcción de la decisión judicial no debe pasarse por alto que, ésta va siempre «precedida de una meta-construcción lógica». Lo que significa que, con anterioridad a la decisión misma, tiene lugar una serie de «operaciones lógicas (...) si se detecta conflicto de principios, de intereses o de derechos, hay que desarrollar una cuidada ponderación, sopesando en el caso concreto qué principio o qué derecho debe primar sobre otro (y motivarlo) y se debe trabajar con conceptos jurídicos indeterminados o conceptos abiertos»[69]. ¿Pueden incluirse todas esas acciones en un algoritmo habida cuenta de que cada caso será único?

En lo que a las consecuencias respecta, ¿puede la matemática resolver cuál es la decisión más adecuada de entre las posibles que el Derecho permite atendiendo a las consecuencias derivadas de una u otras? Si a lo que se aspira es a que las decisiones judiciales no solo sean correctas desde el punto de vista estrictamente formal, sino que además sean consistentes y coherentes, tanto con el sistema jurídico como con la realidad a la que dan respuesta, no puede pasarse por alto la consideración de las consecuencias a que conduce cada decisión[70]. Y todo ello: corrección formal, coherencia sistemática, consistencia jurídica y atención a las consecuencias interactúa en la argumentación que el juez comienza a lleva a cabo en cuanto ha completado su «meta-construcción» lógica.

En el fondo, la premisa que impulsa el desarrollo y aplicación de la I.A. en la esfera judicial es la de concebir la tarea de juzgar como un mero automatismo. Un automatismo proveniente o bien del precedente, en el caso de la *Common Law*, o bien del silogismo jurídico en el del *Civil Law*. Pero esa premisa sobre la que descansa la I.A. «preparada» para el Derecho es una premisa falsa. Y lo es porque, como dice el profesor Martínez García, la decisión judicial y, más allá de ella, la respuesta jurídica «no es una mera cita de un dato normativo sino producto de una operación decisoria. Tiene que ser

68 Surden, H., «Artificial Intelligence and Law: An Overview» en *Georgia State University Law Review*, vol. 35, n.º 4, 2019, pág. 1336.

69 Belloso Martín, N., «Inteligencia artificial en la teoría de la decisión judicial. ¿Del juez-robot al asistente-robot del juez?» en Peralta, A., Torres, L., y Herrera, F. (Coords.), *El derecho y la Inteligencia Artificial*, Editorial Universidad de Granada, Granada, 2022, pág. 353.

70 Maccormick, N., *Legal Reasoning and legal Theory*, Oxford, Clarendon Press, 1978, págs. 80, 118 y 119 entre otras.

«conforme a derecho», «ajustada a derecho», pero no se limita a copiarlo. No es copia, reiteración, sino reproducción. La aplicación del Derecho no corresponde a una estática sino a una dinámica productiva que lo recrea. Hay que inventar la decisión y hacerlo en el marco de la legalidad»[71].

No decimos nada nuevo cuando afirmamos que el recurso al *big data* y a los programas de I.A. que a partir de ellos se elaboran hace que nuestro pensamiento sea más superficial. Del mismo modo que la cultura es más que un agregado de información, la cultura jurídica es mucho más que un repositorio digital de normas y sentencias. Y en todo caso, se trate de pensamiento jurídico o no, el pensamiento humano es cualquier cosa menos lineal pero lineal es el modo en que «trabaja» y decide la I.A. la aplicada en la esfera jurídica y en cualquier otra[72]. Por no hablar de que hemos de tomar conciencia clara sobre la trascendencia de que «siempre que utilizamos una herramienta para ejercer un mayor control sobre el mundo exterior, cambiamos nuestra relación con ese mundo»[73].

Estos hechos y las consideraciones previamente referidas no parecen frenar el impulso que en determinados países está alcanzando la aplicación de I.A. en el ámbito judicial, el de China es el caso más claro, pero no el único[74]. Por lo que se refiere a la Unión Europea, la viabilidad de la implantación de decisiones individuales automatizadas ha de respetar lo previsto en el apartado 1 del artículo 22 del Reglamento General de Protección de datos de 2016[75]:

> «''Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar''.
>
> Sin embargo, en buena medida tal garantía se vacía de contenido en virtud de lo dispuesto en el siguiente apartado
>
> ''El apartado 1 no se aplicará si la decisión:

71 Martínez García, J. I., «La respuesta jurídica», *Anuario de Filosofía del Derecho*, (XXXVI), 2020, pág. 361.

72 Carr, N., *Superficiales ¿Qué está haciendo internet con nuestras mentes?*, tr. P. Cifuentes, Madrid, Taurus, 2010, pág. 84.

73 *Ibídem*, págs. 236, 238 y 255.

74 «Las *Smart Courts* chinas, implementadas en los tribunales superiores de Beijing y Hainan, por ejemplo, resuelven litigios mediante el uso de sistemas predictivos relativos al comercio electrónico e internet, sin que haya habido ningún proceso de distinción entre casos sencillos y complejos, ni ningún debate social sobre la cuestión», De Asis Pulido, M., «La justicia predictiva: tres posibles usos en la práctica jurídica» en Llano Alonso, Fernando (Dir.), *Inteligencia artificial y filosofía del Derecho*, Murcia, Laborum, 2022, pág. 299.

75 Reglamento (UE) 2016/679 del Parlamento europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos

a) es necesaria para la celebración o la ejecución de un contrato entre el interesado y un responsable del tratamiento;

b) está autorizada por el Derecho de la Unión o de los Estados miembros que se aplique al responsable del tratamiento y que establezca asimismo medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado».

En nuestro país, a tenor de lo previsto en el artículo 117 de la Constitución en el que la potestad jurisdiccional es atribuida en exclusividad a jueces y magistrados la asignación de resoluciones judiciales, a programas de I.A. requeriría una previa reforma constitucional, con independencia de cuál fuera cuantía objeto de litigio o el grado de complejidad del proceso en cuestión[76].

Ello, sin embargo, no obsta para que el Proyecto de Ley de Medidas de Eficiencia Digital del Servicio Público de Justicia[77] disponga lo siguiente respecto a las actuaciones automatizadas, proactivas y asistidas:

Artículo 56. Actuaciones automatizadas.

Artículo 57. Actuaciones asistidas.

Artículo 58. Requisitos comunes de las actuaciones automatizadas, proactivas y asistidas.

1. En caso de actuación automatizada, asistida o proactiva podrá realizarse por el Comité técnico estatal de la Administración judicial electrónica la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso, la auditoría del sistema de información y de su código fuente.

2. Los criterios de decisión serán públicos y objetivos, dejando constancia de las decisiones tomadas en cada momento.

3. Los sistemas incluirán los indicadores de gestión que se establezcan por la Comisión Nacional de Estadística Judicial y el Comité técnico estatal de la Administración judicial electrónica, cada uno en el ámbito de sus competencias.

En virtud de ello, podemos concluir que, en España, al menos por el momento, la aportación de la I.A. en la Administración de Justicia sólo podrá ser exclusiva en las labores de trámite, esto es, que no requieran de ningún tipo de interpretación y que, para aquellas otras que sí lo precisen, la última palabra la tendrá siempre el juez, siendo el documento elaborado por la I.A. un mero borrador, no vinculante, que puede servir de apoyo a la resolución judicial pero nunca una resolución judicial definitiva. Sin embargo, sería ingenuo pensar que el problema queda así zanjado pues, en tales supuestos, como en el conjunto de los denominados *decision support systems* «la frontera entre

76 Salvo que en una interpretación extensiva se considerase que un juez-robot es tan juez como el no robot.

77 Boletín Oficial de las Cortes Generales (BOCG) 14-A-116-1, de 12 de septiembre de 2022.

algoritmos de auxilio al juez y algoritmos que sugieren una decisión puede ser muy sutil», tal sutil que, en la práctica, su diferenciación puede resultar imposible[78].

Sea como fuere, coincidimos con De Asís Pulido[79] en que no podemos dar el asunto por cerrado pues podría suceder que, en el futuro, el legislador decidiera aceptar la figura del juez-robot. En tal caso sería de aplicación el artículo 42 de la Ley 18/2011, de 5 de julio, reguladora del uso de las tecnologías de la información y la comunicación en la Administración de Justicia[80]. Lo que significa que:

> «En caso de actuación automatizada, deberá establecerse previamente por el Comité técnico estatal de la Administración judicial electrónica la definición de las especificaciones, programación, mantenimiento, supervisión y control de calidad y, en su caso la auditoría del sistema de información y de su código fuente.
>
> Los sistemas incluirán los indicadores de gestión que se establezcan por la Comisión Nacional de Estadística Judicial y el Comité técnico estatal de la Administración judicial electrónica, cada uno en el ámbito de sus competencias».

Huelga decir que esa ley, como todas, puede ser modificada o derogada, por lo que, en última instancia, como se ha indicado, habrá de estarse a lo previsto en los preceptos constitucionales en vigor, presentes o futuros. Pero, como Cristian Román, somos de la opinión de que el empleo de la IA en la Justicia no supondrá, como los «tecnoptimistas» afirman, el reemplazo del juez-humano por el juez-robot, mas si, una muy importante colaboración y asistencia a aquél, y que englobamos en el concepto asistente-robot del juez»[81]. Y lo creemos porque «manejar solo procedimientos mecánicos, fórmulas y algoritmos, puede ser como juzgar a ciegas»[82], en el sentido de que, por muy sofisticados que sean tales procedimientos, carecen de la perspectiva holística que solo el ser humano puede aportar.

4.1. Inteligencia artificial: una vuelta de tuerca al dilema formalismo-antiformalismo jurídico

Reservamos un pequeño *excursus* para mostrar cómo, a nuestro juicio, la aplicación de la I.A. al ámbito jurídico supone una reactivación del forma-

78 Belloso Martín, N. «Inteligencia artificial en la teoría de la decisión judicial. ¿Del juez-robot al asistente-robot del juez?, *op. cit*., pág. 354.

79 De Asís Pulido, M., «La justicia predictiva: tres posibles usos en la práctica jurídica», *op. cit*., págs. 298. y 299.

80 B.O.E. n.º 160, de 6 de julio de 2011.

81 Román, C., «Inteligencia Artificial en la Justicia (Del juez-robot al asistente-robot del juez)», *Diario Constitucional Santiago de Chile*, 8-11-2021.

82 Martínez García, J. I., «La respuesta jurídica», *op. cit.,* pág. 371.

lismo jurídico. Dicho eso, no creemos, como Bobbio, que un reproche generalizado e indiscriminado al formalismo jurídico esté justificado. En primer lugar, porque, *ab initio*, es necesario precisar a qué formalismo nos referimos cuando afirmamos, con afán crítico, que la I.A. implica un renovado formalismo jurídico. Y, en segundo, porque la forma contribuye con valores propios como el orden, la permanencia o la coherencia a la función estabilizadora del Derecho. Una función nada desdeñable pues sin ella los valores materiales que el Derecho encarna carecerían de una eficaz y legítima protección.

Cuando censuramos que se confíe a la I.A. ciertas competencias en el ámbito jurídico y especialmente judicial es porque la concepción del Derecho que subyace es formalista en un doble sentido. El primero de ellos porque se asume una acepción reduccionista del término «forma» como «continente que no cambia con el cambio de contenido»[83], como un recipiente que permanece ajeno a cuanto contiene. Pero en lo que al Derecho concierne, la forma impregna el fondo y el fondo a aquélla, pues lejos de ser elementos inconexos o estancos, lo que sucede en uno acaba filtrándose al otro, y es que forma y fondo son en el Derecho una sola, aunque compleja cosa. El segundo de los sentidos hace referencia a que, si la idea del Derecho de que se parte es única y exclusivamente normativista, esto es, si estrechándola concibe «la experiencia jurídica como experiencia *sub specie legis*»[84] acabará por construir una ciencia jurídica dogmática «que se resuelve en una reducción de los comportamientos a las estructuras y en una continua constitución y reconstitución de las estructuras mismas»[85]. O, lo que es igual, una investigación científica del Derecho que vive a espaldas de cuanto sucede con la eficacia de las normas que analiza o el valor de las mismas. Una ciencia resignada que se conforma con verse limitada a una parte de lo que realmente es o podría ser.

Confiar a programas «jurídicos» de I.A. algo más que la búsqueda de información implicaría cambios sustantivos en lo que el Derecho es. La alusión a las fases del Derecho —creación, interpretación, argumentación y aplicación— como fases independientes responde, por una parte, a una clara intencionalidad didáctica y, por otra, al deseo de identificación de los responsables de las mismas. A grandes rasgos puede decirse que mientras que la creación es cosa del legislador (por supuesto con permiso de las numerosísimas e importantes normas de derecho privado que provienen del ciudadano), siendo su producto, la ley en sentido amplio la primera y principal fuente del Derecho, el resto corresponde al agente jurídico encargado de dar solución a una petición administrativa, demanda o conflicto judicial. Sin embargo, es preciso matizar y tomar con cautela tan estanca división. Para empezar por-

83 Bobbio, N., *El problema del positivismo jurídico*, tr. E. Garzón Valdés, Fontamara, México D.F., 2004, pág. 22.

84 *Ibídem*, pág. 23.

85 *Ibídem*, pág. 25.

que no sería en absoluto aplicable al *Common Law*, puesto que allí, siendo el precedente judicial la primera fuente del Derecho, autor y aplicador coinciden. Y, para continuar, porque en la familia romano-germánica o de *Civil Law*, el intérprete, omnisciente, más que reproducir la norma en cada resolución, crea nuevas versiones de la misma en atención a lo dispuesto en el artículo 3 del Código Civil, según el cual «las normas se interpretarán según (...) la realidad social del tiempo en que han de ser aplicadas». No es el tiempo y la realidad social del legislador los que cuentan sino cada momento en que, al aplicarse, la norma cobra vida y, habida cuenta de la rapidez con que se suceden los cambios en nuestra sociedad, las interpretaciones no cesan de actualizarse. Actualización que, dadas la independencia y discrecionalidad judicial, no son necesariamente idénticas. Si a ello sumamos lo previsto en el apartado segundo del citado artículo: «la equidad habrá de ponderarse en la aplicación de las normas» nos encontramos ante un panorama abierto, pues la de equidad es más que un concepto una idea indeterminada imposible de anticipar de forma definitiva y, por tanto, imposible de en-cerrar en un algoritmo.

Pero, como se mencionaba, son más los cambios sustantivos a los que se aludía. Entre ellos es el de que el formalismo, o normativismo clausurado, al que responde la I.A. creada para el Derecho supone desandar el camino emprendido por la sociología jurídica. Un camino compartido por la teoría del Derecho y la sociología jurídica que, de la mano, han tratado —y conseguido— superar las limitaciones propias de un positivismo formalista decidido a aislar el hecho jurídico de la realidad social y moral a la que se debe y de la que es parte. Gracias a la teoría de la argumentación, parte ya sustantiva de la Filosofía del Derecho, y al garantismo antiformalista las consideraciones sobre la eficacia y justicia de las decisiones jurídicas pueden ocupar el lugar central que merecen. Siempre, claro está, que no se conciba lo social como agente contaminante de la pureza del Derecho y su ciencia.

Y aún habría de añadirse otro que incumbe a todos demás, la sustantiva repercusión del formalismo ético sobre la Justicia, su teoría y su administración. Como afirma Bobbio, debería llamarse «formalismo ético» a esa modalidad del formalismo jurídico que en particular afecta a la teoría de la justicia y según el cual «acto justo es aquel que es conforme a la ley, e injusto aquél que está en desacuerdo con ella». Y debería denominarse así porque «tiene en común con todas las teorías formalistas de la ética, la afirmación de que el juicio ético consiste en un juicio de conformidad de un acto con la norma, de donde bueno es el acto realizado para cumplir con la ley, y malo aquel que se realiza para transgredirla»[86]. Tal identificación de la norma con el bien, o el Derecho con la Justicia, supera las más complacientes nociones de Derecho y fosiliza la indefectible adaptación histórica de los valores.

86 Bobbio, N., *El problema del positivismo jurídico, op. cit*., pág. 13.

A la vista de lo anteriormente expuesto parece llegado el momento de esbozar unas brevísimas consideraciones finales:

- Creemos que la aplicación indiscriminada de una I.A. indistinta, y falta de rigor en la esfera jurídica, y muy especialmente judicial, conduciría a un empobrecimiento y, en consecuencia, a una cierta precariedad de lo jurídico
- Dada la presencia actual de la I.A. en el ámbito del Derecho, y su más que previsible aumento en el futuro inmediato, se hace preciso negociar nuevos equilibrios entre el hombre y la tecnociencia, huyendo de fascinaciones y catastrofismos, sin más guía que la protección y garantía de los derechos humanos y fundamentales
- En el hipotético caso de que acaben otorgándose a la I.A. competencias, y por tanto poderes, que resten los del ser humano, único agente capaz de autonomía de la voluntad y libertad y, en consecuencia, de responsabilidad, habremos de plantearnos añadir a las «revueltas contra el formalismo» estudiadas por el profesor Treves[87] una nueva revuelta académica contra toda I.A. que vaya más allá de un papel auxiliar.

V. Algoritmos, valores y sesgos: algunos problemas

Volvamos al principio, a la definición misma de lo jurídico. Si la nuestra es una reflexión en torno a la IA desde la filosofía del Derecho conviene establecer, si quiera muy someramente, cuál es lugar desde el que se parte pues no es extraño incurrir en el error de dar por sentadas y obvias consideraciones que van a condicionar tanto la mirada como el objeto observado.

La primera dificultad con que nos topamos y cuyo tratamiento no debe eludirse es la de aclarar qué cosa sea eso del Derecho o si se prefiere de la juridicidad. Desde fuera del ámbito jurídico tiende a pensarse que el Derecho es un objeto conocido y reconocible por todos, un objeto, por tanto, cuya naturaleza no supone ningún misterio ni para los entendidos ni para los ajenos. Una materia cuya única o principal dificultad estribaría en su inabarcable conocimiento tanto por los extenso de sus contenidos como por lo acelerado de sus cambios. Si bien es verdad que «lo jurídico» existe a nivel global no lo es menos que, a diferencia de otras disciplinas (medicina, física, química...),

87 No es esta la ocasión de detenerse en la esencial pugna formalismo-antiformalismo, por lo que nos limitaremos a recordar el de Renato Treves como uno de los análisis más clarificadores al respecto. Como él señala desde un abarcador enfoque histórico, el sempiterno debate formalismo-antiformalismo se ha sustentado hasta ahora a través de las siguientes disyuntivas: el antiformalismo científico francés frente al formalismo jurídico de la Escuela de la Exégesis; el Movimiento del Derecho Libre Alemán frente al formalismo de la Jurisprudencia de conceptos y antiformalismo estadounidense desde Holmes hasta los Critial Legal Studies frente al formalismo jurisprudencial. Treves, R., *La Sociología del Derecho: orígenes, investigaciones, problema*s, trs. de M.ª J. Añon Roig, M. Atienza y J.A. Pérez Lledó, Ariel, Barcelona, 1998, pág. 81 y ss.

ciertos conceptos jurídico-políticos como la soberanía y por ende territorialidad son los que determinan tanto su forma como su fondo. Esta junto con la especialización serían dos condicionantes variables que otorgan al Derecho una especificidad única.

Sobre el complejo y controvertido asunto de la juridicidad, esto es, el elemento que convierte a algo en Derecho, existen varias teorías que podrían agruparse en tres según cuál sea el criterio o conjunto de criterios que conciban como determinantes: los formales (procedimiento, competencia y autoridad), los materiales (sociabilidad) o la obligatoriedad de las reglas. Si bien es verdad que no han de entenderse en términos puros o excluyentes pues los tres se requieren y retroalimentan entre sí, cada uno de los tres pone mucho mayor énfasis en el aspecto que considera determinante. En todo caso, sea como fuere, ninguno prescinde del elemento moral pues la vinculación entre el Derecho y la moral, por mucho que a lo largo de la historia y las diferentes teorías haya sido considerada de un modo u otro, nadie, cualquiera que sea la corriente de pensamiento a la que se pertenezca (iusnaturalismos por supuesto, pero también realismo e iuspositivismo) la ha negado. Cosa diferente es qué valor se le haya dado y cuál el grado de conexión reconocido entre ambos.

Pues bien he ahí un primer grupo de dificultades con el que nos encontramos, las de averiguar si la IA es capaz:

> 1. de comprender el contenido y alcance de los valores morales que todo ordenamiento protege a través de sus reglas, repartidas estas por los diversos sectores del sistema jurídico, aunque de forma especialmente notoria en los textos constitucionales;
>
> 2. de interpretar tales principios y valores con arreglo al momento en que hayan de ser aplicadas las normas, según prescribe el artículo 3 del Código Civil, lo cual exige una actualización dinámica de los mismos;
>
> 3. de problematizar y resolver de forma ponderada los conflictos entre principios y valores morales;
>
> 4. y de comprender la cualidad de la equidad habida cuenta de que el art. 3.2 del Código Civil dispone que «la equidad habrá de ponderarse en la aplicación de las normas, si bien las resoluciones de los Tribunales sólo podrán descansar de manera exclusiva en ella cuando la ley expresamente lo permita».

Otro de los tópicos desde el que suele concebirse simplificadamente el Derecho es el de partir de la premisa de que el caso que ocupa al jurista, el caso que ha de resolverse estará contemplado expresamente por el sistema jurídico en cuestión. En el peor de los supuestos, caso de no estarlo, la interpretación extensiva, la analogía o los precedentes judiciales cubrirán la laguna jurídica, que no legal, de modo que el conflicto objeto de litigio se resolverá una vez conseguida la correspondiente.

Tampoco es esta una verdad incuestionable en el Derecho. De serlo, el camino para la aplicación de la IA se facilitaría de manera considerable. Lo

cierto es que junto a esa situación ideal han de ser tomadas en consideración otras. En esta ocasión nos referiremos fundamentalmente a dos. La primera es la de la especial consideración y desempeño del principio de legalidad en el ámbito penal. En dicha materia en virtud del *principio nullum crimen, nulla poena sine praevia lege poenali*, no cabe ni la aplicación extensiva ni la analógica, tal y como dispone el art 4. 2. del Código Civil «Las leyes penales, las excepcionales y las de ámbito temporal no se aplicarán a supuestos ni en momentos distintos de los comprendidos expresamente en ellas». El Derecho penal es precisamente uno de los sectores del Derecho positivo donde más importancia adquieren los valores constitucionales habida cuenta de que a él se le encomienda la protección y salvaguarda de algunos de los más importantes derechos fundamentales y, por ende, de los principales bienes jurídicos protegidos (vida, libertad, integridad...). La necesidad de realizar complejos juicios de valor para calibrar el alcance y aplicación de los valores y principios en que se inspira el ordenamiento jurídico es, por el momento un coto vedado, desde el punto de vista técnico para la IA. Más adelante se analizará si aun siendo factible en el futuro sería deseable asignarles tal cometido.

La segunda de las situaciones a la que se hacía referencia es de una mayor enjundia filosófico-jurídica. Los supuestos a los que se acaba de hacer mención, aquellos en los que el supuesto de hecho real es subsumible en el supuesto de hecho contenido en una regla jurídica en vigor, son los denominados *casos fáciles*. Y se les califica de ese modo porque la consecuencia jurídica prevista para tal situación se halla específicamente contemplada en el sistema jurídico. Ello puede suceder o bien en la misma norma en que se prevé la situación, en cuyo caso nos encontramos ante una norma completa, o bien en una norma diferente, norma incompleta. La resolución de los casos en que se da dicha circunstancia requiere únicamente de una justificación interna. Esto es la justificación que necesariamente acompaña la respuesta jurídica se produce en el interior del ordenamiento. El jurista no precisa en ningún momento abandonar el Derecho vigente para poner fin al contencioso. En cambio, junto a estos casos, los fáciles, existen otros, a los que se llama, por oposición, *«difíciles»* porque el supuesto con el que el jurista se encuentra no se halla expresa o claramente contemplado en la ley (laguna legal) y, si se trata de un caso penal, tal y como se ha indicado, no cabe ninguna clase de extensión de la letra escrita de la ley (laguna jurídica).

Si a ello se añade que, a tenor del artículo 1.7 del Código Civil, «Los Jueces y Tribunales tienen el deber inexcusable de resolver en todo caso los asuntos de que conozcan, ateniéndose al sistema de fuentes establecido», la situación en la que se halla el juez es que ni hay norma completa ni incompleta de la que partir, ni cabe la posibilidad de dejar el caso sin respuesta. Téngase en cuenta que no es esta una peculiaridad de nuestro ordenamiento, de modo que fuera de él tales dificultades no se dan y por ello la IA podría tener el campo libre. Bien al contrario, encontramos normas en términos muy

similares a los del 1.7 de nuestro C.C. en la práctica totalidad de los Estados. Los casos difíciles se caracterizan por el hecho de que el jurista ha de llevar a cabo una justificación externa, esto es, ha de justificar desde fuera del Derecho, desde una valoración personal, moral motivada por qué selecciona partir de una determinada norma jurídica y no de otra de las posibles para resolverlo. Tal justificación estrictamente moral determinará que se atribuya una u otra consecuencia jurídica al supuesto. Es obvio que tanto la capacidad creativa del jurista como su discrecionalidad han de verse amparadas por los valores morales y principios jurídicos que fundamentan el ordenamiento. No obstante, a nadie se le escapa que no es pequeño el abanico de posibilidades que al agente jurídico se le abren, y mucho por tanto su poder de decisión. Pues bien, tampoco ese parece el marco ideal para la aplicación de sistemas basados en IA y algoritmos.

La cultura de los datos ha colonizado el Derecho. Esa afirmación no debe entenderse en sentido negativo. No nos referimos con ella a que los datos sean despreciables o a que su contribución sea menor. Al contrario, los datos siempre han sido un valor para el Derecho y de manera especialmente relevante cuando de lo que se trata es de conocer la eficacia normativa. Si se prescinde de la información aportada por los datos el Derecho se queda a ciegas. Y sin ver cuál es la realidad de la que parte y a la que se dirige difícilmente podrá llevar de manera satisfactoria su labor de ordenación y transformación social. Lo que de nocivo se intuye del enunciado es que las normas están pasando a ser percibidas y tratadas como si de datos numéricos se tratase, desoyendo y desconociendo su compleja naturaleza. Las normas jurídicas, lejos de ser simples *outputs* manejables a la manera de cifras, han de ser consideradas en su condición de elementos interconectados de un sistema. Las normas son productos del poder, pero también obra de la ciudadanía. No en balde se habla de cultura jurídica o culturas jurídicas para hacer alusión a algo más que la suma de las normas vigentes en un Estado. Siendo el Derecho un sistema ninguna de las piezas que lo componen puede ser interpretada, entendida ni aplicada en el vacío. Todas influyen en todas y toda variación sobre una de ellas produce efecto en el conjunto del que forma parte. La vida de cada norma (creación, modificación, derogación...) no solo le afecta a sí misma, sino que interfiere sobre la del resto. El sistema jurídico está lleno de «interferencias», interferencias que, además, se producen cada vez más rápidamente. Tratar al Derecho, las normas y las resoluciones judiciales, como si ello no existiese, como si fueran «cosas» autónomas, distorsiona la naturaleza del Derecho hasta convertirla en algo que nunca ha sido. Uno de los factores que inciden en ello es la muy común confusión entre Derecho y legalismo. El Derecho es mucho más que el conjunto de leyes, de normas y de resoluciones judiciales de un país. El Derecho es además de ello el vínculo entre todos ellos, un vínculo dinámico, un lazo que aprieta o afloja, pero siempre une. Tomadas una a una, en el vacío, sean leyes, disposiciones o resoluciones incluidas en una base de datos, por grande y completa que ésta

sea, ignorando lo que unas veces es interconexión y otras, interdependencia, especialmente relevante cuando de derechos fundamentales se trata, supone un profundo desconocimiento de la visión sistemática y contextual que el Derecho es y precisa.

A la luz de lo anteriormente expuesto, algunos interrogantes marcan nuestra reflexión: ¿pueden matematizarse los valores superiores de su ordenamiento jurídico, a la sazón, y tal y como establece el artículo 1.1 de la Constitución española, la libertad, la justicia, la igualdad y el pluralismo político? Es claro que no se alude aquí a la Administración de justicia sino a la Justicia con mayúsculas, esa propensión a la que tiende todo Derecho y a cuya búsqueda se consagran todas sus disposiciones. ¿Hay o puede haber algún algoritmo que condense la moral? Y el razonamiento jurídico, ese tipo de razonamiento que la aplicación práctica del Derecho requiere, ¿puede confinarse el razonamiento jurídico en una fórmula matemática por compleja y sofisticada que esta sea? ¿Es subsumible la inteligencia jurídica en la inteligencia artificial? ¿Son racionalidad e inteligencia sinónimos?

Si atendemos a las sustanciales diferencias que entre racionalidad e inteligencia pueden apreciarse, no parece que, ni en lo que respecta al Derecho ni fuera de él, deban identificarse ni confundirse. La inteligencia requiere de diversas formas de racionalidad para expresarse, formas que han sido creadas por ella misma pero que no son ella misma. Son instrumentos de los que la inteligencia se sirve para aplicar y aumentar su inteligencia. Dichas formas de racionalidad son creaciones y la inteligencia su creadora. La inteligencia, por definición compleja, construye racionalidades que aportan resultados que a su vez permiten aumentar una observación lo suficientemente fértil como para crear nuevas herramientas con las que entender la realidad e innovar. El Derecho es una forma evolucionada de inteligencia, la forma de inteligencia intersubjetiva y relacional que creó la racionalidad jurídica pero que no se condensa en ella.

Las normas jurídicas (ninguno de los distintos tipos existentes: reglas, principios o valores), no son fósiles. Son una plasmación, escrita o no, de conceptos, prohibiciones, obligaciones, derechos y libertades, cuyo sentido nunca cristaliza pues por estar vivo es siempre abierto, polémico. El Derecho es, a la vez, sistema y proceso. En tanto que sistema, cualquier alteración que recaiga sobre uno de sus elementos producirá un eco sobre el resto, un eco de distinta entidad si se quiere, según cuál sea el lugar ocupado por dicho elemento en la red normativa e institucional. No siempre pues de idéntico alcance, pero en todo caso con alcance cierto; nada tiene que ver el Derecho con una retahíla de normas enfiladas, sino con un entramado dinámico de piezas que pese a su diferente naturaleza jurídica permanecen interconectadas de manera flexible y abierta. En tanto que proceso, el Derecho no finaliza en su fase de creación, sino que ahí apenas comienza, pues se trata un proceso que en realidad nunca finaliza, ya que en todo momento requiere de selección, interpretación, subsunción, ponderación, valoración y actualización.

A las anteriores consideraciones ha de añadirse que la moral se vehiculiza a través del Derecho, convertido en correa de transmisión. A través de valores y principios jurídicos, aunque no solo por medio de ellos, la moral se propaga. Pero no lo hace a través repeticiones calcadas. De hecho, no puede hacerlo dado, entre otros motivos, lo borroso de la lógica que le es inherente tanto en su interacción como en su metamorfosis[88]. La difícil conjugación de hechos, circunstancias, texto (normativo) y contexto, amén de la potencial concurrencia de sesgos en la elaboración de los algoritmos en los que se basa y depende la IA, no parece ser la mejor tesitura para confiar a programas, por sofisticados y completos que sean desde el punto de vista tecnológico, la resolución de conflictos, especialmente los atinentes a derechos y libertades fundamentales. Las causas que lo desaconsejan en esos supuestos no tienen por qué alcanzar a aquellos otros rutinarios o que únicamente obedecen a plazos de cómputo objetivo. En estos casos, la aplicación de la IA puede ser una buena aliada para ganar celeridad en la búsqueda, tratamiento y gestión de información, así como en la resolución de problemas jurídicos sencillos, contribuyendo de ese modo a aumentar la confianza en la Administración. Esa mayor agilidad que el tratamiento del *big data* procura, hace que la maquinaria administrativa sea más rápida y, en consecuencia, más eficaz. No obstante, velocidad e inteligencia no son sinónimos y, es más, convertida en un fin en sí mismo, tal rapidez puede llegar a ser una presión que interfiere en la justa resolución de los casos más complejos. Algunas de las razones que, a nuestro juicio, no aconsejan atribuir a la IA el conocimiento de los casos de cierta o gran enjundia jurídica han sido ya apuntadas o sugeridas, nos detendremos en aquellas a las que se ha prestado menor atención a fin de completar la justificación de dicha afirmación.

Muchos son los estudios dedicados por la doctrina especializada al análisis y alerta de la existencia de sesgos en el procedimiento de elaboración de los programas de IA a los que se adjudica una labor tradicionalmente asignada a juristas. La cuestión es que tales sesgos, o prejuicios cognitivos, pueden aparecer tanto en los datos con los que se alimenta el sistema; como en el modelo algorítmico empleado y/o en sus variables. Como ya señalara en 2018 la *Commission Nationale de l'Informatique et des libertés* francesa el perjuicio de la concurrencia de tales sesgos resulta especialmente remarcable cuando se refieren a «patrones históricos». En ese caso el daño causado no recae solo sobre sujetos específicos sino sobre grupos o colectivos predeterminados, apuntalando la reproducción de discriminaciones o desigualdades preexistentes. La generación de dichos sesgos discriminatorios no es atribuible a la mala fe, sino que, mayoritariamente, se trata de sesgos subconscientes: datos insuficientes; la insuficiente representación de la diversidad social; el establecimiento de correlación sin causaciones y, en cierta medida,

88 Martínez García, J.I., «Derecho inteligente», *Cuadernos Electrónicos de Filosofía del Derecho*, núm. 37, 2018, pág. 100.

el sesgo de fascinación ante las máquinas. Una fascinación que, salvando las distancias, recuerda a la que los ilustrados profesaran en su día por la idea de progreso encarnada en las máquinas.

En todo caso, cuando por la confirmación estadística se toman medidas cautelares sobre una determinada persona porque por razón de su sexo, raza, religión, dedicación o incluso domicilio la probabilidad de ser reincidente es mayor se deja sin contenido y anula la presunción de inocencia, principio contenido en el artículo 24.2 de nuestra Constitución y en la totalidad de los estados de Derecho, constituyendo una de las piedras angulares de la democracia. «Condenar» con base en una fórmula matemática por un futurible y no por un hecho cierto resquebraja un pilar insustituible de la dignidad humana y la convivencia social. Como es sabido, tales sistemas predictivos son muy utilizados en U.S.A. para fijar las medidas cautelares y si bien es verdad que la última palabra la tiene el juez asistido por el COMPAS *(Correctional Offender Management Profiling for Aternative Santions)* no es menos cierto que éste se sustenta en un algoritmo opaco de predicción, una herramienta que expone la tasa de reincidencia de un individuo habiendo obtenido la información de un cuestionario que el individuo ha de rellenar cuando ingresa en prisión. Ese cuestionario está formado por preguntas cuyas respuestas se vuelcan en el sistema para alimentar el software. De ese modo se genera una puntuación: todo es más sencillo de gestionar expresado en números, y es así como se mide y evalúa la condición humana, prescindiendo de los hechos y conformándose con hipótesis no verificables ni contrastables.

Hasta el momento, ninguno de los notables avances técnicos experimentados por los distintos programas y sistemas diseñados con IA empleados en el ámbito jurídico han sabido o podido resolver ni el grave problema de la impenetrabilidad de los algoritmos empleados ni el de la imprevisibilidad de sus resultados. Lo cierto es que muchos de esos procesos de IA «están diseñados para que una máquina autónomamente y sin asistencia externa, los reprograme, fusionando en base a la experiencia de aprendizaje post-diseño, y actuando de una manera considerada "creativa" o como una manifestación del pensamiento "fuera de la caja". En estos niveles, **los sistemas de IA plantean problemas denominados como de "caja negra": se conocen y controlan los datos con los que el sistema es alimentado, así como los objetivos de la programación; pero no es posible prever con exactitud los resultados que se obtendrán a partir de la herramienta de la IA**»[89]. Y no es ese el único aspecto en que se aprecia opacidad. La inexistencia de un Registro nacional o internacional con los algoritmos implementados, amén de la contraposición entre la defensa de la protección de la propiedad intelectual e industrial de las empresas frente al derecho a la información a la que los ciudadanos

89 Contreras, P. y Trigo, P., «La gobernanza de la IA Esbozo de un mapa entre *Hard Law* y *Soft Law* internacional», en M. Azuaje y P. Contreras, *Inteligencia Artificial y Derecho: desafíos y perspectivas*, Tirant lo Blanch Valencia, 2012, p. 466.

tienen derecho no hace sino incrementar la ya alargada sombra que sobre la IA aplicada a los derechos recae.

Tal opacidad nos exige plantearnos una serie de cuestiones trascendentales desde el punto de vista iusfilosófico: ¿qué pasa con la obligación constitucional de motivar las sentencias judiciales (art. 120.3 C.E.) y por extensión el conjunto de las decisiones que suponen limitación de derechos y libertades? ¿Qué pasa con los deberes de publicidad y transparencia a los que están sometidas las actuaciones de las administraciones públicas (C.E. 105 b y c)? ¿Puede la IA a día de hoy cumplir con arreglo a Derecho con esa exigencia? Y más aún: ¿cuál es la legitimidad de tales herramientas, convertidas bien en depositarias bien en auxiliares de los valores que dirigen la vida individual y social? A día de hoy, ninguna, tan es así que la falta de las garantías jurídicas necesarias produce «agujeros de legitimidad».

En ese sentido, tal vez convenga no pasar por alto que los efectos producidos por la IA trascienden, como es obvio, el ámbito del Derecho, alcanzando a ámbitos extrajurídicos. Es el caso, por ejemplo, de la economía, la ciencia y la tecnología. No obstante, las repercusiones ocasionadas en tales esferas repercuten de forma mediata e inmediata en la esfera jurídica. Desde el punto de vista práctico, porque el entorno digital en que nos desenvolvemos ha llegado a crear una infraestructura de dominación mercantil cuyo desenvolvimiento, con escasos o nulos controles públicos, propicia un tipo de dominación social en el que algunos de los valores superiores del ordenamiento del Estado social y democrático de Derecho (justicia, libertad, igualdad) se ven desplazados por el ascenso de otros valores ajenos a la arquitectura del mismo, la anteposición de la utilidad sobre la igualdad o la intimidad, sería un ejemplo. Y desde el punto de vista teórico, por el panorama científico, que no solo se ve nutrido, sino a veces congestionado por lo que se ha dado en llamar la «ideología dataísta». Esa «dependencia» en virtud de la cual todo es orientado a los datos, por ellos, para ellos y nada sin ellos ha motivado que algunos autores confirmen el «fin de las teorías» (Ch. Anderson 2008) y, por tanto, del método científico como tal, reducido casi por completo a la estadística. Como fácilmente puede inferirse, el Derecho no queda al margen de ello. No sólo la metodología jurídica, sino incluso la propia producción normativa se ve directamente concernida por ello. Algo se ha dicho con anterioridad sobre su incidencia en la metodología, percibidas y tratadas las normas como si fueran simples datos o *inputs* matemáticos, reduciendo la ley a la letra y omitiendo (conscientemente) su espíritu. Por lo que a la creación del Derecho concierne, se observa cómo el «método reputacional» va abriéndose paso, lo que significa que se crean normas específicas para cada situación: «microdirectivas que resultan de los macrodatos ordenados por algoritmos conforme a criterios normativos». No por ser novedosa sino por ser simplificadora y disruptiva, tal forma de generar Derecho prescinde del hecho que el método jurídico es, ante todo, problemático. El jurista trabaja con problemas, que no tienen por qué ser datos y en la mayoría de los casos no lo son, intercambia

puntos de vista jurídicos y busca fuera del Derecho la aportación de otros profesionales para completar su visión. El pensamiento jurídico no es pues lineal sino profundo y complejo en tanto en cuanto para su desarrollo incorpora diversas y heterogéneas perspectivas.

Vemos pues que no son pocos ni triviales los ámbitos y aspectos en los que a resultas de una aplicación no cuidadosa ni restringida de la IA se producen derivaciones no deseadas o, si deseadas, no coherentes con el orden jurídico-político que nos hemos dado. Por proseguir en esa dirección, piénsese que, antes incluso de la consideración normativa comúnmente tomada como el punto de partida del Derecho, el jurista se halla, *prima facie,* ante la necesidad de calificar los hechos condicionándose con tal calificación todo el desarrollo posterior. Pues bien, son muchas las ocasiones en las que es la conciencia del jurista la que procesa y decide sobre la calificación y el tratamiento de los hechos. La conciencia, en el cruce de caminos de la ética, la empatía, el sentido común, el contexto y la razonabilidad, es imposible de formalizar en términos matemáticos.

Además de la erosión, cuando no directamente vulneración, de la presunción de inocencia, y la prescripción de motivación aún habría tres principios más directamente comprometidos si la IA se aplicase a la resolución de conflictos jurídicos no rutinarios: el imperio de la ley, la independencia y la discrecionalidad judicial. Consagrados en los artículos 9.3 y 117.1 de la Constitución (y de forma similar en los países de nuestro entorno) tales principios establecen que el criterio discrecional de los jueces, esto es, el libre margen de apreciación ha de ser independiente y sometido únicamente al imperio de la Ley. Solo la Ley, en tanto que producto de la soberanía popular, otorga legitimidad a las decisiones judiciales y administrativas, y no fórmulas matemáticas, algoritmos o sistemas «inteligentes».

Por supuesto toda inteligencia es bienvenida; sería poco o nada inteligente cerrarle la puerta. Pero, como se ha indicado, la inteligencia de quienes diseñan los sistemas «artificiales» no tiene por qué ser extensible a ellos. Una de las aportaciones más inteligentes hechas por el hombre proviene del campo de la Filosofía del Derecho: los derechos humanos. Pues bien, de lo más inteligente que a día de hoy puede y debe hacerse con esos referentes irrenunciables que son los derechos humanos, es «aliarse» con el desarrollo científico y tecnológico, pero no de cualquier forma ni a cualquier precio. Ni el veto a la IA ni su empleo indiscriminado en todas las instancias administrativas y judiciales sería, a nuestro parecer, muestra de inteligencia. A nadie se le escapa que tan trascendente aportación puede verse amenazada por los avances digitales si no se articulan e implementan garantías que, a la manera de «cortafuegos», impidan ciertos daños y vulneraciones provenientes de un uso «descontrolado»[90].

90 De hecho, esa es una de las preocupaciones de la Alta Comisionada de Naciones Unidas

Lo más acorde con la naturaleza de los derechos humanos, reglado encuentro de la moral y el Derecho, irreductibles a formulaciones algorítmicas, sería, a nuestro parecer, una «reserva de humanidad», esto es, la exclusión de la aplicación de la IA de ciertos ámbitos de decisión humana, habida cuenta de lo inapropiado de su uso en ellos[91]. Dicha reserva de humanidad habría de adoptar la forma de una norma con rango de ley y con plena coherencia podría hacerse al amparo de lo dispuesto en el 18.4 de la C.E.: «la ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos».

El establecimiento de dicho límite no exime de una constante supervisión humana sobre la IA en aquellos otros ámbitos en que, también previa decisión humana, se acepte su empleo. Es lo que se ha dado en llamar *Human in the loop*, o el humano en el sistema. Lo óptimo sería que tal supervisión y control viniese de la mano de las Administraciones públicas que adquieren la práctica totalidad de los sistemas algorítmicos empleados. De ese modo los relevantes problemas jurídicos derivados de la opacidad de los sistemas, especialmente patentes en cuestiones relativas a sesgos y protección de datos, podrían verse si no solucionados al menos sí reducidos. Sin embargo, por el momento las Administraciones no disponen ni de los recursos humanos, ni de los tecnológicos para ello. Incluso puede decirse que, en general, no han mostrado la preocupación o la voluntad de emprender dicha tarea. La laguna jurídica existente en torno al control sea público o privado, de los algoritmos, programas y sistemas de IA, conduce a que las labores de apoyo se externalicen «aceptando las condiciones de uso impuestas por los desarrolladores de manera acrítica. Una de ellas suele ser, además, la no publicación del código fuente»[92], incrementándose así el ya considerable problema de opacidad. La crítica está fundamentalmente dirigida al ámbito privado por prevenir de allí la mayor parte de los desarrollos de IA. Pero, a decir verdad, alcanza también

para los derechos humanos, Michelle Bachelet: «Es esencial que en esta era digital prestemos especial atención a los derechos humanos (...) La revolución digital plantea un considerable problema de derechos humanos a escala mundial. Sus beneficios indudables no anulan sus riesgos evidentes (...) ¿Abordamos estos desafíos mediante la ética o mediante los derechos humanos? (...) Los códigos éticos y el cumplimento voluntario no constituyen por sí mismos una respuesta suficientemente enérgica para la escala del problema que afrontamos. No hay ningún segmento de la revolución digital que no pueda y no deba examinarse desde una perspectiva de derechos humanos». *Cfr.* «Derechos humanos en la era digital ¿Pueden marcar la diferencia?», Discurso de 17 de octubre de 2019: https://www.ohchr.org/es/2019/10/human-rights-digital-age

91 Ponce Solé, J., «Inteligencia artificial, derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», *Revista General de Derecho Administrativo*, n.º 50, 2019, pág. 11.

92 Boix Palop, A., «Transparencia en la utilización de la IA por parte de la Administración», *El Cronista del Estado Social y Democrático de Derecho*, n.º 100, 2022. https://www.iustel.com/v2/revistas/detalle_revista.asp?id_noticia=425521

a los desarrollos llevados a cabo por la propia Administración pública. Sirva como ejemplo el caso del sistema «VioGen» que en lo tocante a la opacidad en nada se distingue de los privados.

Relacionados con el discurso «clásico» de los derechos humanos, pero con entidad propia, mención aparte requiere la consideración de los derechos digitales y neuroderechos. Las opiniones en favor del reconocimiento y salvaguarda de bienes jurídicos hasta ahora desconocidos para el Derecho pero que el empleo de la IA comienza a revelar como de necesaria protección. Designados bien como derechos digitales bien como neuroderechos, el hecho es que habría hasta cinco grupos de derechos que, a raíz del creciente protagonismo de la IA en la gestión, tratamiento y resolución de materias relativas a la esencia misma de ser humano (piénsese, por ejemplo, en el acceso y modificación de la información cerebral), reclamarían un espacio autónomo en el catálogo internacional de derechos humanos. A saber: derechos derivados de la libertad de pensamiento (libertad cognitiva, libertad mental y libertad de pensamiento); derechos derivados de la privacidad (neuroprivacidad y privacidad mental); derechos derivados de la integridad (integridad mental); derechos derivados de la identidad y, por último, el grupo que englobaría el derecho a un acceso igual a la mejora mental y a la protección contra el sesgo algorítmico). Esa es, en lo esencial, la recomendación de especialistas clínicos y bioéticos que, bajo el lema «si esto no es un problema de derechos humanos, ¿qué es un problema de derechos humanos?»[93], propugnan priorizar la reflexión ética y jurídica en lo que a todo ello concierne.

En realidad, de lo que se habla cuando se alude a la «reserva de humanidad» o a la ampliación de la nómina de derechos humanos es de cuál ha de ser el papel que se reserve al Derecho a partir del momento en que el poder conferido a la IA y las neurotecnologías no ha hecho sino empezar a despuntar. La inminente colisión de derechos y libertades fundamentales hacen del asunto una tema trascendente y complejo, no reductible a una cuestión de ritmos. Sin menospreciar el llamado *Pacing problem,* bien sabido que la diferencia de ritmos entre el Derecho y la tecnología no es un asunto baladí, somos de la opinión de que concurren además hondos problemas de naturaleza filosófica, ética y jurídica y, por qué no, política. Y es que determinar de manera clara y precisa las funciones del Derecho en ese nuevo contexto adquiere el rango de desafío intelectual y político a un tiempo, puesto que es la fijación del espacio reservado a la voluntad y decisión humana democráticamente establecidas lo que está detrás.

El potencial conflicto entre la libertad de investigación, económica y de empresa por un lado y la protección de los derechos fundamentales por otro exige que el Derecho tome cartas en el asunto para prevenir, prohibir o sub-

93 Yuste, R., Goering, S, *et al*., «Four ethical priorities for neurotechnologies and A.I.», *Nature*, vol. 551, 2017, pp 159-163. https://www.nature.com/articles/551159a

sanar posibles derivas de la ciencia y tecnología. A tal fin un reducido grupo de principios habrán de actuar como ejes sobre los que gire toda la regulación jurídica:

- principio de neutralidad tecnológica;
- principio de transparencia;
- principio de precaución;
- principio-derecho a una explicación comprensible.

Bibliografia

Arendt, H., *Los orígenes del totalitarismo*, tr. G. Solana, Taurus Madrid, 1998.

Asis Pulido, M., «La justicia predictiva: tres posibles usos en la práctica jurídica» en Llano Alonso, F. (Dir.), *Inteligencia artificial y filosofía del Derecho*, Murcia, Laborum, 2022.

Belloso Martín, N., «Inteligencia artificial en la teoría de la decisión judicial. ¿Del juez-robot al asistente-robot del juez?» en Peralta, A., Torres, L., y Herrera, F. (Coords.), *El derecho y la Inteligencia Artificial*, Editorial Universidad de Granada, Granada, 2022.

Benjamins, R., «Hacia una I.A. sostenible: una perspectiva 360 incluyendo negocio, sociedad, ética y cambio climático» en Peralta, A., Torres, L., y Herrera, F. (Coords.), *El derecho y la Inteligencia Artificial*, Granada, Editorial Universidad de Granda, 2022.

Beñido Añón, J.A, *La antropología dogmática de Pierre Legendre,* Tesis Doctoral, UAM, 2022.

Bobbio, N., *El problema del positivismo jurídico*, tr. E. Garzón Valdés, Fontamara, México D.F., 2004.

Boix Palop, A., «Transparencia en la utilización de la IA por parte de la Administración», *El Cronista del Estado Social y Democrático de Derecho*, n.º 100, 2022.

Carr, N., *Superficiales ¿Qué está haciendo internet con nuestras mentes?*, tr. P. Cifuentes, Madrid, Taurus, 2010.

Contreras, P. y Trigo, P., «La gobernanza de la IA Esbozo de un mapa entre *Hard Law* y *Soft Law* internacional», en M. Azuaje y P. Contreras, *Inteligencia Artificial y Derecho: desafíos y perspectivas*, Tirant lo Blanch Valencia, 2012.

Cortina, A., «Ética de la Inteligencia Artificial», *Anales de la Real Academia de Ciencias Morales y Políticas*, n.º 96, 2019.

Durkheim, E., *De la division du travail social*, 11.ª ed., P.U.F., París, 1986.

Hauriou, M., «La teoría de la institución y de la fundación (Ensayo de vitalismo social)» en *Obra Escogida*, Instituto de Estudios Administrativos, Madrid, 1976.

Heidegger, M., *Essais et Conférences*, Gallimard, París, 1992.

Hölderlin, F.,«Patmos», en *Poesía completa* (Edición bilingüe), 5.ª ed., Ediciones 29, Barcelona, 1995.

Husserl, E., *La crisis de las ciencias europeas y la fenomenología trascendental,* trs. J. Muñoz y S. Más, Crítica, Barcelona, 1991.

Ihering, R. Von, *El espíritu del Derecho Romano en las diversas fases de su desarrollo*, Comares, Granada, 1998.

Jaspers, K.*, La filosofía desde el punto de vista de la existencia*, tr. J. Gaos, F.C.E Buenos Aires, 1973.

Kant, I., «Idea de una historia universal en sentido cosmopolita» *en Filosofía de la historia*, F.C.E, México D.F, 1978.

Latorre, J. I., *Ética para máquinas,* Ariel, Barcelona, 2019.

Legendre, P., *La fabrique de l'homme occidental*, Fayard, París, 1997.

Legendre, P., *Le desir politique de Dieu. Étude sur les montages de l'État du Droit, Leçons* VII, 9.ª ed., Fayard, París, 2005.

Legendre, P., *El Tajo. Discurso a jóvenes estudiantes sobre la ciencia y la ignorancia*, tr. I. Agoff, Amorrortu, Buenos Aires, 2008.

Legendre, P., *Lo que Occidente no ve de Occidente*. *Conferencias en Japón*, tr. I. Agoff, Amorrortu, Buenos Aires, 2013.

Lichtenberg, G. Ch., *Aforismos*, tr. J. del Solar, Edhasa Barcelona, 1990.

Maccormick, N., *Legal Reasoning and legal Theory*, Oxford, Clarendon Press, 1978.

Martínez García, J.I., «Derecho inteligente», *Cuadernos Electrónicos de Filosofía del Derecho*, núm. 37, 2018.

Martínez García, J. I., «La respuesta jurídica», *Anuario de Filosofía del Derecho*, (XXXVI), 2020.

Martínez García, J. I., «Instituciones inteligentes», en **Solar Cayon**, J. I. (Ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho. Cuadernos Democracia y Derechos Humanos*, Alcalá de Henares, Universidad de Alcalá, 2020.

Martínez Zorrilla, D., «El juez artificial: ¿próxima parada?», *Oikonomics*, n.º 12, 2019.

Morin, E., *El Método, 6, Ética*, 2.ª ed., tr. A. Sánchez, Cátedra, Madrid, 2009.

Muñoz Vela, J. M., *Retos, riesgos, responsabilidad y regulación de la I.A. Un enfoque de seguridad física, lógica, moral y jurídica*, Aranzadi, Cizur Menor, 2022.

Pascal, B., *Pensamientos*, 10.ª ed., Espasa Calpe, Madrid, 1995.

Ponce Solé, J., «Inteligencia artificial, derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», *Revista General de Derecho Administrativo*, n.º 50, 2019.

R**oa Avella, Marcela del P., Sanabria Moyano, J. E. y Dinas-Hurtado, K**., «Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos», *Revista Brasileira de Direito Processual Penal*, Vol. 8, n.º 1, 2022.

Rodríguez-Arana Muñoz, J., «La reflexión en la enseñanza del Derecho» en Rodríguez-Arana Muñoz, J. y Palomino Lozano, R., (Dirs.), *Enseñar Derecho en el siglo XXI. Una guía práctica sobre el Grado en Derecho*, Aranzadi, Cizur Menor, 2009.

Román, C., «Inteligencia Artificial en la Justicia (Del juez-robot al asistente-robot del juez)», *Diario Constitucional Santiago de Chile*, 2021.

Savigny, F. C. Von, *Sistema del Derecho Romano actual*, Góngora, Madrid, 1878-1879.

Savigny, F. C. Von, *Metodología jurídica*, Depalma Buenos Aires, 1979.

Sloterdijk, P.*, Normas para el parque humano. Una respuesta a la Carta sobre el Humanismo de Heidegger*, tr. T. Rocha, Siruela, Madrid, 2000.

Solar Cayón, J. I., «Inteligencia artificial en la justicia penal: los sistemas algorítmicos de evaluación de riegos», en Solar Cayón, J.I. (Ed.), *Dimensiones éticas y jurídicas de la inteligencia artificial en el marco del Estado de Derecho*, Alcalá de Henares, Universidad de Alcalá, 2020.

Supiot, A., *Homo juridicus. Essai sur la fonction anthropologique du Droit*, Seuil, París, 2005.

Surden, H., «Artificial Intelligence and Law: An Overview», en *Georgia State University Law Review*, vol. 35, n.º 4, 2019.

Treves, R., *La Sociología del Derecho: orígenes, investigaciones, problemas*, trs. M.º J. Añon Roig, M. Atienza y J.A. Pérez Lledó, Ariel, Barcelona, 1998.

Lasalle, J. M.ª, *Ciberleviatán. El colapso de la democracia liberal frente a la revolución digital*, Arpa, Barcelona, 2019.

Virilio, P., *El cibermundo, la política de lo peor. Entrevista con Philippe Petit*, 3.ª ed., Cátedra, Madrid, 2005.

Yuste, R., Goering, S*, et al*., «Four ethical priorities for neurotechnologies and A.I.», *Nature*, vol. 551, 2017.

9.

INTELIGENCIA ARTIFICIAL ¿SUEÑO O PESADILLA?

Miguel L. Lacruz Mantecón
Profesor Titular de Derecho Civil
Universidad de Zaragoza

SUMARIO: I. La inteligencia artificial y el derecho. II. La IA y la indemnización de daños. III. Las aplicaciones de la IA y el posible daño a los derechos fundamentales. 3.1. Los sistemas de datos biométricos e identificativos, en general. 3.2. Aplicaciones de los datos biométricos. 3.2.1. Sistemas de identificación y localización de personas. 3.2.2. Sistemas de perfilado predictivo del comportamiento. 3.2.3. Sistemas de detección de emociones. 3.2.4. La persuasión y la alteración subliminal de comportamientos. 3.2.5. Sistemas de categorización biométrica. IV. Incidencia de la IA en la vulneración de derechos fundamentales. 4.1. El sistema de crédito social chino. 4.2. La falsificación de la realidad. 4.3. La imitación de resultados humanos inteligentes. V. A modo de conclusión.

I. La inteligencia artificial y el derecho

Cuando nos enfrentamos a una tecnología realmente innovadora, como lo es la de la IA, cualquier intento de predicción resulta simplemente una apuesta, y muy posiblemente esto es lo que ocurra con las afirmaciones que yo pueda hacer aquí hoy. Por eso creo que hay que ser prudentes, y hacer un abundante empleo del tiempo verbal futuro imperfecto, y del modo subjuntivo, que es el que utilizan quienes no están muy seguros de lo que afirman.

La IA es ciertamente una tecnología innovadora, tanto que, como nos dice el profesor de Oxford Luciano Floridi[1], inaugura un nuevo capítulo de la histo-

1 Floridi, L., *The Ethics of Artificial intelligence*, Oxford University Press, Oxford, p. XI., 2023, p. XI.

ria de la humanidad: «La revolución digital sólo sucederá una vez, y esa vez es ahora. Hemos pasado una página en la historia humana y un nuevo capítulo ha comenzado. Las futuras generaciones no sabrán lo que era una realidad analógica, predigital y *off-line*. Somos la última generación en experimentarla». Lo que significa, en definitiva, dos cosas: que somos dinosaurios, y que ya no hay vuelta atrás, que no podemos «desinventar» la IA.

Sin embargo, para otros estamos ante una exageración, o como se dice ahora, un *hype*. En este sentido, considera Bermejo Latre[2] que la IA aparece sobredimensionada en nuestra época, cuando en realidad no es sino una tecnología informática más. Estamos ante la última de las tecnologías de la información y la comunicación (TIC): «la IA es una TIC más, la última versión disponible de una larga serie. Como es lógico en todo proceso evolutivo, la última tecnología disponible ofrece nuevas funcionalidades a las ya existentes, si bien pocas son realmente novedosas o disruptivas, o lo son menos de cuanto se les supone». En definitiva, estima el autor que estamos ante «herramientas informáticas aumentadas, relativamente autónomas», pero «en todo caso, auxiliares». Valora Bermejo sobre todo el papel de la inteligencia artificial como herramienta administrativa.

En contra de las predicciones catastrofistas sobre la IA, opina Rivero Ortega[3] que, como decía Ortega y Gasset, el ser humano no se comprende sin la tecnología, y que las anticipaciones apocalípticas no son sino exageraciones: «Aunque estamos acostumbrados a vivir en un entorno artificial con dependencia plena, creemos que sumar otros mecanismos nos convertirá en algo distinto de lo que somos hace milenios: una mezcla de naturaleza y herramientas».

Sin embargo, hay una diferencia que hace que la IA no sea una mera herramienta, y está en que no viene a potenciar la fuerza humana, sino a sustituir otra capacidad humana, la de dar respuestas inteligentes a los problemas que se le plantean. Viene la IA a ocupar el lugar de la herramienta humana de resolución de problemas, del cerebro, y esto no es un cambio cuantitativo sino cualitativo. Como nos dice Floridi[4], hay un equívoco sobre la IA que es preciso despejar, y es que la IA que consigue resultados no genera inteligencia humana, sino que la remplaza: «Un lavaplatos no limpia los platos como yo lo hago, pero al final del proceso sus platos limpios son indistinguibles de los míos. En realidad, puede que incluso estén más limpios (eficacia),

2 Bermejo Latre, J.L., «La aplicación de la Inteligencia artificial en la actividad formal e informal de la Administración», ponencia en el XVIII CONGRESO DE LA ASOCIACION ESPAÑOLA DE PROFESORES DE DERECHO ADMINISTRATIVO (25-26 de enero de 2024, Vigo), https://www.aepda.es/AEPDAEntrada-3987-XVIII-CONGRESO-DE-LA-ASOCIACION-ESPANOLA-DE-PROFESORES-DE-DERECHO-ADMINISTRATIVO.aspx., p.5.

3 Rivero Ortega, R., *Derecho e Inteligencia Artificial*. Cuatro estudios, Ediciones Olejnik, Santiago de Chile, 2023, p. 9.

4 Floridi, L., *The Ethics of Artificial intelligence*, cit., p. 23.

habiendo utilizado menos recursos (eficiencia). Lo mismo se aplica a la IA». Aunque también nos previene el autor sobre los profetas de la distopía y de los escenarios catastróficos, que colocan siempre en un futuro impreciso: la razón de tales jeremiadas está en que, «como ocurre en las películas de vampiros o zombis, el miedo siempre vende»[5].

Con estas prevenciones, la relación entre IA y Derecho se puede abordar atendiendo a las transformaciones que esta tecnología puede causar en la sociedad, procurando acomodar la regulación jurídica a las nuevas realidades, e intentando que los derechos de las personas no se vean lesionados. Naturalmente, habrá intereses que salgan perjudicados, y frente a esto poco o nada se puede hacer. Cuando se popularizaron los automóviles y los camiones, todo el sector del transporte mediante caballerías y carros sufrió, y miles de personas tuvieron que cambiar de trabajo. En una conferencia sobre IA que tuvo lugar en noviembre 2023 auspiciada por Ibercaja, el investigador del CSIC Pedro Meseguer trajo como icono de la IA una portada de *The Economist*, en la que las letras del acrónimo AI aparecían una con una aureola de ángel y otra con cuernos y rabo de demonio, para reflejar que la nueva tecnología traerá tanto bienes como males. De esto les voy a hablar, pero además intentaré advertir que los mayores males que trae la IA derivan precisamente de su mayor bien, de su funcionamiento perfecto.

Desde el Derecho privado, la incidencia de la IA se advierte en ámbitos como el de la subjetividad y personalidad jurídicas, en el que quieren aparecer los robots como nuevos sujetos o agentes en la vida jurídica, y relacionado con ello, el de la responsabilidad civil por daños. También el ámbito de la contratación electrónica, el de las propiedades intelectual e industrial, en el que aparecen nuevos autores e inventores no humanos, o el de los derechos al honor, intimidad personal y familiar, y otros derechos fundamentales como el de la libertad de opinión e información, en los que se están produciendo alteraciones que no encuentran solución en la ciencia jurídica. A mi juicio, la incidencia más profunda se va a producir en dos ámbitos.

Primero, el de la subjetividad de los sistemas inteligentes, especialmente si se trata de robots androides, subjetividad que jugará en el campo de las interacciones sociales, puesto que estos sistemas y robots inteligentes, adoptarán un papel similar al que hoy tienen las mascotas. Esta subjetividad, pese a lo que se dijo desde la archiconocida Resolución UE de 16 de febrero de 2017, y su tan repetida personalidad electrónica para robots, no creo que venga por la concesión de una personalidad a estos entes, como reconocimiento a su inteligencia. Son máquinas, y aunque realizan tareas inteligentes, su inteligencia no es humana, y cometeremos un grave error si nos confundimos en este punto. Lo que ocurre es que esta inteligencia por un lado produce beneficios, y por otro crea en los humanos que interactúan con los

5 Floridi, L., *The Ethics of Artificial intelligence*, cit., p. 32.

robots un espejismo de humanidad que les impide tratarlos como máquinas. El reconocimiento de una cierta subjetividad de los robots vendrá impuesto por razones de conveniencia, y por nuestra propia ética, no por la inexistente humanidad de las máquinas.

Segundo ámbito de incidencia: el de los derechos de la personalidad. En este ámbito, la IA puede introducir enormes disfunciones, que se relacionan con el campo de los daños y la responsabilidad, pero que va más allá del simple mal funcionamiento del sistema que produce daños que deben ser reparados. Aquí el problema es deslindar qué tipo de sociedad queremos y hasta qué punto podemos tolerar la introducción de los sistemas inteligentes en ámbitos hasta ahora ocupados por humanos.

II. La IA y la indemnización de daños

La cuestión de los daños producidos por la IA se ha abordado al modo clásico, como la determinación e imputación de la obligación de indemnizar derivada de una utilización negligente o malintencionada del sistema, siguiendo una línea de culpa subjetiva, o imponiendo una responsabilidad objetiva o cuasiobjetiva. Estas soluciones clásicas son adaptables a la reparación de daños producidos por el mal funcionamiento de sistemas inteligentes: errores, sesgos, pérdida de beneficios, daños materiales, etc.

En el ámbito europeo, a partir de 2020 la UE reconoce la insuficiencia del marco ético para una protección eficaz de los derechos humanos frente a la IA, y se opta por una defensa propiamente jurídica, proponiendo a la Comisión la aprobación de dos Reglamentos de aplicación en los Estados de la UE. Como nos dice Megías Quirós[6], «El primer Reglamento propone un marco regulador de la IA con la conversión de principios éticos en obligaciones jurídicas, porque «los principios éticos comunes solo son eficaces cuando están también asentados en Derecho» y porque las orientaciones éticas son un buen punto de partida, pero no garantizan que los desarrolladores, desplegadores y usuarios actúen de manera justa ni aseguran la protección eficaz de las personas». El segundo Reglamento es la *Resolución 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial*, que propone el establecimiento de un régimen de responsabilidad civil, objetiva y subjetiva, para que pueda ser reclamado cualquier daño causado por la IA.

Sin embargo, el paso definitivo hacia una regulación de una IA que va más allá de la ética lo da en 2021 la *Propuesta de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia*

6 Megías Quirós, J.J., «Derechos humanos e Inteligencia artificial», *Revista DIKAIOSYNE*, N.º 37, número especial sobre DDHH, en coedición con el Observatorio de Derechos Humanos de la Universidad de Los Andes. Mérida – Venezuela. Enero, 2022, pp. 155 y ss.

de Inteligencia Artificial[7], la llamada *Ley de Inteligencia Artificial* (LIA)*,* que establecería un marco jurídico mediante normas claras para garantizar una IA fiable, segura y respetuosa con los derechos fundamentales. Destaca en este sentido Coca Payeras[8], que, entre los objetivos de la Ley de la IA según su Exposición de motivos, están los de «garantizar que los sistemas de IA introducidos y usados en el mercado de la UE sean seguros y respeten la legislación vigente en materia de derechos fundamentales y valores de la Unión», y «mejorar la gobernanza y la aplicación efectiva de la legislación vigente en materia de derechos fundamentales y los requisitos de seguridad aplicables a los sistemas de IA».

En definitiva, no bastan los principios éticos para la salvaguarda de los derechos fundamentales, y por eso la Exposición de motivos de la LIA dice que, en consecuencia, «...las normas relativas a la IA ...deben estar centradas en las personas, a fin de que la población tenga la seguridad de que la tecnología se usa de un modo seguro y en consonancia con la ley, lo que también implica respetar los derechos fundamentales».

Pero, por otra parte, tenemos los daños causados no por el mal funcionamiento del sistema, sino los que se derivan de un correcto funcionamiento, pero que no fueron tenidos en cuenta. Vengo últimamente repitiendo hasta hacerme pesado que los daños más importantes que puede causar la IA no se producirán porque funcione mal, sino al contrario, porque funciona demasiado bien.

Como ejemplo de esto tenemos la diferenciación recogida en el art. 5 de la llamada Ley de Inteligencia Artificial (LIA, en realidad la *Propuesta de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de IA*, de 2021) entre las técnicas de inteligencia artificial *prohibidas* y las que «solamente» son *de alto riesgo*. Entre las prohibidas se cuentan las técnicas de persuasión y alteración subliminal de comportamientos (como los *nudges*), las de perfilado de sujetos y su calificación social (como el sistema de crédito social chino, o los sistemas de control de fronteras *iBorder ctrl* o *AVATAR*). También los sistemas de identificación biométrica remota «en tiempo real», como los sistemas policiales SARI (italiano) o el sistema inglés AFR *(Automated Facial Recognition)*. Tras

7 Bruselas, 21.4.2021, COM (2021) 206 final, 2021/0106(COD). Finalmente, aprobada como Reglamento (UE) 2024/1689 Del Parlamento Europeo y del Consejo 2024/1689, de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828. Ya no se denomina Ley sino Reglamento de Inteligencia Artificial.

8 Coca Payeras, M., «Las iniciativas de la Unión europea sobre inteligencia artificial: de la persona electrónica, al difícil equilibrio entre la necesidad de impulsarla y evitar sus riesgos», *Revista de Derecho Civil,* vol. X, núm. 2, especial (junio, 2023), p. 24. En http://nreg.es/ojs/index.php/RDC.

estos sistemas prohibidos, la Ley de la IA refiere una serie de sistemas que son calificados como de alto riesgo: sistemas para la gestión de la sanidad, funcionamiento del tráfico rodado y el suministro de agua, gas, calefacción y electricidad, sistemas de selección para el acceso a instituciones educativas y acceso al empleo, etc.

Comparando estos dos tipos de sistemas, vemos que los prohibidos lo son porque su mismo funcionamiento causa daños en derechos fundamentales de las personas, derechos a la intimidad, a la libertad de información y a la libre formación y manifestación de opinión, etc. Y este daño a los derechos se produce, precisamente, porque la perfección del sistema va más allá de la falibilidad e imperfección humanas. En cambio, los sistemas de alto riesgo causan daños cuando su funcionamiento no es todo lo perfecto que debe ser, sobre todo en ámbitos especialmente sensibles como el de la sanidad y medicina, la justicia, el tráfico y las comunicaciones, o los suministros básicos para la vida humana.

La LIA no incide apenas en un tipo de sistemas inteligentes que a mi juicio pueden ser especialmente dañinos. Se trata de los sistemas de IA generalistas, o «modelos fundacionales», aquellos que no tienen una tarea predefinida, sino que asumen los encargos que se les haga. Ejemplo de estos sistemas es el famoso *ChatGpt*, o en el diseño, *Dall-E*.

En nuestra doctrina se señala que esta nueva tecnología implica unos riesgos evidentes para los derechos humanos, y como señala Megías Quirós[9], la limitación de los riesgos derivados de la IA va a ser el objetivo que inspire las más recientes elaboraciones regulatorias en la materia, como son la *Recomendación sobre la Ética de la Inteligencia Artificial* de la UNESCO de noviembre 2021[10] (y el subsiguiente *Report of the International Bioethics Committee on the ethical issues of neurotechnology*, de 15 de diciembre de dicho año), y la europea *Ley de la Inteligencia Artificial*, de 21 de abril 2021[11].

Estos peligros, por otra parte, derivan precisamente de la perfección del funcionamiento de los sistemas inteligentes, que como nos señala Megías Quirós[12], al tiempo que traen muchos beneficios, plantean evidentes ries-

9 Megías Quirós, J.J., «Derechos humanos e Inteligencia artificial», Revista *DIKAIOSYNE*, N.º 37, número especial sobre DDHH, en coedición con el Observatorio de Derechos Humanos de la Universidad de Los Andes. Mérida – Venezuela. Enero, 2022, p. 140.

10 Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), reunida en París del 9 al 24 de noviembre de 2021, en su 41.ª reunión.

11 Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen Normas Armonizadas en Materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión, COM/2021/206 final.

12 Megías Quirós, J.J. *«Derechos humanos e Inteligencia artificial», cit.*, p. 143.

gos para los derechos humanos, «en especial los derechos a la integridad física, la vida privada, la igualdad, la libertad de expresión y de reunión, la tutela judicial efectiva, etc., riesgos que se agudizan cuando afectan a grupos vulnerables (menores, inmigrantes, personas con discapacidad, edad avanzada, etc.). En este sentido, recoge De Asís Roig[13] las palabras de Michelle Bachelet, Alta Comisionada de las Naciones Unidas para los Derechos Humanos, en su discurso «Derechos humanos en la era digital ¿Pueden marcar la diferencia?» de 17 de octubre de 2019, cuando dijo: «Es esencial que en esta era digital prestemos especial atención a los derechos humanos [...]. La revolución digital plantea un considerable problema de derechos humanos a escala mundial. Sus beneficios indudables no anulan sus riesgos evidentes».

Esta idea de la peligrosidad de la IA es compartida por bastantes autores, como Martin Ebers[14], que nos dice que los sistemas de IA pueden dañar de manera impredecible la vida, la salud y la propiedad de las personas, así como «[...] afectar a los valores fundamentales en los que se basan las sociedades occidentales, dando lugar a violaciones de los derechos fundamentales de las personas, incluidos los derechos a la dignidad humana y a la autodeterminación, a la privacidad y a la protección de los datos personales, a la libertad de expresión y de reunión, a la no discriminación o al derecho a la tutela judicial efectiva y a un juez imparcial, así como a la protección de los consumidores».

Pues bien, para tratar de la peligrosidad de la IA, y las posibles actuaciones al respecto, se ha celebrado los días 1 y 2 de noviembre de 2023 una importante cumbre en Bletchley Park, Inglaterra, lugar mítico por ser dónde Alan Turing descifró el código *Enigma*. En esta cumbre, y en un diálogo retransmitido entre el Premier Rishi Sunak y el magnate Elon Musk, éste declaró que la IA era «una de las mayores amenazas para la humanidad», y esto cuando precisamente el lema de la reunión era «Cumbre por la seguridad de la IA».

En la Declaración de esta Cumbre de Bletchley Park se consideran especialmente peligrosos estos sistemas de IA generalistas, o «modelos fundacionales», respecto de los cuales se señala que pueden surgir riesgos sustanciales debido a un posible uso malicioso o a problemas de control. «[...] Existe la posibilidad de que se produzcan daños graves, incluso catastróficos, ya sean deliberados o no».

13 De Asís Roig, R. F., «Ética, Tecnología y Derechos», en *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022, p. 35.

14 Ebers, M., «El futuro marco jurídico europeo de la inteligencia artificial», *Revista General de Legislación y Jurisprudencia*, 2023, número 2, pág. 254.

III. Las aplicaciones de la IA y el posible daño a los derechos fundamentales

3.1. Los sistemas de datos biométricos e identificativos, en general

A continuación, veremos los distintos sistemas inteligentes que hoy se aplican a tareas de vigilancia en calles y comercios, tareas de identificación y control social de la ciudadanía, comenzando por un breve examen de los datos biométricos que sirven de base para su funcionamiento. Los sistemas de reconocimiento de personas por sus datos biométricos son invasivos por sí mismos, e impactan directamente en la protección de datos personales prescrita en el artículo 16 del Tratado de Funcionamiento de la Unión europea.

Se trata de la utilización de sistemas inteligentes que, a través de videocámaras, reconocen los rasgos faciales de las personas y a partir de éstos, la identidad de cada persona en particular. A esto se puede añadir que, mediante diversos periféricos (medidores de tensión arterial, cámaras termográficas, scanner de huellas, micrófonos), se obtienen otros datos corporales como huellas dactilares, temperatura corporal, frecuencia cardiaca, sudoración o respiración del sujeto, tono de voz, movimiento de las pupilas, datos que permiten ampliar la exploración a informes sobre el estado de salud, emociones o incluso pensamientos de la persona, como luego veremos.

La Ley de la IA de 2021 procede a definir los «Datos biométricos» en el art. 3, n.º 33) como *«[...] los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física que permitan o confirmen la identificación única de dicha persona».*

A continuación, pasa el art. 3 LIA a definir los distintos sistemas de datos biométricos, definiciones que nos proporcionan una muestra de las diferentes aplicaciones de esta técnica: *«34) Sistema de reconocimiento de emociones: un sistema de IA destinado a detectar o deducir las emociones o las intenciones de personas físicas a partir de sus datos biométricos. 35) Sistema de categorización biométrica: un sistema de IA destinado a asignar a personas físicas a categorías concretas, como un sexo, edad, color de pelo, color de ojos, tatuajes, origen étnico u orientación sexual o política, en función de sus datos biométricos. 36) Sistema de identificación biométrica remota: un sistema de IA destinado a identificar a personas físicas a distancia comparando sus datos biométricos con los que figuran en una base de datos de referencia [...]».*

Apunta Ebers[15] a que el art. 5.1.d) de la LIA prohíbe los sistemas de IA de identificación biométrica remota (*biometric identification system* o BIS) «en

15 Ebers, M., «El futuro marco jurídico europeo de la inteligencia artificial», *cit.*, p. 267.

tiempo real» en espacios de acceso público (por ejemplo, los sistemas de reconocimiento facial utilizados para identificar a una persona en la calle), pero que este enfoque de prohibir únicamente los sistemas de identificación biométrica utilizados para la aplicación de la ley es demasiado restrictivo, y sus excepciones demasiado amplias.

3.2. Aplicaciones de los datos biométricos

3.2.1. Sistemas de identificación y localización de personas

Examinamos seguidamente las aplicaciones de estos datos biométricos una vez captados por un sistema de IA. Los sistemas biométricos llevan tiempo siendo utilizados para asuntos como desbloquear el teléfono móvil, pero su aplicación estrella consiste en la vigilancia policial, identificación de personas y seguimiento de las mismas. La incidencia de estas técnicas sobre los derechos fundamentales es brutal. Como señala Cotino Hueso[16], nada impide hoy a la policía detectar a un ciudadano en la vía pública o en una manifestación, identificarlo y analizar si está en alguna base de datos específica, georreferenciarlo y reconstruir sus recorridos, interacciones con otras personas y evaluar su comportamiento. Cabe añadir que, si además procesamos los datos de su teléfono móvil, la información puede completarse con todos los perfiles que quepa extraer de dicho dispositivo. En concreto nos señala Dolz Lago[17] que *la Resolución del Parlamento Europeo de 6 octubre 2021 sobre Inteligencia Artificial y su utilización por autoridades policiales y judiciales en asuntos penales* plantea la utilización de sistemas inteligentes por las autoridades policiales para distintas labores de vigilancia, como la identificación biométrica mediante reconocimiento facial, o el reconocimiento automático de matrículas, pero también para «[...] la identificación por voz, el reconocimiento del habla, las tecnologías de lectura de labios, la vigilancia auditiva [...] la investigación y el análisis autónomos de bases de datos identificadas, la predicción (actuación policial predictiva y análisis de puntos críticos de delincuencia), los instrumentos de detección del comportamiento, las herramientas avanzadas de autopsia virtual para ayudar a determinar la causa de la muerte, las herramientas autónomas para detectar fraudes financieros y la financiación del terrorismo, la vigilancia de las redes sociales (rastreo *[scraping]* y recopilación de datos para detectar conexiones) y los sistemas automatizados de vigilancia que incorporan diferentes capacidades de detección (como la detección del latido cardíaco y las cámaras térmicas)» (Considerando M).

16 Cotino Hueso, L., «*Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos...*», *cit.*, p. 72.

17 Dolz Lago, M. J., «Una aproximación jurídica a la Inteligencia Artificial», *Diario La Ley*, cit., p. 4.

Las técnicas de identificación biométrica reciben un tratamiento restrictivo con excepciones. Para empezar, el art. 5.1. h) de la Ley de Inteligencia artificial considera como práctica prohibida *«[...] el uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público*[18] *con fines de garantía del cumplimiento del Derecho».* Ahora bien, esta prohibición es pura fachada, porque inmediatamente a continuación se exceptúa la utilización policial para perseguir el delito, aunque con fuertes exigencias de necesidad: *«[...] salvo y en la medida en que dicho uso sea estrictamente necesario para alcanzar uno o varios de los objetivos siguientes: i) la búsqueda selectiva de víctimas concretas de secuestro, trata de seres humanos o explotación sexual de seres humanos, así como la búsqueda de personas desaparecidas, ii) la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas o de una amenaza real y actual o real y previsible de un atentado terrorista, iii) la localización o identificación de una persona sospechosa de haber cometido un delito a fin de llevar a cabo una investigación o un enjuiciamiento penales o de ejecutar una sanción penal por alguno de los delitos mencionados en el anexo I*[19] *y que en el Estado miembro de que se trate se castigue con una pena o una medida de seguridad privativas de libertad cuya duración máxima sea de al menos cuatro años».* Advierte el art. 5.2.1 que «2*. el uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público [...] para cualquiera de los objetivos mencionados en el apartado 1, párrafo primero, letra h), debe desplegarse para los fines establecidos en dicha letra, únicamente para confirmar la identidad de la persona que constituya el objetivo específico».* Y añade el 5.2.3 que *«el sistema deberá cumplir las garantías y condiciones necesarias que exija el Derecho nacional que autorice dicho uso, ...en particular en lo que respecta a las limitaciones temporales, geográficas y personales. El uso del sistema de identificación*

18 Hace el Considerando 19 una referencia a los espacios públicos: ...*cualquier espacio físico al que pueda acceder un número indeterminado de personas físicas y con independencia de si es de propiedad privada o pública y de la actividad para la que pueda utilizarse el espacio, ya sean actividades comerciales, por ejemplo, tiendas, restaurantes, cafeterías; de prestación de servicios, por ejemplo, bancos, actividades profesionales, hostelería; deportivas, por ejemplo, piscinas, gimnasios, estadios; de transporte, por ejemplo, estaciones de autobús, metro y ferrocarril, aeropuertos, medios de transporte; de entretenimiento, por ejemplo, cines, teatros, museos, salas de conciertos, salas de conferencias; de ocio o de otro tipo, por ejemplo, vías y plazas públicas, parques, bosques, parques infantiles.*

19 Se refiere a delitos de ...*terrorismo, trata de seres humanos, explotación sexual de menores y pornografía infantil, tráfico ilícito de estupefacientes o sustancias psicotrópicas, tráfico ilícito de armas, municiones y explosivos, homicidio voluntario, agresión con lesiones graves, tráfico ilícito de órganos o tejidos humanos, tráfico ilícito de materiales nucleares o radiactivos, secuestro, detención ilegal o toma de rehenes, delitos que son competencia de la Corte Penal Internacional, secuestro de aeronaves o buques, violación, delitos contra el medio ambiente, robo organizado o a mano armada, sabotaje, participación en una organización delictiva*...

biométrica remota «en tiempo real» en espacios de acceso público solo se autorizará si la autoridad garante del cumplimiento del Derecho ha completado una evaluación de impacto relativa a los derechos fundamentales [...] y ha registrado el sistema en la base de datos de la UE». Salvo casos de urgencia debidamente justificados, en que se podrá empezar a utilizar tales sistemas sin el registro en la base de datos de la UE, siempre que dicho registro se complete sin demora indebida.

Además, exige el art. 5.3 que *«el uso policial del sistema se vea amparado por la concesión de una autorización previa por parte de una autoridad judicial o una autoridad administrativa independiente»*, también salvo urgencia, pudiéndose solicitar a posteriori. Aparte, según el art. 5.4, todo uso de un sistema de identificación biométrica remota «en tiempo real» en espacios de acceso público *«[...]se notificará a la autoridad de vigilancia del mercado pertinente y a la autoridad nacional de protección de datos de conformidad con las normas nacionales».*

Aun contando con todas estas prevenciones, la Ley de la IA considera estos sistemas basados en datos biométricos como sistemas de alto riesgo; lo señala así el art. 6.2.

3.2.2. Sistemas de perfilado predictivo del comportamiento

Surge como técnica que parte de la obtención de datos del sujeto, para empezar el de su identidad personal y digital (su dirección IP, número de móvil, dirección de correo electrónico, avatar o *nickname*), y otros acerca de sus intereses vitales (políticos, deportivos, ideológicos o religiosos), nivel económico, preferencias de consumo y de entretenimiento, etc. Su objeto es obtener una imagen del sujeto que permita abordarlo con mensajes de todo tipo, desde publicidad y ofertas de turismo o inmobiliarias a información política o económica, con el objeto de provocar decisiones de compra o actuaciones de cualquier tipo en dicho sujeto-objetivo.

Pero también tienen una polivalente aplicación policial y criminológica. Las herramientas predictivas policiales consisten básicamente en sistemas actuariales, como señala Urruela Mora[20], basados por tanto en la realidad estadística de datos que indican futuras conductas de reincidencia. Su utilización básica se encuentra en el estamento judicial, para la adopción de medidas de libertad provisional (o prisión preventiva) o para medir correctamente la imposición de la pena, siendo el sistema americano COMPAS *(Correctional Offender Mana-*

20 Urruela Mora, A., «Instrumentos de evaluación del riesgo de violencia, justicia algorítmica y Derecho penal. Perspectiva crítica», *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, coord. por Noelia Corral Maraver, Deborah García Magna, Fátima Pérez Jiménez, Bertha Prado, Pablo Rando Casermeiro; Juan Muñoz Sánchez (dir.), Octavio García Pérez (dir.), Ana Isabel Cerezo Domínguez (dir.), Elisa García España (dir.), Tirant lo Blanch, Valencia, 2023, págs. 1877-1888.

gement Profiling for Alternative Sanctions) el más conocido. Estas técnicas son también llamadas de perfilado predictivo, en cuanto operan definiendo un perfil del sujeto y prediciendo su peligrosidad.

En cuanto a su admisibilidad, ya hemos visto que se prohíben salvo que tengan carácter meramente auxiliar. Pero, aun así, el Reglamento o Ley de la IA nos advierte que muchas de ellas pueden ser consideradas de alto riesgo. Así, el artículo 6.3.3 del mismo advierte que *«No obstante lo dispuesto en el párrafo primero, los sistemas de IA a que se refiere el anexo III siempre se considerarán de alto riesgo cuando el sistema de IA efectúe la elaboración de perfiles de personas físicas»*. Por lo tanto, todas las aplicaciones de perfilado predictivo caen bajo esta calificación.

Estas técnicas de policía predictiva tienen un número demasiado alto de falsos positivos y falsas alertas, por lo que están siendo abandonadas por las policías de Nueva York y Cambridge, que las habían adoptado. Da cuenta así Cotino[21] de que en Holanda la sentencia de 5 de febrero de 2020 del Tribunal de Distrito de la Haya declaró contrario al artículo 8 CEDH el sistema *Systeem Risicoindicatie*» («SyRI»), para impedir fraudes a la Seguridad social y tributarios. Ratificando esto, en el ámbito policial y penal, en 2023, la sentencia del Tribunal Constitucional alemán del 16 de febrero (1 BvR 1547/19, 1 BvR 2634/20) estableció requisitos y estándares de calidad legislativa muy altos para el tratamiento automatizado de datos en la prevención del delito. López Riba[22] advierte, en cuanto a los sistemas predictivos zonales, que los algoritmos indican lugares en los que la policía vigilaba ya tradicionalmente, generalmente barrios conflictivos, por lo que se concluye que los sistemas predictivos predicen «el control policial futuro, no los delitos futuros». En España contamos con algunas aplicaciones de este estilo como *Veripol*, la aplicación *VIOGEN*, para analizar el riesgo de violencia de género, unido a sistemas de control biométrico y localizadores para evitar proximidades, o el sistema *Pred-Crime* que es un software para predecir el momento y lugar de ocurrencia de determinados delitos utilizado por la Policía Local de Rivas-Vaciamadrid.

3.2.3. Sistemas de detección de emociones

Aparte tenemos, también basados en datos biométricos, en el n.º 39 los «sistemas de reconocimiento de emociones». Se trata de *«[...] un sistema de IA destinado a distinguir o inferir las emociones o las intenciones de las personas físicas a partir de sus datos biométricos».* Este sistema sirve, fundamentalmente, para, detectando las emociones del sujeto (la felicidad, la

21 Cotino Hueso, L. «El uso jurisdiccional de la inteligencia artificial: habilitación legal...», cit., p. 502.

22 López Riba, J.M., «Inteligencia artificial y control policial. Cuestiones para un debate ...», cit, p. 415.

tristeza, la indignación, la sorpresa, el asco, el apuro, el entusiasmo, la vergüenza, el desprecio, la satisfacción y la diversión —Considerando 16 LIA—) averiguar la verdad o falsedad de sus declaraciones, perfeccionando al polígrafo actual, o identificar otras circunstancias del sujeto.

Naturalmente, esta técnica tiene especial valor para determinar si el sujeto miente o dice la verdad, y por ello la misma Ley de la IA hace expresa referencia a la utilización de los datos biométricos para la labor policial en los interrogatorios, y la averiguación de la veracidad o falsedad de las declaraciones de las personas. Por ello se utiliza en el control de fronteras, como el proyecto europeo *iBorderCtrl*, un sistema inteligente de detección de mentiras que, como dice Coca Payeras[23]: «elabora perfiles de los viajeros a partir de una entrevista automatizada por ordenador realizada a través de la cámara web del viajero antes del viaje y un análisis de 38 microgestos basado en la inteligencia artificial, probado en Hungría, Letonia y Grecia». O el sistema americano para el control de fronteras en EE. UU., que permite leer emociones, detectar la verdad de las manifestaciones, y predecir futuros comportamientos, denominado Agente Virtual Automatizado para la Evaluación de la Verdad en Tiempo Real (AVATAR - *Automated Virtual Agent for Truth Assessments in Real Time*), que analiza el comportamiento no verbal y verbal de los viajeros.

3.2.4. La persuasión y la alteración subliminal de comportamientos

El art. 5 de la Ley de IA considera prohibidas las *«[...] técnicas subliminales que trasciendan la conciencia de una persona para alterar de manera sustancial su comportamiento de un modo que provoque [...] perjuicios físicos o psicológicos a esa persona o a otra»*. Señala Ebers[24] que ni el artículo 5 ni el Considerando 16 de la LIA definen qué es una «técnica subliminal», o en qué consiste alterar *sustancialmente* el comportamiento de una persona. Y además este art. 5 sólo cubre las prácticas de IA que tienen la intención directa de perjudicar a otras personas. En realidad, la mayoría de estas prácticas serán ilegales conforme al Derecho penal en la mayoría de los Estados, si lo que intentan es engañar a las personas (posiblemente para estafarlas), y además la Directiva sobre prácticas comerciales desleales (2005/29/CE) ya prohíbe las prácticas comerciales que distorsionan el comportamiento humano en determinadas condiciones.

Las técnicas subliminales de manipulación de carácter comercial fueron expuestas por la famosa obra del año 1957, *The Hidden Persuaders*, de Vance Packard, libro dedicado a la publicidad y su influencia en los consumidores.

---

23 Coca Payeras, M., (2023), «Las iniciativas de la Unión europea sobre inteligencia artificial: de la persona electrónica, al difícil equilibrio entre la necesidad de impulsarla y evitar sus riesgos», *Revista de Derecho Civil,* vol. X, núm. 2, especial (junio, 2023). En http://nreg.es/ojs/index.php/RDC, p. 24.

24 Ebers, M. «El futuro marco jurídico europeo de la inteligencia artificial», cit., p. 265.

Una de estas técnicas que se usan para influir en los comportamientos es el *nudging*. Señalan Costas Pérez y Tucat[25] que los *nudges* o «empujones» se popularizan a partir de la publicación del libro *Nudge: Improving Decisions About Health, Wealth, and Happiness,* de Cass Thaler y Richard Sunstein, en 2009. Los *nudges* (palabra que significa pequeños «empujones» o «codazos»; también se ha traducido como «acicates») son intervenciones que buscan modificar la toma de decisiones individuales, intentando cambiar el comportamiento de las personas en una dirección concreta, pero sin prohibir ninguna opción, ni alterar en gran medida los incentivos económicos. Es decir, que no son órdenes coercitivas para los ciudadanos, sino que inducen a un comportamiento que se considera adecuado o deseable, como señalaban Thaler y Sunstein. El ejemplo más citado es el de incluir en los documentos médicos la opción por defecto de ser donante de órganos, y estableciendo por tanto que el rechazo ha de ser expreso.

La unión de estas técnicas de *nudging* con la IA produce interesantes resultados. Advierte Ortiz de Zárate-Alcarazo[26] que la aplicación de la IA puede mejorar la implementación de *nudges* en el sector público, e incluso puede llegar a transformar este tipo de intervenciones dando lugar a lo que podríamos denominar «políticas conductuales inteligentes y, más concretamente, *intelligent nudging*. Los *nudges* inteligentes tendrían niveles de eficacia y eficiencia más elevados que los de sus hermanos no inteligentes gracias al uso de tecnologías de reconocimiento facial, detección de objetos, procesamiento del lenguaje natural, predicción, análisis en tiempo real, etc., que permitirían mejorar todos los pasos del proceso».

En este sentido, el investigador de la Universidad de Seikei, Yukari Yamazaki[27] nos anuncia la llegada del *Hypernudge*, como resultado de la unión del nudge y la IA, es decir, el nudge creado por una IA entrenada mediante *machine learning*, y advierte que, sin embargo, las condiciones de autonomía, dignidad y transparencia, que deben regir el empleo de nudges, desaparecen en este caso. En base a un estudio estadístico, concluye que hay que estar alerta en la utilización de los hypernudges por su menor aceptación que los *nudges* originales.

---

25 Costas Pérez, E., y Tucat, P., (2021). «*Nudges: diseño y evaluación», Gestión y Análisis de Políticas Públicas*, (25), 8-22. DOI: https://doi.org/10.24965/gapp.i25.10868., Pág. 9

26 Ortiz de Zárate-Alcarazo, L. (2023). «Las Políticas conductuales inteligentes: Oportunidades y riesgos ético-políticos de la Inteligencia Artificial para el nudging», *Revista Española de Ciencia Política*, 62, 67-93. Doi: https://doi.org/10.21308/recp.62.03. p. 82.

27 Yamazaki, Y., «An Empirical Study for The Acceptance of Original Nudges and Hypernudges», *Societal Challenges in the Smart Society* / coord. por Mario Arias Oliva, Jorge Pelegrín Borondo, Kiyoshi Murata, Ana María Lara Palma, 2020, ISBN 978-84-09-20273-7, Universidad de La Rioja y Universitat Rovira i Virgili, Logroño, 2020, págs. 323-336. En Dialnet: https://dialnet.unirioja.es/servlet/articulo?codigo=7867256

En suma, el *hypernudge* añade a los problemas intrínsecos del *nudge* los problemas específicos derivados de la IA, especialmente la derivada de redes neuronales, como la opacidad algorítmica, los problemas de privacidad, los sesgos, la falta de rendición de cuentas, y sobre todo el ataque a la privacidad e intimidad de las personas, derivando en un paternalismo agobiante.

3.2.5. Sistemas de categorización biométrica

Y finalmente está, en el n.º 40) del art. 3 LIA, el «sistema de categorización biométrica»: *«un sistema de IA destinado a incluir a las personas físicas en categorías específicas en función de sus datos biométricos, a menos que sea accesorio a otro servicio comercial y estrictamente necesario por razones técnicas objetivas».*

Esta categorización se distingue de las anteriores identificación y verificación biométricas, en concreto porque utiliza los datos biométricos *(sexo, edad, el color del pelo, el color de los ojos, los tatuajes, los rasgos conductuales o de la personalidad, la lengua, la religión, la pertenencia a una minoría nacional o la orientación sexual o política* —Considerando 16—) para adscribir al sujeto a una determinada categoría racial, social o ideológica; el sistema es empleado por las autoridades chinas para identificar a las personas de la etnia uigur. Por supuesto, para llegar a ubicar al individuo dentro de una determinada categoría, va a ser precisa una identificación de rasgos biométricos, y si se trata de categorías ideológicas, un perfilado ideológico. Como señala Garriga Domínguez[28], el desarrollo actual de la tecnología de IA permite el perfilado ideológico de los usuarios de cualquier plataforma, dato con el que se les puede enviar información política personalizada. Esto es lo que la autora llama *segmentación psicográfica avanzada*, que permite hacer predicciones precisas de la reacción de la audiencia objetiva a determinados mensajes o estímulos.

IV. Incidencia de la IA en la vulneración de derechos fundamentales

4.1. El sistema de crédito social chino

Ya hemos visto que la IA puede ser lesiva para los derechos fundamentales de las personas. Ahora vamos a examinar algunos ámbitos en los que esta incidencia es particularmente dañina, comenzando por el sistema de clasificación o crédito social chino.

28 Garriga Domínguez, A., «Inteligencia artificial y el fenómeno de la desinformación: el papel del RGPD1 y las garantías recogidas en la propuesta de la ley de servicios digitales», en *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022, p. 452.

Señala Barona Vilar[29] que en algunos países asiáticos se ha acometido un desarrollo tecnológico para favorecer este control a través del reconocimiento facial masivo. El sistema de crédito social o *scoring*, es un instrumento que utiliza el *big data* para calificar el comportamiento de los usuarios o para detectar en las escuelas el absentismo escolar.

Como señalan Roberts, Cowls, Morley y otros[30], esta tecnología del *scoring* viene a ser sancionada por el *Plan de Desarrollo de la Inteligencia Artificial de la Nueva Generación,* aprobado por el supremo órgano del gobierno chino, el Consejo estatal. Este plan tiene tres ámbitos de proyección: Desafío internacional, Desarrollo económico y Gobernabilidad o «Construcción» social, siendo este último el que más nos interesa. El Sistema de Crédito Social todavía no se ha implantado a nivel nacional, pero como nos dicen los citados autores, los ambiciosos objetivos del mismo «[...] ofrecen un convincente ejemplo de la intención del gobierno de confiar en la tecnología digital, no sólo para gobernanza social, sino también para una regulación más detallada del comportamiento»[31]. Añade Rettinger[32] que, en 2018, cuarenta gobiernos municipales y provinciales establecieron planes piloto de sistemas de crédito social, y que la situación actual muestra un sistema fragmentado en tres ámbitos: «sistema de listas negras» a nivel nacional, sistemas de crédito social en determinados municipios, y sistemas de crédito social a efectos financieros pilotados por instituciones financieras.

En nuestra doctrina señala Cotino Hueso[33] que, en China, el uso de las tecnologías biométricas combina las tecnologías de identificación, categorización y reconocimiento de emociones para su famoso sistema de crédito social, pero también para el control policial de la ciudadanía, de la lealtad al partido o para el seguimiento de la atención de los escolares y evaluación y control mental en el ámbito educativo. Naturalmente, la base para el funcionamiento del sistema de crédito social está en la perfecta identificación tanto de personas como empresas, para lo que China cuenta con un sistema

29 Barona Vilar, S., «Tecnología biométrica y datos biométricos. Bondades y peligros. No todo vale», *Actualidad Jurídica Iberoamericana* N.º 21, agosto 2024, págs. 298-331, págs. 315 y 323.

30 Roberts, H., Cowls, J., Morley, J. et al., «The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation». *AI & Society* 36, 59-77 (2021). https://doi.org/10.1007/s00146-020-00992-2

31 Roberts, H. Cowls, J., MORLEY, J. *et al*., «The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation», *AI & Society*, cit. p. 66.

32 Rettinger, L., «The Human Rights Implications of China's Social Credit System», *Journal of High Technology Law*, vol. XXI, n.º 1, (2021), págs 1-33, p.6.

33 Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», *El Cronista del Estado Social y Democrático de Derecho*, N.º 100 (septiembre-octubre), 2022 (Ejemplar dedicado a: Inteligencia artificial y derecho), p. 71.

de seguimiento en Internet, un análisis de grandes datos y 626 millones de cámaras de reconocimiento facial[34].

Las consecuencias pueden ser muy peligrosas para las personas, pues como señala el periodista Serrano Martínez[35], «los ciudadanos pueden entrar en una lista negra con bastante facilidad, lo que tiene graves consecuencias en la vida real. Acciones tan cotidianas como saltarse un semáforo, fumar en lugares prohibidos, tener deudas impagadas o cometer fraude, además de su correspondiente sanción administrativa, conlleva ciertas restricciones: como la prohibición de viajar en avión o en trenes de alta velocidad, y la compra de artículos de lujo. En algunas ciudades, se publicita la información de las personas morosas en pantallas LED de centros comerciales, camiones o paradas del autobús, desvelando datos personales y suponiendo un escarnio social para la persona afectada y su familia.

Por si fuera poco, si el crédito social es negativo, las consecuencias de estar en una lista negra pueden consistir en que no se puede viajar en avión, o en determinados trenes, no se pueden contratar algunos destinos turísticos, o alojarse en ciertos hoteles, o no se puede inscribir a los hijos en los mejores colegios». Pero, sobre todo, se produce un aislamiento social del individuo puesto que el contacto con individuos de baja puntuación repercute en la puntuación de quien interactúa con ellos.

Desde esta perspectiva, está claro que las palabras de Cotino Hueso sobre la devastadora incidencia de esta técnica sobre los derechos fundamentales no son nada exageradas.

En definitiva, se trata del control de los comportamientos sociales, cuya implantación es un proyecto perfectamente establecido, como se deduce del documento «Esquema para el establecimiento de un sistema de crédito social» del Consejo estatal para la implantación del sistema, publicado en 2014.

4.2. La falsificación de la realidad

Hemos visto hasta ahora las amenazas que para los derechos fundamentales de las personas suponen los sistemas de IA basados en la utilización de

34 Rettinger, L. «*The Human Rights Implications of China's Social Credit System*», *cit.*, p. 11.

35 Serrano Martínez, A., «Crédito social chino: el sistema de puntos que ya se exporta a otras sociedades», *El Economista,* 17/06/2023. Añade el autor que también Rusia quiere completar una red de reconocimiento biométrico combinando sus propios algoritmos de IA con su enorme sistema de vigilancia pública: Así, en enero de 2020, Moscú implementó un nuevo sistema de reconocimiento facial en tiempo real en toda la ciudad, con más de 160.000 cámaras: https://www.eleconomista.es/economia/noticias/12325879/06/23/credito-social-chino-el-sistema-de-puntos-que-ya-se-exporta-a-otras-sociedades.html.

datos biométricos, persuasión subliminal y calificación social, todos los cuales aparecen definidos o mencionados en la LIA y en otros textos europeos.

Queda tratar ahora una nueva aplicación de la IA claramente lesiva para los derechos humanos, invasiva de la intimidad, pero también atentatoria contra otros derechos, como el derecho a una información veraz, a la libre formación de la opinión, a la libertad de expresión, lesiones que derivan de la falsificación de la realidad mediante sistemas de IA, y la creación de versiones adulteradas de la misma, generalmente con fines fraudulentos. Los ejemplos más claros son los últimos sistemas inteligentes polivalentes con *ChatGpt, Dall-E, Stable Diffusion*, etc., y su uso es sencillo, así el sistema de producción gráfica *Midjourney* convierte las palabras e instrucciones escritas de sus usuarios en imágenes, más o menos realistas, a demanda del cliente. Son muy comunes para los teléfonos móviles las aplicaciones y sitios web para la creación de fotos y videos falsos, que generalmente repiten el esquema de cambiar la cara del utilizador por la de un famoso o a la inversa.

¿Por qué son estos sistemas generalistas o modelos fundacionales tan peligrosos? La peligrosidad de estos «modelos fundacionales» de propósito general no suele apreciarse en su justa medida, es más, el texto del Parlamento europeo sobre LIA de 14 de junio 2023 todo lo más somete a estos sistemas inteligentes a los requisitos y obligaciones de los sistemas de alto riesgo, como señala Navas Navarro[36]. Ello posiblemente deriva de su carácter general y polivalencia, que hace que las utilizaciones más habituales, como instrumentos de trabajo o de entretenimiento desvíen la atención del verdadero daño.

A mi juicio, la peligrosidad de estos sistemas inteligentes deriva de algo en principio, bueno: de su enorme capacidad creativa, que ya alcanza a la creatividad humana. Los nuevos sistemas de IA pueden mimetizar la capacidad creativa humana y producir obras artísticas, gráficas, pictóricas, obras intelectuales como novelas o monografías de investigación y obras musicales, de manera que en mi opinión hay que hablar ya de un «autor cibernético». En el campo del arte se anuncian producciones de nuevas obras de pintores clásicos, fallecidos hace siglos, cuya obra se crea hoy con su estilo y su arte. En cinematografía se «resucita» a actores fallecidos, y se les hace actuar junto a los vivos, o se rejuvenece a los actores ya ancianos, para hacer efecto de regresión en el tiempo. También mimetizan la capacidad de invención del ser humano, y en investigación genética y biotecnológica se prevén grandes avances con el uso de sistemas inteligentes.

Esta capacidad creativa de la IA produce sin embargo dos consecuencias negativas. La primera deriva de la capacidad de la IA para imitar la realidad

36 Navas Navarro, S. (2023), *Chatgpt y modelos fundacionales. Aspectos jurídicos de presente y de futuro*, Reus, Madrid, 2023, p.18.

hasta el punto de que deja de ser posible distinguir la imagen real de la inventada. Estas creaciones o imitaciones de la realidad pueden incidir en la lesión de algunos de los derechos fundamentales, pero en mi opinión el daño es más importante, porque se falsifica la realidad, produciendo una pérdida de los referentes reales y del marco auténtico de la vida humana. Corremos un serio peligro de perder la certeza acerca de nuestros registros de la realidad, con el resultado de producirse una desinformación general y una inverificabilidad de hechos y pruebas. En medios periodísticos e informativos ya no son novedad las intoxicaciones o *deep fakes* y las falsas noticias o *fake news.*

4.3. La imitación de resultados humanos inteligentes

La segunda consecuencia negativa está en la capacidad de la IA de imitar los resultados inteligentes del esfuerzo humano. Esto a la larga nos puede llevar a delegar en la máquina la realización de todo tipo de esfuerzos intelectuales, con la consecuencia de que perderemos destreza intelectual. Esto ya ha ocurrido: desde los años 70 del pasado siglo con la popularización de las calculadoras de bolsillo, nuestras capacidades de cálculo se han visto desplazadas por las máquinas. En la educación, corremos el riesgo de lo que yo llamo «deslocalizar» nuestra inteligencia, al trasladar la producción de resultados inteligentes de nuestro propio cerebro al de la máquina, y encima engañarnos adjudicándonos unos resultados que en realidad no son nuestros. Cada vez es más difícil distinguir la labor intelectual humana de la de la máquina.

Un dato objetivo que refleja esta pérdida de capacidades es la llamada «reversión del efecto Flynn», que describe un continuo descenso del coeficiente de inteligencia o IQ de los estudiantes en general desde mediados de los años 90, cambio que según el propio James Flynn[37] ha tenido lugar conforme los niños han estado más inmersos en la cultura electrónica visual y auditiva moderna: televisión, videojuegos y teléfonos móviles, que alteran la capacidad de pensar en términos de abstracciones y desarrollar la lógica deductiva y la planificación sistemática.

Para el filósofo Carabantes López[38] la culpable es directamente la IA y la tecnología informática, afirmando que desde 1995: «[...] se ha producido un descenso en la demanda social de habilidades intelectuales [...] Al disminuir la demanda de inteligencia, los individuos, lamarckianamente, han dejado de cultivarla [...] la causa principal ha sido la penetración de la informática en la vida cotidiana». La realidad nos está demostrando que, si la

37 Flynn James R./Shayer, M., «IQ decline and Piaget: Does the rot start at the top?», *Intelligence* (2017), https://doi.org/10.1016/j.intell.2017.11.010

38 Carabantes López, M., «Inteligencia artificial lingüística perfecta: efectos sobre la autopercepción del ser humano», *SCIO. Revista de Filosofía*, n.º 18, Julio de 2020, pp. 224 y ss.

inteligencia se puede obtener de fuentes no humanas a bajo coste, ésta pasa a ser una capacidad menos valorada entre los humanos, igual que la fuerza física de un cargador de puerto es menos valorada cuando se tienen carretillas elevadoras y montacargas: se produce una *devaluación social* de la inteligencia.

Y lo que es peor, se produce lo que el informe *Future Risks of Frontier AI,* que se divulgó en la cumbre de Bletchley Park, denomina «Dependencia humana excesiva de los sistemas de inteligencia artificial», que posteriormente no se puede eliminar, dependencia que ya tenemos con las calculadoras digitales o con Google para recordar cualquier dato. Todos hemos oído la frase. «Para qué te lo vas a aprender de memoria, si lo tienes todo en Internet».

V. A modo de conclusión

Se ha intentado utilizar la vía de la ética, para lograr que los comportamientos de los sistemas inteligentes fueran aceptables, en el sentido de que no causasen daño ni lesionasen derechos humanos, hablándose así de una IA alineada con los valores humanos. Pero claramente se está empleando la ética simplemente como sinónimo de exigencia de seguridad. Una IA ética es simplemente una IA segura. Y para imponer eficazmente esta seguridad no queda otra vía que la regulación jurídica.

Por tanto, frente a estos problemas, ¿qué puede hacer el Derecho? Desde la UE se propone una regulación jurídica que acote el uso de la IA para fines beneficiosos, una especie de principio de no maleficencia. Por supuesto, hay que alejar esta tecnología de las etapas escolares de la educación, y la prohibición de las aplicaciones que sean lesivas para los derechos de las personas y el control de la ciudadanía. Todo esto lo digo, además, con pocas ilusiones, y teniendo muy presente que lo deseable sería un control mundial de la IA, similar al de la energía atómica.

Como señala Presno Linera[39], lo necesario es convertir en auténticas normas jurídicas mandatos y prohibiciones específicos «que garanticen un acceso claro y manejable a los datos personales [...] el establecimiento de mecanismos eficaces para combatir las discriminaciones y los sesgos [...] la regulación del uso de las tecnologías de IA por parte de las autoridades policiales y judiciales [...] el régimen de responsabilidad y rendición de cuentas». En suma, una regulación que imponga las exigencias técnicas antes señaladas y regule el buen uso de esta tecnología.

39 Presno Linera, M.A., *Derechos fundamentales e inteligencia artificial*, Marcial Pons – Fundación Manuel Giménez Abad, Madrid, 2020, p. 87.

Bibliografía

Barona Vilar, S., «Tecnología biométrica y datos biométricos. Bondades y peligros. No todo vale», *Actualidad Jurídica Iberoamericana* N.º 21, agosto 2024.

Bermejo Latre, J. L., «La aplicación de la Inteligencia artificial en la actividad formal e informal de la Administración», *ponencia en el XVIII Congreso de la Asociacion Española de Profesores de Derecho Administrativo* (25-26 de enero de 2024, Vigo). https://www.aepda.es/AEPDAEntrada-3987-XVIII-CONGRESO-DE-LA-ASOCIACION-ESPANOLA-DE-PROFESORES-DE-DERECHO-ADMINISTRATIVO.aspx.

Carabantes López, M., «Inteligencia artificial lingüística perfecta: efectos sobre la autopercepción del ser humano», *SCIO. Revista de Filosofía*, n.º 18, Julio de 2020.

Coca Payeras, M., (2023), «Las iniciativas de la Unión europea sobre inteligencia artificial: de la persona electrónica, al difícil equilibrio entre la necesidad de impulsarla y evitar sus riesgos», *Revista de Derecho Civil*, vol. X, núm. 2, especial (junio, 2023). En http://nreg.es/ojs/index.php/RDC.

Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), reunida en París del 9 al 24 de noviembre de 2021, en su 41.ª reunión.

Costas Pérez, E., y Tucat, P., (2021). «Nudges: diseño y evaluación», *Gestión y Análisis de Políticas Públicas*, (25), 8-22. DOI: https://doi.org/10.24965/gapp.i25.10868

Cotino Hueso, L., «El uso jurisdiccional de la inteligencia artificial: habilitación legal, garantías necesarias y la supervisión por el CGPJ», *Actualidad Jurídica Iberoamericana* N.º 21, agosto 2024.

Cotino Hueso, L., «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», *El Cronista del Estado Social y Democrático de Derecho,* N.º 100 (septiembre-octubre), 2022 (Ejemplar dedicado a: Inteligencia artificial y derecho).

De Asís Roig, R. F., «Ética, Tecnología y Derechos», en *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022.

Ebers, M., «El futuro marco jurídico europeo de la inteligencia artificial», *Revista General de Legislación y Jurisprudencia*, 2023, número 2.

Floridi, L., *The Ethics of Artificial intelligence*, Oxford University Press, Oxford, 2023.

Flynn James R./Shayer, M., «IQ decline and Piaget: Does the rot start at the top?», *Intelligence (2017*), https://doi.org/10.1016/j.intell.2017.11.010.

Garriga Domínguez, A., «Inteligencia artificial y el fenómeno de la desinformación: el papel del RGPD1 y las garantías recogidas en la propuesta de la ley de servicios digitales», en *Inteligencia Artificial y Filosofía del Derecho*, Ediciones Laborum, Murcia, 2022.

López Riba, J. M., «Inteligencia artificial y control policial. Cuestiones para un debate frente al hype», *InDret,* 2.2024. DOI: 10.31009/InDret.2024.i2.10.

Megías Quirós, J. J., «Derechos humanos e Inteligencia artificial», *Revista DIKAIOSYNE*, N.º 37, número especial sobre DDHH, en coedición con el Observatorio de Derechos Humanos de la Universidad de Los Andes. Mérida – Venezuela. Enero, 2022.

Ortiz de Zárate-Alcarazo, L. (2023). «Las Políticas conductuales inteligentes: Oportunidades y riesgos ético-políticos de la Inteligencia Artificial para el nudging», *Revista Española de Ciencia Política*, 62. Doi: https://doi.org/10.21308/recp.62.03.

Rettinger, L., «The Human Rights Implications of China's Social Credit System», *Journal of High Technology Law*, vol. XXI, N.º 1, (2021).

Rivero Ortega, R., *Derecho e Inteligencia Artificial*. *Cuatro estudios*, Ediciones Olejnik, Santiago de Chile, 2023.

Roberts, H., Cowls, J., Morley, J. *Et Al.,* «The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation». *AI & Society* 36, 59-77 (2021). https://doi.org/10.1007/s00146-020-00992-2.

Serrano Martínez, A., «Crédito social chino: el sistema de puntos que ya se exporta a otras sociedades», *El Economista*, 17/06/2023. <https://www.eleconomista.es/economia/noticias/12325879/06/23/credito-social-chino-el-sistema-de-puntos-que-ya-se-exporta-a-otras-sociedades.html>.

Urruela Mora, A., «Instrumentos de evaluación del riesgo de violencia, justicia algorítmica y Derecho penal. Perspectiva crítica», en *Estudios político-criminales, jurídico-penales y criminológicos: libro homenaje al profesor José Luis Díez Ripollés*, coord. por Noelia Corral Maraver, Deborah García Magna, Fátima Pérez Jiménez, Bertha Prado, Pablo Rando Casermeiro; Juan Muñoz Sánchez (dir.), Octavio García Pérez (dir.), Ana Isabel Cerezo Domínguez (dir.), Elisa García España (dir.), Tirant lo Blanch, Valencia 2023.

Yamazaki, Y., «An Empirical Study for The Acceptance of Original Nudges and Hypernudges», en *Societal Challenges in the Smart Society* / coord. por Mario Arias Oliva, Jorge Pelegrín Borondo, Kiyoshi Murata, Ana María Lara Palma, Universidad de La Rioja y Universitat Rovira i Virgili, Logroño, 2020. En Dialnet: <https://dialnet.unirioja.es/servlet/articulo?codigo=7867256>.

10.

LA TRANSPARENCIA ANTE EL NUEVO ENTORNO DIGITAL DE LAS ADMINISTRACIONES PÚBLICAS

Joaquín Meseguer Yebra

Académico Correspondiente de la Real Academia de Jurisprudencia y Legislación[1]

SUMARIO: I. Introducción. II. Los derechos digitales ante la IA. III. Gestión automatizada motivada (y explicable). IV. La realidad: la imprescindible tutela de los órganos de control del acceso a la información pública. V. El futuro: el reglamento de inteligencia artificial.

I. Introducción

Hace apenas un par de años escribía las primeras líneas de un artículo de opinión enunciando algunas de las soluciones de inteligencia artificial (IA, en adelante) que ya utilizaban por aquel entonces las administraciones españolas como apoyo para el ejercicio de sus funciones públicas y citaba, entre

1 Coordinador del Grupo de trabajo de transparencia y acceso a la información pública de la Red de Entidades Locales por la Transparencia y Participación Ciudadana (FEMP), Secretario ejecutivo del capítulo español de la Red Académica de Gobierno Abierto Internacional (RAGAInt) y experto del Foro de Gobierno Abierto de España. Ha sido Subdirector General de Transparencia del Ayuntamiento de Madrid y Director General de Transparencia y Buen Gobierno de la Junta de Castilla y León. Sobre esta temática es autor además del capítulo «Derechos digitales, inteligencia artificial y transparencia» en la obra «Los derechos de la ciudadanía ante la administración digital», coordinada por Medina Guerrero, M. y publicada por el Centro de Estudios Políticos y Constitucionales (2023); de la tribuna «Más transparencia para una Administración algorítmica» publicada en la edición del diario EL PAÍS de 14 de septiembre de 2022 y del post «Sin transparencia, no puede haber inteligencia. Tampoco artificial», publicado en el blog de la Fundación Hay Derecho el 10 de enero de 2024.

otras, Viogén, Saler, Veripol, Bosco, RisCanvi, Hermes o Send@. A la velocidad sideral —fotónica, como dicen algunos— a la que evoluciona en el mundo este nuevo fenómeno ya propio de nuestra realidad cotidiana, es difícil intuir cuál será su devenir y algunos de los retos que nos planteará, amén de los que ya están actualmente encima de la mesa. El futuro empieza a ser esa dimensión del tiempo, también verbal, que se desdibuja cada vez más rápido, consumido por un presente que fagocita todo lo que intuye va a suceder. Solo en los meses que han discurrido desde la intervención de este autor en el Instituto Aragonés de Administración Pública, España ya cuenta con la AESIA, la Agencia Española de Supervisión de la Inteligencia Artificial, con una nueva Estrategia Nacional de Inteligencia Artificial (ENIA)[2], con un Reglamento europeo de Inteligencia Artificial (en adelante, RIA)[3] y una primera regulación nacional para la creación de un entorno controlado de pruebas *(sandbox)* para su ensayo[4].

Sin ahondar en cuestiones tecnológicas más complejas, la administración camina actualmente hacia un funcionamiento cada vez más automatizado para aumentar su eficiencia ante el gran volumen de datos que debe tratar. Los poderes públicos están abandonando poco a poco el mundo analógico para sumergirse en la llamada «revolución industrial 4.0» que nos ha traído el internet de las cosas (IoT) y los sistemas de aprendizaje automatizado que, imitando redes neuronales, pueden convertirse en verdaderas «black boxes» de las que poco o nada se sabe sobre su forma de operar.

Preguntémonos si no es inquietante que un sistema de IA resultara ser el tercer candidato más votado en las elecciones a la alcaldía de un distrito de Tokyo en 2018. El dato nos hace cuestionarnos si, hoy por hoy, confiamos más en la gestión que podría llevar a cabo una máquina que la que podría realizar cualquiera de nuestros congéneres humanos. ¿Qué es lo que mayor o menor confianza nos despierta en quien tiene que tomar una decisión que afecta a nuestra esfera de derechos? Dependiendo de nuestra escala de valores o prejuicios, que en buena medida vendrán dados por nuestra educación, experiencia o contexto social, habrá múltiples factores. Pero seguramente aparecerá uno en todas las respuestas y es la capacidad de explicar por qué razones, por qué motivos la decisión se ha tomado. Está claro: confiamos

2 https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf

3 Reglamento (UE) 2024/1689 del Parlamento europeo y del Consejo de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828.

4 Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial

más en quien puede explicar su conducta, su forma de actuar. O, también, de quien entendemos sin necesidad de ninguna explicación, por su obviedad, la respuesta a una pregunta. Confiamos en lo que entendemos, lisa y llanamente.

Esto mismo, aplicado a la gestión pública, actúa con iguales premisas, si bien cuando la actuación limita o perjudica nuestros derechos e intereses, siempre la recibimos con una natural resistencia o desagrado. Esto es humanamente comprensible, aun cuando la decisión proceda de una máquina, y no de una persona.

Este razonamiento viene al hilo, esencialmente, del intenso debate sobre las posibilidades que tienen o no los sistemas de decisión automatizada de poder explicar cómo operan, como actúan, y de que todos seamos capaces de entenderlo. Nos ha llevado siglos conseguir que la motivación de las decisiones administrativas fuera una de las garantías indispensables para salvaguardar nuestro derecho a defendernos ante ellas, tanto más cuanto más discrecionales son, como para poner ahora en cuestión que este no deba ser un requisito también cuando la decisión la toma un algoritmo o un sistema de IA. No podemos admitir una rebaja de esta exigencia cuando el diseño de estas herramientas corresponde también a personas.

Sacudidos y, casi diría, alarmados por esta vertiginosa evolución, surge una preocupación por incorporar a esta gran transformación digital una visión más humanista y ética de la tecnología. Cobra ahora más sentido aquello que hace tiempo escribía en un documento sobre retos para la administración pública del futuro: «necesitamos sí, más tecnología e innovación en la administración, pero no nos olvidemos de empaparla no solo de inteligencia artificial y algoritmos, sino también de ética, que es a aquellas, como el alma al cuerpo»[5].

II. Los derechos digitales ante la IA

El empleo de la IA por las Administraciones públicas debe ser analizado en el contexto del reconocimiento de los derechos digitales en nuestro país, que va cobrando mayor intensidad con la aprobación de algunos documentos e instrumentos de «soft law» como la Carta de Derechos Digitales (Carta, en lo sucesivo), adoptada por nuestro Gobierno en julio de 2021 con el propósito de servir como marco referencial para garantizar los derechos de la ciudadanía en la nueva realidad digital y reconocer los retos que plantea la adaptación de los derechos actuales al entorno virtual y digital[6].

5 Retos 2021 para una Administración Innovadora. Fundación Novagob: https://novagob.org/wp-content/uploads/2021/02/Publicacion-Retos-NovaGob-2021_compressed.pdf

6 Se puede acceder al texto en la siguiente url: https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

Mucho se ha hablado y se sigue haciendo sobre si nos hallamos ante nuevos derechos o, por el contrario, ante los mismos derechos del mundo analógico ya conocidos, pero en su versión digital, un debate más doctrinal que práctico en la medida que el reconocimiento y garantía efectiva de cualquier derecho pasa necesariamente por la aprobación de normas que así lo contemplen y, en el caso de derechos fundamentales, por el encaje que estos puedan tener en nuestra Constitución a la vista, también, de la interpretación que se desprenda de los tratados y acuerdos internacionales ratificados por España.

Lo que no debería admitir discusión es que el desarrollo y avance tecnológicos deben pivotar en todo caso en torno a las personas y a la necesidad de que, con todos los equilibrios y balances que sean precisos, la tecnología facilite el ejercicio de los derechos, pero no los dañe o arrolle. Debemos recuperar o reforzar la visión humanista y ética de la tecnología, adoptar una gobernanza digital vinculante con este enfoque. No es irrelevante ni debería pasar inadvertido, en absoluto, el uso de la preposición que emplea la Carta para referirse a los derechos «ante» la inteligencia artificial, lo que parece sugerir de alguna manera una voluntad del autor del texto de proteger los derechos de las personas ante el avance de la inteligencia artificial.

Lo cierto es que la Administración ya aplica soluciones algorítmicas «regladas», simples fórmulas para resolver un número muy importante de procedimientos que, en caso de exigir la intervención humana, ralentizarían de una forma significativa, si no fatalmente, una gestión mínimamente eficiente. Y no es menos cierto que de una manera cada vez más creciente empieza a utilizar otro tipo de instrumentos de inteligencia artificial cuyas condiciones de explicabilidad son menores, hasta el punto de poder calificar algunos de ellos de verdaderas cajas negras, lo que razonablemente empieza a suscitar importantes recelos ante el temor de que puedan vulnerar derechos fundamentales o arrojar soluciones injustas o inequitativas. La cuestión es que, como todos los expertos advierten, hay una relación inversa entre explicabilidad y rendimiento de los modelos, tal que cuanto más sofisticados resultan ser —esto es, más valor pueden añadir a lo que la mente humana ya hace—, más difícil es saber cómo operan o actúan.

La desconexión digital en este contexto puede ser definido como el derecho que correspondería a cualquiera de exigir que la decisión que pudiera afectarle se tome por personas, por humanos. La traducción de esto la encontramos en la Carta, que garantiza a las personas que no quieran o puedan utilizar recursos digitales a elegir otras alternativas en el mundo físico que garanticen sus derechos (apartado 3.XVIII). Pero, por otra parte, la Carta también dice que quien resuelva separándose del criterio propuesto por un sistema automatizado o inteligente, deberá razonar o explicar su decisión. Esto es, la carga de la motivación se hace recaer sobre las personas cuando estas consideren que la «máquina» sugiere un modo de proceder que difiere del juicio de quienes son titulares y pueden/deben ejercer las competencias.

Y todo ello en un momento como el actual en el que la norma básica que establece las reglas generales que rigen los procedimientos administrativos, la Ley 39/2015, de 1 de octubre, aún no ha determinado que los algoritmos tengan el valor de precedente administrativo como para tener que justificar por qué decidimos apartarnos de sus predicciones o «decisiones».

El panorama futuro, pues, por mucho que nos resistamos a admitirlo, apunta a que será difícil —o, incluso, poco recomendable— que quien lo desee pueda mantener un estatus analógico. Esto no significa, insistimos, que la gestión automatizada o los sistemas de IA no sean herramientas que presentan indudables ventajas y oportunidades en el campo de la gestión pública. Sería absurdo negar que en el futuro tendremos que apoyarnos necesariamente en este tipo de soluciones, tanto como obviar que estas soluciones también albergan importantes riesgos, como demuestra, solo a título de ejemplo, que la Ley francesa núm. 2019-222 de programación 2018-2022 y de reforma para la justicia, prohíba la reutilización y análisis de datos de jueces y magistrados con el objeto de predecir su actuación al resolver o que ya haya algunos países que se hayan mostrado reticentes al uso de sistemas de reconocimiento facial biométrico en espacios públicos.

Y aquí es donde el pánico puede hacer que nos agarremos a la butaca porque el vértigo no es pequeño. Para empezar: ¿sabemos cuáles de estos sistemas o modelos están utilizando las administraciones actualmente? ¿En qué procedimientos? ¿Resuelven por sí solos, sin intervención humana, o son únicamente predictivos, sirven de ayuda para que la persona tome la decisión? Hoy por hoy sabemos poco o casi nada. Como solemos bromear cuando alguien es el último en enterarse de su propia defunción, solo la prensa nos saca de dudas cuando trae a primera página algún conflicto relacionado con este asunto.

III. Gestión automatizada motivada (y explicable)

Contamos en la actualidad con una variedad importante (y en aumento) de algoritmos y sistemas de IA de todo tipo que tratan datos personales, aunque también conviene advertir que no todos lo hacen. Pensemos, por ejemplo, en los modelos de predicción meteorológica. No huelga hacer esta precisión por cuanto, al igual que sucede en materia de transparencia, la protección de datos personales y los efectos de su tratamiento suele ser uno de los principales escollos que nos encontramos en el abordaje de esta materia, y no siempre con la suficiente justificación.

En las actuaciones automatizadas, el resultado que arroja el algoritmo o la predicción que suministra el sistema de IA resulta ser en muchos casos parte o toda la justificación que requiere una decisión pública. La necesidad de motivar, lo sabemos, era y es un requisito en el mundo analógico según dispone el artículo 35 de la Ley 39/2015, de 1 de octubre, de Procedimiento

Administrativo Común de las Administraciones Públicas. Así lo decían también las leyes que la precedieron en el siglo pasado. Esto significa que las decisiones que hoy en día se apoyan en los razonamientos, juicios o criterios que ofrecen aquellas soluciones tecnológicas deben garantizar que puedan explicarse en iguales condiciones y que se entienda la decisión adoptada. Sería inaceptable dispensar de un requisito así, más aún cuando las decisiones o su fundamentación pueda proceder de una fuente no humana.

Los órganos de control en materia de transparencia ya han empezado a decirlo en «voz alta y clara». El Consejo de Estado Italiano, en su sentencia 08747/2019, de 13 de diciembre, que resuelve el recurso interpuesto contra la sentencia del Tribunal Administrativo del Lazio en el asunto que enfrentaba al Gobierno con un grupo de docentes que había cuestionado la gestión del sistema de concursos de profesorado por un algoritmo, concluye que estos sistemas deben respetar «el pleno conocimiento previo del modelo utilizado y de los criterios aplicados». Y es que, conocer el algoritmo (cómo opera a partir de unos determinados datos para obtener un resultado) permitiría identificar posibles sesgos, vulneraciones de derechos, etc., e, incluso, mejorar su funcionamiento dado que, a la vista de todos, todos podrían también de alguna forma contribuir a la corrección de posibles fallos que se detecten.

Esto no tiene una relevancia anecdótica. Todo lo contrario. Pensemos que en el funcionamiento de los algoritmos y los sistemas de IA puede haber sesgos (bias, en la nomenclatura anglosajona) de todo tipo y con impacto en diferentes momentos:

– En los datos —inputs— que utilizan aquellos o con los que se entrenan para crear patrones. Los conjuntos de datos (datasets) que se utilizan para esta tarea buscan representar el mundo real, pero la mayoría de las veces no consiguen hacerlo. O reflejan estereotipos y prejuicios, o bien, la base o muestra de datos de entrenamiento resulta no ser completa y, por este motivo, no recoge todo el universo de situaciones posibles. Esto determinará irremisiblemente que los resultados de la toma de decisión automatizada replicarán estos sesgos.

  Un buen ejemplo sería el paradigmático caso Estado de Wisconsin contra Loomis, también conocido como caso COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), según se cite el litigio o el algoritmo. COMPAS lo utilizan las autoridades judiciales en el proceso penal norteamericano para valorar el grado de riesgo de reincidencia del sujeto y así ayudar a fijar la condena. Opera sobre la base de un análisis complejo que implica el uso de la información obtenida de una encuesta de 137 preguntas (que versan, entre otros ámbitos, sobre raza, sexo, entorno o pobreza del barrio donde se reside), divididas en varias secciones, así como sobre información relativa a los antecedentes penales individuales. Se realiza una entrevista a la persona procesada con el fin de obtener toda la

información necesaria para efectuar la ponderación y determinar un puntaje del nivel de riesgo de reincidencia. Del caso concreto citado se desprendió como una de las conclusiones que las personas de raza negra tenían el doble de posibilidades que las personas blancas de ser clasificadas por error como de «alto riesgo».

– En el propio código fuente. Los algoritmos están diseñados por personas, y las personas tenemos todas nuestra propia escala de valores, prejuicios, experiencias y opinión.

  Son buenas muestras de esta modalidad de sesgos los casos DELIVEROO o SyRI. El primero aborda el caso de la plataforma de Deliveroo, una de las más populares de envío de comida a domicilio. El rider accede a la plataforma para reservar franjas horarias de forma anticipada, pudiendo elegir en el momento de entrar tanto horarios como área o áreas disponibles. Los riders pueden entrar cada lunes para reservar sus horarios de trabajo según la puntuación que tengan en el ranking de la empresa. Los riders con mejor puntuación pueden acceder a partir de las 11:00 horas, el 25 % siguiente a partir de las 15:00 horas, y el resto a partir de las 17:00 horas.

  La puntuación de los riders depende de la fiabilidad y disponibilidad. El índice de fiabilidad tiene en cuenta el número de ocasiones en que el rider no ha cumplido con una sesión de trabajo previamente reservada, mientras que el índice de disponibilidad tiene en cuenta el número de veces que el rider está disponible en los horarios de mayor demanda (de 20:00 a 22:00 horas de viernes a domingo).

  Pues bien, el Tribunal Ordinario de Bolonia (Italia) consideró que se había producido una discriminación indirecta contra los trabajadores de Deliveroo debido a que la plataforma actúa con una «inconsciencia y ceguera» deliberada respecto de ciertas causas (justificadas) de incumplimiento de la reserva de trabajo realizada por el rider. Entiende el Tribunal que esta discriminación se produce cuando se da el mismo trato a situaciones distintas que merecen un trato diferente. En otras palabras, aprecia la discriminación porque la aplicación valora negativamente tanto la ausencia o falta de puntualidad del rider sin discriminar si esta es caprichosa o se debe a circunstancias tales como el ejercicio del derecho de huelga, un estado de enfermedad o el cuidado de menores a cargo.

  El caso SyRI (System Risk Indication) versa sobre un sistema de detección de fraudes de asistencia social utilizado por el gobierno en los Países Bajos. El resultado del procesado de datos por este sistema son informes de riesgo sobre la probabilidad de que alguien defraude a la seguridad social, pudiendo aconsejar que se lleve a cabo una investigación. Se trata de un instrumento de ayuda para las autoridades públicas encargadas de la investigación del fraude a la

seguridad social, que pueden valerse de las inferencias del sistema para sus tareas de inspección y, en caso de detectar incumplimientos o fraudes, iniciar procedimientos sancionadores.

El modelo de riesgo elaborado por SyRI presentaba efectos no deseados según el Tribunal de Distrito de La Haya, al estigmatizar y discriminar a determinados ciudadanos, no solo por cantidad de información personal que recopilaba, sino también por la falta de transparencia del algoritmo y, especialmente, por el uso sesgado del instrumento, utilizado en barrios donde vivían personas con rentas bajas o zonas donde residían personas pertenecientes a minorías.

Los algoritmos pueden emplearse como fórmula de apoyo o auxilio en la toma de decisiones, pero también, incluso, pueden tomar la propia decisión, lo que denominamos «automatización íntegra». El artículo 41 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público define actuación automatizada como aquel acto o actuación realizada íntegramente a través de medios electrónicos por una Administración en el marco de un procedimiento administrativo y en la que no haya intervenido de forma directa un empleado público.

De hecho, el artículo 22 del Reglamento General de Protección de Datos[7] habla de decisiones basadas íntegramente en tratamientos automatizados que produzcan efectos jurídicos o afecten a la persona de forma significativa. En estos casos, existe el derecho, con algunas excepciones, a no ser objeto de decisiones de este tipo o a exigir la supervisión/intervención humana (apartado 5.XXV.3 de la Carta). Según la Agencia Española de Protección de Datos (AEPD), deben evitarse este tipo de sistemas y dar la opción siempre a que un operador humano pueda ignorar el algoritmo en un momento dado. Adoptar sistemas de este tipo, además, conlleva decidir a quién se imputa el acto, cuál es el régimen de impugnaciones, el régimen de responsabilidad ante eventuales daños, etc. Los sistemas de gestión automatizada ponen a prueba la teoría clásica del órgano, que tendría que adaptarse a las nuevas condiciones que impone un modelo diferente de ejercicio y, en su caso, de atribución de competencias.

Por otra parte, y retomando el debate sobre la explicabilidad, en aquellos casos en los que cualquiera sea objeto de una decisión de este tipo la AEPD considera que las personas deben poder disponer de información significativa sobre la lógica aplicada, la importancia y las consecuencias previstas y, en concreto, sobre:

– El detalle de los datos empleados para la toma de decisión, más allá de la categoría en particular, e información sobre los plazos de uso de los datos.

7 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.

- La ponderación de cada uno de ellos en la toma de decisión.
- La calidad de los datos de entrenamiento y el tipo de patrones utilizados.
- Los perfilados realizados y sus implicaciones.
- Los valores de precisión o de error según la métrica adecuada para medir la bondad de la inferencia.
- La existencia o no de supervisión humana cualificada.
- La referencia a auditorías, así como la existencia de certificaciones realizadas sobre el sistema de IA.
- En el caso de que el sistema de IA contenga información sobre terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo.

Luego está determinar, si es que consideramos que esta cuestión es relevante, en qué tipo de potestades o competencias debería ser más aceptable el uso de este tipo de soluciones. Las facultades que se ejercen en la administración pueden ser del todo (o en buena parte) regladas, esto es, todos o la mayor parte de sus presupuestos y condiciones estar preestablecidos en la norma y, una vez corroborados, solo tener cabida una aplicación automática de aquella. Por el contrario, si la norma puede dejar un margen abierto o libre a la apreciación subjetiva de quien ejerce la competencia, pero siempre dentro del marco jurídico. Esto último es lo que denominamos, potestades o competencias discrecionales.

La pregunta es si puede un algoritmo o un sistema de IA concretar en el caso concreto el contenido de un concepto jurídico indeterminado, resolver qué es o no equitativo o apreciar circunstancias que no estén tipificadas, estandarizadas o mecanizadas con un previo valor asignado. Esta cuestión es importante de resolver para despejar las suspicacias que el uso de estos sistemas puede suscitar y, personalmente, me recuerda a la publicidad de una conocida marca de refrescos donde se representaba una entrevista de trabajo llevada a cabo por un sistema de IA a un candidato humano. La entrevistadora concluía diciendo que no había nada que los seres humanos pudieran aportar que ellos no pudieran hacer, a lo que el joven respondía que ellos solo se actualizan, y versión a versión quedan obsoletos y necesitan de sustitución. Los seres humanos son compatibles con más de siete millones de sistemas operativos y, de una forma poética y sugerente, aseguraba que se reafirman, se reenamoran y renacen a cada instante, y eso tiene más valor que la primera vez que lo hacen. Una manera romántica de expresar los límites a los que puede enfrentarse un modelo de este tipo.

Parece haber cierto consenso en que el ámbito material en el que deben o pueden operar estas soluciones no debe ser incondicionado y que los ámbitos más reglados son los idóneos para que puedan desplegarse. La sentencia 08747/2019, de 13 de diciembre, del Consejo de Estado italiano en el asunto

del algoritmo para los concursos de traslados de profesores así viene a exigirlo y también el § 35 de la Ley alemana de procedimiento administrativo cuando expresa que «un acto administrativo puede ser adoptado enteramente por órganos automáticos, siempre que así lo permita la ley *y que no exista ni discrecionalidad ni margen de apreciación*».

La tendencia que apreciamos en nuestro país, materializada a través de una inflada relación de estrategias y guías, habla de la necesidad de incrementar la transparencia y monitorización en el uso de estas tecnologías. Solo por hacer una referencia a las más importantes, podemos citar:

- Guía sobre adecuación al Reglamento General de Protección de Datos de tratamientos que incorporan IA. Una introducción —AEPD, febrero 2020[8]—.
- España Digital 2026[9].
- Estrategia Nacional de Inteligencia Nacional (ENIA), en su última versión de 2024[10].
- Carta de Derechos Digitales —julio 2021—.
- Información algorítmica en el ámbito laboral. Guía práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral —Ministerio de Trabajo y Economía Social, mayo 2022[11]—.

Detengámonos en el documento de la ENIA 2024, cuyo tercer eje, nada más y nada menos se titula «Desarrollar una IA transparente, responsable y humanística». La primera conclusión a la que podemos llegar tras su lectura y de la comparación con la primera versión de la estrategia de 2020 es un cierto «desvanecimiento» del requisito de explicabilidad y transparencia en el uso de estos modelos por parte de la administración española. Es totalmente sintomático que el primero de aquellos términos solo aparezca en una única ocasión en aquel documento (para hablar de los modelos de lenguaje) y que la transparencia se diluya en cuestiones de ética a las que, indudablemente, está ligada, pero con las que no se identifica «matemáticamente». Da la impresión de que el foco ha virado hacia otros intereses y de que la responsabilidad sobre este atributo a cumplir por parte de la IA se hace reposar estrictamente en la labor que la AESIA debería desplegar en los

8 Se puede acceder al texto en la siguiente url: https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

9 Se puede acceder al texto en la siguiente url: https://espanadigital.gob.es/sites/espanadigital/files/2022-07/Espa%C3%B1aDigital_2026.pdf

10 Se puede acceder al texto en la siguiente url: https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf

11 Se puede acceder al texto en la siguiente url: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/trabajo14/Documents/2022/100622-Guia_algoritmos.pdf

próximos años. Es a ella a quien se le confía la definición del «nivel de autonomía, transparencia y confiabilidad requeridos en los procesos de desarrollo, implantación y uso de estas tecnologías inteligentes». Esto implica establecer, con claridad y con un elevado nivel de consenso social, los límites en los procesos de toma de decisiones y la forma de interactuar entre las personas y los desarrollos realizados por la IA. La confiabilidad y la transparencia en el uso de la IA son elementos esenciales para su utilización y expansión. Para conseguir esa confianza ante usuarios y ciudadanos, es preciso que haya una correcta identificación de los beneficios de la IA, al mismo tiempo que un reconocimiento de sus riesgos y de los factores que los puedan mitigar. Esto manifiesta la ENIA, y yo añadiría, transparencia en cuanto al funcionamiento de los propios sistemas, no solo de sus beneficios o riesgos.

Afortunadamente, el grado de imperatividad de las reglas de transparencia con respecto a los proveedores de soluciones y las administraciones usuarias no se reduce solo a documentos de esta naturaleza, sino que el derecho positivo empieza a acoger también obligaciones en este ámbito:

- El artículo 157.2 de la ya citada Ley 40/2015, de 1 de octubre, prevé que las aplicaciones desarrolladas por las Administraciones o que hayan sido objeto de contratación y de cuyos derechos de propiedad intelectual sean titulares, podrán ser declaradas como fuentes abiertas, cuando de ello se derive una mayor transparencia en el funcionamiento de la Administración Pública o se fomente con ello la incorporación de los ciudadanos a la Sociedad de la información.
- La llamada ley «rider» (Ley 12/2021, de 28 de septiembre, por la que se modifica el texto refundido de la Ley del Estatuto de los Trabajadores, aprobado por el Real Decreto Legislativo 2/2015, de 23 de octubre, para garantizar los derechos laborales de las personas dedicadas al reparto en el ámbito de plataformas digitales).

  Esta ley reconoce el derecho del comité de empresa a ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles.
- El artículo 23 de la Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación, establece que las administraciones favorecerán la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones que se utilicen en las administraciones tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas, siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.

Las administraciones priorizarán la transparencia en el diseño y la implementación y la capacidad de interpretación de las decisiones. Promoverán el uso de una IA ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido, y la creación de un sello de calidad de los algoritmos.

– La Ley 1/2022, de 13 de abril, de Transparencia y Buen Gobierno de la Comunidad Valenciana, obliga ya a publicar la relación de sistemas algorítmicos o de inteligencia artificial que tengan impacto en los procedimientos administrativos o la prestación de los servicios públicos con la descripción de manera comprensible de su diseño y funcionamiento, el nivel de riesgo que implican y el punto de contacto al que poder dirigirse en cada caso, de acuerdo con los principios de transparencia y explicabilidad (entró en vigor en mayo de 2023)[12].

– El Decreto-ley 2/2023, de 8 de marzo, de medidas urgentes de impulso a la inteligencia artificial en Extremadura, anima a que las administraciones públicas favorezcan la puesta en marcha de mecanismos para que los algoritmos involucrados en la toma de decisiones tengan en cuenta criterios de minimización de sesgos, transparencia y rendición de cuentas siempre que sea factible técnicamente. En estos mecanismos se incluirán su diseño y datos de entrenamiento, y abordarán su potencial impacto discriminatorio. Para lograr este fin, se promoverá la realización de evaluaciones de impacto que determinen el posible sesgo discriminatorio.

  Prevé la debida publicidad del mecanismo de decisión, de las prioridades asignadas en el procedimiento de evaluación y de la toma de decisiones, así como de todos los datos que puedan impactar en su contenido. En el Inventario de Información Administrativa se dejará constancia sobre las validaciones realizadas por el órgano responsable de los procedimientos administrativos respecto del proceso lógico diseñado para la realización de los actos, los riesgos que implica y cualesquiera otros aspectos que garanticen los derechos de los interesados.

12 Recomiendo la lectura del informe «La implantación de la transparencia algorítmica Informe para la implantación del registro de algoritmos públicos y el cumplimiento de las obligaciones de publicidad activa de la Ley 1/2022 valenciana de transparencia», coordinado por COTINO HUESO, L. Informe realizado en el marco del convenio de la Cátedra PAGODA específico para esta actividad y Proyecto de Investigación del Plan Nacional de I+D+I «Registro público de algoritmos» (Ref: PDC2022-133890-I00) 2022-20: https://www.uv.es/cotino/publicaciones/informe_gvav3logos.pdf

IV. La realidad: la imprescindible tutela de los órganos de control del acceso a la información pública

Lo curioso es que, a pesar de todas estas referencias normativas, aún no tenemos en nuestro país ni siquiera un mero listado de los sistemas de IA que las administraciones utilizan. Pero más paradójico resulta que los ejemplos de herramientas de esta naturaleza que un documento oficial como la ENIA menciona sean, precisamente, de otros países que sí ofrecen mayor nivel de apertura. Cita desde organismos nacionales como la Poste Italiane o el Servicio de Empleo austriaco hasta ayuntamientos como el de París, Heidelberg (Alemania) o Lidingö (Suecia). Más conocidos y citados son los ejemplos de Amsterdam, Helsinki, la región de Antibes o el ayuntamiento de Nantes, estos dos últimos en Francia, que están publicando los registros de algoritmos que utilizan —la mayor parte de ellos, chatbots— (por ejemplo, multas de aparcamiento, detección de alquileres turísticos ilegales, recogida de residuos, detección de personas en riesgo de pobreza, alerta de aglomeraciones de personas, situaciones de tráfico peligrosas, gestión de los fondos bibliotecarios, etc.). Muchos de estos proyectos se pueden consultar en el Atlas del Observatorio Global de Inteligencia Artificial Urbana[13].

En nuestro país, el Consorcio de Administración Abierta de Cataluña (AOC) es uno de los organismos que está dando más pasos en esta apertura tecnológica con el fin de facilitar información sobre la lógica de los algoritmos empleados así como de sus consecuencias, con el fin de que sea clara y comprensible, haga posible a las personas entender por qué un algoritmo arroja un determinado resultado y permita la impugnación de las decisiones tomadas con uso de estas herramientas, abriendo los algoritmos en el control democrático.

Para conseguir todos estos propósitos, la AOC propuso la publicación en el Portal de Transparencia de fichas comprensibles para cada solución de IA empleada en los servicios de administración digital, a modo de la plantilla elaborada hace unos años por la Oficina Digital y de Datos británica para ayudar a las administraciones a dar una información estándar sobre cada sistema de inteligencia artificial. En esa plantilla —también denominada en el argot técnico «tarjeta modelo»—, de uso no obligatorio, se incluía una descripción no técnica del sistema (el porqué de su utilización, cómo funciona y qué problema trata de resolver) y otras más técnica (quién es el responsable y titular, para qué se ha diseñado, cómo influye en la toma de decisiones, si está supervisada por humanos, su impugnación, datos que utiliza, evaluaciones de impacto realizadas, descripción de riesgos y acciones previstas para mitigarlos).

13 https://gouai.cidob.org/atlas/

Para dar cumplimiento a esta finalidad, la AOC ha ido aprobando las fichas de tres soluciones algorítmicas (Chatbots conversacionales, videoidentificación para obtener el idCat Móvil y el informe de pobreza energética) siguiendo el siguiente modelo: visión general (caso de uso, estado y principales beneficiarios); información de contacto del responsable; información más detallada del servicio (conjuntos de datos de entrada y salida, procesamiento de los datos, supervisión humana, cumplimiento normativo del sistema y gestión de riesgos) y otra información.

Y más recientemente, recién «salidos del horno», tenemos los ejemplos de Euskadi y la Comunitat valenciana. En el primero de los casos, el Departamento de Gobernanza, Administración Digital y Autogobierno ha publicado un Catálogo de algoritmos y sistemas de Inteligencia Artificial del Sector Público[14], que pretende ser un escaparate que aporte una visión general de los algoritmos y sistemas de IA desarrollados por el Gobierno Vasco y su sector público y que arranca con 13 soluciones. De cada una de ellas, se puede conocer el departamento, persona responsable y su contacto, el responsable técnico, el ámbito al que pertenece, el año de implantación y si está en producción y el proveedor. Añaden, además, los conjuntos de daños con los que se entrena, la forma en la que el sistema procesa automáticamente los datos de entrada para generar los resultados del modelo, si ha sido necesario implantar alguna medida para promover la igualdad y la no discriminación por motivos de edad, origen, nacionalidad, idioma, religión, creencia, opinión, actividad política, etc., la forma en que las personas están al tanto o controlan el sistema, asegurándose de que no produzca efectos adversos o no deseados y los riesgos conocidos asociados con el sistema y los métodos utilizados para mitigarlos.

En la Comunitat valenciana, el Registro de Algoritmos ConfIA[15] es una aplicación informática en la que se registran diferentes características de un sistema informático que contiene inteligencia artificial o que, sin contenerla, influye en la toma de decisiones públicas o en la esfera jurídica de la ciudadanía. Es decir, no recoge todos los sistemas algorítmicos implantados en la Administración, sino aquellos que por su relevancia pueda afectar de alguna forma a esferas sensibles de los derechos y libertades. Es algo así como un inventario de sistemas cuyo propósito es centralizar información sobre estas herramientas, que abarca desde características del diseño y funcionamiento del sistema, evaluación del riesgo, uso de datos de carácter personal, características de calidad, etc.

La realidad demuestra, sin embargo, que, a pesar de los compromisos formales o las obligaciones legales, las administraciones siguen arrastrando los pies cuando de transparencia algorítmica hablamos. Ni existe publicidad

14 https://www.euskadi.eus/inteligencia-artificial/

15 https://gvaoberta.gva.es/confia/cerca

activa real sobre los algoritmos y modelos de IA que se utilizan, ni se hacen grandes esfuerzos por facilitar el conocimiento de su funcionamiento. Una buena muestra podemos hallarla en las resoluciones denegatorias al ejercicio del derecho de acceso a la información pública cuando se formulan solicitudes sobre esta materia.

Son los Comisionados y Consejos de Transparencia los que están defendiendo con mayor tesón la apertura de esta información partiendo de experiencias comparadas como la de la *Commission d'accès aux documents administratifs* francesa (CADA). Nuestro Consejo estatal de Transparencia y Buen Gobierno (CTBG) ha resuelto en varias ocasiones que «...mientras no se instauren otros mecanismos que permitan alcanzar los fines señalados con garantías equivalentes —como podrían ser, por ejemplo, auditorías independientes u órganos de supervisión—, el único recurso eficaz a tales efectos es el acceso al algoritmo propiamente dicho, a su código, para su fiscalización tanto por quienes se puedan sentir perjudicados por sus resultados como por la ciudadanía en general en aras de la observancia de principios éticos y de justicia»[16]. En sentido parecido se ha pronunciado la Comisión de Garantía del Derecho de Acceso a la Información Pública catalana (GAIP)[17].

La defensa de la transparencia algorítmica es absolutamente compatible con la aceptación de límites, algo consustancial a cualquier derecho. Mucho se ha hablado de la propiedad intelectual e industrial o de los secretos comerciales, límites difícilmente comprensibles cuando de productos generados in-house por la propia administración o sobre los que posee los derechos de explotación se trata. Más justificado está invocar la seguridad pública, pero muy restringida su aplicación a un puñado de casos donde los riesgos son claramente identificables. Hay decisiones de la CADA francesa que deniegan el acceso al código fuente del Sistema de Alerta y de Información de la Población, que permite notificar mensajes de alerta ante sospecha de eventuales

16 Resolución RT/0379/2020, de 10 de noviembre, del CTBG: documentos sobre el sistema de cálculo de la asignación para el funcionamiento a los centros públicos docentes de Madrid y sobre los criterios objetivos utilizados para valorar las necesidades de gasto. Resolución R/0058/2021, de 20 de mayo, del CTBG: algoritmo con el que el programa de cálculo de la pensión realiza los porcentajes de cálculo correspondientes a los años completos de servicio efectivo. Resolución RT/0253/2021, de 19 de noviembre, del CTBG: código fuente de la aplicación utilizada para el sorteo de tribunales asociado a procesos selectivos de personal docente en la Comunidad de Madrid. Resolución RT/0748/2021, del CTBG: código fuente asociado al algoritmo de la aplicación utilizada para la admisión de alumnos para el curso 2021/2022 en la Comunidad de Castilla-La Mancha.

17 Resoluciones de 21/9/2016 y 200/2017, de 21 de junio, de la GAIP: acceso al algoritmo para la selección de miembros de tribunales de las pruebas de acceso a la universidad. Resolución 93/2019, de 22 de febrero, de la GAIP: reafirma el carácter de información pública de los algoritmos o bases de datos informáticas. Resolución 1023/2021, de 18 de noviembre, de la GAIP: acceso al informe técnico que describía el análisis estadístico realizado para establecer la fórmula de cálculo de la base imponible del impuesto sobre las emisiones de dióxido de carbono de vehículos de tracción mecánica.

atentados o eventos excepcionales de seguridad civil, o de la aplicación ALICEM, que permite el acceso a servicios públicos en línea mediante la autentificación certificada de identidad por reconocimiento facial biométrico.

En el caso español, no se ha venido apreciando riesgo para el límite de la seguridad púbica en los casos en los que se ha planteado el acceso a algoritmos o códigos fuente, básicamente por falta de argumentación de las resoluciones denegatorias, pero en el caso BOSCO, promovido ante los tribunales por la Fundación Ciudadana CIVIO, donde se juzga el acceso al código fuente del algoritmo que sirve para tramitar y resolver las solicitudes de bono social (precio bonificado por el consumo de energía eléctrica), sí lo ha apreciado el juzgado central de lo contencioso-administrativo haciendo suyas las consideraciones del Centro Criptológico Nacional[18].

Precisamente este caso ha generado un interesante debate doctrinal del que son máximos exponentes Huergo Lora[19] y Ponce Solé[20] a raíz de la sentencia de la Audiencia Nacional de 30 de abril de 2024, dictada en apelación a la sentencia del Juzgado Central de lo Contencioso-Administrativo núm. 8, referida a pie de página y citada en el párrafo anterior. En mi opinión, el debate interesante es el de la prevalencia o no de los límites o intereses invocados sobre el acceso a la información, y no tanto sobre si el algoritmo decide directamente sobre la concesión o no de la ayuda, o solo actúa como apoyo a la toma de la decisión como el profesor Huergo Lora sugiere. En cualquier caso, personalmente, considero que conocer cómo opera la aplicación forma parte clave del elemento motivacional de la resolución de concesión, al presuponer que quien finalmente la toma, no revisa ni prejuzga el resultado de la operación. Si lo hiciera con el detalle preciso como para ser garantía suficiente de la corrección de la decisión adoptada, de poca ayuda serviría valerse de un algoritmo. La hace suya y la «refrenda», como no podría ser de otra forma para que puedan imputarse los efectos del acto administrativo a un órgano cuyo titular actúa las competencias que están atribuidas a aquel.

Más recientemente, hemos conocido que a través de una solicitud de información pública Lighthouse Reports y El Confidencial han accedido a varios detalles técnicos, pero no a todos, de una solución utilizada por la Seguridad social para deducir la probabilidad de que una persona de baja esté lista o no para volver a su trabajo. El algoritmo determina qué expedientes de baja revisarán primero los médicos inspectores del Instituto Nacional de la Seguridad Social y cuáles dejarán para el final. Se trata de un sofisticado programa

18 Sentencia núm. 143/2021, de 30 de diciembre, del Juzgado Central de lo Contencioso-Administrativo núm. 8.

19 https://almacendederecho.org/por-que-aciertan-las-sentencias-sobre-el-algoritmo-del-bono-social-electrico

20 https://almacendederecho.org/por-que-se-equivocan-las-sentencias-sobre-el-algoritmo-del-bono-social-electrico

informático que define el orden en que cada trabajador deberá visitar a un facultativo para saber si puede seguir cobrando el subsidio. El sistema identifica quién debería estar de vuelta en su empleo, y si no se ha tramitado ya el alta, señala el expediente como un posible caso de fraude. Los periodistas han obtenido algunos de los factores que los algoritmos tienen en cuenta en su cálculo, pero se desconocen otros aspectos clave, como la importancia que tiene cada uno en la obtención de las puntuaciones finales y el tipo de datos con que el sistema ha sido entrenado[21].

El debate que sí me parece interesante es poder justificar o que resulte imposible hacerlo que pueda sustraerse del conocimiento ciudadano cómo actúa un algoritmo con un peso determinante en una decisión final, en este caso, con un impacto económico, social y distributivo no poco importantes. Más aún cuando hay sospechas de que pueda no actuar en todos los casos como debería, como Ponce Solé pone de manifiesto en su post echando mano de ejemplos llamativos[22].

V. El futuro: el reglamento de inteligencia artificial

Aunque ya es presente, tenemos que dedicar algunas líneas al RIA, aunque en noviembre de 2023 en el seminario celebrado en el IAAP no podíamos más que citarlo dado que estaba a expensas de los trámites finales de aprobación. Desde entonces han tenido lugar también otros hitos con relevancia en esta materia, tal como decíamos al inicio de este trabajo. El Consejo de Europa, por citar uno de ellos, ha aprobado el pasado 17 de mayo el Convenio Marco sobre Inteligencia Artificial y derechos humanos, democracia y Estado de Derecho, primer tratado internacional de la organización sobre IA con participación de 11 países no miembros y representantes del sector privado y sociedad civil, al que podrán adherirse los terceros estados que lo deseen. El Convenio introduce un marco legal que abarca todo el ciclo de vida de los sistemas de IA y aborda cuestiones que puedan poner en riesgo los derechos humanos, la democracia y el Estado de Derecho. Proporciona criterios de transparencia y control para evaluar la necesidad de establecer prohibiciones o pautas en uso de determinadas herramientas, exige adoptar un mecanismo de control independiente para asegurar su implementación efectiva, que se informe a los usuarios sobre el uso de aplicaciones de IA y del deber de garantizar la rendición de cuentas y la responsabilidad por posibles impactos negativos. Es cierto que todo ello se recoge en términos muy concisos y genéricos, seguramente con el ánimo de conseguir mayor adhesión.

21 https://pablojimenezarandia.com/la-seguridad-social-usa-una-ia-secreta-para-rastrear-bajas-laborales-y-cazar-fraudes/

22 Recomiendo también la lectura de Cotino Hueso, L. «Caso Bosco, a la tercera tampoco va la vencida. Mal camino en el acceso los algoritmos públicos», en *Diario LA LEY*, núm. 84, Sección Ciberderecho de 17 de mayo de 2024.

En concreto sobre transparencia, su art. 8 dispone que cada parte adoptará o mantendrá medidas para garantizar que existan requisitos adecuados de transparencia y supervisión adaptados a los contextos y riesgos específicos en relación con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, también en lo que respecta a la identificación de contenidos generados por los sistemas de inteligencia artificial.

Ya en nuestro país, la AESIA va tomando cuerpo con el nombramiento de su director, Ignasi Belda, y la determinación de su sede en A Coruña, pero aún queda por ponerla efectivamente en marcha. Sus funciones principales serán:

- Actuar como think tank, anticipando tendencias, divulgando conocimiento y dinamizando el debate social sobre las posibilidades y límites de la IA.
- Desarrollar las funciones que le asigna el RIA, supervisando los sistemas de IA de alto riesgo, coordinando la supervisión con las autoridades de vigilancia del mercado, promoviendo estándares y buenas prácticas y evaluando los modelos de IA.
- Coordinar la aplicación del Reglamento con la Oficina Europea de IA y el resto de autoridades de supervisión nacionales de la Unión Europea para garantizar su aplicación uniforme y su adaptación a los cambios tecnológicos y de mercado.
- Promover la innovación en IA tanto en el sector público como en el privado, facilitando el cumplimiento de la normativa y desarrollando mejores prácticas; en este sentido, uno de los cometidos principales de la institución será el desarrollo de entornos de prueba *(sandboxes)* sectoriales.
- Participar en el debate global sobre IA, interviniendo en los diferentes foros multinacionales en los que se establezcan estándares planetarios mínimos para evitar la fragmentación regulatoria, y abogando por una IA confiable, responsable y humanista, en línea con la visión sobre esta tecnología plasmada en la ENIA.

En junio de 2022 nuestro país presentó en Bruselas junto con autoridades de la Comisión Europea un proyecto piloto para la creación de un *sandbox* regulatorio para el despliegue del RIA. Es, finalmente, el Real Decreto 817/2023, de 8 de noviembre, el que regula este entorno controlado de pruebas o experiencias en materia de IA. Mediante una convocatoria se elegirá un conjunto de empresas participantes cuyos sistemas de IA tendrán que cumplir los requisitos del RIA, contando para ello con soporte y asistencia personalizados de la Administración. A los participantes se les dará información, formación y asesoramiento sobre cómo aplicar el reglamento.

La Administración recogerá de estas empresas información muy valiosa sobre sus dudas, problemas y soluciones que servirá para perfeccionar las

guías prácticas elaboradas como punto de partida, guías que, finalmente, también se publicarán como recurso de máxima utilidad para todos. El informe final, que será remitido a la Comisión Europea, recopilará todos los aprendizajes, sintetizando y documentando lo aprendido, las fortalezas y debilidades, riesgos y oportunidades. Este informe se hará llegar a la Comisión Europea y a otras entidades interesadas para transmitir este aprendizaje normativo.

El 3 de octubre de 2023, con el objeto de conocer y motivar el interés de participación en el *sandbox* de IA antes de la publicación de la convocatoria, el entonces Ministerio de Asuntos Económicos y Transformación Digital abrió el pre-registro para entidades y expertos interesados para participar, por un lado, desde la perspectiva de usuario como de proveedor de los sistemas de IA, y, por otro lado, para formar parte del Consejo Asesor de Expertos en IA que brindará apoyo al instrumento.

Y finalmente, el RIA. Los considerandos de la norma expresan que, con el fin de establecer un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo y contenido de las normas a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA de que se trate. Por ello, es necesario prohibir determinadas prácticas que no son aceptables, definir los requisitos que deben cumplir los sistemas de alto riesgo y las obligaciones aplicables a los operadores, así como imponer obligaciones de transparencia a determinados sistemas de IA. Por «transparencia», el RIA entiende que los sistemas deben desarrollarse y utilizarse de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema y a las personas afectadas acerca de sus derechos.

El art. 13 del Reglamento ordena que los sistemas de IA de alto riesgo se diseñen y desarrollen de un modo que se garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente sus resultados de salida. Debe garantizarse un tipo y nivel de transparencia adecuados para que el proveedor y el responsable del despliegue cumplan las obligaciones previstas. Estos sistemas de alto riesgo irán acompañados de las instrucciones de uso correspondientes, que incluirán información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue sobre determinados datos como la identidad y los datos de contacto del proveedor, las características, capacidades y limitaciones del funcionamiento del sistema de IA de alto riesgo, los cambios en el sistema de IA de alto riesgo y su funcionamiento predeterminados por el proveedor en el momento de efectuar la evaluación de la conformidad inicial, las medidas de supervisión humana y los recursos informáticos y de hardware necesarios, la vida útil prevista del

sistema y las medidas de mantenimiento y cuidado necesarias para garantizar su correcto funcionamiento, entre otros contenidos.

El art. 50 desgrana las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA. Estos aspectos y la variedad de sistemas son tal en número que no pueden abordarse en este trabajo que tiene como vocación dibujar mínimamente la evolución de la transparencia de estos sistemas en los últimos tiempos.

Al margen de estas iniciativas, debe trabajarse en auditorías independientes que monitoricen periódicamente los sistemas para garantizar la inexistencia de sesgos, auditorías que validen los modelos y documenten los métodos y resultados. También la creación de certificaciones y sistemas de autorregulación que fijen códigos éticos o de conducta, similares a los sistemas de *compliance*, pero en este ámbito. En octubre de 2023 saltaba la noticia de que Adigital (Asociación Española de la Economía Digital), en colaboración con la Fundación Eticas y con el apoyo de Blablacar, Cabify, Fintonic y Holaluz, había comenzado el diseño de una certificación de transparencia algorítmica para empresas. El objetivo es diseñar mecanismos que garanticen que las empresas certificadas se comprometen a aportar información sobre los sistemas de IA que utilizan, para que su uso sea más transparente y responsable.

En el sustrato de todos estos mecanismos, hay otras acciones previas, de mayor calado y que exigen un compromiso de todas las áreas. Una de ellas es la educación digital de toda la población, pero especialmente de las personas y colectivos en situación de riesgo de exclusión, menores y mayores, pero también la mejora del marco normativo de la reutilización de la información pública, los datos abiertos. Como ya sabemos, llegamos nuevamente tarde a la transposición de la última directiva sobre esta materia que, finalmente, ha tenido que realizarse por vía urgente (Real Decreto-ley 24/2021, de 2 de noviembre). La creación de repositorios de datos compartidos y la mejora del acceso a ellos permitirá que la tarea de entrenamiento de los modelos de IA y sus resultados sean cada vez mejores.

Bibliografía

Arellano Toledo, W. «El derecho a la transparencia algorítmica en big data e inteligencia artificial», en *Revista General de Derecho Administrativo (Iustel)*, núm. 50, enero 2019.

Boix Palop, A. «Transparencia en la utilización de inteligencia artificial por parte de la Administración», en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, septiembre 2022.

Capdeferro Villagrasa, O. «La inteligencia artificial del sector público: desarrollo y regulación de la actuación administrativa inteligente en la cuarta revolución industrial», *en Revista de Internet, Derecho y Política* núm. 30, marzo 2020.

Consejo para la Transparencia, *Transparencia algorítmica buenas prácticas y estándares de transparencia en el proceso de toma de decisiones automatizadas*, octubre 2020.

Cotino Hueso, L. «Sistemas de inteligencia artificial con reconocimiento facial y datos biométricos. Mejor regular bien que prohibir mal», en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, septiembre 2022.

Fernández Aller, C. «La Carta de Derechos Digitales de España. desafíos pendientes», en *Revista de Privacidad y Derecho Digital*, núm. 25, enero-marzo 2022.

Fuertes López, M. «Reflexiones ante la acelerada automatización de actuaciones administrativas», en *Revista Jurídica de Asturias*, núm. 45/2022.

González López, E. «Los derechos digitales fundamentales: ¿es necesaria su reconfiguración en el ordenamiento jurídico?», en *Revista de Derecho Administrativo*, núm. 20, 2021.

Gutiérrez David, M. E. «Administraciones inteligentes y acceso al código fuente y los algoritmos públicos. Conjurando riesgos de cajas negras decisionales», en *Revista Derecom*, núm. 30, marzo-septiembre 2021.

Huergo Lora, A. «El uso de algoritmos y su impacto en los datos personales», en *Revista de Derecho Administrativo*, núm. 20, 2021.

Huergo Lora, A. «Gobernar con algoritmos, gobernar los algoritmos», en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, septiembre 2022.

Mendilibar Navarro, P. «La aplicación de sistemas algorítmicos en el sector público: la actuación administrativa automatizada y las predicciones algorítmicas», en *Derecho Digital e Innovación*, núm. 13, julio-septiembre 2022.

Moreno Brenes, P. «La dictadura del algoritmo: el derecho como respuesta», en *Diario La Ley*, núm. 10166, 9 de noviembre de 2022.

Ponce Solé, J. «Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», en *Revista General de Derecho Administrativo (Iustel)*, núm. 50, enero 2019.

Ponce Solé, J. «Reserva de humanidad y supervisión humana de la Inteligencia artificial», en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, septiembre 2022.

Presno Linera, M. A. «Fundamentales e inteligencia artificial: una aproximación», en *Revista Jurídica de Asturias*, núm. 45/2022.

Presno Linera, M. A. «Derechos fundamentales e inteligencia artificial en el estado social, democrático y digital de Derecho», en *El Cronista del Estado Social y Democrático de Derecho*, núm. 100, septiembre 2022.

Soto Bernabeu, L. «La importancia de la transparencia algorítmica en el uso de la inteligencia artificial por la administración tributaria», en *Crónica Tributaria*, núm. 179/2021.

Vestri, G. «La inteligencia artificial ante el desafío de la transparencia algorítmica. Una aproximación desde la perspectiva jurídico-administrativa», en *Revista Aragonesa de Administración Pública* núm. 56, 2021.

Villanueva Turnes, A. «La inteligencia artificial: Regulación y límites en el sector público», en *Revista española de Derecho Administrativo*, núm. 221, julio-septiembre 2022.

11.

REFLEXIONES SOBRE LOS NEURODERECHOS[1]

Elisa Moreu Carbonell
Catedrática de Derecho Administrativo
Universidad de Zaragoza

SUMARIO: I. ¿Qué son las neurotecnologías? II. ¿Qué relación existe entre la inteligencia artificial y las neurotecnologías? III. ¿Qué retos plantea el avance de estas neurotecnologías? IV. ¿Las neurotecnologías pueden manipular los cerebros y controlar los pensamientos? V. ¿Qué son los neuroderechos? VI. ¿Quién debería regular estos neuroderechos? VII. ¿Qué país(es) reconocen los neuroderechos? VIII. ¿Ha reconocido España los neuroderechos? IX. En conclusión, ¿protege el derecho nuestros cerebros?

Este trabajo, de carácter divulgativo e introductorio, aborda la forma en que el Derecho se acerca a las neurotecnologías. O mejor aún: la forma en que los avances de las neurotecnologías ponen en jaque al Derecho y cómo el Derecho *debería afrontar* su regulación.

Me ha parecido adecuado articular mi estudio en forma de sencillas preguntas y respuestas que abren el debate sobre los retos jurídicos de las neurotecnologías.

1 Este trabajo se ha realizado en el marco del grupo de investigación consolidado ADESTER, *Derecho Administrativo de la Economía, la Sociedad y el Territorio* y del proyecto IN-SALVAN, *Innovación para una salud de vanguardia*, PID2021-127828NB-I00. Parte de mi estudio «La regulación de los neuroderechos», *Revista General de Legislación y Jurisprudencia* 01/2022. Agradezco a Gonzalo Castro Marquina su apoyo y el interés que ha manifestado siempre en la difusión de este debate sobre las implicaciones jurídicas de las neurotecnologías.

I. ¿Qué son las neurotecnologías?

Las neurotecnologías son un conjunto de herramientas, técnicas y dispositivos diseñados para interactuar con el sistema nervioso de los seres humanos, ya sea para estudiarlo, monitorearlo, influir en su funcionamiento o incluso potenciarlo mediante técnicas de estimulación cerebral, tanto invasivas como no invasivas. La neurotecnología es un concepto interdisciplinar en el que confluyen la inteligencia artificial, la informática y las neurociencias.

Sus aplicaciones van más allá de la medicina asistencial y se extienden a distintas áreas como la neurociencia, la educación y el aprendizaje (dispositivos para estimular la memoria), la pediatría (neurodesarrollo infantil), la psicología y, por supuesto, el entretenimiento (realidad aumentada y virtual)[2]. La denominada mejora cognitiva o *neurocognición* estimula y potencia las capacidades intelectuales de las personas (Moreu Carbonell, 2023).

El avance de las neurotecnologías puede marcar un antes y un después en el estudio de la actividad cerebral y en el diseño de modelos para entender

2 Las neurotecnologías mejoran la capacidad de diagnóstico clínico y permiten personalizar los tratamientos. Desde hace décadas, la estimulación cerebral mediante electrodos se conoce para estudiar y tratar ciertas enfermedades. A nadie sorprende ya el uso de implantes cocleares en personas sordas, basados en impulsos eléctricos que llegan al cerebro, o la resonancia magnética para captar imágenes cerebrales. Existen infinidad de técnicas que permiten la intervención directa sobre el cerebro: estimulación profunda (marcapasos cerebral), terapia con radioisótopos, estimulación magnética transcraneal, terapia electroconvulsiva y otros, que se utilizan en el ámbito clínico para tratar afecciones (Vincent, 2014 y Yuste, 2019).

Las empresas tecnológicas (entre otras: https://www.starlab.es/team, https://www.bitbrain.com/es, https://neuralink.com y https://visionintegral.org) desarrollan interfaces cerebro-máquina para conectar a los humanos directamente con la tecnología. Desde biosensores para monitorizar variables fisiológicas y de movimiento, hasta dispositivos capaces de analizar nuestra respuesta visual a ciertos estímulos. Las interfaces cerebro-computadora (BCI, por sus siglas en inglés) permiten la comunicación directa entre el cerebro y dispositivos externos, como computadoras o prótesis, mediante pensamientos o señales neuronales. Los «sistemas ciberfísicos», SCF, aplicados al cuerpo humano, son mecanismos controlados por algoritmos e IA, de modo que los componentes físicos y de software están entrelazados.

En medicina se utilizan para el tratamiento de trastornos neurológicos como el Parkinson, la epilepsia o la depresión. En 2014, un experimento llevado a cabo por científicos de la Universidad de Pittsburg permitió a un joven tetrapléjico palpar o presionar un objeto, introduciéndole un implante cerebral conectado a un brazo robótico. Mediante interfaces cerebro-máquina (microtransmisores) es posible observar y dirigir los procesos cerebrales para tratar a pacientes que han sufrido ictus o enfermedades neurológicas como alzheimer, parkinson, trastornos mentales y adictivos. Se ensayan dispositivos para la estimulación cerebral profunda para el tratamiento de múltiples enfermedades e interfaces cerebro-ordenador que registran la actividad del cerebro y la trasladan para controlar prótesis o sistemas robóticos periféricos. Estos y otros ejemplos a nivel divulgativo «Los retos de la neurotecnología en tiempos de inteligencia artificial», *National Geographic*, 2020 https://www.nationalgeographic.com.es/ciencia/retos-neurotecnologia-tiempos-inteligencia-artificial_15289 y «Los secretos del cerebro», *Revista Uciencia* (2012).

los procesos cognitivos. En combinación con la inteligencia artificial (IA) las posibilidades se multiplican.

II. ¿Qué relación existe entre la inteligencia artificial y las neurotecnologías?

Los sistemas de IA son tecnologías de procesamiento de la información que funcionan con autonomía y operan con técnicas de aprendizaje automático mediante la integración de modelos y algoritmos capaces de realizar tareas cognitivas (Tahirí Moreno, 2024).

La IA y las neurotecnologías están estrechamente conectadas; ambas van a revolucionar la comprensión de nuestro cerebro y las posibilidades para interactuar con él. El científico y divulgador Ray Kurzweil (2019) se atrevió a fijar una fecha en la que seremos capaces de incorporar la IA a nuestros cuerpos, cuando la inteligencia de las máquinas convergerá con la de los humanos: el año 2029.

La IA aplicada a las neurotecnologías permite analizar y extraer patrones de datos sobre la actividad neuronal; facilita el desarrollo de interfaces cerebro-ordenador; optimiza la estimulación cerebral profunda y la estimulación transcraneal ajustando parámetros de nuestra actividad cerebral en tiempo real; permite hacer diagnósticos más precisos y tratamientos médicos personalizados; puede inspirar modelos computacionales del cerebro que ayudan a simular procesos cognitivos; facilita el desarrollo de terapias de rehabilitación cerebral para recuperar funciones tras un accidente cerebrovascular o lesión cerebral; puede interpretar señales cerebrales y traducir pensamientos en comandos para dispositivos externos.

III. ¿Qué retos plantea el avance de estas neurotecnologías?

Los avances de la neurotecnología prometen bienestar para la humanidad, pero, sobre todo, los juristas debemos valorar sus riesgos jurídicos, éticos y morales (Monasterio Astobiza y otros, 2019, 43).

En efecto, el impacto transversal de las neurotecnologías no deja indemne al Derecho: desde el Derecho procesal (Taruffo, Nieva, 2013)[3], al Derecho

3 Por ejemplo, la «prueba neurocientífica» impacta decisivamente en el Derecho procesal y penal. El conocido como «test de la verdad p-300» es un método potencial de evocación cognitiva que se basa en análisis neurofisiológicos de las respuestas sensoriales captadas mediante electrodos. Se trata de una prueba que se practica sobre el sujeto mediante la colocación de terminales que detectan las ondas cerebrales y permite obtener los potenciales evocados, como respuestas eléctricas del sistema nervioso central ante la aplicación de un estímulo, que puede ser auditivo o visual. Un auto de la Sala de lo Civil y Penal del TSJ de Aragón, de fecha 20/07/2015, declaró nula la prueba del P-300 a la que se sometió un imputado por homicidio porque se realizó contra su voluntad. Fue la primera vez

administrativo (Amoedo-Souto, 2018), pasando por el Derecho civil (Donnelly-Lazarov, 2018), el Derecho laboral (Iglesias Cáceres, 2021) y, por supuesto, el Derecho penal (Crespo, 2013)[4].

¿Habría que poner límites severos al desarrollo de las neurotecnologías? La libertad de investigación en este ámbito debe quedar limitada por la regulación, como lo está la investigación general en materia bioética (Rivas García, 2016), pero tampoco podemos asfixiarla. Buscar ese equilibrio es muy complicado.

Ocurre que el ritmo regulador es más lento que el ritmo al que avanzan las neurotecnologías. El marco jurídico se ve desbordado por la tecnología y no sabe cómo afrontar los nuevos retos. Ante situaciones desconocidas, el Derecho suele refugiarse en el puro formalismo (Moreu Carbonell, 2022).

IV. ¿Las neurotecnologías pueden manipular los cerebros y controlar los pensamientos?

Por el momento, esto no es posible. Psicólogos, magos, políticos y publicitarios saben que la mente humana es fácilmente manipulable y aplican sus conocimientos para orientar nuestra voluntad (Macknik, Martínez-Conde, 2010). También se conoce que, sin necesidad de usar técnicas invasivas, el cerebro humano está sufriendo cambios importantes por el simple uso de las tecnologías (Bueno i Torrens, 2016).

En realidad, no es necesario esperar a que las neurotecnologías invasivas se instalen en nuestros cerebros para replantear la personalidad y el libre albedrío en el Derecho. La psicología conductual nos demuestra que el ser humano no es racional en la toma de decisiones, que somos seres complejos, de una racionalidad limitada, sumamente emocionales, influidos por las interacciones sociales, pero también por el contexto y el entorno en el que

que un tribunal se pronunciaba sobre una prueba que afecta a la neurotecnología. El TSJ concluyó que no se puede equiparar a otras pruebas periciales de intervención corporal, sino que equivale a una declaración o confesión del acusado y por eso solo se puede realizar con la voluntad del sujeto. Más ampliamente, Sánchez Rubio (2016). También las llamadas técnicas de «neuroimagen» tiene aplicación al proceso (Desmoulin-Canselier, 2019; Martínez Sánchez, 2013). La sentencia de la sala de lo contencioso-administrativo de la Audiencia Nacional, SAN 2350/2020, de 30/09/2020 (recurso n.º 2187/2019), nos ilustra del impacto que puede alcanzar la utilización de sistemas de IA para indagar en la conducta humana (aquí el llamado sistema Viogén que utiliza la Guardia Civil para analizar el riesgo de violencia machista) en un caso de responsabilidad patrimonial de la Administración.

4 Ariano (2016, 103) da cuenta de dos sentencias dictadas por la Corte D'Assise di Appelo de Trieste (Italia) que se refieren al impacto de las técnicas de *neuro-imaging* como método de prueba en el proceso penal (sentencias n.º 5, de 18/09/2009, y n.º 40, de 20/07/2011).

tomamos las decisiones, llenos de sesgos[5]. Un jurista sensible a las evidencias debe tomar buena nota de las limitaciones de nuestro cerebro.

Las investigaciones del cerebro pueden ya medir, registrar, alterar y manipular la actividad cerebral; es lo que se conoce como *neuromodulación* o alteración de la actividad cerebral por medio de la introducción de estímulos (Ausín, Morte, Monasterio, 2020).

Algunos científicos alertan de que los avances de las neurotecnologías, junto con el desarrollo de la IA y de la computación, permitirán un conocimiento milimétrico del cerebro humano y, con ello, un eventual control del mismo que podría escapar a la propia conciencia humana (Rafael Yuste, 2019, 23). En este escenario distópico podrían producirse ataques a nuestros cerebros por parte de *hackers* que quisieran controlar nuestras decisiones y pensamientos, incluso sin nuestro consentimiento y sin ser conscientes de ello.

Los mismos científicos que estudian cómo controlar nuestro cerebro, —capitaneados por Rafael Yuste de la *NeuroRights Foundation*, quizás la voz más conocida— son los impulsores del reconocimiento de unos nuevos derechos humanos que denominan «neuroderechos». Por eso se ha abierto un jugoso debate sobre la necesidad de ampliar el marco legal existente para hacer frente al uso y abuso de las neurotecnologías. Una de esas propuestas es la relativa a los neuroderechos, que propone la creación de nuevos derechos humanos que protejan nuestra intimidad cerebral y nuestros pensamientos frente a ataques externos.

De momento los seres humanos no sentimos que nuestro cerebro ni nuestra libre voluntad se encuentren realmente amenazados; lo vemos casi como algo de ciencia ficción. Pero, con independencia del ritmo al que avance la neurotecnología, no nos puede pillar por sorpresa y el Derecho tiene que estar preparado para afrontar estos retos. Como ocurre en otros ámbitos sectoriales, como la prevención de los riesgos ambientales, la función del Derecho no puede ser meramente reactiva.

---

5 La Academia Nacional de Ciencia de EEUU publicó los resultados del experimento de tres científicos de la Universidad Ben-Gurión del Néguev titulado «Factores externos a las decisiones judiciales». Los científicos analizaron durante meses las decisiones de una comisión judicial encargada de conocer las peticiones de libertad condicional de los/as presos/as. El estudio llegaba a una simple conclusión: la justicia se ve influenciada por lo que haya desayunado el Juez. Los resultados revelaban que, al comienzo del día, las probabilidades de que el/la acusado/a obtuviera una sentencia favorable podían llegar hasta 65 %, descendiendo paulatinamente hasta el 0 % a lo largo de la mañana. Tras la pausa del almuerzo, las probabilidades de conseguir la libertad remontaban de nuevo hasta el 65 %.
En definitiva, jueces fatigados por el trabajo y con hambre tienden a tomar la decisión más fácil para el cerebro humano y que consiste, en éste y en otros casos, en denegar peticiones de libertad condicional (Kahneman, 2011). Otros estudios demuestran el efecto de los factores externos en las oposiciones o exámenes. Una visión general sobre la manipulación de nuestros sesgos cognitivos en Ponce Solé (2021).

V. ¿Qué son los neuroderechos?

La autoría del término *neuroderecho* se atribuye a Tovino (2008), que alertó sobre el alcance que tendría los avances de los estudios empíricos sobre el cerebro, en la medida en que el cerebro se considera fundamental para explicar la conducta humana[6].

Como digo, el neologismo *neuroderechos* designa a los nuevos derechos humanos que protegen la privacidad e integridad de las personas frente al uso abusivo de las neurotecnologías. Podríamos definir los neuroderechos como los derechos que protegen a las personas frente al conocimiento del cerebro. Los neuroderechos son principios de libertad éticos, legales y sociales relativos a la propiedad del cerebro y del pensamiento de las personas; es decir, los derechos que protegen el pensamiento humano y su cerebro.

La ley parte de la premisa de que las personas tenemos la capacidad para elegir libremente (autonomía de la voluntad o libre albedrío). Pero ¿qué pasará cuando los seres humanos estemos expuestos a que otros conozcan nuestras emociones o a que introduzcan pensamientos en nuestro cerebro? Por eso hay que prestar atención a los neuroderechos. La resolución del Parlamento Europeo, de 16/02/2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL), ya señala los riesgos de la manipulación o la desconexión de los dispositivos integrados en el cuerpo humano o incluso el borrado de su memoria, que podría poner en peligro la salud humana.

La creación de los neuroderechos se basa en dos pilares: la privacidad mental (consciente o inconsciente) y la identidad personal. Ienca y Andorno (2017) diseñan la creación de cuatro neuroderechos, que nos protejan frente a las amenazas de las neurotecnologías:

1. *Derecho a la libertad cognitiva o «autodeterminación mental»*, una actualización del Derecho a la libertad de pensamiento y de conciencia.
2. *Derecho a la privacidad mental*, que protege frente a cualquier información que pueda obtenerse de nuestros cerebros por medio de neurotecnologías y distribuirse por medios digitales, para evitar filtraciones de estos datos.
3. *Derecho a la integridad mental*, en la idea de que las intrusiones en el cerebro no sólo afectan a la privacidad, sino que también pueden suponer un cambio en el estatuto neuronal de cada persona.
4. *Derecho a la continuidad psicológica*, que garantiza la percepción de la propia identidad como seres humanos.

6 Al respecto, García Deltell (2013, 920), Petoft y Abbasi (2020).

Por su parte, Yuste (2019) propone diseñar un marco regulatorio que reconozca cinco nuevos derechos humanos básicos:

1. *Derecho a la privacidad mental*, que garantiza que nadie pueda utilizar, sin consentimiento del sujeto, los datos relativos a la actividad de las neuronas que son el soporte de su pensamiento.
2. *Derecho a la identidad y autonomía personal*, para garantizar que las personas no pierdan su identidad como seres humanos cuando las neurotecnologías permitan el control sobre nuestro cerebro.
3. *Derecho a la libertad cognitiva* (libre albedrío o autodeterminación), para garantizar que las personas tomen libremente sus decisiones. Si nuestro cerebro está conectado a través de lectores de actividad cerebral a una computadora, el riesgo es que no sea libre para tomar decisiones o que un tercero podrá invadir nuestro cerebro, igual que ahora un hacker puede usurpar nuestro ordenador[7].
4. *Derecho al acceso equitativo al aumento de la neurocognición*, que permite al ser humano mejorar su actividad cerebral. Se trata de que este aumento cognitivo sea accesible a todos equitativamente y que no quede reservado a un sector de la sociedad (el que tenga más recursos).
5. *Derecho a la protección contra los sesgos de la inteligencia artificial,* para que los conocimientos de la neurotecnología no discriminen a las personas por raza, color, sexo, idioma, religión, opinión, origen nacional o social, posición económica, nacimiento o cualquier otra condición.

La propuesta para elaborar una carta de (nuevos) neuroderechos es sugestiva. Igual que los avances informáticos obligan a extremar la seguridad de nuestros datos, la neurotecnología hace que debamos prestar atención a la privacidad de nuestro cerebro. La humanidad tiene que estar en condiciones de protegerse ante los avances de la neurotecnología, para que no se cometan violaciones de derechos o manipulaciones no deseadas del cerebro. Este objetivo se plasma en la expresión, entre expectante y dudosa, del conocido trabajo de Bublitz (2013), «My Mind is Mine!?», que nos pone en alerta.

VI. ¿Quién debería regular estos neuroderechos?

Una vez que se han diseñado los neuroderechos, surge la pregunta de cómo regularlos. Yuste y Goering (2017:162) recomiendan incorporar cláusulas que protejan los neuroderechos en instrumentos internacionales del más alto nivel, como la Declaración Universal de los Derechos Humanos de 1948 o el Conve-

7 Es una posibilidad ya tratada por el Derecho penal; véase «¿Se puede "hackear" el cuerpo humano?», blog *OpenMind*, en https://bbvaopenmind.com.

nio Europeo de Derechos Humanos de 1950. Sin embargo, conscientes de que es complicado el acuerdo entre Estados, abogan por un acuerdo internacional que defina qué acciones se consideran prohibidas en relación con las neurotecnologías y la IA, similar a las prohibiciones enumeradas en la convención internacional para la protección de las personas contra las desapariciones forzosas de 2010, que no recoge derechos, pero sí acciones comunes.

También Muñoz Machado (2000) y Piñar Mañas (2017) coinciden en la necesidad de regular a nivel internacional, para que la tutela de los derechos frente a las tecnologías no quede en meras palabras o buenas intenciones. De la Quadra-Salcedo (2018) plantea, además, la necesidad de reflexionar sobre si la regulación de estos derechos debe hacerse de forma global o en la legislación sectorial.

Estas organizaciones internacionales, la ONU y el Consejo de Europa, ya han formulado declaraciones y recomendaciones en las que se deduce la preocupación por en proteger los derechos humanos frente a las tecnologías.

El alto comisionado de la ONU para los derechos humanos ha elaborado distintos informes sobre la privacidad en la era digital en los que ya se mencionan los neuroderechos. Su informe «Impacto de las nuevas tecnologías en la promoción y protección de los derechos humanos en el contexto de las reuniones, incluidas las protestas pacíficas» (A/HRC/44/24, de 24/06/2020) se centra en las prácticas de vigilancia y reconocimiento facial y «El Derecho a la privacidad en la era digital» (A/HRC/39/29, de 3/08/2018) en el impacto de las redes y de internet en la privacidad de los datos.

Por su parte, el comité de expertos en derechos humanos del Consejo de Europa aprobó en 2018 un borrador de recomendaciones para los Estados sobre el impacto de los algoritmos en los derechos humanos (Adressing the impacts of Algorithms on Human Rights, MSI-AUT(2018)06rev3), que aconseja revisar las regulaciones nacionales y ofrece pautas para llevarlo a cabo.

En octubre de 2024 se aprobó el informe del Comité Asesor del Consejo de Derechos Humanos de la ONU titulado *Efectos, oportunidades y retos de la neurotecnología en relación con la promoción y la protección de todos los derechos humanos.* Este informe concluye que para afrontar los importantes retos de esos avances habrá que desarrollar y adaptar el marco de derechos humanos existente. A nivel conceptual, las normas de derechos humanos deberán reforzarse para garantizar una protección más eficaz. Además, habrá que fortalecer la arquitectura institucional y el acceso a la justicia. En los casos en que las neurotecnologías puedan dar lugar a vulneraciones de los derechos humanos habría que aplicar a nivel mundial limitaciones a su desarrollo y despliegue, incluida su prohibición.

La Unión Europea también ha revelado su preocupación por el impacto que internet y los avances de la IA tienen en derechos como la intimidad, la libertad de expresión o la protección de datos. La comunicación de la Comisión titulada «Inteligencia artificial para Europa» (COM(2018) 237 final) apunta los proble-

mas éticos y jurídicos de las nuevas tecnologías. Afirma, en concreto, que la «UE debe asegurarse de que la IA se desarrolle y aplique en un marco adecuado, que promueva la innovación y respete los valores y derechos fundamentales de la Unión, así como principios éticos tales como la obligación de rendir cuentas y la transparencia». La comunicación asume el objetivo de «garantizar el establecimiento de un marco ético y jurídico apropiado basado en los valores de la Unión y en consonancia con la Carta de los Derechos Fundamentales de la UE».

La Comisión europea promueve una «Declaración de principios digitales» para que los Estados miembros afronten los retos jurídicos y éticos de la sociedad digital, en desarrollo de la conocida como «Brújula digital 2030: el enfoque de Europa para el decenio digital» (COM(2021) 118 final). Esta iniciativa sitúa el respeto a los derechos de los ciudadanos —en concreto la intimidad, la protección de datos, la libertad de expresión, los derechos del niño y de los consumidores—, en el centro de la ciudadanía digital.

El Reglamento 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial, insiste en la conveniencia de garantizar los derechos fundamentales de la UE, pero no alude específicamente a los neuroderechos.

VII. ¿Qué país(es) reconocen los neuroderechos?

Algunos Estados, aisladamente, han movido ficha en este empeño regulatorio. Aparte de Chile, que fue pionero, otros países como Brasil[8], Italia[9], Alemania[10]

8 El congreso brasileño aprobó en 2014 el denominado «Marco Civil de Internet» para garantizar, entre otros, el Derecho a la privacidad de las comunicaciones

9 Italia fue el primer estado europeo en aprobar una declaración de derechos en internet. En 2015, la Cámara de los Diputados italiana aprobó una *Declaración de Derechos de Internet*, que, aunque menciona a los neuroderechos, opta por extender la garantía de los derechos humanos ya reconocidos a este nuevo entorno tecnológico. Y así, el artículo 1 de la Declaración reconoce que: «1. Quedarán garantizados en Internet los derechos fundamentales del individuo reconocidos por la Declaración universal de Derechos Humanos de las Naciones Unidas, por la Carta de los Derechos Fundamentales de la Unión Europea, por las Constituciones nacionales y por las declaraciones internacionales en materia. 2. Tales derechos deberán interpretarse de manera que quede garantizada su efectividad dentro de la dimensión de la Red.23. El reconocimiento de los derechos en Internet deberá fundarse en el pleno respeto de la dignidad, la libertad, la igualdad y la diversidad del individuo, esto es, en los principios sobre cuya base deberán compaginarse de forma equilibrada con los demás derechos».

10 El Gobierno alemán constituyó en 2019 una Comisión Ética de Datos *(Daten Ethik Kommission)* que defiende una nueva regulación europea sobre sistemas algorítmicos y fija unos requisitos generales horizontales que deberían ser desarrollados por normas sectoriales, entre ellos, la garantía de la dignidad humana, la auto-determinación, la privacidad y la seguridad. Esta comisión ha diseñado una propuesta de regulación basada en niveles de riesgo, en función de las variables de probabilidad y severidad del daño causados por algoritmos.

o España[11] han puesto en marcha distintas iniciativas encaminadas a proteger los derechos humanos frente al abuso de las neurotecnologías o, incluso, piden su reconocimiento internacional.

Chile ha sido, hasta el momento, el único país que inició hace cuatro años un procedimiento de reforma constitucional para ampliar su marco de derechos fundamentales. Lo hizo a través de dos vías que se tramitaron paralelamente: una reforma constitucional y un proyecto de ley complementario sobre protección de los neuroderechos y la integridad mental, que no llegó a buen puerto.

El 7 de octubre de 2020 un grupo de senadores chilenos encabezados por Guido Girardi, miembro de la Comisión de Desafíos del Futuro, Ciencias, Tecnología e Innovación, presentó una iniciativa para regular por ley la protección de los neuroderechos y la integridad mental, y el desarrollo de la investigación y las neurotecnologías. En la propuesta había participado activamente el neurobiólogo español Rafael Yuste, que entusiasmó al rector de la Pontificia Universidad Católica y al senador Girardi[12].

El proyecto tiene una amplísima parte expositiva y una parte dispositiva muy reducida, estructurada en tres títulos y diez artículos. Lo primero que llama la atención es que el proyecto no recoge un listado de neuroderechos, ni tampoco se califican como derechos subjetivos. Tratándose de nuevos derechos, lo deseable hubiera sido nominarlos y definirlos, regulando también los mecanismos para hacerlos efectivos. Pero el proyecto se limita a describir una serie de prohibiciones relativas a la investigación y el avance de las neurotecnologías.

Esta deficiente técnica legislativa ha sido criticada por generar inseguridad jurídica.

Como digo, ha habido ya voces discrepantes con la deficiente técnica normativa empleada y con la inconsistencia entre el mensaje de la parte expositiva del proyecto y su articulado. La técnica jurídica que el proyecto acoge para brindar la prometida protección a la integridad y a la privacidad mentales es la inclusión de preceptos prohibitivos y unos pocos imperativos, pero ningún artículo que *defina un derecho subjetivo*. ¿Cómo es posible que un proyecto que buscaba defender los neuroderechos no contemple derechos subjetivos? Da la impresión de que se está legislando sin entender bien lo que se quería proteger, o que, simplemente se trata de anunciar al mundo que Chile está aprobando una legislación pionera en este ámbito.

---

11 El Gobierno de España aprobó en 2021 una Carta de Derechos Digitales que incluye los «derechos digitales en el empleo de las neurotecnologías».

12 Su denominación es «Proyecto de ley, iniciado en moción de los Honorables Senadores señor Girardi, señora Goic y señores Chahuán, Coloma y De Urresti, sobre protección de los neuroderechos y la integridad mental y el desarrollo de la investigación y las neurotecnologías», Boletín n.º 13.828-19, https://www.diarioconstitucional.cl.

En realidad, el proyecto no incluye un reconocimiento de neuroderechos, sino «medidas para proteger la integridad y privacidad mental» bajo la forma de límites o prohibiciones a la investigación en neurociencia y neurotecnología. Muchos preceptos son ambiguos, como la distinción entre la «integridad mental» y la «integridad psíquica» o el concepto mismo de neurotecnologías.

La tramitación del proyecto de ley sobre neuroderechos fue precedida de una moción para reformar la Constitución Política de la República de Chile de 1980, impulsada también por los senadores Girardi y Goic, de fecha 7 de octubre de 2020. La moción pretendía introducir un nuevo párrafo en el número 1 del artículo 19 de la de la Constitución Política de la República de Chile, que regula el derecho a la vida y a la integridad física y psíquica de la persona, para proteger la integridad y la indemnidad mental ante los avances de las neurotecnologías.

Tras la oportuna tramitación parlamentaria[13] se aprobó la Ley n.º 21383, publicada en el diario oficial del 25 de octubre de 2021, que modifica el número 1 del artículo 19 y añade un nuevo apartado relativo a las limitaciones del desarrollo científico y tecnológico:

> «El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella».

La modificación operada por la Ley n.º 21383 supone un reconocimiento constitucional expreso de que los avances de la ciencia y la tecnología están al servicio de las personas y que deben respetar el derecho a la vida y a la integridad física y psíquica. La mención —seguramente superflua— se completa con una remisión a la ley ordinaria para que regule los requisitos, condiciones y restricciones del uso de la neurotecnología cuando afecte a las personas y pueda poner en peligro «la actividad cerebral». En cualquier caso, la Ley n.º 21.383 evita cualquier mención a los neuroderechos.

VIII. ¿Ha reconocido España los neuroderechos?

No, nuestro país no ha aprobado ninguna ley ni emprendido una reforma constitucional que reconozca los neuroderechos.

En julio de 2021, el Gobierno de España presentó la llamada Carta *de Derechos Digitales con el objetivo de proteger los derechos de los ciudadanos en la nueva era de internet y la inteligencia artificial* (en adelante, simplemente,

13 Es muy interesante conocer el iter parlamentario, documento «Historia de la Ley N.º 21.383», Biblioteca del Congreso Nacional de Chile, <www.bcn.cl/historiadelaley.

«la Carta»). Se trata de propuestas que adoptan la forma de recomendaciones o principios orientadores, esto es, que eligen la fórmula del *soft law* o Derecho blando[14].

La Carta recoge seis categorías principales de derechos, y entre ellas, un elenco de «derechos digitales en entornos específicos» (apartado 5) que incluye un conjunto de «derechos digitales en el empleo de las neurotecnologías» (apartado 5.XXVI). Reproduzco aquí el texto de la Carta:

> 1. Las condiciones, límites y garantías de implantación y empleo en las personas de las neurotecnologías podrán ser reguladas por la ley con la finalidad de:
>
> a) Garantizar el control de cada persona sobre su propia identidad.
>
> b) Garantizar la autodeterminación individual, soberanía y libertad en la toma de decisiones.
>
> c) Asegurar la confidencialidad y seguridad de los datos obtenidos o relativos a sus procesos cerebrales y el pleno dominio y disposición sobre los mismos.
>
> d) Regular el uso de interfaces persona-máquina susceptibles de afectar a la integridad física o psíquica.
>
> e) Asegurar que las decisiones y procesos basados en neurotecnologías no sean condicionadas por el suministro de datos, programas o informaciones incompletos, no deseados, desconocidos o sesgados.
>
> 2. Para garantizar la dignidad de la persona, la igualdad y la no discriminación, y de acuerdo en su caso con los tratados y convenios internacionales, la ley podrá regular aquellos supuestos y condiciones de empleo de las neurotecnologías que, más allá de su aplicación terapéutica, pretendan el aumento cognitivo o la estimulación o potenciación de las capacidades de las personas.

Los llamados *derechos digitales*, correlato de la idea de la *ciudadanía digital*, comprenden genéricamente los derechos de los ciudadanos en el entorno digital. Esta categoría incluye no sólo la actualización de derechos tradicionales para identificar nuevos contenidos en el marco de la sociedad digital (por ejemplo, el Derecho al olvido dentro del Derecho fundamental a la protección de datos, admitido por la STC 58/2018, de 4 de junio), sino también nuevos derechos fundamentales, como serían el acceso universal a internet, la ciberseguridad o el Derecho a la verdad para luchar contra las noticias falsas *(fake news)*.

La Carta no tiene carácter normativo, sino que se limita a ser un marco de referencia para los poderes públicos que sirva de guía o molde para futuros proyectos legislativos. En palabras de uno de sus promotores, Tomás de la Quadra-Salcedo, la Carta actúa como un faro para los navegantes de la sociedad digital. Insiste en que no hay una «duplicidad» de derechos porque

14 En la dicotomía entre normas jurídicas y Derecho blando, Mozo Seoane (2021, 107) parece preferir la segunda opción en el campo de las emergentes tecnologías. Sin embargo, también destaca la necesidad de encontrar los límites precisos para ambos espacios normativos.

su fundamento último es el marco constitucional de derechos fundamentales[15]. Sin embargo, Reche Tello (2021) pone en duda que «frente a los riesgos que nos vienen» resulten suficientes esta propuesta de autorregulación o el «derecho blando».

En cualquier caso, esos *derechos digitales en el empleo de las neurotecnologías* que enumera el apartado 5.XXVI de la Carta se identifican fácilmente con el diseño original de los neuroderechos: el Derecho a la privacidad mental (letras a y c), el Derecho a la identidad y autonomía personal (letra b), el Derecho al libre albedrío y la autodeterminación (letra c), el Derecho de acceso equitativo a la aumentación cognitiva (punto 2) y el Derecho a la protección de los sesgos de los algoritmos (letra e).

La preocupación por los neuroderechos ha llevado incluso a que algunas comunidades autónomas, como la valenciana, a través del Consejo Valenciano de Cultura y bajo el auspicio de Rafael Yuste, impulsara en 2023 una denominada «Declaración de Valencia» para defender la incorporación de los neuroderechos en la declaración internacional de Derechos Humanos[16]. Interesantes también las reflexiones del documento titulado «Avances en neurociencia: aplicaciones e implicaciones éticas. La disrupción de las neurotecnologías», de la Oficina de Ciencia y Tecnología (oficina C) del Congreso de los Diputados[17].

IX. En conclusión, ¿protege el derecho nuestros cerebros?

De momento sí, y no hacen falta neuroderechos. Yo creo que *no es necesaria una regulación específica de los neuroderechos,* por la simple razón de que los derechos humanos reconocidos en los textos internacionales y constitucionales, como el derecho a la igualdad, a la integridad física y psíquica, a la intimidad, a la protección de datos o a la libertad de expresión, *ya protegen frente a los abusos de las neurotecnologías*[18].

15 En https://digitalfuturesociety.com/es/agenda/carta-derechos-digitales-streaming/ se pueden encontrar intervenciones de los ponentes de los grupos de trabajo encargados de redactar la Carta (recuperado: 20/11/2021), entre otros, las de Tomás de la Quadra-Salcedo y Susana de la Sierra. En concreto, sobre los neuroderechos, no hay que perderse la intervención de Rafael Yuste (minuto 1:47).

16 Disponible en la siguiente dirección: https://cvc.gva.es/wp-content/uploads/Declaracio-de-Valencia-Neurodrets_bilingue_Signar.pdf.

17 En https://oficinac.es/es/informes-c/neurociencia-aplicaciones.

18 La Constitución española de 1978 hace incluso mención expresa a que «la Ley limitará el uso de la informática para garantizar el honor, la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos». Además, los derechos al honor, la intimidad y a la propia imagen están desarrollados en España por la Ley Orgánica 1/1982,

Como mucho, sería necesario adaptar o extender el ámbito material de estos derechos, pero no crear una categoría nueva. Los derechos humanos son permeables a las nuevas necesidades humanas y se van adaptando a ellas.

Por ejemplo, el Derecho fundamental a la inviolabilidad del domicilio y el Derecho a la integridad física y psíquica del individuo se han configurado como escudo protector frente a la contaminación acústica. Así lo ha reconocido la jurisprudencia del Tribunal Europeo de Derechos Humanos (STEDH de 16/11/2004, asunto Moreno Gómez c. España) y el Tribunal Constitucional español (entre otras, en STC 119/2001 o STC 150/2011).

En definitiva, nada hay en los neuroderechos que permita identificarlos como nuevos derechos, a diferencia de lo que ocurrió cuando a los derechos *de primera generación* (civiles y políticos) se sumaron los conocidos como *de segunda o tercera generación* (económicos, sociales y culturales). Los neuroderechos se quieren encajar en una *cuarta generación de derechos humanos,* pero en realidad no tienen la suficiente autonomía y se identifican con derechos que ya están asegurados por los tratados internacionales y textos constitucionales.

Para evitar que nuevas amenazas limiten viejos derechos humanos, el mejor camino no es crear más derechos. Los derechos humanos están constantemente expuestos a amenazas inéditas. El derecho a la privacidad, por ejemplo, protege frente tanto frente a la prensa escrita como frente a la videovigilancia o las tecnologías de reconocimiento facial. El sistema universal de protección de los derechos humanos, que ha llevado 75 años construir, instalar y fortalecer, no puede verse de repente puesto en duda ante una nueva construcción jurídica que debe valorarse con mayor rigor. Resulta inapropiado crear nuevas normas para un campo con hallazgos tan incipientes[19].

---

de 5 de mayo, posteriormente actualizada en la LO. 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
Sin embargo, Reche Tello (2021: 442) llama la atención sobre el hecho de que los datos cerebrales no pueden considerarse propiamente como un dato personal y que, por tanto, el marco jurídico vigente no sería una herramienta eficaz frente a los riesgos de las neurotecnologías.

19 Se ha llegado a poner en entredicho el éxito de los proyectos neurocientíficos europeos y norteamericanos para mapear el cerebro, muy lejos de las aspiraciones iniciales por razones burocráticas, entre otras. Un apunte crítico en E. Mullin, «El fracaso de los proyectos de la UE y EE. UU. para entender el cerebro», https://www.technologyreview.es, de 14/09/2021.
Por otro lado, los hallazgos sobre nuestro cerebro son todavía limitados y no permiten entender el alcance de las decisiones humanas, de momento. Supongamos que debemos firmar un documento. Al hacerlo, nuestro cerebro activa una serie de disparos neuronales. Podríamos estar firmando un cheque o un autógrafo, legitimando un testamento o autorizando una compraventa o un certificado de defunción. En cada caso, el disparo neuronal es el mismo. Y, sin embargo, el significado de lo que hemos hecho al firmar es completamente diferente. Hoy, la neurociencia cognitiva solo es capaz de identificar de dónde

La sentencia pionera del Tribunal de Distrito de La Haya sobre los algoritmos predictivos, es un buen ejemplo de ello. En efecto, en Holanda se planteó un interesante conflicto en relación con la privacidad de los datos contra los riesgos de la vigilancia gubernamental, del que pueden extraerse conclusiones interesantes para la construcción de los neuroderechos. El gobierno puso en marcha en 2014 un el sistema de rastreo predictivo basado en algoritmos conocido como SyRI, con el objetivo de determinar qué ciudadanos eran supuestamente más proclives a defraudar al Estado. La sentencia del Tribunal de Distrito de La Haya, de 5 de febrero de 2020 (caso NJCM c.s./De Staat der Nederlanden (SyRI case), n.º C/09/550982/HA ZA 18/388) estimó el recurso al concluir que el algoritmo predictivo vulnera el artículo 8 de la Convención Europea de Derechos Humanos, relativo al Derecho a la vida privada.

El asunto llamó la atención del relator especial de la ONU sobre la pobreza y derechos humanos, que remitió un informe al tribunal criticando la desigualdad que la vigilancia estatal genera para los colectivos más desfavorecidos[20], y pone el foco en las dificultades de la IA y de las neurotecnologías para alcanzar un conocimiento del comportamiento del ser humano a través del estudio de su cerebro[21].

Por eso me he mostrado escéptica ante la necesidad imperiosa de crear nuevos derechos; al menos, no ahora, pues la propuesta exige un debate reposado, y, en definitiva, más tiempo. Es necesario asumir una posición crítica y cauta ante el entusiasmo que ha generado la propuesta de crear neuroderechos (Moreu Carbonell, 2022)[22].

Ante el avance de las neurotecnologías y la IA, es verdad que nuestros derechos son cada vez más vulnerables y que eso desembocará en la formación de nuevos conceptos y en la implementación de herramientas jurídicas innovadoras. Pero, desde un enfoque dogmático-sistemático, incluso si aceptamos que la manipulación neuronal va a ser posible en el futuro, el intento por introducir conceptos nacidos de la euforia o de la moda es peligroso por dos razones.

provienen esos disparos neuronales, pero no tiene más idea de su significado que el que tienen las redes sociales cuando hacemos «like» a una imagen

20 Al respecto, Lazcoz y Castillo (2020).

21 De hecho, hay voces críticas de carácter ético-filosófico, para quienes la defensa de los neuroderechos se adscribe a una desacertada teoría reduccionista de la neurociencia. El reduccionismo nace de la antigua confusión cartesiana que defiende la identidad entre el ser humano y su cerebro; por eso defienden la necesidad de construir nuevos derechos para proteger el cerebro. Sin embargo, la identidad del ser humano no se identifica con el cerebro.

22 En esta línea escéptica, también Borbón Rodríguez y otros (2020), López Silva y Madrid (2021), Walker Silva (2021), Hertz (2023), Herrán (2023) y Bublitz (2024).). Desde una posición pragmática, López Silva y Madrid (2021) defienden que la creación de nuevos neuroderechos no es *necesaria*, porque ya están protegidos por los derechos humanos, aunque sí es c*onveniente* para hacer conscientes a los ciudadanos de la relevancia del tema.

En primer lugar, desvía energía y recursos que serían mejor aprovechados si los invirtiéramos en garantizar mejor los derechos humanos hasta ahora reconocidos frente a los avances científicos y tecnológicos. En segundo lugar, conduce a una espiral inflacionaria en la que cada nuevo desarrollo cuestiona la vigencia de la protección que ofrece el sistema universal de los derechos humanos.

Por otro lado, el Derecho no es la panacea contra los abusos. Todo el mundo sabe que los derechos humanos básicos (la vida, la libertad de pensamiento, la educación...) se reconocen en la Declaración Universal de Derechos Humanos de la ONU desde 1948 y en otros textos internacionales, pero no se cumplen en muchos países del mundo. Así que, si ya es difícil que las declaraciones internacionales de derechos humanos sean respetadas con las herramientas jurídicas disponibles, debemos acoger con cautela estas declaraciones formalistas de nuevos derechos sin saber exactamente qué es lo que van a proteger, ni si su reconocimiento implicará mayores y más eficaces garantías.

En el ámbito del Derecho administrativo, podemos extender principios ya explorados para justificar una acción pública en la protección de los derechos fundamentales vinculados con la neurotecnología (intimidad, libertad de pensamiento, protección de datos y otros). La protección que supuestamente aportan los neuroderechos se relaciona con principios ya asentados como la buena regulación, el buen gobierno, la transparencia, y, por supuesto, la protección de datos y de los consumidores. Se ha dicho, con razón, que la vulnerabilidad humana ante las neurotecnologías que recurren a la IA para el tratamiento masivo de datos cerebrales queda reconducida a la problemática general sobre la protección de datos personales (Rollnert Liern, 2024).

Ponce Solé (2019) propone una reserva de humanidad para que determinadas decisiones queden reservadas a los humanos, que podría extenderse a la protección de la intimidad cerebral. Por su parte, el principio de precaución tecnológica de Cotino Hueso (2018) conduciría a rechazar aquellas neurotecnologías cuya implantación conlleve un riesgo de daños irreversibles o que sean opacas al control.

Pero habrá que estar muy atentos y modificar este marco preexistente, por si es necesario cambiar las categorías jurídicas (personalidad, imputación, integridad, libre albedrío). Para alcanzar este objetivo resulta más adecuada una regulación internacional que una nacional, como defiende Piñar Mañas (2020). La Unión Europea debe jugar aquí un papel protagonista, igual que en el ámbito de la protección de datos.

El marco normativo debería completarse con la aprobación de códigos éticos y medidas educativas *(soft law)*, que reflejen las implicaciones sociales, médicas y bioéticas de la neurotecnología elaborados con el acuerdo de todas las partes interesadas a escala europea (autoridades públicas, ciudadanos, empresas, asociaciones, etc.).

Los obstáculos organizativos son otro factor a valorar. Los legisladores, los gobiernos y las administraciones deben estar en condiciones de aden-

trarse en otras ciencias sociales (como la economía, la psicología conductual o la sociología) y científicas (las matemáticas biología evolutiva, las neurociencias, e incluso la física). La inter-trans-disciplinariedad es una dificultad añadida, pero es un reto insoslayable si pretendemos una protección jurídica eficaz de nuestros datos cerebrales (Garrigues Walker, 2024).

En conclusión, más apremiante que inventar nuevos (neuro)derechos —que corren el riesgo de quedar en un simple formalismo—, es reforzar, adaptar o moldear los derechos ya existentes ante los riesgos de los avances tecnológicos. Como en otros ámbitos vinculados a la IA y a la revolución tecnológica, quedan muchas aristas jurídicas que perfilar. Pero el solo hecho de que se plantee este debate es una señal clara de que la dignidad de la persona, su libre albedrío y el respeto de los derechos humanos deben estar por encima de cualquier otro interés.

Bibliografía

Amoedo-Souto, C.A., «De las neurociencias al neuroderecho: perspectivas de regulación jurídica del cerebro», *El Cronista del Estado Social y Democrático de Derecho n.º 78,* 2018, págs. 42-51.

Ariano, C., «Reflexiones sobre el neuroderecho», Vox Juris Lima (Perú), n.º 32 (2), 2016, págs. 101-106.

Ausín, T., Morte, R. y **Monasterio, A**., «Neuroderechos: derechos humanos para las neurotecnologías», *Diario La Ley,* n.º 43, Sección ciberderecho.

Borbón Rodríguez, D. y otros, «Análisis crítico de los Neuro-Derechos Humanos al libre albedrío y al acceso equitativo a tecnologías de mejora», *Ius et scientia,* n.º 6, Universidad de Sevilla, 2020, págs. 135-161.

Bublitz, J.C., «Neurotechnologies and human rights: restating and reaffirming the multi-layered protection of the person», en *The International Journal of Human Rights*, Vol 28, n.º 5, 2024, disponible en https://doi.org/10.1080/13642987.2024.2310830.

Bublitz, J.C., «My Mind is Mine!? Cognitive Liberty as a Legal Concept», en Hildt, E. y Francke, A. (eds), *Cognitive Enhancement: An Interdisciplinary Perspective,* Springer, 2013, págs. 233-264.

Bueno I Torrens, D., *Cerebroflexia. El arte de construir el cerebro*, Plataforma Editorial, Barcelona, 2016.

Cotino Hueso, L., «La necesaria actualización de los derechos fundamentales como derechos digitales ante el desarrollo de internet y las nuevas tecnologías», en González, E. y, Rubio, R. (coords), y Pendás García, B. (dir), *España constitucional (1978-2018): trayectorias y perspectivas*, vol. 3, Centro de Estudios Políticos y Constitucionales, 2018, págs. 2347-2361.

De la Quadra Salcedo, T., «Retos, riesgos y oportunidades de la sociedad digital», en de la Quadra-Salcedo, T. y Piñar Mañas, J.L. (dirs), *Sociedad Digital y Derecho,* BOE, Madrid, 2018, págs. 21 y ss.

De la Sierra, S., «Digitalización y derechos fundamentales. Parte II: Tutela judicial efectiva en la era digital», 2021.

Accesible en: https://www.ibericonnect.blog/2021/11/simposio-digitalizacion-y-derechos-fundamentales-parte-ii-tutela-judicial-efectiva-en-la-era-digital/

Desmoulin-Canselier, S., «La France à l'ère du neurodroit? La neuro-imagerie dans le contentieux civil français», *Droit et societé,* n.º 101, 2019, págs. 115-135.

Donnelly-Lazarov, B., *Neurolaw and Responsibility for Action*, Cambridge University Press, Cambridge, 2018.

García Deltell, J.J., «El neuroderecho y el impacto de las neurociencias como nuevo paradigma para la filosofía del Derecho», en *Bioética, neuroética, libertad y justicia*, Comares, Granada, págs. 919-931, 2013.

Garrigues Walker, A. y González de la Garza, L.M., *Qué son los neuroderechos y cuál es su importancia para la evolución de la naturaleza humana,* Aranzadi, Cizur Menor, 2024.

Herran Ortiz, A. I., «A vueltas con los neuroderechos. ¿Es tiempo de configurar nuevos derechos constitucionales en España?», en *Revista General de Derecho Constitucional,* n.º 38, 2023.

Hertz, N., «Neurorights–Do We Need New Human Rights? A Reconsideration of the Right to Freedom of Thought», en *Neuroethics,* Vol 16, n.º 1, 2023.

Ienca, M. y Andorno, R., «Towards new human rights in the age of neuroscience andneurotechnology», *Life Sciences, Society and Policy 13,* 5, 2017, disponible en https://lsspjournal.biomedcentral.com. Traducción española en *Análisis filosófico* n.º 41, Chile, 2017, págs. 141-185.

Iglesias Cáceres, J.A., «La Carta de Derechos Digitales de España y su incidencia en los neuroderechos laborales», en *Revista Jurídica del Trabajo* vol.2 n.º 5, 2021, págs. 92-115.

Kahneman, D., *Pensar rápido, pensar despacio*, Debolsillo, 2011, reimpresión mayo 2017.

Kurzweil, R., *Cómo crear una mente. El secreto del pensamiento humano*, Lola Books, 2019.

Lazcoz, G. y Castillo Parrilla, J. A., «Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos: el caso SyRI», en *Revista Chilena de Derecho y Tecnología,* n.º 9, 2020.

López-Silva, P. y Madrid, R., «Sobre la conveniencia de incluir los neuroderechos en la Constitución o en la Ley», en *Revista Chilena de Derecho y Tecnología* n.º 10, 2021, págs. 49-72.

Macknik Stephen, L. y Martínez-Conde, S., *Los engaños de la mente*, Destino, 2010.

Martínez Sánchez, A., «Neuroimágenes y neurodisciplinas: sobre ciertas limitaciones de la utilización de la imagen por resonancia magnética funcional», em *Revista Internacional de Filosofía,* n.º 59, 2013, págs. 115-123.

Monasterio, A. y otros, «Traducir el pensamiento en acción: interfaces cerebro-máquina y el problema ético de la agencia», en *Revista bioética y Derecho,* n.º 46, 2019, págs. 29-46, 2019.

Moreu Carbonell, E., «El derecho a la igualdad de acceso a la mejora cognitiva», en Bayod López, C. (dir), *Persona y derecho civil, los retos del siglo XXI (persona, género, transgénero, inteligencia artificial y animales sensibles),* Aranzadi, Cizur Menor, 2023, págs. 141-149.

Moreu Carbonell, E., «La regulación de los neuroderechos», en *Revista General de Legislación y Jurisprudencia,* n.º 1, 2022, págs. 69-98.

Mozo Seoane, A., *Los límites de la tecnología. Marco ético y regulación jurídica*, Editorial Reus, 2021.

Muñoz Machado, S., *La regulación de la Red, poder y Derecho en internet*, Madrid, Taurus, 2020.

Petoft, A. y Abbasi, M., «Current limits of neurolaw: A brief overview», en *Médecine & Droit,* 2019, págs. 29-34.

Piñar Mañas, J.L., «Notas sobre las propuestas normativas para la regulación de los neuroderechos en Chile*»,* en *Derecho Digital e Innovación,* n.º *7*, 2020.

Piñar Mañas, J.L., «Sociedad, innovación y privacidad», en *Información Comercial Española ICE. El cambio digital en la economía. Un proceso disruptivo,* n.º 897, 2017, págs. 67-75.

Ponce Solé, J., «El riesgo de las Tecnologías de Incentivos de la Compulsión (TIC): arrojando algo (más) de luz a la oscuridad», *red temática* «El nudging aplicado a la mejora de la regulación», 2021. Disponible en https://rednmr.wordpress.com.

Ponce Solé, J., «Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico», en *Revista General de Derecho Administrativo*, n.º 50, Iustel, 2019.

Reche Tello, N., «Nuevos derechos frente a la neurotecnología: la experiencia chilena», en *Revista de Derecho Político UNED,* n.º 112, 2021, págs. 415-446. Disponible en https://doi.org/10.5944/rdp.112.2021.32235.

Rivas García, F., «Límites éticos y jurídicos de la investigación biomédica», en *Revista de Derecho UNED* n.º 18, 2016, págs. 293-309.

Röllnert Liern, G., *Los neuroderechos y la libertad de pensamiento*, Dykinson, 2024.

Sánchez Rubio, A., «El uso del test p300 en el proceso penal español: algunos aspectos controvertidos», en *Revista electrónica de ciencia penal y criminología,* n.º 18, 2016. Disponible en http://criminet.ugr.es/recpc.

Tahirí Moreno, J., «Una panorámica de los sistemas de inteligencia artificial desde la perspectiva del Derecho administrativo», en *Revista Aragonesa de Administración Pública*, n.º 61, 2024, págs. 137-168.

Taruffo, M. y Nieva, J. (dirs), *Neurociencia y proceso judicial*, Marcial Pons, 2013.

Tovino, S., «The impact of Neuroscience on Health Law», en *Neuroethics*, vol 1, 2008, págs. 101-117.

Vincent, N. A., «Neurolaw and direct brain intervention», *Criminal Law and Philosophy* vol. 8-1, 2014, págs. 43-50.

Yuste, R., «Las nuevas neurotecnologías y su impacto en la ciencia, medicina y sociedad», *Lecciones Cajal 1*, Prensas de la Universidad de Zaragoza, 2019.

Yuste, R., Goering, S. y otros, «Four ethical priorities for neurotechnologies and AI», Nature, 551, 2017, págs. 159-163.

12.

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EUROPA: EL REGLAMENTO DE LA UNIÓN EUROPEA Y CONVENIO MARCO DEL CONSEJO DE EUROPA

Miguel Ángel Presno Linera[1]
Catedrático Derecho constitucional
Universidad de Oviedo

SUMARIO: I. Introducción. II. Las iniciativas para la regulación supranacional de la inteligencia artificial. 2.1. En la Unión Europea. 2.2. En el Consejo de Europa. III. ¿De qué se habla cuando se habla de inteligencia artificial en la normativa europea? IV. Los principios que inspiran la regulación de la inteligencia artificial en Europa. V. Un enfoque de la regulación de la inteligencia artificial basado en los riesgos. VI. Las diferencias estructurales entre el reglamento de la Unión Europea y el convenio marco del Consejo de Europa. VII. Algunas conclusiones sobre la normativa europea de inteligencia artificial y los derechos fundamentales.

I. Introducción

En el mes de marzo de 2024, y con muy pocos días de diferencia, culminaron los trabajos de elaboración tanto del texto del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifi-

1 Este trabajo es uno de los resultados del Proyecto de investigación PID2022-136548NB-I00 FEDER, UE Los retos de la inteligencia artificial para el Estado social y democrático de Derecho financiado por el Ministerio de Ciencia e Innovación.

can determinados actos legislativos de la Unión Europea (en adelante, RIA), como del Convenio Marco del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho (en lo sucesivo, CIA).

Ambas normas, sobre las que nos ocuparemos en las líneas siguientes, son el resultado presente de un proceso de debate legislativo, institucional, social, tecnológico y económico que se remonta a varios años atrás y que parte del convencimiento, como se dice en el primer párrafo del Libro Blanco sobre la inteligencia artificial de la Comisión Europea, de 19 de febrero de 2020, de que la inteligencia artificial (en lo sucesivo, IA) «se está desarrollando rápido. Cambiará nuestras vidas, pues mejorará la atención sanitaria (por ejemplo, incrementando la precisión de los diagnósticos y permitiendo una mejor prevención de las enfermedades), aumentará la eficiencia de la agricultura, contribuirá a la mitigación del cambio climático y a la correspondiente adaptación, mejorará la eficiencia de los sistemas de producción a través de un mantenimiento predictivo, aumentará la seguridad de los europeos y nos aportará otros muchos cambios que de momento solo podemos intuir. Al mismo tiempo, la IA conlleva una serie de riesgos potenciales, como la opacidad en la toma de decisiones, la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas o su uso con fines delictivos»[2].

II. Las iniciativas para la regulación supranacional de la inteligencia artificial

2.1. En la Unión Europea

En su reunión de 19 de octubre de 2017, el Consejo Europeo concluyó que, para construir con éxito una Europa digital, la Unión Europea (UE) necesita, en particular, «concienciarse de la urgencia de hacer frente a las nuevas tendencias, lo que comprende cuestiones como la inteligencia artificial y las tecnologías de cadena de bloques, garantizando al mismo tiempo un elevado nivel de protección de los datos, así como los derechos digitales y las normas éticas. El Consejo Europeo ruega a la Comisión que, a principios de 2018, proponga un planteamiento europeo respecto de la inteligencia artificial y le pide que presente las iniciativas necesarias para reforzar las condiciones marco con el fin de que la UE pueda buscar nuevos mercados gracias a innovaciones radicales basadas en el riesgo y reafirmar el liderazgo de su industria»[3].

2 https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1 [Consulta: 18/02/2025]

3 https://www.consilium.europa.eu/media/21604/19-euco-final-conclusions-es.pdf [Consulta: 18/02/2025]

Se comienza a evidenciar así la preocupación de las instituciones de la UE a propósito, por lo que aquí interesa, de la regulación jurídica de la IA, de manera que se pueda aprovechar todo lo que supone en materia de innovación y desarrollo tecnológico y, al mismo tiempo, queden garantizados de manera adecuada los derechos fundamentales y el propio Estado social y democrático de Derecho.

Transcurrido poco más de un año, el 7 de diciembre de 2018, la Comisión Europea presentó la Comunicación titulada «Plan Coordinado sobre la Inteligencia Artificial», junto con el Plan Coordinado sobre el Desarrollo y Uso de la Inteligencia Artificial «Made in Europe»-2018, preparado por los Estados miembros (como parte del Grupo sobre la Digitalización de la Industria Europea y la Inteligencia Artificial), Noruega, Suiza y la Comisión[4]. Cabe destacar que aquí se ofrece un concepto de IA que, como iremos viendo, cambiará a lo largo de este proceso: «el término "inteligencia artificial" se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos». Se apunta aquí a la idea de «cierto grado de autonomía», que será esencial para la conceptualización de los sistemas de IA.

La preocupación por la garantía de los derechos fundamentales frente a los riesgos que plantea la IA se exteriorizó de manera evidente en el Consejo de la Unión Europea de 11 de febrero de 2019, donde se destacó la importancia de garantizar el pleno respeto de los derechos de los ciudadanos europeos mediante la aplicación de directrices éticas para el desarrollo y el uso de la inteligencia artificial dentro de la Unión Europea y a nivel mundial, haciendo de la ética de la inteligencia artificial una ventaja competitiva para la industria europea[5].

Poco más de un año después, el 19 de febrero de 2020, la Comisión publicó el ya citado *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza*[6], donde se afirma que «la Comisión respalda un enfoque basado en la regulación y en la inversión, que tiene el doble objetivo de promover la adopción de la inteligencia artificial y de abordar los riesgos vinculados a determinados usos de esta nueva tecnología. La finalidad del presente Libro Blanco es formular alternativas políticas para alcanzar estos objetivos...».

4 https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52018DC0795 [Consulta: 18/02/2025]

5 https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/es/pdf [Consulta: 18/02/2025]

6 https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1 [Consulta: 18/02/2025]

En el mes de octubre del mismo año 2020, el Parlamento Europeo aprobó diversas resoluciones en materia de IA en el ámbito de la ética[7], la responsabilidad civil[8] y los derechos de propiedad intelectual[9], a las que siguieron, ya en 2021, resoluciones sobre el uso de la IA y en los sectores educativo, cultural y audiovisual[10] y en materia penal[11].

Antes de esta última Resolución ya se había publicado el documento con el que formalmente se abrió uno de los procedimientos normativos que nos ocupan: el 21 de abril de 2021 se conoció la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecer normas armonizadas en materia de inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión, elaborada por la Comisión[12].

Cabe recordar ahora, por lo que luego veremos, que la IA se definió entonces (artículo 3.1) como «el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa». También que en esa propuesta no se incluyó referencia alguna a los luego llamados «modelos fundacionales», es decir, los sistemas de IA entrenados con una cantidad ingente de datos y que son capaces de realizar una gran variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural, los que el Reglamento europeo ha acabado denominando sistemas de inteligencia artificial de uso general y a los que nos referiremos más adelante.

---

7 Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, 2020/2012(INL) [Consulta: 18/02/2025]

8 Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un régimen de responsabilidad civil en materia de inteligencia artificial, 2020/2014(INL) [Consulta: 18/02/2025]

9 Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial, 2020/2015(INI) [Consulta: 18/02/2025]

10 Resolución del Parlamento Europeo, de 19 de mayo de 2021, sobre la inteligencia artificial en la educación, la cultura y el sector audiovisual, https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_ES.html [Consulta: 18/02/2025]

11 Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en materia penal, https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_ES.html [Consulta: 18/02/2025]

12 https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52021PC0206 [Consulta: 18/02/2025]

A propósito de esta propuesta, el 6 de diciembre de 2022, el Consejo de la Unión Europea hizo pública su Orientación general de 25 de noviembre[13], donde señaló que «para garantizar que la definición de los sistemas de IA proporcione criterios suficientemente claros para distinguirlos de otros sistemas de software más clásicos, el texto transaccional restringe la definición del artículo 3, apartado 1, a los sistemas desarrollados a través de estrategias de aprendizaje automático y estrategias basadas en la lógica y el conocimiento», es decir, se introduce el criterio del aprendizaje automático como una de las características de los sistemas de IA, algo que no estaba previsto así en la propuesta de la Comisión; también se amplían los sistemas que se pretende prohibir y se introducen cambios en los considerados de «alto riesgo».

A continuación, cabe mencionar las enmiendas introducidas por el Parlamento Europeo en el texto de la Comisión y aprobadas el 14 de junio de 2023[14], que incorporaron una nueva definición de sistema de IA: «un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida —como predicciones, recomendaciones o decisiones— que influya en entornos reales o virtuales»; también de lo que se entendió por un modelo fundacional —«un modelo de sistema de IA entrenado con un gran volumen de datos, diseñado para producir información de salida de carácter general y capaz de adaptarse a una amplia variedad de tareas diferentes»—, al tiempo que, entre otras cosas. se ampliaron los sistemas de IA prohibidos.

En el mes de diciembre de 2023 tuvieron lugar los *trílogos* entre las instituciones europeas implicadas[15] (Parlamento, Comisión, Consejo) para limar las diferencias en cuestiones tan relevantes como la prohibición, o no, del uso de los sistemas de identificación biométrica remota en tiempo real en espacios de acceso público o el alcance de la regulación de los entonces denominados «modelos fundacionales», que pasaron a llamarse «modelos de IA de uso general».

Finalmente, se aprobó la Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión y se publicó en el Diario Oficial de la UE, el 12 de julio de 2024, el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifi-

13 https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf [Consulta: 18/02/2025]

14 https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html [Consulta: 18/02/2025]

15 https://www.consilium.europa.eu/es/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai [Consulta: 18/02/2025]

can los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)

2.2. En el Consejo de Europa

En un contexto más amplio como es el de los 46 Estados que componen el Consejo de Europa, incluidos los 27 de la Unión Europea, cabe destacar, en primer lugar, el trabajo de investigación sobre algoritmos y derechos humanos, según el cual la IA afectará a un gran número, sino a la práctica totalidad, de nuestros derechos fundamentales[16]; así, al derecho a la libertad personal y, muy relacionado con él, al derecho a un juicio justo y a la tutela de los tribunales; en segundo lugar, a los derechos de las personas en su dimensión más privada, como el derecho a la intimidad y a la protección de datos; en tercer lugar, a los derechos vinculados a la dimensión pública y relacional de las personas, como las libertades de expresión, información, creación artística e investigación pero también a las libertades de reunión y asociación, tanto en el plano meramente ciudadano como en lo que se refiere, por ejemplo, al ámbito laboral (libertad sindical, derecho de huelga); en cuarto lugar, y a su vez vinculado a muchos otros derechos, al de no sufrir discriminación por raza, género, edad, orientación sexual...; en quinto lugar, a los derechos dependientes del acceso a los servicios públicos (educación, sanidad...) y, en general, a los derechos sociales (prestaciones por desempleo, enfermedad, jubilación...); finalmente, y por no extendernos mucho más, al derecho a intervenir en procesos participativos de índole política (elecciones, referendos, iniciativas legislativas populares...) y en, general, a las libertades en el ámbito ideológico (de pensamiento, conciencia y religión).

En segundo lugar, la Asamblea Parlamentaria del Consejo de Europa aprobó, el 22 de octubre de 2020, un conjunto de principios éticos básicos que deberían respetarse al elaborar y establecer aplicaciones de IA, incluida la transparencia, la justicia y la equidad, la responsabilidad humana de la toma de decisiones, la seguridad, la privacidad y la protección de datos. Ha identificado la necesidad de crear un marco normativo transversal para la IA, con principios específicos basados en la protección de los derechos humanos, la democracia y el Estado de derecho e instó al Comité de Ministros a elaborar un instrumento jurídicamente vinculante que regule la IA.

En tercer lugar, el Comité de Ministros adoptó, el 20 de mayo de 2022, un enfoque transversal de la inteligencia artificial en los diversos sectores del

16 Consejo de Europa, *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, 2018, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 [Consulta: 18/02/2025]

Consejo de Europa, estableciendo el Comité sobre Inteligencia Artificial (CAI) y encomendándole la elaboración de un Convenio [marco] jurídicamente vinculante sobre el desarrollo, diseño y aplicación de sistemas de IA, basado en las normas del Consejo de Europa en materia de derechos humanos, democracia y estado de derecho, sobre la base de estos principios fundamentales.

El Comité de Ministros también decidió permitir la inclusión en las negociaciones de la Unión Europea y de los Estados no europeos interesados que compartan los valores y objetivos del Consejo de Europa; a saber, Argentina, Australia, Canadá, Costa Rica, la Santa Sede, Israel, Japón, México, Perú, los Estados Unidos de América y Uruguay. El Consejo de Europa también hizo partícipes a actores no estatales en las negociaciones: un total de 68 representantes de la sociedad civil y de la industria participaron en calidad de observadores, interviniendo junto con Estados y representantes de otras organizaciones internacionales, como la Organización para la Seguridad y la Cooperación en Europa (OSCE), la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y los órganos y comités pertinentes del Consejo de Europa. La Unión Europea también participó en las negociaciones representada por la Comisión Europea, incluyendo en su delegación también a representantes de la Agencia de los Derechos Fundamentales de la Unión Europea (FRA) y del Supervisor Europeo de Protección de Datos (SEPD). Aunque se formularon acusaciones de que se excluía a las ONG de las negociaciones, supuestamente porque los Estados Unidos se negaban a compartir cierta información con participantes no estatales, desde el Consejo de Europa se justificó el proceso argumentando que la manera de desarrollar el borrador (discutirlo primero a puerta cerrada entre los participantes estatales y solo publicarlo y discutirlo en el plenario después) era conforme con las normas y prácticas vigentes.

El resultado final ha sido el citado Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho, que se abrió a la firma durante la conferencia de Ministros de Justicia del Consejo de Europa celebrada en Vilna el 5 de septiembre de 2024 y ha sido firmado por Andorra, Georgia, Islandia, Noruega, Moldavia, San Marino, el Reino Unido, Israel, Estados Unidos y la Unión Europea.

III. ¿De qué se habla cuando se habla de inteligencia artificial en la normativa europea?

En la corta historia de la IA[17] se han proporcionado distintas definiciones que, en general, aluden al desarrollo de sistemas que imitan o reproducen el

17 Existe acuerdo en ubicar el nacimiento del nombre IA en un taller científico que, en el verano de 1956, reunió, entre otros, a John McCarthy, Marvin Minsky, Claude Shannon, Herbert Simon, Allan Nevell... en el Dartmouth College y en que esa denominación la

pensamiento y obrar humanos, actuando racionalmente —en el sentido de hacer lo «correcto» en función de su conocimiento— e interactuando con el medio. La IA pretende sintetizar o reproducir los procesos cognitivos humanos, tales como la percepción, la creatividad, la comprensión, el lenguaje o el aprendizaje. Para ello, la IA utiliza todas las herramientas a su alcance, entre las que destacan las proporcionadas por la computación, incluidos los algoritmos. No obstante, los sistemas de IA no usan cualquier algoritmo sino, esencialmente, los que «aprenden» a base del procesamiento de datos.

Por otro lado, en ocasiones se habla de IA cuando en realidad estamos hablando de un subcampo, el aprendizaje automático (o *machine learning* en inglés, AA en lo sucesivo). El AA trata de encontrar patrones en datos para construir sistemas predictivos o explicativos; por tanto, puede considerarse una rama de la IA ya que a partir de la experiencia (los datos) toma decisiones o detecta patrones significativos y eso es una característica fundamental de la inteligencia humana. Es importante resaltar que para que un sistema de AA tenga éxito es tan necesario utilizar los algoritmos adecuados como realizar una correcta gestión y tratamiento de los datos utilizados para desarrollar el sistema.

Profundizando un poco más, nos encontramos con las redes neuronales, también llamadas redes neuronales artificiales, que son un modelo computacional de aprendizaje automático que procesa la información a través de un conjunto de unidades llamadas neuronas, o neuronas artificiales, que están conectadas entre sí y organizadas por capas, formando una red. Los datos de entrada atraviesan la red neuronal, donde son procesados mediante operaciones matemáticas, generando una salida. Por su parte, el concepto de aprendizaje profundo *(Deep learning)* hace referencia a las redes neuronales de un gran número de capas. No existe un criterio claro en cuanto a partir de qué número de capas ocultas podemos considerar una red neuronal como profunda y, por tanto, aprendizaje profundo, pero hay una opinión cada vez más extendida entre los expertos que afirma que cualquier red con más de 2 capas ocultas puede considerarse «profunda».

La dificultad de ofrecer una definición «acabada» de la IA se presenta también en el ámbito jurídico, como se puede comprobar leyendo las diferentes versiones que se han ido ofreciendo durante el proceso de aprobación del RIA: así, en el texto que presentó la Comisión el 21 de abril de 2021 se entendía como «el software que se desarrolla empleando una o varias de técnicas y estrategias que figuran en el Anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de

propuso John McCarthy; pueden verse al respecto el libro de Mitchell (2024: 27 y ss.) y el discurso de ingreso de la profesora Asunción Gómez-Pérez en la Real Academia Española con el título Inteligencia artificial y lengua española, https://www.rae.es/sites/default/files/2023-05/Discurso%20de%20ingreso%20de%20Asuncion%20Gomez-Perez.pdf [Consulta: 18/02/2025]

salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa» (artículo 3).

Tras las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023, se definió como «un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida —como predicciones, recomendaciones o decisiones— que influya en entornos reales o virtuales».

En el texto definitivo se ha conceptualizado como un sistema basado en máquinas diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de las entradas que recibe, salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales (artículo 3).

En el Preámbulo del RIA se explica que una característica clave de los sistemas de IA es su capacidad para inferir. Esta inferencia se refiere al proceso de obtención de resultados, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos y virtuales, y a una capacidad de los sistemas de IA para derivar modelos y/o algoritmos a partir de entradas/datos. Las técnicas que permiten la inferencia al construir un sistema de IA incluyen enfoques de aprendizaje automático que aprenden, a partir de los datos, cómo alcanzar determinados objetivos; y enfoques basados en la lógica y el conocimiento que infieren a partir del conocimiento codificado o la representación simbólica de la tarea que debe resolverse. La capacidad de un sistema de IA para inferir va más allá del procesamiento básico de datos, permitiendo el aprendizaje, el razonamiento o el modelado. El término «basado en máquinas» se refiere al hecho de que los sistemas de IA funcionan en máquinas. A efectos del presente Reglamento, los entornos deben entenderse como los contextos en los que operan los sistemas de IA, mientras que los resultados generados por el sistema de IA reflejan diferentes funciones realizadas por los sistemas de IA e incluyen predicciones, contenidos, recomendaciones o decisiones. Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que tienen cierto grado de independencia de las acciones de la intervención humana y capacidades para funcionar sin intervención humana. La capacidad de adaptación que puede mostrar un sistema de IA tras su despliegue se refiere a las capacidades de autoaprendizaje, que permiten al sistema cambiar mientras se utiliza.

Esta definición coincide con la que asume el CIA, en cuyo artículo 2 se dice que, a los efectos del presente Convenio, se entenderá por «sistema de inteligencia artificial» un sistema basado en máquinas que, con objetivos explícitos o implícitos, infiere, a partir de los datos que recibe, cómo generar resultados, como predicciones, contenidos, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales. Los diferentes sistemas de inteligencia artificial varían en sus niveles de autonomía y adaptabilidad después de la implementación».

Como se expone en el documento explicativo sobre el Convenio, esta definición de sistema de inteligencia artificial se extrae de la última definición revisada adoptada por la OCDE el 8 de noviembre de 2023. La elección de los redactores de utilizar este texto en particular es significativa no sólo por la alta calidad del trabajo realizado por la OCDE y sus expertos, sino también por la necesidad de mejorar la cooperación internacional en el tema de la inteligencia artificial y facilitar los esfuerzos destinados a armonizar la gobernanza de la inteligencia artificial a nivel mundial, en particular mediante la armonización de la terminología pertinente. La definición refleja una amplia comprensión de lo que son los sistemas de inteligencia artificial, específicamente en comparación con otros tipos de sistemas de software tradicionales más simples basados en las reglas definidas únicamente por personas físicas para ejecutar operaciones automáticamente. Su objetivo es garantizar la precisión y la seguridad jurídicas, sin dejar de ser lo suficientemente abstracto y flexible como para seguir siendo válido a pesar de los futuros avances tecnológicos. La definición se redactó a los efectos de la Convención Marco y no tiene por objeto dar un significado universal al término pertinente. Los redactores tomaron nota de la exposición de motivos que acompaña a la definición actualizada de sistema de inteligencia artificial que figura en la Recomendación de la OCDE sobre Inteligencia Artificial (OCDE/LEGAL/0449, 2019, modificada en 2023) para una explicación más detallada de los distintos elementos de la definición. Si bien esta definición proporciona un entendimiento común entre las Partes en cuanto a lo que son los sistemas de inteligencia artificial, las Partes pueden especificarlo en sus ordenamientos jurídicos nacionales para mayor seguridad jurídica y precisión, sin limitar su alcance.

IV. Los principios que inspiran la regulación de la inteligencia artificial en Europa

El Grupo independiente de expertos de alto nivel sobre IA creado por la Comisión Europea hizo públicos en abril de 2019 una serie de principios con el objeto de contribuir a garantizar la fiabilidad y el fundamento ético de la IA: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas[18].

Y, como se recuerda en el considerando 27 del RIA, por acción y supervisión humanas se entiende que los sistemas de IA se desarrollan y utilizan como una herramienta al servicio de las personas, que respeta la dignidad

18 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai [Consulta: 18/02/2025]

humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos; por solidez técnica y seguridad se entiende que los sistemas de IA se desarrollan y utilizan de manera que sean sólidos en caso de problemas y resilientes frente a los intentos de alterar el uso o el funcionamiento del sistema de IA para permitir su uso ilícito por terceros y reducir al mínimo los daños no deseados; por gestión de la privacidad y de los datos se entiende que los sistemas de IA se desarrollan y utilizan de conformidad con normas en materia de protección de la intimidad y de los datos, al tiempo que tratan datos que cumplen normas estrictas en términos de calidad e integridad; por transparencia se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos; por diversidad, no discriminación y equidad se entiende que los sistemas de IA se desarrollan y utilizan de un modo que incluya a diversos agentes y promueve la igualdad de acceso, la igualdad de género y la diversidad cultural, al tiempo que se evitan los efectos discriminatorios y los sesgos injustos prohibidos por el Derecho nacional o de la Unión; por bienestar social y ambiental se entiende que los sistemas de IA se desarrollan y utilizan de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia.

Estos principios se concretan a lo largo del articulado del Reglamento y, en especial, cuando se regulan las condiciones de funcionamiento de los sistemas de alto riesgo; así, por ejemplo, estos sistemas (artículo 14) se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas. El objetivo de la supervisión humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos. Las medidas de supervisión serán proporcionales a los riesgos, al nivel de autonomía y al contexto de uso del sistema de IA de alto riesgo. Entre otras, las personas responsables de la supervisión podrán decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere; también, intervenir en el funcionamiento del sistema o interrumpirlo pulsando un botón de parada o mediante un procedimiento similar que permita que se detenga de forma segura.

En segundo lugar, a estos sistemas de alto riesgo se les exige (artículo 15) que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida, algo a lo que pueden contribuir soluciones de redundancia técnica, tales como copias de seguridad o planes de prevención contra fallos. Entre las soluciones técnicas destinadas a subsanar vulnerabilidades específicas de la IA figurarán medidas para prevenir, detectar, combatir, resolver y controlar los ataques que traten de manipular el conjunto de datos de entrenamiento («envenenamiento de datos»), o los componentes entrenados previamente utilizados en el entrenamiento («envenenamiento de modelos»), la información de entrada diseñada para hacer que el modelo de IA cometa un error («ejemplos adversarios» o «evasión de modelos»), los ataques a la confidencialidad o los defectos en el modelo.

En tercer lugar, los sistemas de alto riesgo (artículo 13) se diseñarán y desarrollarán de un modo que se garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente sus resultados de salida. Irán acompañados de las instrucciones de uso correspondientes en un formato digital o de otro tipo adecuado, las cuales incluirán información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue. Además (artículo 50), los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA, excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización. Por su parte, los proveedores de sistemas de IA, entre los que se incluyen los sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto, velarán por que los resultados de salida del sistema estén marcados en un formato legible por máquina y que sea posible detectar que han sido generados o manipulados de manera artificial.

En una línea similar, aunque de manera más genérica, el Convenio del Consejo de Europa prevé, entre las obligaciones generales para los Estados parte en relación con los sistemas de IA, las de supervisión[19], igualdad y no

19 Artículo 8 – Transparencia y supervisión. Cada Parte adoptará o mantendrá medidas para garantizar que existan requisitos adecuados de transparencia y supervisión adaptados a los contextos y riesgos específicos con respecto a las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, incluida la identificación de los contenidos generados por los sistemas de inteligencia artificial.

discriminación[20], privacidad y protección de datos personales[21]; fiabilidad[22] y transparencia[23].

V. Un enfoque de la regulación de la inteligencia artificial basado en los riesgos

La regulación de la IA, tal y como se concibe en Europa, aunque no solo en este espacio jurídico, requiere la aplicación de un enfoque basado en los riesgos, que adapte el tipo de las normas y su contenido a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA. Se trata de una de las concreciones del bien conocido «principio de precaución», que guía la actuación de la Unión Europea[24]; así, y por citar únicamente dos ejemplos en

20 Artículo 10 – Igualdad y no discriminación. 1. Cada Parte adoptará o mantendrá medidas con vistas a garantizar que las actividades relacionadas con el ciclo de vida de los sistemas de inteligencia artificial respeten la igualdad, incluida la igualdad de género, y la prohibición de la discriminación, según lo dispuesto en la legislación internacional y nacional aplicable. 2.Cada Parte se compromete a adoptar o mantener medidas destinadas a superar las desigualdades para lograr resultados justos, equitativos y equitativos, en consonancia con sus obligaciones nacionales e internacionales aplicables en materia de derechos humanos, en relación con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial.

21 Artículo 11 – Privacidad y protección de datos personales. Cada Parte adoptará o mantendrá medidas para garantizar que, en lo que respecta a las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial: a. los derechos de privacidad de las personas y los datos personales están protegidos, incluso a través de leyes, normas y marcos nacionales e internacionales aplicables; b. Se han establecido garantías y salvaguardias efectivas para las personas, de conformidad con las obligaciones jurídicas nacionales e internacionales aplicables.

22 Artículo 12 – Fiabilidad. Cada Parte adoptará, según proceda, medidas para promover la fiabilidad de los sistemas de inteligencia artificial y la confianza en sus resultados, que podrían incluir requisitos relacionados con la calidad y la seguridad adecuadas a lo largo de todo el ciclo de vida de los sistemas de inteligencia artificial.

23 Artículo 15 – Garantías procesales. 2. Cada Parte procurará garantizar que, según corresponda al contexto, se notifique a las personas que interactúen con sistemas de inteligencia artificial que están interactuando con dichos sistemas y no con un ser humano.

24 Como se explica en la comunicación de la Comisión Europea, de 2 de febrero de 2000, sobre el recurso al principio de precaución, el Tratado de la Comunidad Europea solo contiene una referencia explícita al principio de precaución, a saber, en el título dedicado a la protección del medio ambiente. No obstante, en la práctica, su ámbito de aplicación es mucho más amplio y se extiende asimismo a la política de los consumidores y a la salud humana, animal o vegetal. A falta de una definición del principio de precaución en el Tratado o en otros textos comunitarios, el Consejo solicitó a la Comisión, en su Resolución de 13 de abril de 1999, que elaborase líneas directrices claras y eficaces con vistas a la aplicación de este principio. La comunicación de la Comisión es una respuesta a esta solicitud. El establecimiento de líneas directrices comunes acerca de la aplicación del principio de precaución tendrá asimismo repercusiones positivas a escala internacional.
La Comisión subraya que el principio de precaución solo puede invocarse en la hipótesis

el ámbito que nos ocupa, en la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, se dice que las actividades de investigación en el ámbito de la robótica deben llevarse a cabo de conformidad con el principio de precaución, anticipándose a los posibles impactos de sus resultados sobre la seguridad y adoptando las precauciones debidas, en función del nivel de protección, al tiempo que se fomenta el progreso en beneficio de la sociedad y del medio ambiente; por su parte, en la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas se recuerda que tal enfoque debe estar en consonancia con el principio de precaución que guía la legislación de la Unión y debe ocupar un lugar central en cualquier marco regulador para la inteligencia artificial.

El RIA define el riesgo como la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio (artículo 3.2) y, como consecuencia, en algunos casos el riesgo resultante será inaceptable y eso conducirá a la prohibición de los sistemas de IA que lo generen. Otros sistemas se caracterizan de alto riesgo porque son capaces de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas y a ellos se les aplicarán las cautelas que ya hemos visto en el apartado anterior y se les someterá (artículo 9) a un proceso iterativo continuo, planificado y ejecutado durante todo el ciclo de vida, que requerirá revisio-

de un riesgo potencial y que en ningún caso puede justificar una toma de decisión arbitraria. El recurso al principio de precaución solo está justificado cuando se cumplen las tres condiciones previas, a saber: identificación de los efectos potencialmente negativos, evaluación de los datos científicos disponibles y determinación del grado de incertidumbre científica.

Medidas que se derivan del recurso al principio de precaución. El recurso al principio de precaución debe guiarse por tres principios específicos: 1) la aplicación del principio debe basarse en una evaluación científica lo más completa posible; en cada etapa esta evaluación debe determinar, en la medida de lo posible, el grado de incertidumbre científica; 2) toda decisión de actuar o de no actuar en virtud del principio de precaución debe ir precedida de una determinación del riesgo y de las consecuencias potenciales de la inacción; 3) tan pronto como se disponga de los resultados de la evaluación científica o de la determinación del riesgo, todas las partes interesadas deben tener la posibilidad de participar, con la máxima transparencia, en el estudio de las diferentes acciones que pueden preverse.

Aparte de estos principios específicos, siguen siendo aplicables los principios generales de una buena gestión de los riesgos cuando se invoca el principio de precaución. Se trata de los cinco principios siguientes: la proporcionalidad entre las medidas adoptadas y el nivel de protección elegido; la no discriminación en la aplicación de las medidas; la coherencia de las medidas con las ya adoptadas en situaciones similares o utilizando planteamientos similares; el análisis de las ventajas y los inconvenientes que se derivan de la acción o de la inacción y la revisión de las medidas a la luz de la evolución científica.

https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2000:0001:FIN:es:PDF [Consulta: 18/02/2025]

nes y actualizaciones sistemáticas periódicas. Constará de las siguientes etapas: a) la determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda plantear para la salud, la seguridad o los derechos fundamentales cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista; b) la estimación y la evaluación de los riesgos que podrían surgir cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista y cuando se le dé un uso indebido razonablemente previsible; c) la evaluación de otros riesgos que podrían surgir, a partir del análisis de los datos recogidos con el sistema de vigilancia post-comercialización; d) la adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados.

Con vistas a eliminar o reducir los riesgos asociados a la utilización del sistema de IA de alto riesgo, se tendrán debidamente en cuenta los conocimientos técnicos, la experiencia, la educación y la formación que se espera que posea el responsable del despliegue, así como el contexto en el que está previsto que se utilice el sistema.

También el Convenio Marco contempla un marco jurídico de gestión de riesgos e impactos (artículo 16), imponiendo a los Estados parte medidas para la identificación, evaluación, prevención y mitigación de los riesgos planteados por los sistemas de inteligencia artificial, teniendo en cuenta los impactos reales y potenciales en los derechos humanos, la democracia y el Estado de Derecho. Dichas medidas se graduarán y diferenciarán, según proceda, y deberán: a). tener debidamente en cuenta el contexto y el uso previsto de los sistemas de inteligencia artificial, en particular en lo que respecta a los riesgos para los derechos humanos, la democracia y el Estado de Derecho; b). tener debidamente en cuenta la gravedad y la probabilidad de los posibles impactos; c). considerar, cuando proceda, las perspectivas de las partes interesadas pertinentes, en particular de las personas cuyos derechos puedan verse afectados; d). aplicar de forma iterativa a lo largo de las actividades dentro del ciclo de vida del sistema de inteligencia artificial; e). incluir el monitoreo de los riesgos e impactos adversos para los derechos humanos, la democracia y el estado de derecho; f). incluir documentación de los riesgos, los impactos reales y potenciales, y el enfoque de gestión de riesgos; g). exigir, cuando proceda, que se sometan a ensayo los sistemas de inteligencia artificial antes de ponerlos a disposición para su primer uso y cuando se modifiquen significativamente. Finalmente, cada Parte evaluará la necesidad de una moratoria o prohibición u otras medidas adecuadas con respecto a determinados usos de los sistemas de inteligencia artificial cuando considere que dichos usos son incompatibles con el respeto de los derechos humanos, el funcionamiento de la democracia o el Estado de Derecho.

VI. Las diferencias estructurales entre el Reglamento de la Unión Europea y el Convenio Marco del Consejo de Europa

Ya se ha visto que el Reglamento y el Convenio Marco se terminaron de elaborar en fechas muy próximas y siendo conscientes en cada institución de lo que estaba haciendo la otra; de hecho, el Comité de Ministros del Consejo de Europa decidió permitir la inclusión en las negociaciones de la Unión Europea, que participó representada por la Comisión Europea, incluyendo en su delegación también a representantes de la Agencia de los Derechos Fundamentales de la Unión Europea y del Supervisor Europeo de Protección de Datos. Incluso el artículo 27.2 del Convenio Marco prevé que las Partes que sean miembros de la Unión Europea aplicarán, en sus relaciones mutuas, las propias normas de la Unión Europea que regulen las materias comprendidas en el ámbito de aplicación del Convenio, sin perjuicio de su objeto y fin y de su plena aplicación con las demás Partes.

Entrando en las diferencias estructurales más destacadas entre ambas normas cabe destacar, en primer lugar, que el ámbito de aplicación del Reglamento es regional, el territorio de la Unión Europea[25], al margen de que pueda generar, como veremos más adelante, un «efecto Bruselas», mientras que el Convenio Marco ya nace con una cierta vocación de globalidad; así, en su Preámbulo se invoca la necesidad de establecer, con carácter prioritario, un marco jurídico aplicable a escala mundial que establezca principios generales y normas comunes que rijan las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, preservando eficazmente los valores compartidos y aprovechando los beneficios de la inteligencia artificial para la promoción de esos valores de manera que favorezca la innovación responsable, y el artículo 30.1 prevé que el Convenio estará abierto a la firma de los Estados miembros del Consejo de Europa, de los Estados no miembros que hayan participado en su elaboración y de la Unión Europea.

En segundo lugar, el Reglamento, de conformidad con lo dispuesto en el artículo 288 del Tratado de Funcionamiento de la Unión Europea, «será obligatorio en todos sus elementos y directamente aplicable en cada Estado miembro»; el Convenio Marco será obligatorio en la medida en los Estados se incorporen a él. De acuerdo con esta característica, los destinatarios del CIA son los Estados, mientras que el RIA se dirige directamente a los proveedores y los responsables del despliegue.

25 De acuerdo con el artículo 1, «el objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, «sistemas de IA») en la Unión...»

En tercer término, el Reglamento incluye una normativa extensa (113 artículos y XIII Anexos) y muy prolija, con preceptos muy detallados, mientras que el Convenio Marco se compone de un articulado mucho más reducido (36 artículos) y, desde luego, menos detallado. El propio Convenio destaca su «carácter marco..., que podrá complementarse con otros instrumentos para abordar cuestiones específicas relacionadas con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial» (número 11 del Preámbulo).

En cuarto lugar, el Reglamento, en términos generales, contiene más reglas, es decir, incluye comportamientos precisos de lo que puede, o no, hacerse en materia de IA; el Convenio Marco, por su parte, adopta una configuración mucho más principialista, esto es, contiene mandatos de optimización, caracterizados por el hecho de que pueden ser cumplidos en diferente grado. Así, por ejemplo, el Reglamento establece una serie de prácticas de IA que estarán prohibidas (artículo 5)[26] e impone unas obligaciones que deben cumplir los sistemas de alto riesgo[27] y, entre otros, los proveedores de modelos de IA de uso general (artículo 51). Por su parte, el Convenio Marco dispone (artículo 4) que «cada Parte adoptará o mantendrá medidas para garantizar que las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial sean compatibles con las obligaciones de protección de los derechos humanos, consagradas en el derecho internacional aplicable y en su legislación nacional».

En quinto lugar, y también con carácter general pero no absoluto y en la línea de lo dicho en el punto anterior, el Reglamento impone obligaciones de medios y de resultado mientras que el Convenio Marco incluye, esencialmente, obligaciones de resultado, dejando a los Estados la concreción de las medidas adecuadas para alcanzarlos. Así, por ejemplo, el Reglamento prevé que los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo

26 Así, por citar algunas, las que se sirvan de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas; las que exploten alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoquen, o sea razonablemente probable que provoquen, perjuicios considerables a esa persona o a otra; las que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión...

27 Un sistema de gestión de riesgos (artículo 9), la gobernanza de los datos (artículo 10), la documentación técnica (artículo 11), supervisión humana (artículo 14).

que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas (artículo 14.1); más adelante, dispone que los proveedores de sistemas de IA de alto riesgo que consideren o tengan motivos para considerar que un sistema de IA de alto riesgo que han introducido en el mercado o puesto en servicio no es conforme con el presente Reglamento adoptarán inmediatamente las medidas correctoras necesarias para que sea conforme, para retirarlo del mercado, desactivarlo o recuperarlo, según proceda (artículo 20.1).

Por su parte, el Convenio Marco establece (artículo 1.1 y 1.2) que sus disposiciones «tienen por objeto garantizar que las actividades realizadas durante el ciclo de vida de los sistemas de inteligencia artificial sean plenamente compatibles con los derechos humanos, la democracia y el Estado de Derecho. Cada Parte adoptará o mantendrá medidas legislativas, administrativas o de otra índole apropiadas para dar efecto a las disposiciones establecidas en el presente Convenio. Estas medidas se graduarán y diferenciarán según sea necesario en función de la gravedad y la probabilidad de que se produzcan efectos adversos para los derechos humanos, la democracia y el Estado de Derecho a lo largo del ciclo de vida de los sistemas de inteligencia artificial»; además, y por mencionar otro supuesto, conforme al artículo 5, «1. Cada Parte adoptará o mantendrá medidas que tengan por objeto garantizar que los sistemas de inteligencia artificial no se utilicen para socavar la integridad, independencia y eficacia de las instituciones y procesos democráticos, incluido el principio de separación de poderes, el respeto de la independencia judicial y el acceso a la justicia. 2. Cada Parte adoptará o mantendrá medidas que tengan por objeto proteger sus procesos democráticos en el contexto de las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, incluido el acceso equitativo de las personas al debate público y su participación en él, así como su capacidad para formarse opiniones libremente)».

Finalmente, el Reglamento contiene un sistema de sanciones: «los Estados miembros establecerán el régimen de sanciones y otras medidas de ejecución, como advertencias o medidas no pecuniarias, aplicable a las infracciones del presente Reglamento que cometan los operadores y adoptarán todas las medidas necesarias para garantizar que se aplican de forma adecuada y efectiva... Tales sanciones serán efectivas, proporcionadas y disuasorias» (artículo 99.1). A continuación, el artículo especifica el importe de las multas administrativas que se deben imponer, que son significativas[28].

28 Así, el no respeto de la prohibición de las prácticas de IA a que se refiere el artículo 5 estará sujeto a multas administrativas de hasta 35.000.000 de euros o, si el infractor es una empresa, de hasta el 7 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior. El incumplimiento de cualquiera de las disposiciones que figuran a continuación en relación con los operadores o los organismos notificados, distintas de los mencionados en el artículo 5, estará sujeto a multas administrativas de hasta 15.000.000 de euros o, si el infractor es una empresa,

Por su parte, el Convenio Marco se limita a disponer que «cada Parte establecerá o designará uno o más mecanismos eficaces para supervisar el cumplimiento de las obligaciones establecidas en el Convenio» (artículo 26.1). Sin embargo, al tratarse de un tratado, en función de las normas del sistema jurídico nacional con respecto a la aplicación del derecho internacional, los tribunales nacionales podrán declarar las violaciones de determinados artículos. Para muchos signatarios, esto significará que los tribunales u otras autoridades tendrán que determinar si las normas de la CIA son suficientemente precisas para ser consideradas autoejecutables.

De todas maneras, el CIA probablemente funcionara como documento interpretativo para el Tribunal Europeo de Derechos Humanos (TEDH), que no está mencionado en el texto del CIA, que prevé una «Conferencia de las partes» como mecanismo de resolución de disputas (artículo 23).

VII. Algunas conclusiones sobre la normativa europea de inteligencia artificial y los derechos fundamentales

Los derechos fundamentales están muy presentes en el Reglamento Europeo, al menos en teoría, pues encontramos 96 referencias a esa expresión (la mayoría, 51, en el Preámbulo) pero, ya desde el primer párrafo del Preámbulo, se deja claro que «el objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior mediante el establecimiento de un marco jurídico uniforme, en particular para el desarrollo, la introducción en el mercado, la puesta en servicio y la utilización de sistemas de inteligencia artificial en la Unión», es decir, el Reglamento está orientado al «mercado», palabra que aparece en 356 ocasiones, y a la seguridad de los productos de IA, lo que ocurre es que dicha orientación se pretende que sea compatible con «un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente».

de hasta el 3 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior: a) las obligaciones de los proveedores con arreglo al artículo 16; b) las obligaciones de los representantes autorizados con arreglo al artículo 22; c) las obligaciones de los importadores con arreglo al artículo 23; las obligaciones de los distribuidores con arreglo al artículo 24; e) las obligaciones de los responsables del despliegue con arreglo al artículo 26... La presentación de información inexacta, incompleta o engañosa a organismos notificados o a las autoridades nacionales competentes en respuesta a una solicitud estará sujeta a multas administrativas de hasta 7.500.000 de euros o, si el infractor es una empresa, de hasta el 1 % del volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior (artículo 99.3, 4 y 5).

Dichos condicionamientos no son, obviamente, triviales y, siguiendo el análisis de Anu Bradford, son los que permiten diferenciar el modelo regulador europeo, que atiende a los derechos fundamentales, de los modelos norteamericano —basado en el mercado y en los incentivos a la innovación— y el chino —centrado en el desarrollo económico y en la «estabilidad del país» a costa de los derechos de la ciudadanía—, pero ninguno de dichos modelos es «puro» y, como ya se apuntado, el RIA se orienta claramente al mercado y a «brindar apoyo a la innovación» a través de unas normas armonizadas.

El RIA no crea nuevos «derechos fundamentales de la UE», sino que, conforme al artículo 1, trata de proteger a los «consagrados en la Carta» y lo hace prohibiendo ciertas «prácticas de IA» y sometiendo a condiciones a los sistemas de alto riesgo para los derechos fundamentales. Nos parece razonable el tratamiento de los sistemas de alto riesgo y un tanto decepcionante el apartado de las prohibiciones, en el que la tutela de los derechos se somete en ocasiones a políticas de seguridad o, en apariencia, al deseo de no poner demasiados obstáculos a las posibilidades de avance técnico de la Unión, un avance que no era precisamente rápido antes del Reglamento; además, varias de las prohibiciones —por ejemplo, las relativas a los sistemas que discriminan— ya se podrían amparar en normas vigentes y otras se han quedado cortas; por ejemplo, limitando la prohibición de los sistemas que infieren emociones al ámbito laboral y educativo y en ellos, además, con matices.

No cabe olvidar, además, que el RIA «no se aplicará a los sistemas de IA que, y en la medida en que, se introduzcan en el mercado, se pongan en servicio o se utilicen, con o sin modificaciones, exclusivamente con fines militares, de defensa o de seguridad nacional, independientemente del tipo de entidad que lleve a cabo estas actividades» (artículo 2.3).

Además, y esa es otra prueba —y no menor— de la importancia del mercado, el RIA permite la exportación de sistemas de IA cuyo uso estaría prohibido en la Unión por ser lesivo de los derechos fundamentales.

Finalmente, nos parece oportuna la necesaria evaluación previa del impacto de los sistemas de IA antes de desplegar tecnologías de alto riesgo que repercutan de forma directa y significativa en los derechos fundamentales, tal y como prevé el artículo 27 del RIA para los sistemas de alto riesgo y «que consistirá en: a) una descripción de los procesos en los que se utilizará el sistema de IA de alto riesgo en consonancia con su finalidad prevista; b) una descripción del período de tiempo durante el cual se prevé utilizar cada sistema de IA de alto riesgo y la frecuencia con la que está previsto utilizarlo; c) las categorías de personas físicas y colectivos que puedan verse afectados por su utilización en el contexto específico; d) los riesgos de perjuicio específicos que puedan afectar a las categorías de personas físicas y colectivos; e) una descripción de la aplicación de medidas de supervisión humana, de acuerdo con las instrucciones de uso; f) las medidas que deben adoptarse en caso de que dichos riesgos se materialicen, incluidos los acuerdos de gobernanza interna y los mecanismos de reclamación».

Por su parte, el Convenio Marco, como ya se ha dicho más arriba, adopta una configuración mucho más principialista que el Reglamento, esto es, contiene mandatos de optimización, caracterizados por el hecho de que pueden ser cumplidos en diferente grado.

En segundo lugar, y al igual que el Reglamento, tampoco el Convenio incluye derechos que no estén reconocidos en otros instrumentos internacionales.

Finalmente, el Convenio supone un intento de armonización y construcción de un marco estándar y mínimo para Europa, con alcance mundial; un esfuerzo centrado en los derechos humanos, la democracia y el Estado de derecho y no tanto en la economía y el mercado.

Bibliografía

Barrio Andrés, M. (dir.), *Comentarios al Reglamento Europeo de Inteligencia Artificial*, La Ley, Madrid, 2024.

Bradford, A., *Imperios digitales*, Shackleton books, 2024.

Cotino Hueso, L, «El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa»*, Revista Administración & Ciudadanía*, EGAP, 2024.

Cotino Hueso, L. y Simón Castellano, P. (dirs.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea*, Aranzadi, 2024

Gascón Marcén, A., «Nuevo Convenio Marco sobre Inteligencia Artificial del Consejo de Europa», *Derecho Digital e Innovación. Digital Law and Innovation Review*, 20, 2024.

Gómez-Pérez, A., *Inteligencia artificial y lengua española*, Real Academia Española, Madrid, 2023.

González Cabanes, F. y Díaz Díez, N., «¿Qué es la Inteligencia Artificial?» en Gamero Casado, E. y Pérez Guerrero, F. (dirs.), *Inteligencia artificial y sector público: retos, límites y medios* (37-72). Tirant lo Blanch, 2023, pp. 37-72.

Pérez-Ugena, M., «Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU., China e Iberoamérica», *Anuario Iberoamericano de Justicia Constitucional*, 28(1), 2024, pp. 129-156.

Presno Linera, M. Á., «La propuesta de "Ley de Inteligencia Artificial" europea», *Revista de las Cortes Generales*, 116, 2023, pp. 81-133.

Presno Linera, M. Á. y Meuwese, A.C.M., «La regulación de la inteligencia artificial en Europa», *Teoría y Realidad Constitucional*, 54, 2024, pp. 131-161.

Russel, S. y Norvig, P., *Inteligencia Artificial: un enfoque moderno*, Pearson Education, Madrid, 2008.

13.

CONSIDERACIONES POLÍTICO CRIMINALES SOBRE LA PROTECCIÓN PENAL DE LA VIDA PRIVADA EN LA ERA DIGITAL

M.ª Ángeles Rueda Martín
Catedrática de Derecho penal
Universidad de Zaragoza

SUMARIO: I. Vida privada e intimidad: dos conceptos estrechamente vinculados. II. Elementos comunes que integran la protección penal dispensada a la intimidad en el Código Penal de 1995. III. Sobre la necesidad político-criminal de ampliar el objeto de protección en los delitos contra la intimidad y la propia imagen. IV. Algunas observaciones críticas a la regulación penal en los delitos contra la intimidad y la propia imagen.

I. Vida privada e intimidad: dos conceptos estrechamente vinculados

Dentro de los derechos primarios esenciales que posee la persona individual como persona física, considerada de forma aislada, destacan los denominados derechos de la personalidad, que son aquellos que conceden un poder a las personas para proteger la esencia del ser humano y sus más importantes cualidades[1], y cuya extensión y contenido son diversos. La doctrina reconoce como tales, entre otros, el derecho a la intimidad personal y

1 Este trabajo desarrolla uno de los objetivos de investigación del Laboratorio de Sociología Jurídica de la Universidad de Zaragoza, reconocido como grupo de investigación de referencia por el Departamento de Innovación, Investigación y Universidad del Gobierno de Aragón (BOA 28/04/2023).
Véanse Serrano Alberca, J.A., *Comentarios a la Constitución*, 3.ª ed., p. 399, que los define de esta manera acogiendo la formulación de de Castro y Bravo; Bajo Fernández, «Protección del honor y de la intimidad», p. 99.

a la propia imagen que afectan, en general, «a lo que se ha dado en llamar su *vida privada»* y en cuya esfera se encuentran «el derecho a la imagen, el derecho a la inviolabilidad del domicilio y de las comunicaciones (postales, telegráficas y telefónicas) y el derecho al honor»[2]. La protección del derecho —fundamental— a la intimidad y la propia imagen se recoge en el art. 18.1 de la CE, que establece que *«se garantiza el derecho al honor, a la intimidad personal y familiar y a la propia imagen»*[3]. Además, en el art. 18.3 de la CE, donde se dispone que *«se garantiza el secreto de las comunicaciones y en especial, de las postales, telegráficas y telefónicas, salvo resolución judicial»*[4], y en el art. 18.4 de nuestra Carta Magna, que establece que *«la ley limitará el uso de la informática para garantizar el honor y la identidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos»*, se recogen expresamente dos manifestaciones del derecho a la intimidad personal y familiar[5]. Con razón, señalan Doval País y Anarte Borrallo que «la intimidad alude a la *vida privada* y esta puede manifestarse en espacios privados (lo más frecuente) o públicos. Es en el marco de la vida privada, en el que las perso-

2 Véase Serrano Alberca, *Comentarios a la Constitución*, 3.ª ed., p. 399. Espín indica también que los derechos recogidos en el art. 18 de la CE, aun con su respectiva especificidad, versan «todos ellos sobre la protección de un ámbito privado reservado para la propia persona y del que quedan excluidos los demás ... Así, todos estos derechos pueden considerarse destinados a la protección de un bien constitucionalmente relevante que abarca el ámbito protegido de todos ellos, la vida privada. Son, asimismo, derechos vinculados a la propia personalidad, derivados por ello de la dignidad de la persona reconocida en el art. 10 de la CE (STC 231/88, caso *Pantoja-Paquirri*)»; véase Espín, *Manual de Derecho Constitucional, I*, 1.ª ed., p. 221.

3 Véase el comentario de este precepto en Bajo Fernández, «Protección del honor y de la intimidad», pp. 97 y ss. El art. 18.1 de la CE garantiza el derecho «a la propia imagen» y constituye una manifestación concreta del derecho a la intimidad (véase Bajo Fernández, ob. cit. p. 104), que consiste en la facultad de la persona de decidir respecto al empleo de su imagen, como medio de garantizar la capacidad del individuo de controlar, en la medida de lo posible, la difusión de un elemento tan personal como la propia efigie, de tal forma que no pueda emplearse ésta, con o sin finalidad de lucro, sin su propio consentimiento. Véase Espín, *Manual de Derecho Constitucional*, I, 1.ª ed., p. 221. Cuerda Riezu enfatiza, con razón, la libre configuración de la apariencia física dentro del derecho a la propia imagen; véase Cuerda Riezu, «El velo islámico y el derecho a la propia imagen», pp. 31, 32, 33, 34 y 35.

4 El derecho fundamental consagrado en el art. 18.3 de la CE relativo al secreto de las comunicaciones, se suspenderá en los supuestos de estados de excepción y de sitio con arreglo a lo previsto en los arts. 55.1 y 116 de la CE. Véanse, asimismo, los arts. 18 y 32.3 de la LO 4/1981, de 1 de junio, de los Estados de Alarma, Excepción y Sitio. Véase al respecto Alonso Pérez, *Intervención de las comunicaciones postales, telegráficas y telefónicas*, pp. 147 y ss.

5 En la STC n.º 34/1996 se afirmó que «entre los derechos fundamentales que la norma constitucional enumera a título enunciativo y nunca como *numerus clausus*, se encuentra el que pone a buen recaudo la intimidad personal y familiar (art. 20.4 CE), con el que guarda un estrecho parentesco, por ser una de sus manifestaciones fenoménicas, *el secreto de las comunicaciones y, en especial, de las postales, telegráficas y telefónicas, salvo resolución judicial* (art. 18.1 y 3 CE)». Véanse, además, Huerta Tocildo/Andrés Domínguez, «Intimidad e informática», p. 19.

nas encuentran un contexto *para sí* de extraordinarias posibilidades para su libre manifestación y desarrollo como seres humanos, fin último al que sirve dicha protección. De este ámbito deberán excluirse, con carácter general, en consecuencia, los hechos, las informaciones, los datos, etc., que no afectan a dicho contexto íntimo de las personas por referirse a hechos y relaciones que se proyectan inevitablemente hacia el exterior con alguna —pero evidente— notoriedad, como los que tienen lugar en ámbitos profesionales, laborales, patrimoniales o *sociales* ... y, desde luego, aquellos que la propia persona opta voluntariamente por extraer del ámbito de su propia intimidad»[6].

Desde luego, no es una tarea sencilla delimitar lo que se entiende por intimidad[7]. Por una parte, una característica común de tal concepto que se puede deducir de la Constitución y de las leyes pone de relieve, como señaló Bajo Fernández, la existencia de una esfera íntima que alude a *ese ámbito personal donde cada uno, preservado del mundo exterior, encuentra las posibilidades de desarrollo y fomento de su personalidad*[8]. También el TC español ha reconocido en numerosas ocasiones —SsTC n.º 231/1988; n.º 197/1991; n.º 20/1992; n.º 219/1992; n.º 142/1993; n.º 57/1994; n.º 117/1994; n.º 143/1994; n.º 207/1996; n.º 170/1997; n.º 202/1999; n.º 98/2000; n.º 186/2000; n.º 70/2002; n.º 127/2003; n.º 233/2005; n.º 89/2006; n.º 77/2009, n.º 60/2010,

6 Véanse Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 497 (cursivas en el original). Véase una exposición del origen del concepto de vida privada realizada por de la Mata Barranco/Barina Ubiñas, «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», pp. 14 y ss., 19 y ss.

7 Véanse de la Mata Barranco/Barina Ubiñas, «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», pp. 28 y ss., 40 y ss.
En relación con el concepto de intimidad subyacente en el CP anterior, véase, Bajo Fernández, «Protección del honor y de la intimidad», p. 100. Sobre dicho concepto en el CP de 1995, véase, entre otros, Romeo Casabona, *Comentarios al Código penal*, Art. 197, marginal n.º 15.

8 Véase Bajo Fernández, «Protección del honor y de la intimidad», p. 101. Esta definición se encuentra expresamente en el Auto de la Audiencia Provincial de Burgos de 15 de junio de 2001: «*en dicho precepto —art. 197 del CP— el bien jurídico tutelado es la intimidad personal, es decir, el ámbito personal, donde cada uno, preservado del mundo exterior, encuentra las posibilidades de desarrollo y fomento de su personalidad*». Paredes Castañón ha señalado, asimismo que, en los delitos contra la intimidad, el Derecho penal interviene en la faceta del derecho al libre desarrollo de la personalidad, «para garantizar un determinado ámbito de exclusión de terceros, en el que esa libre personalidad del individuo pueda ser elaborada sin intromisiones ajenas *(privacy)*»; véase Paredes Castañón, *Enciclopedia Penal Básica*, p. 410.
Desde un punto de vista constitucional, véanse, por ejemplo, Serrano Alberca, *Comentarios a la Constitución*, 3.ª ed., p. 400: «la vida privada, la intimidad, aparece como un derecho a la soledad, a la reserva y al aislamiento reflejada en la máxima inglesa "*my home is my castle*" que se va configurando como "*the rigth of privacy*"»; Alzaga Villaamil, *Derecho Político Español según la Constitución de 1978, II, Derechos Fundamentales y Órganos del Estado,* 4.ª ed., p. 99 también indica que «la garantía del derecho a la intimidad implica el reconocimiento de un *espacio íntimo*, de cada persona —no jurídica— al que no tienen derecho a penetrar las demás».

n.º 173/2011, n.º 12/2012 y n.º 66/2022 y n.º 92/2023[9]—, que el derecho a la intimidad personal implica *«la existencia de un ámbito propio y reservado frente a la acción y el conocimiento de los demás, necesario, según las pautas de nuestra cultura, para mantener una calidad mínima de la vida humana»*. Esta concepción objetiva —y tradicional— de la intimidad personal se halla muy influenciada por la realidad social, económica y cultural en la que se desenvuelve[10], tal y como se reconoce en el art. 2.1 de la LO 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, en el que se establece que *«la protección civil del honor, de la intimidad y de la propia imagen quedará delimitada por las leyes y por los usos sociales atendiendo al ámbito que, por sus propios actos, mantenga cada persona reservado para sí misma o su familia»*.

Junto a esta concepción expuesta y, posiblemente, por influencia de la realidad social, económica y cultural se ha impuesto paulatinamente, por otra parte, un concepto subjetivo o formal del derecho a la intimidad que abarca asimismo el control sobre la publicidad de la información relativa a la persona y su familia, tal y como estima la STC n.º 134/1999 —además, entre otras, las SsTC n.º 115/2000, n.º 83/2002, n.º 185/2002; n.º 89/2006; n.º 236/2007; n.º 60/2010, n.º 173/2011, n.º 12/2012, n.º 176/2013 y n.º 64/2019—: *«el art. 18.1 CE no garantiza una "intimidad" determinada, sino el derecho a poseerla, a tener vida privada, disponiendo de un poder de control sobre la publicidad de la información relativa a la persona y su familia, con independencia del contenido de aquello que se desea mantener al abrigo del conocimiento público. Lo que el art. 18.1 garantiza es un derecho al secreto, a ser desconocido, a que los demás no sepan qué somos o lo que hacemos, vedando que terceros, sean particulares o poderes públicos, decidan cuáles sean los lindes de nuestra vida privada pudiendo cada persona reservarse un espacio resguardado de la curiosidad ajena, sea cual sea lo contenido en ese espacio»*[11]. La

9 Nuestro TS también ha conceptuado de este modo el derecho a la intimidad. Véanse entre otras, las SsTS n.º 574/2001 de 4 de abril; n.º 782/2007 de 3 de octubre; n.º 1148/2010 de 12 de diciembre; n.º 97/2015 de 24 febrero; n.º 360/2017 de 19 de mayo; n.º 1399/2024 de 23 de octubre; n.º 1144/2024 de 12 de diciembre; n.º 240/2025, de 13 de marzo.

10 Véanse DE LA MATA BARRANCO/BARINA UBIÑAS, «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», p. 14, quienes afirman que «el reconocimiento de un derecho no es ajeno a la realidad social, económica y cultural del tiempo en que pretende hacerse valer. Así ocurre también con el reconocimiento de la vida privada como interés a proteger».

11 El Tribunal Constitucional alemán en su sentencia de 15 de diciembre de 1983 sobre la Ley del Censo formuló la tesis de la autodeterminación informativa. En esta sentencia se resolvió una cuestión de inconstitucionalidad interpuesta contra algunos arts. relacionados con la regulación del tratamiento automatizado de determinados datos personales de la Ley del Censo. La mencionada tesis de la autodeterminación informativa se centra, según el Tribunal Constitucional alemán, en «la facultad del individuo, derivada de la idea de la autodeterminación, de decidir básicamente por sí mismo cuándo y dentro de qué límites procede revelar situaciones referentes a la propia vida». Se puede observar

STC n.º 170/2013, de 7 de octubre, se ha pronunciado sobre la delimitación de ese ámbito reservado: «la esfera de la intimidad personal está en relación con la acotación que de la misma realice su titular, habiendo reiterado este Tribunal que cada persona puede reservarse un espacio resguardado de la curiosidad ajena; en consecuencia, corresponde a cada persona acotar el ámbito de intimidad personal y familiar que reserva al conocimiento ajeno (STC n.º 241/2012, de 17 de diciembre), de tal manera que el consentimiento eficaz del sujeto particular permitirá la inmisión en su derecho a la intimidad (STC n.º 173/2011, de 7 de noviembre). Asimismo, también hemos declarado que la intimidad protegida por el art. 18.1 CE no se reduce a la que se desarrolla en un ámbito doméstico o privado; existen también otros ámbitos, en particular el relacionado con el trabajo o la profesión, en que se generan relaciones interpersonales, vínculos o actuaciones que pueden constituir manifestación de la vida privada (STC n.º 12/2012, de 30 de enero). Por ello expresamente hemos afirmado que el derecho a la intimidad es aplicable al ámbito de las relaciones laborales (SsTC n.º 98/2000, de 10 de abril; y n.º 186/2000, de 10 de julio)». Esta delimitación de la intimidad planteada por el TC con los elementos apuntados abre una importante brecha en la concepción objetiva y tradicional de la intimidad, que se basa en entender que las intromisiones ilegítimas a la intimidad se determinan en función de lo que la sociedad en cada momento entiende como tal, para pasar a otra con un componente subjetivo sin un contenido socialmente preestablecido[12] y a la que es consustancial el poder de control sobre la información que afecta a una persona y su familia. Como han concluido de la Mata Barranco y Barinas Ubiñas «la concepción primaria del derecho, surgido —en la plasmación de su reconocimiento— desde la preocupación por proteger la esfera privada frente a la injerencia de los medios de comunicación desde un punto de vista negativo, materializada en la concepción del «ser dejado en paz», ha dado pie a un amplio desarrollo del mismo que en lo que aquí se quiere destacar ha conducido a una visión positiva del mismo que permite reconocer no sólo un derecho a la no injerencia o intromisión arbitraria en la esfera privada sino el derecho a determinar y controlar esta esfera del modo deseado en aras al favorecimiento de un desarrollo definido desde la autodeterminación de lo que se quiere ser. Tenemos derecho a que nos dejen en paz, pero también tenemos derecho a callar, a que no se sepa de nosotros, a que nadie nos defina a partir de datos

un paralelismo entre el poder de control sobre la publicidad de la información relativa a la persona y su familia expuesto por el Tribunal Constitucional español, y la aludida autodeterminación informativa del Tribunal Constitucional alemán. Sobre el concepto de la autodeterminación informativa, véanse las exposiciones de Piñar Mañas, «Protección de datos: origen, situación actual y retos de futuro», pp. 81 y ss.; Nicolás Jiménez, *La protección jurídica de los datos genéticos de carácter personal*, pp. 164, 165, 166, 167, 168, 169 y 170.

12 Véanse al respecto las reflexiones de Pardo Falcón, *Comentarios a la Constitución española*, Tomo I, p. 520.

a que pueda accederse sobre nuestros gustos o modo de vivir, a que no se nos clasifique, a que no se nos perfile, a que no se nos interrogue a partir de códigos configurados sin que sepamos cómo»[13]. En síntesis, nos encontramos con un concepto de intimidad que abarca la vida privada de una persona en la que confluyen numerosos derechos vinculados a la propia personalidad, como el derecho a la imagen, el derecho a la inviolabilidad del domicilio y de las comunicaciones (postales, telegráficas y telefónicas), el derecho al honor, etc., derivados de la dignidad de la persona reconocida en el art. 10 de la CE[14]. Dicho concepto de intimidad que abarca la vida privada de una persona se define por dos elementos estrechamente vinculados: el ámbito personal donde cada uno, preservado del mundo exterior, encuentra las posibilidades de desarrollo y fomento de su personalidad, recordando la definición acuñada por Bajo Fernández[15], del que se deriva otro elemento adicional en el que prima ante todo el control de la persona sobre la información relativa a su vida privada[16].

II. Elementos comunes que integran la protección penal dispensada a la intimidad en el código penal de 1995

La protección de la intimidad con las características y elementos señalados se articula en nuestro ordenamiento jurídico a través de diversas normas jurídicas[17], pero, sobre todo, nos interesa la protección penal dispensada a la intimidad y la propia imagen en el Capítulo I del Título X del CP español de 1995, que tiene como rúbrica *«Del descubrimiento y revelación de secretos»*. El objeto de protección abarca ya desde sus inicios determinados elementos que han definido el derecho a la intimidad y la vida privada de las personas, de modo que en el mencionado capítulo I se protege penalmente en líneas generales, por una parte, el secreto de las comunicaciones como una manifesta-

13 Véanse, De la Mata Barranco/Barina Ubiñas, «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», p. 41; De la Mata Barranco, «Reflexiones sobre el bien jurídico a proteger en el delito de acceso informático ilícito (art. 197 bis CP). El concepto de privacidad informática y la tutela del buen funcionamiento de los sistemas de información y comunicación», p. 67.

14 Véase Espín, *Manual de Derecho Constitucional, I*, 1.ª ed., 221.

15 Véase Bajo Fernández, «Protección del honor y de la intimidad», p. 101.

16 Véase ampliamente Rueda Martín, *La nueva protección de la vida privada y de los sistemas de información en el Código penal*, pp. 34 y ss.

17 Véanse las exposiciones de estas normas jurídicas realizada por Espín, *Manual de Derecho Constitucional, I*, 1.ª ed., 224 y ss.; Romeo Casabona, *Comentarios al Código penal*, Art. 197, marginal n.º s. 85 y ss.; Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 494; Serrano Alberca, *Comentarios a la Constitución*, 3.ª ed., pp. 407 y ss., 414, 415 y ss., 435 y ss., 442 y ss.

ción de aquel ámbito propio y reservado que una persona configura frente a la acción y el conocimiento de los demás, heredero de una concepción objetiva —y tradicional— de la intimidad personal; y, por otra parte, se protege el poder de control sobre la publicidad de la información relativa a la persona y su familia, que subyace al concepto subjetivo o formal de dicha intimidad personal[18]. Como afirma Muñoz Conde, la nota en común de los delitos objeto de nuestro interés es que en ellos se protege la voluntad de una persona de que no sean conocidos determinados hechos que sólo son conocidos por ella o por un círculo reducido de personas, es decir, que pueden ser calificados de *secretos*, y también el derecho de la persona a controlar cualquier información o hecho que afecte a su vida privada y, por tanto, a su *intimidad*[19].

Si nos centramos en la evolución de la protección penal de la intimidad constataremos que la indicada protección penal ha sufrido con el transcurso de los años unos cambios que no son intrascendentes[20]. El objetivo fundamental de las reformas penales acaecidas en este ámbito hasta el CP de 1995 fue suplir la falta de adecuación de dicha protección con las exigencias contempladas el art. 18 de la CE[21]. Ya con el CP de 1995 se otorgó un tratamiento global y más coherente sistemáticamente a la tutela penal de la intimidad que en el CP de 1973[22], y una muestra es que los *«Delitos contra la intimidad, el derecho contra la propia imagen y la inviolabilidad del domicilio»* ganaron

18 Serrano Alberca indica que, a su juicio, el derecho a la intimidad tiene un tratamiento penal adecuado en el Código penal de 1995; véase Serrano Alberca, *Comentarios a la Constitución*, 3.ª ed., p. 411.

19 Véase Muñoz Conde, *Derecho penal, Parte Especial,* 25.ª ed., p. 302.

20 Véase en este sentido Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., p. 1430. Sin embargo, determinados autores han reducido la relevancia de estos cambios como, por ejemplo, Paredes Castañón, *Enciclopedia Jurídica Básica*, p. 412 al afirmar que son importantes pero no esenciales; Díaz-Maroto y Villarejo, «Los delitos contra la intimidad, la propia imagen y la inviolabilidad del domicilio», p. 1190.

21 Véase la completa exposición efectuada por Luzón Peña, «Protección penal de la intimidad y derecho a la información», pp. 42 y ss. sobre tales deficiencias de la protección penal de la intimidad que trataron de superarse con algunas reformas del CP de 1973, con el proyecto de reforma del CP de 1980 y con el anteproyecto de 1983.

22 Véanse en este sentido Carbonell Mateu/González Cussac, *Comentarios al CP de 1995, I,* pp. 991 y 992; González Cussac, *Derecho penal, Parte Especial*, 7.ª ed., p. 300; Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 494; Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., pp. 1430, 1431, 1432 y 1433; el mismo, Comentarios a la PE del Derecho penal, 10.ª ed., pp. 432 y 433; el mismo, «La reforma de los delitos contra la intimidad artículo 197 CP», pp. 460 y 461; Agustín Jorge Barreiro, *Comentarios al Código penal*, p. 563; Lozano Miralles, *Compendio de Derecho penal, PE, II*, p. 193.
Sobre la protección penal dispensada a la intimidad en el CP de 1973, véase Sierra López, «Los delitos de descubrimiento y revelación de secretos en el Código penal de 2015: artículos 197, 197 bis, 197 ter, 197 quáter, 197 quinques y 198», pp. 175 y ss.

«nomen iuris»[23]. Las reformas penales que se han producido posteriormente tras la entrada en vigor del CP de 1995 también han incorporado como delictivas nuevas modalidades de ataque sobre la intimidad, con el fin de adaptar la regulación penal a las nuevas exigencias de protección derivadas de la aparición de peligros que proceden de nuevos contextos sociales, económicos y culturales, sobre todo en el marco de las tecnologías de la información y de la comunicación (TIC). Ello ha obligado a modificar sustancialmente el objeto de protección de los delitos contra la intimidad y la propia imagen, ampliándolo o, incluso superándolo, tal y como se confirma en las reformas del CP operadas por la LO 5/2010, de 22 de junio, y por la LO 1/2015, de 30 de marzo[24]. Por una parte, la LO 5/2010, de 22 de junio, introdujo un nuevo delito en el apartado 3.º del art. 197: *«el que por cualquier medio o procedimiento y vulnerando las medidas de seguridad establecidas para impedirlo, acceda sin autorización a datos o programas informáticos contenidos en un sistema informático o en parte del mismo o se mantenga dentro del mismo en contra de la voluntad de quien tenga el legítimo derecho a excluirlo, será castigado con pena de prisión de seis meses a dos años»*. Por otra parte, la LO 1/2015, de 30 de marzo, también ha incorporado importantes novedades. En primer lugar, se ha previsto una nueva agravación en el art. 197.4 b) al castigarse con una pena de prisión de tres a cinco años los hechos descritos en los apartados 1.º y 2.º de este mismo artículo cuando *«se lleven a cabo mediante la utilización no autorizada de datos personales de la víctima»*. Asimismo, se añade como agravación ulterior que *«si los datos reservados se hubieran difundido, cedido o revelado a terceros, se impondrán las penas en su mitad superior»*. En segundo lugar, se contemplan nuevas conductas delictivas. En el art. 197.7 se establece que *«será castigado con una pena de prisión de tres meses a un año o multa de seis a doce meses el que, sin autorización de la persona afectada, difunda, revele o ceda a terceros imágenes o grabaciones audiovisuales de aquélla que hubiera obtenido con su anuencia en un domicilio o en cualquier otro lugar fuera del alcance de la mirada de terceros, cuando la divulgación menoscabe gravemente la intimidad personal de esa persona. Se impondrá la pena de multa de uno a tres meses a quien habiendo recibido las imágenes o grabaciones audiovisuales a las que se refiere el párrafo anterior las difunda, revele o ceda a terceros sin el consentimiento de la persona afectada. En los supuestos de los párrafos anteriores, la pena se impondrá en su mitad superior cuando los hechos hubieran sido cometidos por el cónyuge o por persona que esté o haya estado unida a él por análoga relación de afectividad, aun sin convivencia, la víctima fuera menor de edad o una persona con discapacidad necesitada de especial protección, o los hechos se*

23 Véase en este sentido Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., p. 1430; el mismo, *Comentarios a la PE del Derecho penal*, 10.ª ed., p. 433.

24 Sobre esta reforma del CP y su tramitación parlamentaria, véase la exposición efectuada por Castelló Nicas, «Delitos contra la intimidad, el derecho a la propia imagen y la inviolabilidad del domicilio, y delitos contra el honor», pp. 487 y ss.

hubieran cometido con una finalidad lucrativa»[25]. En tercer lugar, se ha trasladado al apartado 1.º del art. 197 bis del CP el delito consistente en acceder o facilitar a otro el acceso, vulnerando las medidas de seguridad establecidas para impedirlo, y sin estar debidamente autorizado al conjunto o una parte de un sistema de información o mantenerse dentro del mismo en contra de la voluntad de quien tenga el legítimo derecho a excluirlo. En su apartado 2.º se penalizan por primera vez ataques contra los sistemas de información que giran en torno a la interceptación de transmisiones no públicas de datos electrónicos cuando no se trata de una comunicación personal. Finalmente, en el art. 197 ter del CP se tipifican determinados actos preparatorios que faciliten la comisión de los delitos recogidos en los apartados 1.º y 2.º del art. 197 o del art. 197 bis del CP. Nos encontramos, en consecuencia, ante una modificación muy significativa de los límites de la protección penal de la intimidad personal en el CP, y la primera pregunta que debemos plantearnos es si era necesario desde un punto de vista político criminal ampliar la protección penal de la intimidad personal centrada, inicialmente, en el secreto de las comunicaciones como una manifestación de aquel ámbito propio y reservado que una persona configura frente a la acción y el conocimiento de los demás, heredero de una concepción objetiva —y tradicional— de la intimidad personal, y en el poder de control sobre la publicidad de la información relativa a la persona y su familia, que subyace al concepto subjetivo o formal de dicha intimidad personal.

Para un importante sector de nuestra doctrina la respuesta a la pregunta planteada es negativa, de modo que estas reformas no merecen una valoración positiva[26]. Por ejemplo, Morales Prats concluye que la introducción del delito de acceso ilícito a sistemas informáticos se ha realizado «sin reparar en criterio sistemático alguno, y con desprecio incluso del bien jurídico protegido por el que optó el legislador de 1995... Con posterioridad, la Reforma de 2015 (LO 1/2015) ha ahondado en el camino abierto por la Reforma de 2010. Al transponer las exigencias de la Directiva de la UE de 2013 sobre los ataques a los sistemas de información (arts. 197 bis y 197 ter CP), se verifica una confusión de planos pues lo que realmente se buscaba era la protección de la seguridad de los sistemas; su inserción en los delitos contra la intimidad provoca problemas, incluso si se utiliza un concepto amplio de *privacy*, el intrusismo informático, que pasa a incriminarse, ya no se verifica sobre los datos

25 Este precepto fue modificado por la disposición final 4.17 de la LO 10/2022, de 6 de septiembre, de garantía integral de la libertad sexual.

26 Véanse, por ejemplo, Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., p. 298; Bolea Bardon, *Comentarios al Código penal*, p. 732; Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., pp. 494, 498 y 499; Morales Prats, «La reforma de los delitos contra la intimidad artículo 197 CP», pp. 459 y ss.; Alonso de Escamilla, *Delitos,* 3.ª ed., p. 223 estima una decisión político criminal desacertada la introducción de estos nuevos tipos penales, *«pues el legislador encamina el Derecho penal en esta materia a una intervención excesiva»*, con cursivas en el original.

o los programas informáticos en que se hallan los mismos (art. 197 bis); se trata de la vulneración de las medidas de seguridad de los propios sistemas. El encaje de estas conductas entre los delitos de la *privacy* genera problemas interpretativos... La opción del legislador de 2015 culmina con la introducción de un nuevo delito de revelación ilícita de imágenes obtenidas con consentimiento del titular, verificadas en domicilio o fuera de la mirada de terceros. Esta decisión pone en cuestión los postulados de partida del modelo de incriminación por el que optó el CP de 1995 y, además, supone una cercenación del principio de intervención mínima, al convertir en confidentes necesarios (obligados penales al sigilo), artificialmente, a ciudadanos, con desprecio de otros instrumentos jurídicos de resolución de problemas del problema...»[27]. También Bolea Bardon estima que el hecho de que la LO 1/1982, de protección civil del derecho al honor, a la intimidad personal y familiar, y a la propia imagen «recoja como intromisión ilegítima a la intimidad la captación, reproducción o publicación de fotografías o imágenes que afecten a la vida privada, siempre que falte el consentimiento de la persona afectada, lleva a cuestionar la necesidad político-criminal de crear nuevos delitos como el introducido por LO 1/2015 en el art. 197.7»[28]. O Romeo Casabona en relación con la agravación introducida en el nuevo delito que tipifica la difusión no autorizada de imágenes o grabaciones audiovisuales en el art. 197.7, segundo párrafo del CP[29], concluye que «de nuevo aparece aquí un tipo agravado, con un sinfín de variantes comisivas, en la línea de escasa continencia intervencionista de la que ha hecho gala el legislador, probablemente rebasando el principio de mínima intervención y de proporcionalidad»[30].

27 Véase Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., pp. 1433 y 1434. En relación con la introdución del nuevo apartado b) del n.º 4 del art. 197 del CP afirma también que «la ausencia de criterios político-criminales certeros es notable»; véase el mismo, ob. cit, p. 1463.

28 Véase Bolea Bardon, *Comentarios al Código penal. Reformas LLOO 1/2023, 3/2023 y 4/2023,* 2.ª ed., p. 955. Estima, asimismo, «la innecesaridad de la regulación penal de la conducta» Durán seco, *El delito de difusión no consentida de captaciones íntimas de la imagen consentida*s, p. 84.

29 El art. 197.7, segundo párrafo según la LO 1/2015, de 30 de marzo, y tercer párrafo de acuerdo con la reforma operada por la LO 10/2022, de 6 de septiembre, dispone que *«en los supuestos de los párrafos anteriores, la pena se impondrá en su mitad superior cuando los hechos hubieran sido cometidos por el cónyuge o por persona que esté o haya estado unida a él por análoga relación de afectividad, aun sin convivencia, la víctima fuera menor de edad o una persona con discapacidad necesitada de especial protección, o los hechos se hubieran cometido con una finalidad lucrativa»*.

30 Véase Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., p. 314.

III. Sobre la necesidad político-criminal de ampliar el objeto de protección en los delitos contra la intimidad y la propia imagen

Sin embargo, las novedosas figuras delictivas incorporadas en las últimas reformas del CP dentro del Título X y que se centran en los delitos contra la intimidad y el derecho a la propia imagen, a mi juicio, deben ser valoradas positivamente desde un punto de vista político-criminal, por la innegable importancia que han adquirido las nuevas tecnologías de la información y de la comunicación (TIC) en el funcionamiento del sistema social en la actualidad, con la utilización de redes y sistemas de tratamiento de la información, como medio de crecimiento económico y desarrollo social[31]. No puede ponerse en duda que el contexto en el que se encuentra involucrada la intimidad con una expansión de las telecomunicaciones con la informática o con el desarrollo de tecnologías en diversos ámbitos, condiciona la protección de aquélla por nuevas amenazas[32], que comprometen no sólo la posibilidad de reservar a una persona un ámbito propio frente a la acción y el conocimiento de los demás, sino también el poder de control sobre la publicidad de la información relativa a la persona y su familia. Las TIC se han extendido y se han enraizado en nuestras modernas sociedades de tal manera que han conformado unas estructuras y unas relaciones comerciales, administrativas, laborales, formativas, sociales, etc., que trascienden el ámbito estrictamente económico y que son radicalmente nuevas[33]. La generalización de las TIC ha permitido la aparición de nuevos escenarios como, por ejemplo, el comercio electrónico *(e-commerce)*, el acercamiento de los bancos a los clientes *(home-banking),* la gestión electrónica de los recursos de las empresas *(e-management)*, la

31 Pionero en subrayar la relevancia de las tecnologías de la información y comunicación en el funcionamiento de nuestra vida social actual fue Romeo Casabona, *Poder informático y seguridad jurídica. La función tutelar del derecho penal ante las Nuevas Tecnologías de la Información,* pp. 19 y ss.; el mismo, «Derecho penal y libertades de expresión y comunicación en Internet», pp. 299 y ss. Véanse, además, Rueda Martín, «Los ataques contra los sistemas informáticos: conductas de hacking. Cuestiones político-criminales», pp. 371, 372 y 373; Miró Llinares, *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio,* pp. 25 y ss.; de la Mata Barranco, «Reflexiones sobre el bien jurídico a proteger en el delito de acceso informático ilícito (art. 197 bis CP). El concepto de privacidad informática y la tutela del buen funcionamiento de los sistemas de información y comunicación», pp. 63 y ss.

32 Véanse sobre estas cuestiones Romeo Casabona, *Comentarios al Código penal*, Art. 197, marginal n.º s. 18 y ss.; Huerta Tocildo/Andrés Domínguez, «Intimidad e informática», pp. 12 y ss.; Valls Prieto, «Nuevas formas de protección penal de la intimidad», pp. 175 y ss.

33 Véanse Lucena Cid, «El concepto de la intimidad en los nuevos contextos tecnológicos», pp. 33 y ss.; Ribagorna Garnacho, «Seguridad de las tecnologías de la información», p. 310. Estas estructuras y relaciones se pueden mantener mediante el ordenador e internet, dispositivos móviles, etc. En cualquier caso en un futuro más o menos inmediato pueden aparecer otros canales que aún no están disponibles hoy en día.

gestión doméstica *(domótica)*[34], o la denominada transformación digital de la Administración General del Estado y sus organismos públicos. Las TIC han posibilitado asimismo una nueva forma de relacionarse entre las personas desconocida hasta hace unos pocos años, a través de determinadas redes sociales como, por ejemplo, *Facebook*, *Twitter*, *Tuenti* o *Youtube* que se componen de grupos de personas conectadas entre sí porque tienen intereses comunes, con independencia de que se conozcan personalmente[35]. En todos estos escenarios novedosos se involucran bienes jurídicos tales como, por ejemplo, el patrimonio, la intimidad personal y familiar, los derechos de autor o la capacidad competitiva de la empresa, de manera que los sistemas de información y comunicación permiten también su desarrollo en las modernas sociedades[36]. Nuestra organización social (la Administración pública, el sistema financiero, el sistema sanitario, las infraestructuras básicas de transporte, las empresas, los particulares, etc.) ha pasado a depender de forma extraordinaria de unos sistemas y redes informáticos, de los que surgen unas amenazas para un amplio conjunto de bienes jurídicos por la realización de determinados comportamientos[37].

La expansión de las comunicaciones a través de las TIC ha supuesto una gran vulnerabilidad en el control de la publicidad de la información relativa a la persona y su familia lo que ha requerido, por una parte, una detenida reflexión sobre la oportunidad de establecer una protección más amplia y específica de las comunicaciones de forma que se garantice, incluso penalmente, que

34 Véase Salom Clotet, «Delito informático y su investigación», pp. 93 y ss.

35 Sobre el concepto de red social *on-line* y sus características, véanse Soler Presas, «Am I in Facebook? Sobre la responsabilidad civil de las redes sociales *on-line* por la lesión de los derechos de la personalidad, en particular por usos no consentidos de la imagen de un sujeto», pp. 3 y ss.; Moreno Navarrete, «Aspectos jurídico privados de las tecnologías Web 2.0 y su repercusión en el derecho a la intimidad», pp. 339 y ss; Baym, «Social networks 2.0», p. 384 y ss.; Boyd/Ellison, «Social networks sites: definition, history, and scholarship», pp. 1 y ss.; Vázquez de Castro, «Protección de datos personales, redes sociales y menores», pp. 28-31. Sobre redes sociales, desarrollo de la personalidad en el ciberespacio y nuevos cibercrímenes, véase Miró Llinares, *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio,* pp. 122 y ss. En la Exposición de Motivos de la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos (BOE de 23 de junio de 2007), ya derogada, se indicaba que «las tecnologías de la información y las comunicaciones están afectando también muy profundamente a la forma e incluso al contenido de las relaciones de los seres humanos entre sí y de las sociedades en que se integran».

36 Véase, por ejemplo, Rueda Martín, «Los ataques contra los sistemas informáticos: conductas de hacking. Cuestiones político-criminales», pp. 369-374.

37 Véase una exposición de estos comportamientos en Romeo Casabona, «De los delitos informáticos al cibercrimen. Una aproximación conceptual y político-criminal», pp. 1 y ss., y en Miró Llinares, *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio,* pp. 47 y ss., especialmente pp. 84 y ss.

aquéllas puedan realizarse de manera pacífica[38]. Por otra parte, la mencionada expansión de las telecomunicaciones ha obligado a adoptar también unas medidas extraordinarias para controlar la difusión de los contenidos de las comunicaciones o el tratamiento de datos personales en el marco de las TIC. Podemos afirmar que la reforma del CP operada por la LO 1/2015 ha procurado superar las deficiencias que nuestra regulación penal presentaba en relación con la intimidad que abarca la vida privada de una persona, ante el avance de las TIC en nuestro desarrollo social y económico. En efecto, en el CP se ha incorporado la punición de una serie de comportamientos como la utilización no autorizada de datos personales de la víctima obtenidos para realizar las conductas de los apartados 1.º y 2.º del art. 197 del CP —apartado 4 b) de dicho precepto—, y la difusión, revelación o cesión no autorizada de grabaciones o imágenes íntimas obtenidas con el consentimiento de la víctima —apartado 7.º del art. 197 del CP—, que atentan de un modo directo o indirecto contra el control de la publicidad o difusión de la información relativa a la persona y su familia o del tratamiento de datos personales en el marco de las TIC. La necesidad de establecer unas medidas extraordinarias para verificar la difusión legítima de los contenidos de las comunicaciones o el tratamiento de datos personales en el marco de las TIC, se encuentra también en la criminalización de los comportamientos aludidos. La expansión de las telecomunicaciones con la informática o con el desarrollo de tecnologías en diversos ámbitos conlleva un conjunto de amenazas, entre las que destacan por su gravedad aquellas que giran en torno a la intimidad o el derecho a la propia imagen, dada la facilidad de crear y recrear datos e imágenes a través del ciberespacio y por la rapidez en su difusión a cualquier parte del mundo. Así podemos recordar la noticia comentada por el diario digital *El País* el 22 de septiembre de 2012 bajo el título «*Sexting, sexo inseguro*», cuya persona afectada es mayor de edad[39]: «El último caso sonado ha sido el de la concejal de Los Yébenes, ..., cuyo vídeo sexual —difundido por un amigo— ha dado la vuelta al mundo y puede verse en páginas web extranjeras. «Esa foto te va a perseguir el resto de tu vida. Seguramente quede en Internet para los siglos», afirma Pérez San-José. El experto en nuevas tecnologías apunta la dificultad de borrar una imagen de la Red, que puede acabar alojada en un servidor en cualquier país. Eso sin contar las copias en sus dispositivos que pueden guardar millones de usuarios, aunque se haya retirado la información inicial»[40]. Con el fin de ofrecer una respuesta penal adecuada a este tipo de

38 Véase Romeo Casabona, *Comentarios al Código penal*, Art. 197, marginal n.º 15.

39 Véase sobre este caso Rueda Martín, «La relevancia penal del consentimiento del menor de edad en relación con los delitos contra la intimidad y la propia imagen. (Especial consideración a la disponibilidad de la propia imagen del menor de edad en el ciberespacio)», pp. 4, 5 y 6.

40 Conductas como las descritas en las que un menor se fotografía o se graba por cualquier medio de forma desnuda o semidesnuda y su transmisión a otras personas —sobre todo a menores de edad— mediante las TIC, reciben actualmente el nombre de *Sexting.*

comportamientos, que suponen una grave intromisión en la intimidad personal, se introdujo en la LO 1/2015, de 30 de marzo, una nueva figura delictiva en el art. 197.7 del CP que, como se señala en la Exposición de Motivos de dicha LO, se centra en supuestos «en los que las imágenes o grabaciones de otra persona se obtienen con su consentimiento, pero son luego divulgados contra su voluntad, cuando la imagen o grabación se haya producido en un ámbito personal y su difusión, sin el consentimiento de la persona afectada, lesione gravemente su intimidad». Las características de este escenario agravan cualquier lesión del bien jurídico aludido por los siguientes factores. Por un lado, por las dimensiones colosales del número de usuarios, las frecuencias de acceso y uso por parte de los mismos que refleja el carácter asequible del ciberespacio, así como la libre circulación y navegación, tanto para emitir, transferir y difundir información como para acceder a ella por medio de la red, de forma anónima y cómoda[41]. El control sobre la publicidad de la información relativa a la persona desaparece absolutamente en este contexto, sin que sea posible establecer órganos o instituciones de control de la información que circula por el ciberespacio[42]. Esta falta de control produce una la lesión del bien jurídico intimidad personal y familiar porque el ciberespacio multiplica exponencialmente las posibilidades de emisión, transferencia y difusión de la información de forma extraordinariamente rápida a cualquier parte del mundo. Una vez que se ha puesto en circulación una imagen o una grabación a través del ciberespacio es prácticamente imposible frenar su acceso y difusión. Resulta más grave, entonces, la lesión del bien jurídico indicado si se utilizan las TIC, ya que el menoscabo de la intimidad personal tiene mayor permanencia en el tiempo y llega a todo el mundo, por lo que su dañosidad es más intensa[43]. Por la incidencia de las TIC en nuestra forma de comunicarnos actualmente el legislador se ha replanteado los límites de la protección penal de la intimidad personal en el CP, dada la vulnerabilidad en la que se encuentra el bien jurídico de la intimidad personal y la gravedad de

Sobre las características de estos comportamientos, véase el estudio de Agustina, «¿Menores infractores o víctimas de pornografía infantil? Respuestas legales e hipótesis criminológicas ante el *Sexting*», pp. 4 y ss. Sin embargo, esta denominación parece que se aplica ahora también a los mismos comportamientos realizados por adultos.

41 Véanse expresamente Romeo Casabona, «De los delitos informáticos al cibercrimen. Una aproximación conceptual y político-criminal», p. 3 y Livingstone/Millwood, «Harmful to Children. Drawing Conclusions from Empirical Research on Media Effects», p. 26.

42 Véanse Romeo Casabona, «De los delitos informáticos al cibercrimen. Una aproximación conceptual y político-criminal», p. 3; Hernández Fernández/Ramón Fernández, «El derecho a la propia imagen de los menores en los medios de comunicación y redes sociales», p. 31.

43 Véanse Valeije Álvarez, «Intimidad y difusión de imágenes sin consentimiento», pp. 1890 y 1891; Tomás y Valiente Lanuza, *Comentarios prácticos al Código penal, art. 197*, p. 670. Miró Llinares, *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio*, p. 124 apunta también que el catálogo de comportamientos criminales en la Red que pueden afectar a las esferas más personales del individuo aumenta cuantitativamente y, en lo cualitativo, su dañosidad es significativamente superior.

su lesión en este contexto. Los motivos que se han indicado explican y fundamentan la introducción del novedoso delito autónomo contra la intimidad y el derecho a la propia imagen en el art. 197.7 del CP[44].

Por otro lado, sin abandonar el marco de las TIC se han ampliado los límites sobre el objeto de protección de los delitos de descubrimiento y revelación de secretos con la incorporación de figuras delictivas adicionales, al transponerse tanto la Decisión Marco 2005/222/JAI, de 24 de febrero de 2005, relativa a los ataques contra los sistemas de información, como la Directiva 2013/40/UE del Parlamento Europeo y del Consejo de la UE, de 12 de agosto de 2013, relativa a los ataques contra los sistemas de información, que penalizan determinados ataques contra los sistemas de información y la interceptación de datos electrónicos cuando no se trata de una comunicación personal. En concreto en la Exposición de Motivos de la LO 1/2015 se concluye que «de acuerdo con el planteamiento recogido en la Directiva, se introduce una separación nítida entre los supuestos de revelación de datos que afectan directamente a la intimidad personal, y el acceso a otros datos o informaciones que pueden afectar a la privacidad pero que no están referidos directamente a la intimidad personal: no es lo mismo el acceso al listado personal de contactos, que recabar datos relativos a la versión de software empleado o a la situación de los puertos de entrada a un sistema. Por ello, se opta por una tipificación separada y diferenciada del mero acceso a los sistemas informáticos. Con el mismo planteamiento, y de acuerdo con las exigencias de la Directiva, se incluye la tipificación de la interceptación de transmisiones entre sistemas, cuando no se trata de transmisiones personales: la interceptación de comunicaciones personales ya estaba tipificada en el Código Penal; ahora se trata de tipificar las transmisiones automáticas —no personales— entre equipos. Se tipifica la facilitación o la producción de programas informáticos o equipos específicamente diseñados o adaptados para la comisión de estos delitos». La introducción de diversos tipos delictivos en los arts. 197 bis[45] y ter[46] ha generado una crítica muy negativa por los

44 Señala también el carácter autónomo de este nuevo delito Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., pp. 313 y ss.

45 En este precepto se dispone que *«1. El que por cualquier medio o procedimiento, vulnerando las medidas de seguridad establecidas para impedirlo, y sin estar debidamente autorizado, acceda o facilite a otro el acceso al conjunto o una parte de un sistema de información o se mantenga en él en contra de la voluntad de quien tenga el legítimo derecho a excluirlo, será castigado con pena de prisión de seis meses a dos años. 2. El que mediante la utilización de artificios o instrumentos técnicos, y sin estar debidamente autorizado, intercepte transmisiones no públicas de datos informáticos que se produzcan desde, hacia o dentro de un sistema de información, incluidas las emisiones electromagnéticas de los mismos, será castigado con una pena de prisión de tres meses a dos años o multa de tres a doce meses»*.

46 En el art. 197 ter del CP se penalizan unos actos preparatorios específicos aplicables a los arts. 197.1 y 2 y 197 bis: *«Será castigado con una pena de prisión de seis meses a dos años o multa de tres a dieciocho meses el que, sin estar debidamente autorizado,*

problemas interpretativos que ocasionan y por vulnerar el principio básico político-criminal relativo a la intervención mínima del Derecho penal[47]. Por ejemplo, Anarte Borrallo/Doval País han señalado, en relación con el art. 197 bis 1 que «además de por razones de legitimidad (teniendo en cuenta el mero peligro que comportan semejantes hechos *desde la perspectiva de la intimidad*), se ha cuestionado tanto su necesidad técnica (ante las posibilidades que podían ofrecer modalidades del art. 197.1 para el castigo de estos hechos), como su idoneidad político-criminal (considerando las escasas posibilidades preventivas de la estrategia penal frente a estas acciones)», concluyendo que «esta disposición es, más bien, una consecuencia más de la imparable —pero inútil y contraproducente— escalada del Derecho penal hacia el «riesgo cero», que la reforma de 2015 confirma también en esta materia»[48]. Desde mi punto de vista, sin embargo, existen argumentos que justifican la necesidad de revisar los límites en la protección penal de la intimidad personal en el CP también en este ámbito centrado en los sistemas de información, lo que debe ser valorado como un progreso o un avance acorde a la adaptación del Derecho penal al desarrollo social y tecnológico. Como se ha apuntado anteriormente, nuestra organización social (la Administración pública, el sistema financiero, el sistema sanitario, las infraestructuras básicas de transporte, las empresas, los particulares, etc.) ha pasado a depender de forma extraordinaria de unos sistemas y redes informáticos, de los que surgen unas amenazas para un amplio conjunto de bienes jurídicos por la realización de determinados comportamientos. Para controlar estas amenazas se criminalizan tales comportamientos a través de un conjunto de figuras delictivas recogidas en los arts. 197 bis y 197 ter del CP, lo que ha conducido, además, a constatar la existencia de un bien jurídico diferente a la intimidad personal, pero estrechamente vinculados. Este nuevo bien jurídico definido como la confidencialidad, integridad y disponibilidad de los sistemas informáticos[49], por una parte, ejerce una función de barrera de contención

produzca, adquiera para su uso, importe o, de cualquier modo, facilite a terceros, con la intención de facilitar la comisión de alguno de los delitos a que se refieren los apartados 1 y 2 del artículo 197 o el artículo 197 bis: a) un programa informático, concebido o adaptado principalmente para cometer dichos delitos; o b) una contraseña de ordenador, un código de acceso o datos similares que permitan acceder a la totalidad o a una parte de un sistema de información».

47 Véase, a mero título de ejemplo, Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197)*, 7.ª ed., p. 1434.

48 Véanse Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 516.

49 Véase sobre este bien jurídico Rueda Martín, «Los ataques contra los sistemas informáticos: conductas de hacking. Cuestiones político-criminales», pp. 362 y ss, asumiendo y desarrollando la formulación planteada por Rodríguez Mourullo/Alonso Gallo/Lascuraín Sánchez, «Derecho penal e internet», p. 260, 261, 262 y 269. Véanse también Sieber, «Legal Aspects of Computer-Related Crime in the Information Society —Comcrime-Study—, prepared for the European Commission by Frof. Dr. Ulrich Sieber, versión de enero de 1998, p. 42, se refiere también a la integridad del sistema informático

de riesgos respecto de aquellos bienes jurídicos que se involucran en los sistemas y redes informáticos, como sucede con la intimidad personal y familiar; y, por otra parte, establece las condiciones necesarias y vinculantes para un correcto funcionamiento del sistema social en el marco de las TIC, lo que legitima la intervención del Derecho penal en su protección, así como en la represión de aquellos comportamientos que lo lesionen. El Derecho cuando protege las TIC reconoce su valor social positivo como necesario y vinculante para un correcto funcionamiento del sistema social. La necesidad de proteger los sistemas de comunicación e información por dicho valor social se ha reconocido también en la STC alemán de 27 de febrero de 2008 que establece en su parágrafo n.º 181 que: «c) De la importancia de la utilización de los sistemas tecnológicos de información para el desarrollo de la personalidad y de los peligros para la misma unidos a dicha utilización, resulta una importante necesidad de protección desde el punto de vista de los derechos fundamentales. Los particulares exigen que el Estado atienda las expectativas de confidencialidad e integridad de tales sistemas justificadas en el marco del libre desarrollo de la personalidad»[50]. Y, posteriormente, en el parágrafo 203 afirma que: «Por otra parte, el derecho fundamental a la garantía de la integridad y confidencialidad de los sistemas tecnológicos de información hay que esgrimirlo cuando la facultad de injerencia afecte a un sistema, que por sí solo o en conexión con redes tecnológicas, en un contorno cerrado o amplio, pueda contener datos del afectado referentes a su persona, de modo que el acceso al sistema permitiría formarse una idea sobre aspectos esenciales de la vida de una persona o incluso obtener una imagen representativa de su personalidad»[51].

---

que resulta vulnerada con estas conductas de hacking. Carrasco Andrino, *Tratado de Derecho Penal. Parte Especial (I). Delitos contras las personas, 4.ª ed.,* pp. 1700 y 1701; Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197)*, 7.ª ed., p. 1478. De forma similar también González Cussac, *Derecho penal, Parte Especial*, 7.ª ed., p. 326; Castelló Nicas, «Delitos contra la intimidad, el derecho a la propia imagen y la inviolabilidad del domicilio, y delitos contra el honor», p. 505.

50 Véase BVerfG, Urteil des Ersten Senats vom 27. Februar 2008 - 1 BvR 370/07 - Rn. (1-333), disponible en la dirección http://www.bverfg.de/e/rs20080227_1bvr037007.html. La redacción original del parágrafo traducido es la siguiente: «c) Aus der Bedeutung der Nutzung informationstechnischer Systeme für die Persönlichkeitsentfaltung und aus den Persönlichkeitsgefährdungen, die mit dieser Nutzung verbunden sind, folgt ein grundrechtlich erhebliches Schutzbedürfnis. Der Einzelne ist darauf angewiesen, dass der Staat die mit Blick auf die ungehinderte Persönlichkeitsentfaltung berechtigten Erwartungen an die Integrität und Vertraulichkeit derartiger Systeme achtet».

51 La redacción original del parágrafo traducido es la siguiente: «Das Grundrecht auf Gewährleistung der Integrität und Vertraulichkeit informationstechnischer Systeme ist hingegen anzuwenden, wenn die Eingriffsermächtigung Systeme erfasst, die allein oder in ihren technischen Vernetzungen personenbezogene Daten des Betroffenen in einem Umfang und in einer Vielfalt enthalten können, dass ein Zugriff auf das System es ermöglicht, einen Einblick in wesentliche Teile der Lebensgestaltung einer Person zu gewinnen oder gar ein aussagekräftiges Bild der Persönlichkeit zu erhalten».

Una vez fundamentada la existencia y la autonomía del bien jurídico relativo a la confidencialidad, integridad y disponibilidad de los sistemas informáticos, debemos exponer argumentos que justifiquen la necesidad político-criminal de criminalizar aquellas conductas que supongan una lesión del bien jurídico señalado. Por una parte, elevar a la categoría de delito en nuestro CP esta clase comportamientos que suponen un ataque contra los sistemas de comunicación o información, supone una obligada armonización penal en este ámbito de nuestra legislación con lo dispuesto en otros estados de la Unión Europea, en consonancia con lo establecido tanto en la Directiva 2013/40/UE del Parlamento Europeo y del Consejo de la Unión Europea, de 12 de agosto de 2013, relativa a los ataques contra los sistemas de información, que derogó la Decisión Marco 2005/222/JAI, como en el Convenio del Consejo de Europa sobre Cibercriminalidad de 23 de noviembre de 2001. Dicha armonización es necesaria además porque en esta clase de ataques podemos encontrar una nota que le añade un especial grado de peligrosidad: su conexión internacional o transfronteriza, de modo que sus actuaciones pueden ir más allá de un ámbito geográfico concreto, y resulta sorprendente que en algún territorio un acceso ilegal a sistemas informáticos con independencia de la finalidad que haya tenido quien accede, resulte impune[52]. Por otra parte, hay que tener en cuenta que, como afirma Romeo Casabona, el ciberespacio presenta unos perfiles de gran interés para el Derecho penal entre los que destaca la potencialidad multiplicadora de las acciones ilícitas y de sus efectos lesivos para los bienes jurídicos afectados[53]. Esta característica se puede apreciar con especial intensidad en las conductas de acceso ilícito a sistemas informáticos, que como ha puesto de relieve un sector doctrinal tienen un efecto criminógeno[54]. Por ello y con carácter general, el ciberespacio se presenta en las sociedades modernas como una de las posibles fuentes de riesgos necesitados de control, y dada la gravedad de sus repercusiones sobre diferentes bienes jurídicos se legitima la intervención del Derecho penal[55].

52 Véase sobre esta necesidad de la Mata Barranco, *Derecho penal europeo y legislación española: las reformas del Código penal. Actualizado a la reforma penal 2015*, pp. 71 y ss.

53 Véase Romeo Casabona, «De los delitos informáticos al cibercrimen. Una aproximación conceptual y político-criminal», p. 4. También destacan, con carácter general, el factor criminógeno del procesamiento electrónico de datos, Sieber, *Computerkriminalität*, 1.ª ed., pp. 158 y ss.; Mata y Martín, *Delincuencia informática y Derecho penal,* p. 17, 24 y ss.; Morón Lerma, *Internet y Derecho penal: Hacking y otras conductas ilícitas en la red*, p. 75.

54 Véase respecto de las conductas de hacking, Gutiérrez Francés, «El intrusismo informático (Hacking): ¿Represión penal autónoma?», pp. 1179 y ss.; la misma, «Notas sobre la delincuencia informática: atentados contra la "información" como valor económico de empresa», p. 206.

55 Véase ampliamente Rueda Martín, *La nueva protección de la vida privada y de los sistemas de información en el Código penal*, pp. 47 y ss.

En suma, podemos concluir en relación con la decisión político-criminal adoptada por nuestro legislador en las últimas reformas del CP que la tutela de la intimidad ha adaptado su regulación a los avances de las modernas tecnologías[56].

IV. Algunas observaciones críticas a la regulación penal en los delitos contra la intimidad y la propia imagen

No obstante, la regulación penal de la intimidad personal y familiar en el CP de 1995 adolece de deficiencias que deberían revisarse para que fuese más satisfactoria. Se apuntan las siguientes. En primer lugar, se evidencia en el legislador una grave carencia —o prácticamente inexistencia— de criterios sistemáticos, que permitan incardinar todas las figuras delictivas agrupadas en una rúbrica más acorde con su objeto de protección[57]. La rúbrica *«del descubrimiento y revelación de secretos»* resulta absolutamente inadecuada, porque los delitos que engloba penalizan comportamientos que van más allá del simple descubrimiento y revelación de secretos. En relación con la criminalización de los ataques contra los sistemas de información el legislador puede operar de dos maneras distintas[58]. Por un lado, mediante tipos específicos o «tipos de equivalencia» que contemplen la incriminación de determinados ataques contra los sistemas de información y comunicación en cada figura de delito para suplir las posibles lagunas de punibilidad que pudieran evidenciarse. De esta forma se podría tipificar, por ejemplo, el comportamiento consistente en acceder de manera ilícita a sistemas informáticos en relación con las conductas de descubrimiento y revelación de secretos (art. 197 CP), de descubrimiento, revelación o cesión de datos reservados de personas jurídicas (art. 200 CP), estafas (art. 249.2 a) CP), de utilización de cualquier equipo terminal de telecomunicación (art. 256 CP), daños infor-

56 Véanse Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., p. 1432; Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 494; Agustín Jorge Barreiro, *Comentarios al Código penal*, p. 563; Díaz-Maroto y Villarejo, «Los delitos contra la intimidad, la propia imagen y la inviolabilidad del domicilio», p. 1190. Queralt Jiménez, *Derecho penal español, Parte Especial*, 7.ª ed., p. 294 señala también respecto de la reforma del CP del 2015 que «son en general cambios que abundan en la protección de esta parcela de la intimidad que tan expuesta queda a las agresiones provenientes de las TIC's».

57 Véanse Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., p. 1433; Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., p. 298; Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 511; Colás Turégano, «Nuevas conductas delictivas contra la intimidad (arts. 197; 197 bis; 197 ter)», p. 663.

58 Véase, con carácter general, sobre la tipificación de conductas relacionadas con agresiones en conexión con sistemas informáticos, Romeo Casabona, «Tendencias actuales sobre las formas de protección jurídica ante las nuevas tecnologías», pp. 180 y 181.

máticos (arts. 264 y 264 bis CP), delitos relativos a la propiedad intelectual (arts. 270 y ss. CP), de descubrimiento de secretos de empresa (art. 278 CP) o de descubrimiento y revelación de secretos e informaciones relativas a la defensa nacional (art. 598 CP). Esta opción, sin embargo, plantea inconvenientes como el excesivo casuismo que conllevaría y la falta de adaptación a la rapidez de los avances tecnológicos que no se previeran en un determinado momento y que impediría aplicar este delito consistente en acceder de manera ilícita a sistemas informáticos a nuevos ámbitos[59]. Por otro lado, el sistema de criminalización de estas conductas se puede llevar a cabo a través del establecimiento de nuevos tipos penales genéricos que tipificaran como delito determinados ataques contra los sistemas de información y comunicación. Esta opción parece que resulta más idónea al adaptarse a las nuevas formas de criminalidad que pudieran surgir en el futuro, si bien es cierto que implica buscar y encontrar un adecuado lugar sistemático en nuestro CP. Desde un punto de vista sistemático decidirse por esta segunda opción supone plantearse, incluso, la necesidad de formular un nuevo título en el CP dedicado a la delincuencia informática[60], con una rúbrica, por ejemplo, sobre los *«Delitos contra los sistemas de información»*, en el que se contemple la protección penal de los sistemas informáticos y que aglutinaría aquellas conductas que lesionan el bien jurídico relativo a la confidencialidad, integridad y disponibilidad de los sistemas informáticos[61]. ¿Qué sistema de criminalización ha asumido nuestro legislador? La penalización de determinados ataques contra los sistemas de información se ha introducido dentro del Título X del CP sobre los *«Delitos contra la intimidad, el derecho a la propia imagen y la inviolabilidad del domicilio»*, por lo que el legislador los ha vinculado a la protección penal de la intimidad[62] en el sentido de que los sistemas de información deben poder albergar información relevante para la intimidad personal y familiar que abarca la vida privada de una persona en la que confluyen numerosos derechos vinculados a la propia personali-

59 Véanse Romeo Casabona, «Tendencias actuales sobre las formas de protección jurídica ante las nuevas tecnologías», p. 181 y sobre la posible intervención del Derecho penal en la red, con carácter general, Álvarez Vizcaya, «Consideraciones político-criminales sobre la delincuencia informática: el papel del Derecho penal en la red», pp. 268 y ss.

60 Opción por la que se decanta Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., p. 315. Entiendo que se muestra de acuerdo de la Mata Barranco, *Derecho penal europeo y legislación española: las reformas del Código penal. Actualizado a la reforma penal 2015*, p. 88.

61 Véase también en un sentido similar la opinión de Morón Lerma que estimaba que si se introdujera un tipo penal que castigara los accesos ilícitos a sistemas informáticos, habría que crear un título autónomo que castigue los atentados a los sistemas informáticos, en el que se ubicara éste y otros incidentes relativos a los mismos, como los daños a los datos y a los sistemas; véase Morón Lerma, «Delitos contra la confidencialidad, integridad y disponibilidad de datos y sistemas informáticos», p. 107.

62 Véanse Alonso de Escamilla, *Delitos,* 3.ª ed., p. 231; Bolea Bardon, *Comentarios al Código penal. Reformas LLOO 1/2023, 3/2023 y 4/2023,* 2.ª ed., pp. 970 y 971.

dad[63], aunque sea de manera tangencial y no principal, de modo que en ellos se manifieste la pretensión de valor de este bien jurídico[64]. Se puede concluir con Piñar Mañas que «el derecho a la privacidad alcanza también a los dispositivos informáticos que utilizamos y que forman parte ya de nuestra propia vida, que contienen información que nos identifica y que puede dar una imagen de nuestra personalidad. Lo que supone para tal derecho, para la protección de datos, un desarrollo espectacular»[65].

En segundo lugar, con carácter general, la regulación penal de la intimidad personal y familiar del Código penal recogida en el art. 197 del CP resulta casuística, excesivamente prolija[66], y en torno a la que surgen insalvables dificultades interpretativas, lo que no ha sido valorado positivamente por la doctrina[67]. Las posibles causas de esta deficiencia pueden ser, por una parte,

63 Véase *supra*.

64 Como han afirmado en nuestra doctrina también Miró Llinares, «Delitos informáticos. Hacking. Daños», marginal n.ºs. 1436, 1437 y 1438; Anarte Borrallo/Doval País, «Efectos de la reforma de 2015 en los delitos contra la intimidad», p. 1257; De la Mata Barranco/Barina Ubiñas, «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», p. 82; Sierra López, «Los delitos de descubrimiento y revelación de secretos en el Código penal de 2015: artículos 197, 197 bis, 197 ter, 197 quáter, 197 quinques y 198», p. 188. Aunque Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197 bis)*, 7.ª ed., p. 1481 se ha manifestado en contra de esta tesis sí estima que «en el art. 197 bis 1 basta con que el acceso se produzca en sistemas de información, que, a mi juicio, potencialmente pueden albergar datos personales, aunque sea de manera tangencial y no principal por respeto a la ubicación sistemática del precepto entre los delitos contra la intimidad». En contra de esta restricción Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197 bis)*, 7.ª ed., p. 1481 quien concluye que se determina «una ampliación del ámbito de conductas tipificadas, pues en el nuevo precepto basta con la concurrencia de un acceso ilícito sin más matizaciones, que las propias exigencias de la conducta dolosa, en el bien entendido de que ya no se exige un plus intencional en este precepto. Pero es que, además, tras la Reforma de 2015, el tipo proyecta el acceso a los sistemas de información; desaparece la referencia a los datos, de modo que la órbita del tipo se ha agrandado aún más». No obstante, finaliza con la exigencia apuntada en el párrafo anterior de esta misma nota; Morales García, «Delincuencia informática: intrusismo, sabotaje informático y uso ilícito de tarjetas (arts. 197.3 y 8, 264 y 248», pp. 186 y 187; Romeo Casabona, *Derecho penal, Parte Especial*, 3.ª ed., p. 316; Colás Turégano, «Nuevas conductas delictivas contra la intimidad (arts. 197; 197 bis; 197 ter)», p. 677.

65 Véase Piñar Mañas, «Protección de datos: origen, situación actual y retos de futuro», p. 101.

66 Véanse Muñoz Conde, *Derecho penal, Parte Especial,* 25.ª ed., pp. 302 y 303; Sáinz-Cantero Caparrós, *Sistema de Derecho penal, Parte Especial*, 2.ª ed., p. 306; Carrasco Andrino, *Tratado de Derecho Penal. Parte Especial (I). Delitos contras las personas, 4.ª ed.,* p. 1627; Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., 1436 quien considera que en este precepto se contemplan varios tipos «expresados a través de una atormentada e inacabable redacción»: Anarte Borrallo/Doval País, *Derecho penal, Parte Especial,* 2.ª ed., p. 494.

67 Véanse Morales Prats, *Comentarios al Código Penal Español, Tomo I (artículo 197*), 7.ª ed., p. 1434; Lozano Miralles*, Compendio de Derecho penal, PE, II*, p. 193; Romeo Casabona, *Derecho penal, Parte Especial*, p. 254.

la falta de comprensión por parte del legislador tanto del concepto de intimidad y los diversos elementos que la integran como de sus modalidades de ataque; y, por otra parte, la falta de una política criminal cuyo objetivo persiga una adecuada sistematización de los ataques contra la intimidad y la propia imagen en el marco de las TIC.

Bibliografía

Agustina Sanllehi, J.R., «¿Menores infractores o víctimas de pornografía infantil? Respuestas legales e hipótesis criminológicas ante el *Sexting*», en *Revista Electrónica de Ciencia Penal y Criminología,* RECPC 12-11, 2010.

Alonso de Escamilla, A., en Lamarca Pérez, C. (coord.), *Delitos. La parte especial del Derecho penal*, 3.ª ed., Colex, Madrid, 2015.

Alonso Pérez, F., *Intervención de las comunicaciones postales, telegráficas y telefónicas,* Dykinson, 2001.

Álvarez Vizcaya, M., «Consideraciones político criminales sobre la delincuencia informática: el papel del Derecho penal en la red», en *Internet y Derecho penal*, Cuadernos de Derecho Judicial, Consejo General del Poder Judicial, 2002.

Anarte Borrallo, E. y Doval País, A., en Boix Reig, J. (Dir), *Derecho penal, Parte Especial, Volumen I, La protección penal de los intereses jurídicos personales. (Adaptado a la reforma de 2015 del Código penal*), 2.ª ed., Iustel, Madrid, 2016.

Anarte Borrallo, E. y Doval País, A., «Efectos de la reforma de 2015 en los delitos contra la intimidad», en Diario La Ley, n.º 8744, D-163, 2016.

Bajo Fernández, M., «Protección del honor y de la intimidad», en *Comentarios a la legislación penal*, *Tomo I, Derecho penal y Constitución*, Cobo del Rosal (Dir.), Edersa, Madrid, 1982.

Baym, N., «Social networks 2.0», en *The Handbook of Internet Studies,* Burnet, Consalvo, Ess Eds., 2010, Wiley-Blackwell.

Bolea Bardon, C., en Corcoy Bidasolo, M., y Mir Puig, S. (Dirs)., y Vera Sánchez, J.S. (Coord), *Comentarios al Código penal. Reformas LLOO 1/2023, 3/2023 y 3/2023*, 2.ª ed., Tirant lo blanch, Valencia, 2024.

Boyd, D. y Ellison. N., «Social networks sites: definition, history, and scholarship», en *Journal of Computer-Mediated Communication,* volumen 13, número 1, 2007.

Carbonell Mateu, J.C. y González Cussac, J.L., en Vives Antón, T. (Coord), *Comentarios al Código Penal de 1995, Volumen I, (Arts. 1 a 233),* Vives Antón (coord.), Tirant lo blanch, Valencia, 1996.

Carrasco Andrino, M.M, en Álvarez García, F.J. (Dir) y Ventura Püschel, A. (Coord.), T*ratado de Derecho Penal. Parte Especial (I). Delitos contras las personas*, 4.ª Edición, Tirant lo blanch, Valencia, 2024.

Castelló Nicas, N., *«Delitos contra la intimidad, el derecho a la propia imagen y la inviolabilidad del domicilio, y delitos contra el honor»,* en Estudios sobre el Código penal reformado. (Leyes Orgánicas 1/2015 y 2/2015), Dykinson, Madrid, 2015.

Colás Turégano, M.A., «Nuevas conductas delictivas contra la intimidad (arts. 197; 197 bis; 197 ter)», en González Cussac, J.L. (Dir) y Matallín Evangelio, A., y Górriz Royo, E., (Coords), *Comentarios a la reforma del Código penal de 2015,* González Cussac Dir., Matallín Evangelio/Górriz Royo Coords., Tirant lo blanch, Valencia, 2015.

Cuerda Riezu, A., «El velo islámico y el derecho a la propia imagen», en Boix Reig, J. (Dir) y Jareño Leal, M.A., *La protección jurídica de la intimidad,* Iustel, Madrid, 2010.

De la Mata Barranco, N.J. y **Barina Ubiñas, d.,** «La protección penal de la vida privada en nuestro tiempo social: ¿necesidad de redefinir el objeto de tutela», en *Revista de Derecho penal y criminología,* 3.ª época, n.º 11 (enero de 2014).

De la Mata Barranco, N.J. *Derecho penal europeo y legislación española: las reformas del Código penal. Actualizado a la reforma penal 2015*, Tirant lo blanch, Valencia, 2015.

Díaz-Maroto y Villarejo, J., «Los delitos contra la intimidad, la propia imagen y la inviolabilidad del domicilio», en *La Ley*, D-228, 1996.

Espín, E, *Manual de Derecho Constitucional, Volumen I, La Constitución y las fuentes del Derecho. Derechos fundamentales y garantías*, 1.ª ed., López Guerra/Espín Dirs., Tirant lo blanch, Valencia, 2022.

González Cussac, J.L., *Derecho penal, Parte Especial*, 7.ª ed. actualizada a la Ley Orgánica 1/2015, González Cussac (Coord), Tirant lo blanch, Valencia, 2022.

Gutiérrez Francés, M.L., «El intrusismo informático (Hacking): ¿Represión penal autónoma?», en *Informática y Derecho, II Congreso Internacional de Informática y Derecho, Actas, Volumen II*, números 12-15, 1996.

Gutiérrez Francés, M.L., «Notas sobre la delincuencia informática: atentados contra la "información" como valor económico de empresa», en Tiedemann, K. y Arroyo Zapatero, L.A. (editores), *Estudios de Derecho penal económico*, Ediciones de la Universidad de Castilla-La Mancha, 1994.

Hernández Fernández, A. y Ramón Fernández, F., «El derecho a la propia imagen de los menores en los medios de comunicación y redes sociales», en *Revista Aranzadi de Derecho y Nuevas Tecnologías,* 2009-2, número 20.

Huerta Tocildo, S. y Andrés Domínguez, A.C., «Intimidad e informática», en *Revista de Derecho penal*, n.º 6, 2002.

Jorge Barreiro, A., *Comentarios al Código penal*, Rodríguez Mourullo (Dir.), Civitas, Madrid, 1997.

Livingstone, S. y Millwood, H., «Harmful to Children. Drawing Conclusions from Empirical Research on Media Effects», en Carlsson, U. (Ed)., *Regulation, Awareness, Empowerment. Young people and harmful media content in the digital age*,., 1.ª ed., Nordicom, Göteborg University, 2006.

Lozano Miralles, J., *Compendio de Derecho penal, (Parte Especial), Volumen II*, Bajo Fernández (Dir), Editorial Centro de Estudios Ramón Areces, Madrid, 1998.

Lucena Cid, I. V., «El concepto de la intimidad en los nuevos contextos tecnológicos», en Galán Muñóz, A. (Coord), *La protección jurídica de la intimidad y de los datos de carácter personal frente a las nuevas tecnologías de la información y comunicación*, Tirant lo blanch, Valencia, 2014.

Luzón Peña, D.M., «Protección penal de la intimidad y derecho a la información», en ADPCP, Vol. I, enero-abril, 1988.

Mata y Martín, R.M., *Delincuencia informática y Derecho penal*, Edisofer, Madrid, 2001.

Miró LLinares, F.M., *El cibercrimen. Fenomenología y criminología de la delincuencia en el ciberespacio,* Marcial Pons, Madrid, Barcelona, Buenos Aires, São Paolo, 2012.

Miró LLinares, F.M., «Delitos informáticos. Hacking. Daños», en Ortiz de Urbina Gimeno, I. (Coord), *Memento Experto. Reforma Penal 2010, Ley Orgánica 5/2010*, Ediciones Francis Lefebvre, Madrid, 2010.

Morales García, O., «Delincuencia informática: intrusismo, sabotaje informático y uso ilícito de tarjetas (arts. 197.3 y 8, 264 y 248», en *La reforma penal de 2010: análisis y comentarios,* Thomson Reuters, 2010.

Morales Prats, F., *Comentarios a la Parte Especial del Derecho penal*, Quintero Olivares (Dir.), Morales Prats (Coord.), 10.ª ed., Aranzadi Thomson Reuters, 2016.

Morales Prats, F., *Comentarios al Código Penal Español*, Tomo I (Artículos 1 a 233), Quintero Olivares (Dir.)/Morales Prats (coord.), Thomson Reuters Aranzadi, Pamplona, 7.ª ed., 2016.

Morales Prats, F., «La reforma de los delitos contra la intimidad artículo 197 CP», EN *Comentario a la reforma penal de 2015,* Thomson Reuters Aranzadi, Pamplona, 2015.

Moreno Navarrete, M.A., «Aspectos jurídico-privados de las tecnologías Web 2.0 y su repercusión en el derecho a la intimidad», en Boix Reig, J (Dir) y Jareño Leal, M.A. (Coord) *La protección jurídica de la intimidad,* Iustel, Madrid, 2010.

Morón Lerma, E., *Internet y Derecho penal: Hacking y otras conductas ilícitas en la red*, 2.ª ed., Aranzadi, 2002.

Morón Lerma, E., «Delitos contra la confidencialidad, integridad y disponibilidad de datos y sistemas informáticos», en *Delito e informática: algunos aspectos, Cuadernos penales José María Lidón*, número 4, 2007, Bilbao, Universidad de Deusto.

Muñoz Conde, F., *Derecho penal, Parte Especial,* 25.ª ed., revisada y puesta al día con la colaboración de LÓPEZ PEREGÍN, C. conforme a las LLOO 13/2022, 14/2022, 1/2023 y 4/2023, Tirant lo blanch, Valencia, 2023.

Nicolás Jiménez, P., La protección jurídica de los datos genéticos de carácter personal*, Cátedra Interuniversitaria Fundación BBVA-Diputación Foral de Bizkaia de Derecho y Genoma HumanoComares, Granada, 1996.*

Pardo Falcón, J., en Rodríguez-Piñero y Bravo Ferrer, M. y Casas Baamonde, M.E. (Dirs), Arnaldo Alcubilla, E. y Remón Peñalver, J. (Eds), y Peréz Manzano, M. y Borrajo Iniesta, I. (Coord), *Comentarios a la Constitución española, Tomo I, Conmemoración del XL aniversario de la Constitución*, Boletín Oficial del Estado, Ministerio de Justicia, Fundación Wolters Kluver, Madrid, 2018.

Paredes Castañón, J.M., *Enciclopedia Jurídica Básica*, Luzón Peña (Dir.), Comares, Granada, 2002.

Piñar Mañas, J.L., «Protección de datos: origen, situación actual y retos de futuro», *El derecho a la autodeterminación informativa,* Fundación Coloquio Jurídico Europeo, Madrid, 2009.

Queralt Jiménez, J.J., *Derecho penal español, Parte Especial*, 7.ª ed. revisada y actualizada con las Leyes Orgánicas 1/2015 y 2/2015, de 30 de marzo, 1.ª ed. en la Editorial Tirant lo blanch, Tirant lo blanch, Valencia, 2015.

Ribagorna Garnacho, A., «Seguridad de las tecnologías de la información», en *Ámbito jurídico de las tecnologías de la información,* Consejo General del Poder Judicial, 1996.

Rodríguez Mourullo, G., Alonso Gallo, J., y Lascuraín Sánchez, J.A., «Derecho penal e internet», en Fernández-Ordoñez, M., Cremades García, J. y Illescas Ortíz, R. (Coords), *Régimen jurídico de internet,* Cremades, Editorial La Ley, Madrid, 2002.

Romeo Casabona, C.M., «De los delitos informáticos al cibercrimen. Una aproximación conceptual y político-criminal», en Romeo Casabona, C.M. (Coord), *El cibercrimen: Nuevos retos jurídico-penales, nuevas respuestas político-criminales,* Comares, Granada, 2006.

Romeo Casabona, C.M., «Derecho penal y libertades de expresión y comunicación en Internet», en Romeo Casabona, C.M y Sánchez Lázaro, F.G. (Eds) y Armaza Armaza, E.J. (Coord), *La adaptación del Derecho penal al desarrollo social y tecnológico*, Comares, Granada, 2010.

Romeo Casabona, C.M., *Comentarios al Código penal*, Arts. 197, 198 y 201, Díez Ripollés/Romeo Casabona coords., Tirant lo blanch, Valencia, 2004.

Romeo Casabona, C.M, *Derecho penal, Parte Especial*, Romeo Casabona/Sola Reche/Boldova Pasamar coords., Comares, Granada, 2023.

Romeo Casabona, C.M., «Tendencias actuales sobre las formas de protección jurídica ante las nuevas tecnologías», Poder Judicial, n.º 31, 1993.

Romeo Casabona, C.M., *Poder informático y seguridad jurídica*. *«La función tutelar del derecho penal ante las Nuevas Tecnologías de la Información»,* Fundesco, Madrid, 1988.

Rueda martín, M.A., «La relevancia penal del consentimiento del menor de edad en relación con los delitos contra la intimidad y la propia imagen. (Especial consideración a la disponibilidad de la propia imagen del menor de edad en el ciberespacio)», en *InDret Penal,* n.º 4/2013.

Rueda martín, M.A., en Romeo Cabsona, C.M. y Sánchez Lazaro, F.G. (Eds) y Armaza Armaza, E.J. (Coord), «Los ataques contra los sistemas informáticos: conductas de hacking. Cuestiones político-criminales», en *La adaptación del Derecho penal al desarrollo social y tecnológico,* Comares, Granada, 2010.

Rueda martín, M.A., *Protección penal de la intimidad personal e informática. (Los delitos de descubrimiento y revelación de secretos de los artículos 197 y 198 del Código penal,* Atelier, Barcelona, 2004.

Rueda martín, M.A., *La nueva protección de la vida privada y de los sistemas de información en el Código penal,* Atelier, Barcelona, 2018.

Sáinz-Cantero Caparrós, J.A. *Sistema de Derecho penal, Parte Especial,* 2.ª ed. revisada y puesta al día conforme a las leyes Orgánicas 1/2015 y 2/2015, Morillas Cueva (Dir)., Dykinson, Madrid, 2016.

Salom Clotet, J.S., «Delito informático y su investigación», en Velasco Núñez, E. (Dir), *Delitos contra y a través de las nuevas tecnologías. ¿Cómo reducir su impunidad?,* Cuadernos de Derecho Judicial, Consejo General del Poder Judicial, 2006.

Serrano Alberca, J.M., *Comentarios a la Constitución*, Garrido Falla (Dir), 3.ª ed., Civitas, Madrid, 2001.

Sieber, U., «Legal Aspects of Computer-Related Crime in the Information Society —Comcrime-Study— prepared for the European Commission», versión de enero de 1998.

Sieber, U. *Computerkriminalität und Strafrecht*, 1.ª ed., Heymann, Munich, 1977.

Sierra López, M.V., «Los delitos de descubrimiento y revelación de secretos en el Código penal de 2015: artículos 197, 197 bis, 197 ter, 197 quáter, 197 quinques y 198», en Revista Penal, n.º 39, enero 2017.

Soler Presas, A., «Am I in Facebook? Sobre la responsabilidad civil de las redes sociales *on-line* por la lesión de los derechos de la personalidad, en particular por usos no consentidos de la imagen de un sujeto», en *InDret Derecho privado*, 3/2011.

Tomás y Valiente Lanuza, M.C., *Comentarios prácticos al Código penal. Los delitos contra las personas, artículo 138-233,* Gómez Tomillo (Dir)., Tomo II, 1.ª ed., Thomson Reuters, Madrid, 2015.

Valeije Álvarez, I., «Intimidad y difusión de imágenes sin consentimiento», en *Constitución, Derechos fundamentales y Sistema penal. (Semblanzas y estudios con motivo del setenta aniversario del profesor Tomás Salvador Vives Antón),* Tomo II, Tirant lo blanch, Valencia, 2009.

Vázquez de Castro, E., «Protección de datos personales, redes sociales y menores», en *Revista Aranzadi de Derecho y Nuevas Tecnologías,* mayo-agosto 2012, núm. 29.

14.

EL MARCO NORMATIVO DE LA DIRECTIVA (UE) 2022/2555: DE LA SEGURIDAD DE REDES Y SISTEMAS A LA CIBERSEGURIDAD INTEGRAL

Margarita Robles Carrillo[1]

Catedrática de Derecho Internacional Público y Derecho de la Unión Europea
Universidad de Granada

SUMARIO: I. Objeto y marco general. II. Ámbito de aplicación. 2.1. Criterios de aplicación. 2.2. Tipología. 2.3. Integración sectorial. III. Régimen jurídico. 3.1. Medidas de gestión. 3.2. Obligaciones de notificación. 3.3. Medidas adicionales. IV. Control y garantía de cumplimiento. 4.1 Jurisdicción. 4.2. Registro y bases de datos. 4.3. Supervisión y ejecución. V. Implementación. 5.1. Directrices relativas a las obligaciones del art. 3.4. 5.2. Directrices relativas a las obligaciones de los art. 4.1 y 2. 5.3. Reglamento de Ejecución 2024/2690. VI. Conclusiones.

Antecedentes

La seguridad de las redes y los sistemas de información constituye, desde hace años, una prioridad en el marco nacional, internacional y europeo. La UIT considera que se trata de un componente crítico no solo para los propios sistemas de información y comunicación, sino para el conjunto de una sociedad y de una economía del conocimiento sustentadas en el avance de las tecnologías de la información y la comunicaciones (TIC)[2]. En esa misma línea,

1 Miembro del Grupo Network Engineering & Security Group (NESG). Este trabajo es parte del Proyecto de I+D+i con referencia PID2020-114495RB-I00, financiado por el Ministerio de Ciencia e Innovación de España.

2 La actividad de la UIT en materia de seguridad de redes y sistemas de información se desarrolla principalmente en los sectores de telecomunicaciones y de normalización. Disponible en: https://www.itu.int/itunews/issue/2002/06/security-es.html

la OCDE hace públicas sus primeras Directrices sobre seguridad de redes y sistemas en 1992, que son actualizadas, una década después, abundando en la doble necesidad de generar una mayor conciencia y entendimiento de los aspectos de seguridad y desarrollar una cultura de seguridad[3]. El proceso de regulación en la UE no ha sido rápido ni tampoco sencillo pero, finalmente, ha desembocado en la adopción de una normativa que resuelve muchos de los problemas existentes en este ámbito y, en particular, la fragmentación normativa, material, funcional y subjetiva que había caracterizado el sistema anterior[4]. Se trata de la Directiva (UE) 2022/2555 relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión por la que se deroga la anterior Directiva (UE) 2016/1148[5].

La Directiva 2022/2555 supone un cambio de paradigma esencial porque implica la superación del modelo sectorial de seguridad de redes y sistemas de información de su predecesora mediante la adopción de un enfoque integral y global de la misma más coherente, racional y operativo que el diseñado con anterioridad. A pesar de ser la sucesora de la Directiva NIS 1, en su propia denominación desaparece la referencia a redes y sistemas de información que se sustituye por el término ciberseguridad. Ese cambio de paradigma se advierte en su objeto y marco general (1), el ámbito de aplicación, (2), el régimen jurídico (3), el control y garantía de cumplimiento (4) y las disposiciones sobre la implementación (5).

I. Objeto y marco general

En un mismo día, el 14 de diciembre de 2022, se aprueban tres normas esenciales para la UE: a) La Directiva (UE) 2022/2555 relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión[6]; b) La Directiva (UE) relativa a la resiliencia de las entidades críticas

3 OCDE (2002). *Directrices de la OCDE para la seguridad de sistemas y redes de información: hacia una cultura de seguridad*. Disponible en: https://www.oecd.org/sti/ieconomy/34912912.pdf

4 El sistema anterior es el diseñado por la Directiva (UE) 2016/1148 del Parlamento Europeo y del Consejo, de 6 de julio de 2016, relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información en la Unión, DO L 194,19.7.2916, 1.

5 Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión, por la que se modifican el Reglamento (UE) 910/2014 y la Directiva (UE) 2018/1972 y por la que se deroga la Directiva (UE) 2016/1148 (Directiva SRI 2) de 6 de julio de 2016, relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información en la Unión, DO L 333, 27.12.2022, 80.

6 Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión, por la que se modifican el Reglamento (UE) 910/2014 y la

por la que se deroga la anterior Directiva 2008/114/CE[7]; y c) El Reglamento (UE) 2022/2554 del Parlamento Europeo y del Consejo sobre la resiliencia operativa digital del sector financiero[8]. El régimen jurídico resultante constituye, por diversos motivos, un avance en términos de seguridad, coherencia y transparencia al situar la normativa NIS como núcleo vertebrador de la ciberseguridad en el conjunto de la UE y establecer marcos jurídicos específicos cuando son necesarios en determinados ámbitos[9].

La Directiva 2022/2555, conocida como NIS 2, tiene como objeto establecer medidas para alcanzar un elevado nivel común de ciberseguridad en toda la UE con el propósito de mejorar el funcionamiento del mercado interior. Esta norma no se refiere a seguridad de redes y sistema de información como su predecesora, sino a un elevado nivel comun de ciberseguridad. Esta nueva normativa reorganiza, reformula y amplía significativamente aquellas medidas pero, sobre todo, introduce mayor racionalidad, coherencia y transparencia, aunque el resultado final no esté exento de cierta complejidad. Con una extensa exposición de motivos, un mayor número de disposiciones y un contenido considerablemente más amplio que el de la Directiva 2026/1148, la Directiva NIS 2 se estructura en nueve capítulos con 46 artículos y tres anexos en los que se relacionan los sectores de criticidad y se incluye una tabla de correspondencias de sus respectivas disposiciones. Esta normativa introduce novedades y cambios significativos, aunque el más importante consiste en esa opción por un enfoque integral de la ciberseguridad que, incluso a pesar de algunas imperfecciones, constituye un avance sustancial respecto del sistema anterior diseñado en la Directiva NIS 1.

La Directiva NIS 2 establece tres categorías de normas: a) Un conjunto de obligaciones a cargo de los Estados miembros; b) Medidas para la gestión de riesgos y obligaciones de notificación; y c) Normas y obligaciones relativas al intercambio de información. La cooperación a escala nacional se regula en el

Directiva (UE) 2018/1972 y por la que se deroga la Directiva (UE) 2016/1148 (Directiva SRI 2) de 6 de julio de 2016, relativa a las medidas destinadas a garantizar un elevado nivel común de seguridad de las redes y sistemas de información en la Unión, DO L 333, 27.12.2022, 80.

7 Directiva (UE) 2022/2557 del Parlamento Europeo y del Consejo de 14 de diciembre de 2022 relativa a la resiliencia de las entidades críticas y por la que se deroga la Directiva 2008/114/CE del Consejo, DO L 333, 27.12.2022, 164.

8 Reglamento (UE) 2022/2554 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre la resiliencia operativa digital del sector financiero y por el que se modifican los Reglamentos (CE) 1060/2009, (UE) 648/2012, (UE) 600/2014, (UE) 909/2014 y (UE) 2016/1011, DO L 333, 27.12.2022, 1.

9 Un año después se adopta el Reglamento (UE, Euratom) 2023/2841 del Parlamento Europeo y del Consejo, de 13 de diciembre de 2023, por el que se establecen medidas destinadas a garantizar un elevado nivel común de ciberseguridad en las instituciones, los órganos y los organismos de la Unión. DO L de 18.12.2023, 1.

art. 13. La cooperación internacional y europea es objeto del capítulo III. El art. 17 se ocupa del ámbito internacional previendo la posibilidad de celebrar acuerdos con terceros países y organizaciones. Por su parte, para la cooperación estrictamente europea, prevé dos tipos de acciones específicas: los informes sobre la situación de la ciberseguridad en la UE[10] y la revisión interpares[11]. En el plano orgánico establece el Grupo de Cooperación[12], la Red de CSIRT[13] y la Red europea de organizaciones de enlace para las crisis de ciberseguridad EU-CyCLONe, que se crea con la finalidad de «respaldar la gestión coordinada de los incidentes y las crisis de ciberseguridad a gran escala en el ámbito operativo y de garantizar el intercambio regular de información relevante entre los Estados miembros y las instituciones, los órganos y los organismos de la Unión». De acuerdo con el art. 16, EU-CyCLONe está formada por representantes de las autoridades de gestión de crisis de ciberseguridad de los Estados miembros y la secretaría corresponde a ENISA. La Comisión participa en calidad de observador con carácter general y como parte de la red en caso de incidentes de ciberseguridad a gran escala con un impacto significativo en los servicios y actividades incluidos en el ámbito de aplicación de la directiva. EU-CyCLONe informará periódicamente al Grupo de Cooperación, cooperará con la red de CSIRT y presentará periódicamente al Parlamento Europeo y al Consejo informes sobre la evaluación de su actividad. Junto

10 Conforme al art. 18, la responsabilidad corresponde a ENISA, en cooperación con la Comisión y el Grupo de Cooperación. Este informe bienal, que se ha de presentar al Parlamento Europeo, ha de incluir evaluaciones de riesgos, desarrollo de capacidades, sensibilización y ciberhigiene, resultados de las revisiones interpares y nivel de madurez de capacidades y recursos en el ámbito de la ciberseguridad, así como recomendaciones políticas concretas para su mejora.

11 El art. 19 atribuye al Grupo de Cooperación, con la ayuda de la Comisión, ENISA y, cuando proceda, de la red de CSIRT, la función de determinar la metodología y los aspectos organizativos de las revisiones interpares, si bien la mayoría de esas cuestiones vienen determinadas por la propia directiva en esta disposición. El objetivo de estas revisiones interpares es aprender de las experiencias compartidas, reforzar la confianza mutua, lograr un elevado nivel común de ciberseguridad y mejorar las capacidades y políticas de ciberseguridad de los Estados miembros necesarias para la aplicación de la Directiva.

12 El Grupo de Cooperación estará formado por representantes de los Estados miembros, la Comisión que ejerce la secretaría y ENISA. El Servicio Europeo de Acción Exterior participa en las actividades del Grupo de Cooperación en calidad de observador. Las Autoridades Europeas de Supervisión y las autoridades competentes con arreglo al Reglamento (UE) 2022/2554 podrán participar en las actividades del Grupo de Cooperación de conformidad con el art. 47.1 de dicho Reglamento. La función de este grupo consiste en apoyar y facilitar la cooperación estratégica y el intercambio de información entre los Estados miembros y fortalecer la confianza y la colaboración desarrollando el extenso abanico de cometidos establecidos en el art. 14.4.

13 Con el mismo propósito del Grupo de Cooperación, la Red de CSIRT se establece para contribuir al refuerzo de la confianza y la seguridad y la promoción de una cooperación operativa rápida y eficaz entre los Estados miembros. Está formada por representantes de los CSIRT con la misión de desarrollar el amplio listado de tareas incluido en del art. 15.3 de la directiva.

con este ampliado y reforzado marco orgánico, la Directiva NIS 2 modifica y extiende el ámbito de acción de su predecesora.

II. Ámbito de aplicación

El ámbito de aplicación de la Directiva NIS 1 se encuentra determinado por dos premisas básicas: por una parte, la identificación de dos categorías de sujetos —los operadores de servicios esenciales y los proveedores de servicios digitales— con regímenes distintos basados en el tipo de servicio que presta cada entidad; y, por otra parte, la exclusión de proveedores de redes y servicios de comunicaciones electrónicas, infraestructuras críticas y proveedores de servicios de confianza. El modelo NIS 1 se encuentra doblemente fragmentado, desde el punto de vista subjetivo, por la exclusión de estos últimos y por la diferenciación entre los dos primeros. La fragmentación se manifiesta, en el plano normativo, por la existencia de normas diferentes en cada caso y, desde una perspectiva funcional, porque son diferentes respecto de cada servicio o ámbito de actividad.

La Directiva NIS 2 opta por un sistema distinto para acabar con la extrema fragmentación que caracteriza la determinación de las entidades sometidas a esta normativa y con las considerables divergencias existentes a este respecto entre los Estados. Para ello establece una serie de criterios en el art. 2 (2.1), una tipología que distingue entre entidades esenciales e importantes en el art. 3 (2.2) y un sistema de integración sectorial en el art. 4 (2.3) que, a pesar de no estar exento de una cierta complejidad, permite racionalizar el modelo en su conjunto.

2.1. Criterios de aplicación

La Directiva NIS 2 cuenta con un ámbito de aplicación más extenso, más coherente y, también, más complejo porque combina varios criterios. En primer lugar, siguiendo un *criterio material*, los anexos I y II relacionan los sectores de alta criticidad y otros sectores críticos, respectivamente, respecto de los cuales se aplica esta normativa[14].

En segundo lugar, siguiendo el criterio del *tamaño*, esta directiva se aplica a todas las entidades que, prestando servicios o ejerciendo actividades en

14 El Anexo I incluye los sectores alta criticidad que son: Energía; Transporte; Banca; Infraestructuras de los mercados financieros; Sector sanitario; Agua potable; Aguas residuales; Infraestructura digital; Gestión de servicios de TIC; Entidades de la Administración pública; y Espacio. El Anexo II relaciona los otros sectores críticos: Servicios postales y de mensajería; Gestión de residuos; Fabricación, producción y distribución de sustancias y mezclas químicas; Producción, transformación y distribución de alimentos; Fabricación; Proveedores de servicios digitales; e Investigación.

cualquiera de esos ámbitos, se consideren empresas medianas o superen los límites máximos previstos para dichas empresas en la normativa europea, en concreto, en la Recomendación 2003/361/CE de la Comisión[15]. Conforme a la misma, una empresa mediana es aquella que ocupa entre 50 y 250 personas y cuyo volumen de negocios anual supera los 10 millones de euros y no excede de 50 millones de euros o cuyo balance general anual no excede de 43 millones de euros. Las empresas con más de 50 empleados y más de 10 millones de volumen anual de negocios están, en consecuencia, sometidas a la Directiva NIS 2.

En tercer lugar, con independencia de su tamaño, siguiendo un criterio que atiende a la *importancia clave* para la sociedad, la economía o determinados sectores o servicios, esta directiva se aplica a: a) Las entidades identificadas como críticas con arreglo a la Directiva 2022/2557 (art. 2.3); b) Los proveedores de redes públicas de comunicaciones o servicios electrónicos, los prestadores de servicios de confianza, los servicios de registros de nombres de dominio de primer nivel y los proveedores de servicios de sistema de nombres de dominio (art. 2.2); c) Las entidades que presten servicios de registro de nombres de dominio (art. 2.4) y d) En los siguientes supuestos concretos: a) Cuando la entidad sea una Administración pública central o regional; b) Cuando la entidad sea crítica por su importancia específica a nivel nacional o regional para el sector o tipo de servicio; c) Cuando la entidad sea el único proveedor de un Estado miembro de un servicio esencial; y d) Cuando una perturbación del servicio prestado pudiera tener repercusiones significativas para la seguridad pública, el orden público o la salud pública o pudiera inducir riesgos sistémicos significativos (art. 2.2).

Para terminar, siguiendo el criterio de su *voluntad*, los Estados miembros podrán decidir la aplicación de la Directiva NIS 2 a entidades de la Administración pública a nivel local y a centros de enseñanza, en particular, cuando lleven a cabo actividades críticas de investigación (art. 2.5).

2.2. Tipología

Una vez determinados los proveedores y las entidades a los que se aplica la normativa conforme a esos diversos criterios, la Directiva NIS 2 establece una tipología que determina el régimen jurídico al que están sujetos en cada caso, aunque las diferencias son realmente mínimas. Mientras que la Directiva NIS 1 establecía dos tipos de sujetos —operadores y proveedores— y dos regímenes normativos claramente distintos para cada uno de ellos, la Directiva NIS 2 determina, primero, en el art. 2, las entidades y los proveedores incluidos en su ámbito conforme a los criterios explicados en el punto ante-

15 Recomendación de la Comisión, de 6 de mayo de 2003, sobre la definición de microempresas, pequeñas y medianas empresas, DO L 124 de 20.5.2003, 36.

rior y, a continuación, en el art. 3, los clasifica en dos categorías —esenciales o importantes— y establece normas algo diferentes. Aunque las diferencias se concentran en el ámbito del control y la supervisión de las entidades, las sanciones y multas, la normativa es más estricta respecto de las entidades esenciales que en relación con las importantes. Como explica la directiva, esa diferenciación de los regímenes de supervisión y de garantía del cumplimiento tiene como objetivo «garantizar un equilibrio justo entre los requisitos y las obligaciones en función del riesgo, por un lado, y la carga administrativa derivada de la supervisión del cumplimiento, por el otro»[16].

Las entidades importantes se definen por exclusión: son aquellas que no son esenciales. Para la determinación de las entidades esenciales, el art. 3 de la Directiva NIS 2 utiliza varios criterios: a) Por el *tipo de entidad* se consideran esenciales los siguientes: los prestadores cualificados de servicios de confianza y registros de nombres de dominio de primer nivel; los proveedores de servicios de DNS; las entidades identificadas como entidades críticas con arreglo a la Directiva (UE) 2022/2557; y las entidades de la administración pública central; b) Por el *tipo de entidad y el tamaño* se definen como esenciales: los proveedores de redes públicas de comunicaciones electrónicas o de servicios de comunicaciones electrónicas que sean consideradas medianas empresas; y las entidades recogidas en el anexo I de la directiva que superen los límites establecidos para las medianas empresas; y c) Por *decisión del Estado*, se considerarán esenciales: por una parte, cualquier entidad de los anexos I y II que un Estado identifique como esencial porque sea el único proveedor de un servicio esencial para el mantenimiento de actividades sociales o económicas críticas, o porque sea crítica por su importancia específica a nivel nacional o regional para el sector o tipo de servicio, o porque una perturbación en el servicio que presta pudiera tener repercusiones significativas sobre la seguridad pública, el orden público o la salud pública o pudiera inducir riesgos sistémicos significativos; y, por otra parte, las entidades que un Estado haya identificado como operador de servicios esenciales antes del 16 de enero de 2023 conforme a la Directiva NIS 1.

Siguiendo este modelo, las entidades importantes serán, por exclusión, aquellas comprendidas en los anexos 1 y 2 que no puedan considerarse esenciales y, por decisión estatal, aquellas que un Estado identifique como importantes dentro de las incluidas en las letras b) a e) del art. 2.2. Ello significa que cuando se trata de un único proveedor o una entidad crítica por su importancia o cuando la perturbación del servicio puede tener repercusiones significativas o sistémicas en los términos previstos en esa disposición, cada Estado es competente para proceder a su calificación como esenciales o importantes. En realidad, los criterios que atienden al tipo de entidad o a

16 Considerando 15.

su tamaño ofrecen una objetividad que se pierde cuando depende de una decisión de cada Estado la calificación de las entidades como esenciales o importantes[17].

En definitiva, mediante los criterios del art. 2 se determinan las entidades y proveedores sometidos a la normativa NIS 2 y mediante la distinción del art. 3 entre entidades esenciales e importantes se establecen normas más o menos estrictas en función de cada tipo. El hecho de que tengan distintos regímenes normativos está justificado por la diferente importancia de las entidades y proveedores en cada caso, que responde, en general y salvo los supuestos en los que resulta de una decisión del Estado, a criterios objetivos.

2.3. Integración sectorial

Junto con las medidas establecidas en los arts. 2 y 3 extendiendo y racionalizando el ámbito de aplicación, el artículo 4 de la Directiva NIS 2 trata de resolver el problema de la fragmentación mediante la regla de integración sectorial. El art. 4.1 distingue dos supuestos. Por una parte, los casos en los que la normativa sectorial establece requisitos en materia de gestión de riesgos de ciberseguridad o notificación de incidentes que son, al menos, equivalentes a las obligaciones establecidas en la Directiva NIS 2. En estos casos no es preciso aplicar esta directiva porque la normativa sectorial es igual o equivalente. Por otra parte, en los supuestos en los que la normativa sectorial no cubre todos los requisitos, será de aplicación esta directiva[18]. Esta regla sirve para respetar la particularidad sin atentar contra la homogeneidad general.

III. Régimen jurídico

El capítulo IV de la Directiva NIS 2 establece el régimen jurídico en materia de ciberseguridad con las medidas para la gestión de los riesgos (3.1), las

17 Una vez realizada esta clasificación, el art. 3 de la Directiva estipula una serie de obligaciones relativas a la información sobre estas entidades y la posibilidad de establecer sistemas nacionales de registro. Con un plazo que se extiende hasta el 17 de abril de 2025, los Estados han de elaborar un listado de esas entidades que habrán de actualizar periódicamente con la información requerida y las entidades, por su parte, habrán de tener asimismo actualizada. Las autoridades competentes notificarán a la Comisión y al Grupo de Cooperación el número de entidades esenciales e importantes respecto de cada sector y subsector. Aunque supondrá una carga significativa en términos de gestión, con estas medidas se garantiza una mayor transparencia del modelo de ciberseguridad. En este sentido, puede verse la Comunicación de la Comisión. *Directrices de la Comisión sobre la aplicación del artículo 3, apartado 4, de la Directiva (UE) 2022/2555*. C (2023) 6070 final, Bruselas, 13.9.2023.

18 El concepto de efecto equivalente se precisa en el apartado 2 del art. 4 y en la Comunicación de la Comisión. *Directrices de la Comisión sobre la aplicación del artículo 4, apartados 1 y 2, de la Directiva (UE) 2022/2555*. C (2023) 6068 final, Bruselas, 13.9.2023.

obligaciones de notificación (3.2) y la medidas adicionales entre las que se incluyen las evaluaciones coordinadas de riesgos, la posibilidad de utilizar esquemas europeos de certificación y la normalización (3.3).

En términos de gobernanza se arbitran dos medidas generales que introducen un interesante cambio de paradigma respecto del enfoque anterior: por una parte, conforme a la directiva, hay que velar para que los órganos de dirección de las entidades aprueben medidas para la gestión de riesgos de ciberseguridad, supervisen su puesta en práctica y respondan por los incumplimientos; y, por otra parte, hay que garantizar que los miembros de los órganos de dirección asistan a formaciones e impulsan formaciones similares periódicas para sus empleados con objeto de adquirir conocimientos y destrezas suficientes para detectar amenazas y evaluar las prácticas de gestión de riesgos de ciberseguridad. Ambas medidas, al identificar una responsibilidad individual en cuanto a la aplicación de esta normativa y promover activamente la formación, tienen una utilidad mayor de lo que en principio cabría esperar tanto en términos de exigencia de responsabilidad como de promoción de la cultura de ciberseguridad.

3.1. Medidas de gestión

Las medidas para la gestión de los riesgos de ciberseguridad, que se establecen en el art. 21, están dirigidas a los Estados —porque se trata de una directiva— para que las impongan a las entidades esenciales e importantes mediante la normativa de transposición. Ambos tipos de entidades: a) Han de tomar las medidas técnicas, operativas y de organización adecuadas y proporcionadas para gestionar los riesgos de seguridad y para prevenir o minimizar las repercusiones de los incidentes; b) Han de tener en cuenta las vulnerabilidades específicas de cada proveedor y prestador de servicios y la calidad general de los productos y las prácticas en materia de ciberseguridad de sus proveedores y prestadores de servicios; c) Han de tener en cuenta los resultados de las evaluaciones coordinadas de los riesgos de seguridad de las cadenas de suministro críticas; y d) Han de adoptar, en su caso, todas las medidas correctoras apropiadas y proporcionadas necesarias.

Al adoptar esas medidas, los proveedores han de seguir un enfoque basado en todos los peligros y el entorno físico de dichos sistemas frente a incidentes incluyendo, al menos, los elementos prescritos en la directiva.

3.2. Obligaciones de notificación

Las obligaciones de notificación se regulan en un extenso art. 23 donde destacan las dos siguientes: por una parte, las entidades esenciales e importantes han de notificar al CSIRT o a la autoridad competente, según corresponda, cualquier incidente que tenga un impacto significativo en la

prestación de sus servicios; y, por otra parte, notificarán a los destinatarios de sus servicios los incidentes significativos susceptibles de afectar negativamente a la prestación de dichos servicios.

El objetivo es conseguir una notificación ágil al tiempo que minuciosa. Los tiempos están marcados por la fórmula recurrente «sin demora indebida» desde que hay constancia del incidente. La información ha de transmitirse al CSIRT directamente o a través de la autoridad competente y a los puntos de contacto único si el incidente tiene carácter transfronterizo o intersectorial. Sin demora indebida, el CSIRT o la autoridad competente: a) Ofrecerá una respuesta a la entidad notificante con sus comentarios y un asesoramiento operativo sobre las posibles medidas paliativas; b) Prestará apoyo técnico si lo solicita la entidad afectada; c) Proporcionará orientación para la denuncia si el incidente muestra tintes delictivos; y d) Podrá informar al público cuando sea necesario para hacer frente a un incidente o por motivos de interés público.

3.3. Medidas adicionales

El capítulo IV establece un conjunto de medidas dirigidas a completar, reforzar o perfeccionar el modelo de gestión de riesgos: a) Las evaluaciones coordinadas de los riesgos de seguridad de las cadenas de suministro críticas a escala de la UE se realizarán por el Grupo de Cooperación, en colaboración con la Comisión y ENISA (art. 22); b) La certificación podrá ser exigida a las entidades esenciales e importantes para productos, servicios y procesos TIC mediante un esquema europeo de certificación de la ciberseguridad adoptado conforme al artículo 49 del Reglamento (UE) 2019/881(art. 24); y c) La normalización prevista en el art. 25 está dirigida al fomento de la utilización de normas y especificaciones técnicas europeas e internacionales que sean pertinentes en materia de seguridad de los sistemas de redes y de información para promover una aplicación convergente de las medidas de gestión de los riesgos. ENISA ha de elaborar directrices y orientaciones al respecto (art. 25).

IV. Control y garantía de cumplimiento

La Directiva 2022/2555 prevé distintos tipos de medidas sobre control y garantía del cumplimiento de sus disposiciones. En primer lugar, establece un régimen de jurisdicción más amplio y preciso que el de su predecesora que se aplica a todas las entidades (4.1). En segundo lugar, incorpora un sistema de registro (4.2). Por último, dedica un capítulo completo, el VII, a la supervisión y ejecución (4.3). Todo ello compone un sistema sólido y garantista para el cumplimiento de las obligaciones establecidas en la Directiva NIS 2.

4.1. Jurisdicción

Las cuestiones relativas a la jurisdicción se regulan en el art. 26.1 mediante una norma general y unas reglas específicas. Con carácter general, las entidades comprendidas en el ámbito de aplicación de la Directiva NIS 2 están sometidas a la jurisdicción del Estado miembro en el que están establecidas. Existen, sin embargo, tres reglas específicas en función del tipo de entidad: a) Las entidades de la Administración pública se considerarán sometidas a la jurisdicción del Estado miembro que las establece, esto es, su propio Estado; b) Los proveedores de redes públicas o de servicios de comunicaciones electrónicas públicos se considerarán sometidos a la jurisdicción del Estado miembro donde prestan sus servicios; y c) Los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, las entidades que prestan servicios de registro de nombres de dominio, los proveedores de servicios de computación en nube, los proveedores de servicios de centro de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados y los proveedores de servicios de seguridad gestionados, así como los proveedores de mercados en línea, de motores de búsqueda en línea o de plataformas de servicios de redes sociales, se considerarán sometidos a la jurisdicción del Estado miembro en el que se encuentre su establecimiento principal en la Unión. Para terminar, en el caso de que la entidad que ofrezca servicios en la UE no esté establecida en ella, el art. 26.3 exige la designación de un representante que se ha de establecer en el Estado miembro en el que preste servicios de manera que la entidad estará sometida a la jurisdicción de ese Estado[19].

4.2. Registro y bases de datos

El registro del art. 27 y las bases de datos del art. 28 son mecanismos de control y supervisión que no se aplican a todas las entidades. Según los términos de la directiva, mantener bases de datos precisas y completas con los datos de registro de los nombres de dominio y proporcionar un acceso lícito a tales datos «es fundamental para garantizar la seguridad, estabilidad

19 En caso de no fijarse esa representación, cualquier Estado en el que se presten servicios podrá ejercer su jurisdicción. Está prevista, asimismo, la asistencia mutua en estos casos. Para determinar este extremo, el art. 26.2 establece tres posibilidades: a) El establecimiento principal se encuentra en el Estado miembro en el que se adopten de forma predominante las decisiones relativas a las medidas para la gestión de riesgos de ciberseguridad; b) Si esa opción no funciona porque no puede determinarse dicho Estado miembro o porque aquellas decisiones no se toman en la Unión, el establecimiento principal se situará en el Estado miembro en el que se lleven a cabo las operaciones de ciberseguridad; y c) Si tampoco así se pudiese fijar la jurisdicción, se considerará que el establecimiento principal se encuentra en el Estado miembro en el que la entidad tenga el establecimiento con mayor número de trabajadores en la Unión.

y resiliencia del DNS, lo que a su vez contribuye a garantizar un elevado nivel común de ciberseguridad en toda la Unión»[20]. En concreto, la base de datos sobre el registro de nombres de dominio está diseñada con el objetivo de contribuir a la seguridad, estabilidad y resiliencia del DNS. Con ese propósito, se establecen una serie de obligaciones que los Estados miembros exigirán a los registros de nombres de dominio de primer nivel y a las entidades que prestan servicios de registro de nombres de dominio[21].

El registro de entidades, creado y mantenido por ENISA, está previsto para los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, las entidades que prestan servicios de registro de nombres de dominio, los proveedores de servicios de computación en nube, los proveedores de servicios de centro de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados y los proveedores de servicios de seguridad, así como los proveedores de mercados en línea, de motores de búsqueda en línea y de plataformas de servicios de redes sociales.

4.3. Supervisión y ejecución

El capítulo VII de la Directiva 2022/2555 relativo a la supervisión y ejecución define un marco general (art. 31) y, después, una variedad de acciones que incluyen un régimen diferente para las entidades esenciales (art. 32) y las importantes (art. 33). Junto con ello, establece las condiciones para la imposición de multas (art. 34), la actuación en caso de incumplimientos que implican violación de la seguridad de datos personales (art. 35), las sanciones (art. 36) y la asistencia mutua (art. 37).

La regla general y principal del art. 31 dispone que las autoridades competentes de los Estados han de adoptar las medidas necesarias para garantizar el cumplimiento de la directiva y supervisar dicho cumplimiento, aplicando metodologías que permitan priorizar esta actividad de control y adoptando un enfoque basado en el riesgo. Para ello, han de disponer de las competencias necesarias para cumplir dichas funciones con independencia operativa. Junto con ello, las medidas de supervisión y ejecución (A) y las multas y sanciones (B) merecen una atención especial por su relevancia, alcance y contenido.

20 Considerando 109.

21 Esas obligaciones son: a) Recopilar y mantener datos precisos y completos sobre el registro de nombres de dominio; b) Establecer políticas y procedimientos para garantizar que las bases de datos incluyan información precisa y completa; c) Publicar sin demoras los datos de registro que no sean de carácter personal; d) Conceder acceso a datos específicos sobre el registro de nombres de dominio, previa solicitud lícita y debidamente justificada, a los solicitantes de acceso legítimo; y e) Cooperar entre sí para evitar duplicidades.

A) Las medidas de supervisión y ejecución han de ser adecuadas, proporcionadas y eficaces. Algunas de esas medidas son similares para las entidades esenciales y para las importantes[22], mientras que otras son diferentes. Hay diferencias más significativas en el ámbito de la supervisión[23] que en el de la ejecución o la cooperación[24]. Entre las medidas previstas solo en rel-

22 Entre las medidas similares o equivalentes para entidades esenciales e importantes en materia de supervisión se incluyen las siguientes: a) En ambos casos se atribuye a los Estados la obligación de velar por que las autoridades competentes dispongan de las competencias necesarias para ejercer sus funciones; b) Estas autoridades podrán realizar análisis de seguridad basados en criterios de evaluación del riesgo objetivos, no discriminatorios, justos y transparentes; c) Las autoridades podrán solicitar acceso a datos, documentos o información, así como pruebas de la aplicación de la política de ciberseguridad; y d) Las auditorías de seguridad se basarán en evaluaciones de riesgo, los resultados se pondrán a disposición de la autoridad competentes y los costes serán sufragados por la entidad auditada, salvo decisión en contrario de la autoridad competente.
En el ámbito de las competencias de ejecución, las autoridades competentes tendrán las mismas atribuciones respecto de las entidades esenciales y las importantes en los siguientes aspectos: a) Apercibir por los incumplimientos; b) Adoptar instrucciones vinculantes o una orden de requerimiento para subsanar deficiencias o incumplimientos; c) Exigir a las entidades que pongan fin a los incumplimientos y se abstengan de repetirlos; d) Exigir a las entidades que garanticen que sus medidas para la gestión de riesgos de ciberseguridad y las obligaciones de notificación son conformes con lo previsto de manera específica y en un plazo concreto; e) Ordenar a las entidades que informen a las personas a las prestan servicios sobre la posibilidad de que se vean afectadas por una ciberamenaza significativa, así como sobre la naturaleza de la misma y las posibles medidas de protección; f) Ordenar la aplicación de recomendaciones; g) Ordenar que se hagan públicos los incumplimientos; y h) Imponer o solicitar la imposición de una multa administrativa.

23 En el ámbito de la supervisión se advierten las siguientes diferencias. En primer lugar, en el caso de las entidades importantes, las autoridades competentes actuarán cuando dispongan de pruebas, indicios o información de que presuntamente una entidad no cumple la directiva (art. 33.1), mientras que esas medidas se aplicarán directamente a las entidades esenciales sin necesidad de que se produzca aquel desencadenante o justificación (art. 32.1). En segundo lugar, las inspecciones in situ y la supervisión a distancia serán a posteriori en las importantes y podrán ser previas e incluir controles aleatorios en las esenciales. En tercer lugar, las auditorías de seguridad serán periódicas y específicas en las esenciales donde también podrá haber auditorías ad hoc en determinados supuestos y solo podrán ser específicas en las entidades importantes. Finalmente, las autoridades podrán realizar solicitudes de información para evaluar las medidas para la gestión de riesgos de ciberseguridad adoptadas por estas entidades, pero solo a posteriori en el caso de las importantes.

24 En el ámbito de la ejecución, hay dos diferencias. Por un lado, en cuanto a las funciones de las autoridades competentes, solo en el caso de las entidades esenciales está previsto designar un responsable de supervisión de la ejecución (art. 32.4.g). Por otro lado, en los supuestos de ineficacia de las medidas de ejecución, el art. 32.5 establece la posibilidad de subsanación de las deficiencias o cumplimiento de los requisitos solo en el caso de las esenciales. Si no se subsana en plazo, las autoridades competentes estarán facultadas para: a) Suspender temporalmente o solicitar la suspensión temporal de una certificación o autorización referente a una parte o la totalidad de los servicios prestados por la entidad en cuestión; y b) Solicitar la prohibición temporal de ejercicio de responsabilidades de dirección a cualquier persona que ejerza dichas responsabilidades a nivel de director general o representante legal.

ación con las entidades esenciales, se incluye la cooperación entre las autoridades competentes conforme a esta directiva y las autoridades competentes de la Directiva 2022/2557 cuando ejerzan su actividad de supervisión sobre una entidad crítica. En cambio, la cooperación entre las autoridades competentes conforme a esta directiva y las designadas por el Reglamento 2022/2554 sobre resiliencia operativa digital del sector financiero está prevista tanto para las entidades esenciales como para las importantes (art. 32.10 y art. 33.6).

En este punto, hay tres disposiciones del art. 32 relativas a las entidades esenciales que se aplicarán asimismo *mutatis mutandis* a las entidades importantes (art. 33.5), constituyendo una significativa novedad respecto de la Directiva NIS 1, a saber:

a) Los Estados miembros garantizarán que cualquier persona física responsable de una entidad esencial o que actúe como representante de ella con facultades para representarla, la autoridad para tomar decisiones en su nombre o la autoridad para ejercer control sobre ella tenga competencias para velar por que se cumpla esta normativa y pueda considerarse responsable por el incumplimiento de su deber de garantizar el cumplimiento de la directiva (art. 32.6);

b) Al adoptar las medidas de ejecución, las autoridades competentes respetarán los derechos de defensa y tendrán en cuenta aspectos como la gravedad o la duración del incumplimiento, la importancia de las disposiciones infringidas, los eventuales precedentes, los perjuicios materiales e inmateriales, la intencionalidad o negligencia, así como medidas preventivas y protectoras y el grado de cooperación de las personas con las autoridades competentes (art. 32.7); y

c) Esas autoridades notificarán a las entidades afectadas sus constataciones, les concederán un plazo para presentar observaciones y argumentarán detalladamente sus medidas de ejecución (art. 32.8).

B) Las multas y sanciones habrán de ser efectivas, proporcionadas y disuasorias, teniendo en cuenta las circunstancias de cada caso individual conforme a los arts. 34.1 y 36 de la Directiva NIS 2. El régimen de sanciones se establece por los Estados que han de comunicarlo a la Comisión antes del 17 de enero de 2025. La Directiva NIS 2 contempla la posibilidad de multas coercitivas, el caso especial de las multas administrativas a las entidades de la administración, la solución en los casos en los que el ordenamiento de un país no regula las multas administrativas y el supuesto específico de las multas en el caso de incumplimientos que conllevan una violación de la seguridad de los datos personales[25]. Si en ese supuesto ya se ha dispuesto una

25 Las multas administrativas, los supuestos y los criterios a seguir en su determinación están establecidos en el art. 34. En el caso de las entidades esenciales, se prevé un máximo de, al menos, 10.000.000 EUR o un máximo de, al menos, el 2 % del volumen de

multa conforme al Reglamento General de Protección de Datos, no se impondrá una nueva multa, pero podrán adoptarse otras medidas de ejecución.

V. Implementación

Además de ser necesaria su transposición al derecho interno de los Estados miembros, que debería haberse culminado a más tardar el 17 de octubre de 2024, la implementación de la Directiva 2022/2555 requiere la adopción de actos de ejecución por parte de la Comisión en diversos ámbitos. En septiembre de 2023, esta institución adopta las directrices relativas a la aplicación de los arts. 3.4 (5.1) y 4.1 y 2 (5.2) y, en octubre de 2024, publica el reglamento de ejecución relativo a los requisitos técnicos y metodológicos para la gestión de riesgos de ciberseguridad y los casos en que un incidente se considera significativo (5.3).

5.1. Directrices relativas a las obligaciones del art. 3.4

El art. 3 de la Directiva 2022/2555 regula las entidades esenciales e importantes disponiendo que los Estados miembros han de elaborar un listado de las mismas, antes del 17 de abril de 2025. Para ello, la Comisión, asistida ENISA, debe proporcionar directrices y plantillas. Cumpliendo ese mandato, la Comisión publica las Directrices sobre la aplicación del artículo 3, apartado 4, de la Directiva (UE) 2022/2555[26]. Conforme a las mismas, en primer lugar, los Estados miembros han de exigir a las entidades que presenten a las autoridades competentes la información siguiente: el nombre, la dirección y los datos de contacto actualizados, incluidas las direcciones de correo electrónico, los rangos de IP y los números de teléfono de la entidad, el sector y el subsector pertinentes conforme a los anexos y, en su caso, una lista de los Estados miembros en los que prestan los servicios comprendidos en el ámbito de aplicación de la directiva. En segundo lugar, han de establecer mecanismos nacionales para que las entidades puedan inscribirse. En tercer lugar, se establecen obligaciones adicionales respeto de determinados tipos de entidades. Para articular esas obligaciones, en las directrices se incluye un apéndice con la plantilla y las diferentes exigencias de información.

negocios anual total a nivel mundial de la empresa a la que pertenece la entidad esencial durante el ejercicio financiero anterior, optándose por la de mayor cuantía. En el caso de las entidades importantes, las multas alcanzarán un máximo de, al menos, 7.000.000 EUR o de un máximo de, al menos, el 1,4 % del volumen de negocios anual total a nivel mundial de la empresa a la que pertenece la entidad importante durante el ejercicio financiero anterior, optándose por la de mayor cuantía.

26 Comunicación de la Comisión. Directrices de la Comisión sobre la aplicación del artículo 3, apartado 4, de la Directiva (UE) 2022/2555, DO C de 14.9.2023, 2.

5.2. Directrices relativas a las obligaciones de los art. 4.1 y 2

El art. 4.1 de la Directiva 2022/2555 se ocupa de los actos jurídicos sectoriales para determinar los supuestos de aplicación de esta directiva. En los casos en los que los actos jurídicos de carácter sectorial dispongan requisitos de efecto al menos equivalente a las obligaciones de la directiva, no se aplica esta norma sino el acto sectorial. En los casos en los que no ocurra así, se aplica la directiva a las entidades no cubiertas por los actos jurídicos sectoriales. El concepto de efecto equivalente se precisa en el apartado 2 indicando que se produce cuando las medidas de gestión de los riesgos sean equivalentes a la prevista en el art. 21 o cuando el acto jurídico sectorial prevé el acceso inmediato, automático y directo a las notificaciones de incidentes por parte de los CSIRT, las autoridades competentes o los puntos de contacto únicos y los requisitos de notificación de incidentes significativos tienen un efecto al menos equivalente al establecido en el art. 23, apartados 1 a 6 de la directiva. Para ello, la Comisión, teniendo en cuenta las observaciones del Grupo de Cooperación y ENISA, publica las Directrices sobre la aplicación del artículo 4, apartados 1 y 2, de la Directiva 2022/2555[27]. En su apéndice solo identifica como acto sectorial el Reglamento 2022/2554 del Parlamento Europeo y del Consejo sobre la resiliencia operativa digital del sector financiero[28]. Estas directrices se estructuran en torno a dos aspectos clave: la equivalencia de los requisitos de ciberseguridad de los actos jurídicos sectoriales y los efectos de la equivalencia.

A) La equivalencia de los requisitos se determina atendiendo a dos parámetros principales. Por una parte, en cuanto al cumplimiento de los requisitos de gestión de riesgos de ciberseguridad, hay que valorar tres aspectos en particular: a) si las obligaciones de seguridad del acto jurídico sectorial incluyen medidas destinadas a garantizar la seguridad de las redes y los sistemas de información; b) si las medidas de gestión de riesgos de ciberseguridad se basan en un planteamiento que abarque todos los riesgos; y c) si las medidas de gestión de riesgos de ciberseguridad incluyen los requisitos específicos de seguridad enumerados en el art. 21, esto es, medidas como políticas de seguridad de los sistemas de información y análisis de riesgos, la gestión de incidentes, la continuidad de la actividad, la gestión de crisis, la seguridad de la cadena de suministro y las políticas y procedimientos relativos al uso de criptografía y, en su caso, de cifrado. Por otra parte, en relación con los requisitos de comunicación de información, se establecen

27 Comunicación de la Comisión. Directrices de la Comisión sobre la aplicación del artículo 4, apartados 1 y 2, de la Directiva (UE) 2022/2555, DO C de 18.9.2023, 2.

28 Reglamento (UE) 2022/2554 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre la resiliencia operativa digital del sector financiero y por el que se modifican los Reglamentos (CE) 1060/2009, (UE) 648/2012, (UE) 600/2014, (UE) 909/2014 y (UE) 2016/1011, DO L 333, 27.12.2022, 1.

reglas precisas sobre notificación de incidentes significativos a los CSIRT, las autoridades competentes y los destinatarios, la necesidad de un enfoque en varias etapas y plazos, el contenido de la obligación de notificación y el acceso inmediato a la misma.

B) Los efectos de la equivalencia no solo han de comprender las medidas de gestión de riesgos y notificación de incidentes significativos sino, también, las medidas de supervisión y ejecución de las obligaciones, aunque se opera sobre la base de la cooperación entre las autoridades competentes en el marco de la Directiva 2022/2555 y las autoridades competentes de los actos sectoriales de la UE. Las directrices se ocupan de explicar la situación específica en relación con la estrategia nacional de ciberseguridad, la designación de los CSIRT, los marcos nacionales de gestión de crisis y la EU-CyCLONe y la exclusión de la aplcación de algunas disposiciones.

5.3. Reglamento de Ejecución 2024/2690

La Directiva 2022/2555 prevé en su art. 21.5 que la Comisión adoptará o podrá adoptar actos de ejecución para establecer los requisitos técnicos, metodológicos y sectoriales de las medidas de gestión de los riesgos de ciberseguridad a más tardar el 17 de octubre de 2024[29]. Para ello habrá de seguir, en la mayor medida posible, las normas europeas e internacionales y las especificaciones técnicas pertinentes en colaboración con el Grupo de Cooperación y ENISA. Por su parte el art. 23.11 relativo a las obligaciones de notificación dispone que la Comisión podrá adoptar actos de ejecución para especificar el tipo de información, el formato y el procedimiento de las notificaciones y comunicaciones realizadas conforme a esta disposición. El Reglamento de Ejecución 2024/2690 de la Comisión es la consecuencia de ambos mandatos[30].

29 Los proveedores respecto de los cuales ha de adoptar esos actos son: los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, los proveedores de servicios de computación en nube, los proveedores de servicios de centros de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados, los proveedores de servicios de seguridad gestionados, así como los proveedores de mercados en línea, de motores de búsqueda en línea y de plataformas de servicios de redes sociales, y los prestadores de servicios de confianza. Podrá adoptarlos en relación con el resto de los sujetos incluidos dentro del ámbito de aplicación de la directiva.

30 Reglamento de Ejecución (UE) 2024/2690 de la Comisión, de 17 de octubre de 2024, por el que se establecen las disposiciones de aplicación de la Directiva (UE) 2022/2555 en lo que respecta a los requisitos técnicos y metodológicos de las medidas para la gestión de riesgos de ciberseguridad y en el que se detallan los casos en que un incidente se considera significativo con respecto a los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, los proveedores de servicios de computación en nube, los proveedores de servicios de centro de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados, los proveedores

Atendiendo a la distinción realizada en los arts. 21.2 y 23.3, el objeto del Reglamento de Ejecución 2024/2690 es establecer los requisitos técnicos y metodológicos de las medidas de gestión de los riesgos de ciberseguridad y determinar los casos en los que un incidente se considerará significativo en relación con los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, los proveedores de servicios de computación en nube, los proveedores de servicios de centros de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados, los proveedores de servicios de seguridad gestionados, así como los proveedores de mercados en línea, de motores de búsqueda en línea y de plataformas de servicios de redes sociales,y los prestadores de servicios de confianza (art. 1). El anexo detalla esos requisitos técnicos y metodológicos contemplados en el artículo 2 del presente Reglamento incluyendo disposiciones precisas en relación con la política sobre la seguridad de los sistemas de redes y de información, la política de gestión de riesgos, la gestión de incidentes, la continuidad de las actividades y la gestión de las crisis, la seguridad de las cadenas de suministro, la seguridad en la adquisición, el desarrollo y el mantenimiento de sistemas de redes y de información, las orientaciones y procedimientos para evaluar la eficacia de las medidas de gestión de riesgos de ciberseguridad, la formación en seguridad y prácticas básicas de ciberhigiene, la criptografía, la seguridad de los recursos humanos, el control de accesos, la gestión de activos y la seguridad medioambiental y física.

El art. 2 exige que las entidades en cuestión garanticen un nivel de seguridad de los sistemas de redes y de información adecuado a los riesgos y apliquen los requisitos técnicos y metodológicos de las medidas para la gestión de riesgos de ciberseguridad teniendo debidamente en cuenta tres aspectos principales: a) el grado de exposición a los riesgos; b) su tamaño; y c) la probabilidad de que se produzcan incidentes y la gravedad de los mismos, especialmente, sus repercusiones sociales y económicas. El concepto de incidente significativo se define en el art. 3 en función de la presencia de uno o varios de los criterios especificados en esa disposición[31]. Los incidentes no considerados significativos individualmente pueden adquirir esa calificación considerados conjuntamente como incidente recurrente y, por tanto, como

de servicios de seguridad gestionados, los proveedores de mercados en línea, motores de búsqueda en línea y plataformas de servicios de redes sociales, y los proveedores de servicios de confianza, DO L de 18.10.2024, 1.

31 Esos criterios son: a) que el incidente haya causado o pueda causar a la entidad pertinente pérdidas financieras directas superiores a 500.000 EUR o al 5 % de su volumen de negocios total anual en el ejercicio financiero anterior; b) que el incidente haya causado o pueda causar la exfiltración de secretos comerciales; c) que el incidente haya causado o pueda causar la muerte de una persona física; d) que el incidente haya causado o pueda causar daños considerables a la salud de una persona física; e) que se haya producido efectivamente el acceso presuntamente malicioso y no autorizado a los sistemas de redes y de información; f) que el incidente cumpla los criterios establecidos en el artículo 4; y g) que el incidente cumpla uno o varios de los criterios establecidos en los artículos 5 a 14.

un incidente significativo, cuando cumplan todos los criterios siguientes: a) que se hayan producido al menos dos veces en un plazo de seis meses; b) que tengan la misma causa fundamental aparente; y c) que cumplan colectivamente los criterios establecidos en el art. 3 (art 4).

Sobre esa base general, el resto de las disposiciones de este reglamento de ejecución concretan el significado del concepto de incidente significativo respecto de los proveedores de servicios del sistema de nombres de dominio (art. 5)[32], los registros de nombres de dominio de primer nivel (art. 6)[33], los proveedores de servicios de computación en nube (art. 7)[34], los proveedores de servicios de centros de datos (art. 8)[35], los proveedores de redes de distribución de contenidos (art. 9)[36], los proveedores de servicios gestionados y los

32 Un incidente será significativo cuando: a) un servicio de resolución de nombres de dominio recursivo o autoritativo esté completamente indisponible durante más de treinta minutos o, durante más de una hora, el tiempo medio de respuesta de un servicio de resolución de nombres de dominio recursivo o autoritativo a las solicitudes de DNS sea superior a diez segundos; o b) la integridad, confidencialidad o autenticidad de los datos se vean comprometidas, excepto en los casos en que los datos de menos de 1.000 nombres de dominio gestionados por el proveedor de servicios DNS, que representen como máximo el 1 % de los nombres de dominio gestionados por el proveedor de servicios DNS, no sean correctos debido a fallos de configuración.

33 Los criterios son: a) que un servicio autoritativo de resolución de nombres de dominio esté totalmente indisponible; b) que, durante más de una hora, el tiempo medio de respuesta de un servicio autoritativo de resolución de nombres de dominio a las solicitudes de DNS sea superior a diez segundos; o c) que la integridad, confidencialidad o autenticidad de los datos almacenados, transmitidos o tratados en relación con el funcionamiento técnico del dominio de primer nivel se vean comprometidas.

34 Para ser un incidente significativo, se requiere que: a) el servicio de computación en la nube esté completamente indisponible durante más de treinta minutos o con disponibilidad limitada para más del 5 % de los usuarios o de más de un millón de sus usuarios durante más de una hora; o b) que la integridad, confidencialidad o autenticidad de los datos se vean comprometidas como consecuencia de una acción presuntamente malintencionada o afecte a más del 5 % de los usuarios o a más de un millón de usuarios de estos servicios en la UE.

35 Para los proveedores de servicios de centros de datos, un incidente se considerará significativo cuando cumpla alguno de los siguientes criterios: a) que el servicio del centro de datos esté totalmente indisponible o la disponibilidad esté limitada durante más de una hora; b) que la integridad, confidencialidad o autenticidad de los datos se vean comprometidas como consecuencia de una acción presuntamente malintencionada; o c) que se encuentre en riesgo el acceso físico a un centro de datos gestionado por el proveedor en cuestión.

36 En este caso, para ser significativo, se requiere que: a) una red de distribución de contenidos esté completamente indisponible durante más de treinta minutos o la disponibilidad limitada para más del 5 % de los usuarios o de más de un millón de sus usuarios en la Unión durante más de una hora; o b) que la integridad, confidencialidad o autenticidad de los datos se vean comprometidas como consecuencia de una acción presuntamente malintencionada o que afecte a más del 5 % de los usuarios o a más de un millón de usuarios de estos servicios en la UE.

proveedores de servicios de seguridad gestionados (art. 10)[37], los proveedores de mercados en línea (art. 11)[38], los proveedores de motores de búsqueda en línea (art. 12)[39], los proveedores de plataformas de servicios de redes sociales (art. 13)[40] y los prestadores de servicios de confianza (art. 14)[41]. Hay dos criterios básicos: la afectación en cuanto a la disponibilidad del servicio y el hecho de que puedan verse comprometidas la integridad, confidencialidad o autenticidad de los datos almacenados, transmitidos o tratados por estos proveedores. El Reglamento de Ejecución 2024/2690 establece los requisitos técnicos y metodológicos de las medidas para la gestión de riesgos de ciberseguridad sobre la base de normas internacionales y europeas —como las normas ISO/IEC 27001, ISO/IEC 27002 y ETSI EN 319401— o especificaciones técnicas, como CEN/TS 18026: 2024, pertinentes para la seguridad de las redes y los sistemas de información.

37 Los criterios son: a) que un servicio esté completamente indisponible durante más de treinta minutos o con la disponibilidad limitada para más del 5 % de los usuarios o de más de un millón de sus usuarios en la Unión, si esta cifra es menor, durante más de una hora; o b) que la integridad, confidencialidad o autenticidad de los datos se vean comprometidas como consecuencia de una acción presuntamente malintencionada o que afecte a más del 5 % de los usuarios o a más de un millón de usuarios de estos servicios en la UE.

38 En este caso se trata de un incidente significativo si: a) un mercado en línea esté completamente indisponible o con disponibilidad limitada para más del 5 % de los usuarios o para más de un millón de sus usuarios en la Unión, si esta cifra es menor; o b) la integridad, confidencialidad o autenticidad de los datos se vea comprometida como consecuencia de una acción presuntamente malintencionada o afecte a más del 5 % de los usuarios o a más de un millón de usuarios del mercado en línea en la Unión

39 Esto ocurre si un motor de búsqueda en línea: a) está completamente indisponible o con disponibilidad limitada para más del 5 % de los usuarios o para más de un millón de sus usuarios en la Unión, si esta cifra es menor; o b) se ven comprometidos la integridad, confidencialidad o autenticidad de los datos como consecuencia de una acción presuntamente malintencionada o afectados más del 5 % de los usuarios o a más de un millón de usuarios de los servicios del motor de búsqueda en línea en la Unión.

40 En este caso, ha de cumplir alguno de estos criterios para ser considerado significativo: a) que la plataforma de servicios de redes sociales esté completamente indisponible o con disponibilidad limitada para más del 5 % de los usuarios o para más de un millón de sus usuarios en la Unión; o b) que se vean comprometidas la integridad, confidencialidad o autenticidad de los datos por una acción presuntamente malintencionada o afecte a más del 5 % de los usuarios o a más de un millón de sus usuarios en la Unión.

41 Un incidente será significativo si: a) el servicio de confianza está completamente indisponible durante más de veinte minutos o está indisponible para los usuarios o las partes usuarias durante más de una hora; b) más del 1 % de los usuarios o partes usuarias o más de 200.000 usuarios o partes usuarias en la Unión se ven afectados por la disponibilidad limitada; c) el acceso físico a una zona en las que estén situados los sistemas de redes y de información se ve comprometido; o d) si se ven comprometidos la integridad, confidencialidad o autenticidad de los datos almacenados, transmitidos o tratados en relación con la prestación de un servicio de confianza.

Para terminar, en España, el Centro Criptológico Nacional publica en agosto de 2024 el Perfil de Cumplimiento Específico CCN-STIC 892[42]. Este Perfil de Cumplimiento Específico para organizaciones en el ámbito de aplicación de la Directiva NIS2 (PCE-NIS2) pretende dar respuesta a la necesidad de cumplir esta normativa por parte de aquellas organizaciones incluidas en el ámbito de aplicación del Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad (ENS)[43]. El ámbito de aplicación de este perfil de cumplimiento se extiende a todo el sector público nacional, las entidades del sector privado que presten servicios o provean soluciones a las entidades del sector público y las entidades titulares de sistemas que tratan información clasificada en base a la Ley 9/1968, de 5 de abril, sobre secretos oficiales y demás normativa derivada. Este perfil de cumplimiento, que se alinea con las directrices de ENISA y tiene en cuenta los controles de otras normas internacionales sobre gestión de la seguridad de la información, cumple varias funcionalidades: a) describe los principales requisitos de la Directiva NIS2; b) analiza esos requisitos en relación con los vigentes en materia de ciberseguridad del ENS, con el fin de identificar las medidas de seguridad necesarias del Anexo II, así como aquellas en las que se requieren determinados refuerzos; y c) establece una metodología para alinear otros marcos de seguridad con la Directiva NIS2.

VI. Conclusiones

La UE aspira a convertirse en un referente a nivel mundial mediante la adopción de políticas y normativas obligatorias, claras y rigurosas en materia de ciberseguridad. Aunque ya existían la Directiva 2016/1148 y el Reglamento 2019/881, denominado Reglamento de Ciberseguridad[44], a partir de la

42 El perfil de cumplimiento específico se define como «un conjunto de medidas de seguridad, comprendidas o no en el Real Decreto 311/2022, de 3 de mayo, que, como consecuencia del preceptivo análisis de riesgos, resulten de aplicación a una entidad o sector de actividad concreto y para una determinada categoría de seguridad». En concreto, las Guías CCN-STIC, del Centro Criptológico Nacional, podrán establecer perfiles de cumplimiento específicos para entidades o sectores concretos, que incluirán la relación de requisitos, medidas y refuerzos que en cada caso resulten aplicables, o los criterios para su determinación (https://www.ccn-cert.cni.es/es/800-guia-esquema-nacional-de-seguridad/7235-ccn-stic-892-perfil-de-cumplimiento-especifico-para-organizaciones-en-el-ambito-de-aplicacion-de-la-directiva-nis2-pce-nis2/file.html).

43 https://www.ccn-cert.cni.es/es/800-guia-esquema-nacional-de-seguridad/7235-ccn-stic-892-perfil-de-cumplimiento-especifico-para-organizaciones-en-el-ambito-de-aplicacion-de-la-directiva-nis2-pce-nis2/file.html

44 Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) n526/2013 (Reglamento sobre la Ciberseguridad), DO L 151, 7.6.2019, 15

Estrategia Digital Europea[45] y de la Estrategia Europea de Ciberseguridad[46] se acelera el proceso de digitalización europea en cuyo núcleo se sitúa la seguridad digital y la resiliencia. La Directiva 2022/2555 comparte protagonismo con la Directiva 2022/2557 relativa a la resiliencia de las entidades críticas[47] y con el Reglamento 2022/2554 sobre la resiliencia operativa digital del sector financiero[48]. Ese marco normativo se completa en 2024 con Reglamento (UE) 2024/2747 del Parlamento Europeo y del Consejo, de 9 de octubre de 2024, por el que se establece un marco de medidas relativas a una emergencia del mercado interior y a la resiliencia de dicho mercado[49] y con el Reglamento (UE) 2024/2847 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, relativo a los requisitos horizontales de ciberseguridad para los productos con elementos digitales, conocido como Reglamento de Ciberresiliencia[50]. En ese contexto general, la Directiva 2022/2555 constituye una reforma y un avance sustancial, en términos subjetivos, materiales y funcionales respecto de su predecesora porque pone fin a la fragmentaciçon característica del modelo anterior optando por un enfoque integral y global de la ciberseguridad. Entre las novedades introducidas por la Directiva NIS 2 merecen una atención particular las disposiciones que afectan a responsabilidades individuales de los encargados de la dirección o representación de las entidades o proveedores.

Entre los aspectos críticos hay que destacar, en primer lugar, la propia complejidad técnica de la directiva, sus disposiciones y el modelo resultante. No es un problema carente de solución, pero requiere la realización de estu-

45 Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Configurar el futuro digital de Europa. COM (2020) 67 final, Bruselas 19.2.2020.

46 Comunicación Conjunta al Parlamento Europeo y al Consejo. La Estrategia de Ciberseguridad de la UE para la Década Digital. JOIN (2020) 18 final, Bruselas 16.12.2020, 4.

47 Directiva (UE) 2022/2557 del Parlamento Europeo y del Consejo de 14 de diciembre de 2022 relativa a la resiliencia de las entidades críticas y por la que se deroga la Directiva 2008/114/CE del Consejo, DO L 333, 27.12.2022, 164

48 Reglamento (UE) 2022/2554 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, sobre la resiliencia operativa digital del sector financiero y por el que se modifican los Reglamentos (CE) 1060/2009, (UE) 648/2012, (UE) 600/2014, (UE) 909/2014 y (UE) 2016/1011, DO L 333, 27.12.2022, 1.

49 Reglamento (UE) 2024/2747 del Parlamento Europeo y del Consejo, de 9 de octubre de 2024, por el que se establece un marco de medidas relativas a una emergencia del mercado interior y a la resiliencia de dicho mercado y se modifica el Reglamento (CE) 2679/98 del Consejo (Reglamento de Emergencia y Resiliencia del Mercado Interior), DO L de 8.11.2024.

50 Reglamento (UE) 2024/2847 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, relativo a los requisitos horizontales de ciberseguridad para los productos con elementos digitales y por el que se modifica el Reglamento (UE) 168/2013 y el Reglamento (UE) 2019/1020 y la Directiva (UE) 2020/1828 (Reglamento de Ciberresiliencia), DO L de 20.11.2024, 1).

dios que permitan clarificar el alcance y el contenido de las medidas, obligaciones y requisitos impuestos en cada caso. En segundo término, la brevedad de los plazos establecidos para poner en marcha las distintas medidas y proceder a la transposición obliga a los Estados, la Comisión y el resto de los organismos implicados a operar con extrema diligencia so pena de incurrir en incumplimientos en la ejecución de la directiva susceptibles de sanción. La fecha de transposición se fija para el 17 de octubre de 2024 con objeto de que los Estados apliquen las medidas adoptadas a esos efectos, a más tardar, el día 18, fecha en la que deroga la Directiva NIS 1. Pocos países han culminado con éxito esa operación en la fecha determinada por la directiva. En general, la Directiva NIS 2 está trufada de plazos exiguos que requieren una intensa actividad por parte de aquellos distintos actores y aumentan las posibilidades de retrasos en su aplicación. No obstante, dada la extensión y la precisión de sus normas, su práctica reproducción habría agilizado el proceso y permitido a los Estados concentrarse en el cumplimiento de las restantes obligaciones operativas de esta normativa. A ello hay que sumar el hecho de que la Comisión ha adoptado actos de ejecución que clarifican sus disposiciones desde las Directrices sobre la aplicación del art. 3.4[51] y el art. 4.1 y 2[52], al propio Reglamento de Ejecución (UE) 2024/2690 de la Comisión, de 17 de octubre de 2024, por el que se establecen las disposiciones de aplicación de la Directiva (UE) 2022/2555 en lo que respecta a los requisitos técnicos y metodológicos de las medidas para la gestión de riesgos de ciberseguridad y en el que se detallan los casos en que un incidente se considera significativo[53].

El modelo NIS 2 resulta más coherente y racional que el establecido en la Directiva NIS 1, acercándose en mayor medida a la idea de un enfoque integral de la ciberseguridad. Este enfoque integral se manifiesta desde un punto de vista subjetivo, material y normativo. En primer lugar, la Directiva NIS 1 excluía expresamente de su ámbito de aplicación a los proveedores de redes públicas o de servicios de comunicaciones electrónicas, los proveedores de

51 Comunicación de la Comisión. Directrices de la Comisión sobre la aplicación del artículo 3, apartado 4, de la Directiva (UE) 2022/2555. C (2023) 6070 final, Bruselas, 13.9.2023.

52 Comunicación de la Comisión. Directrices de la Comisión sobre la aplicación del artículo 4, apartados 1 y 2, de la Directiva (UE) 2022/2555. C (2023) 6068 final, Bruselas, 13.9.2023.

53 Reglamento de Ejecución (UE) 2024/2690 de la Comisión, de 17 de octubre de 2024, por el que se establecen las disposiciones de aplicación de la Directiva (UE) 2022/2555 en lo que respecta a los requisitos técnicos y metodológicos de las medidas para la gestión de riesgos de ciberseguridad y en el que se detallan los casos en que un incidente se considera significativo con respecto a los proveedores de servicios de DNS, los registros de nombres de dominio de primer nivel, los proveedores de servicios de computación en nube, los proveedores de servicios de centro de datos, los proveedores de redes de distribución de contenidos, los proveedores de servicios gestionados, los proveedores de servicios de seguridad gestionados, los proveedores de mercados en línea, motores de búsqueda en línea y plataformas de servicios de redes sociales, y los proveedores de servicios de confianza, DO L de 18.10.2024, 1.

servicios de confianza y las infraestructuras críticas respecto de las cuales se aplicaban las disposiciones específicas recogidas en sus respectivas normativas. Como consecuencia de ello, no solo se producía una fragmentación normativa, sino también subjetiva y funcional porque existían normas diferentes en función del sujeto y de la naturaleza de los servicios prestados en cada caso. La Directiva NIS 2 incluye ahora en su ámbito de aplicación a todos estos proveedores, junto con otros igualmente implicados en materia de ciberseguridad.

En segundo lugar, abundando aún más en el problema de fragmentación, como se ha visto, la Directiva NIS 1 se aplicaba a los operadores y a ciertos proveedores y de modo diferente en cada caso. Al optar por un criterio material, importando más el tipo de servicio que su titularidad, un mismo operador o proveedor había de ajustarse a modelos distintos en función de la actividad o el servicio en cuestión. La Directiva NIS 2 establece un modelo de aplicación subjetiva en el que importa tanto el tipo como el tamaño de la entidad, que permite corregir los problemas derivados de esa situación previa.

En tercer lugar, la fragmentación normativa resultante de la existencia de regímenes jurídicos diversos se corrige mediante tres medidas principales: a) La definición de NIS 2 como la norma general de aplicación; b) La previsión en su art. 4 de un sistema para la integración de los actos sectoriales; y c) La reforma de las disposiciones de su predecesora homogeneizando su alcance y contenidos. Por todo ello, es previsible que la Directiva NIS 2 suponga un punto de inflexión definitivo en materia de ciberseguridad.

Un enfoque integral en materia de ciberseguridad es, realmente, fundamental para los titulares de esos diferentes tipos de prestaciones —porque han de cumplir los diferentes requisitos establecidos en cada caso—, para los usuarios —porque han de conocer las obligaciones exigidas en cada supuesto y sus correspondientes derechos—, para las administraciones —que han de gestionar los procedimientos y velar por su cumplimiento— y, en definitiva, para el conjunto de una sociedad y una economía basadas en el conocimiento.

15.

EL DEBATE SOBRE LOS DERECHOS DE LA INTELIGENCIA ARTIFICIAL

Manuel Jesús Rodríguez Puerto
Catedrático de Filosofía del Derecho
Universidad de Cádiz

SUMARIO: I. La personalidad electrónica. II. La realidad de la IA. III. La imposibilidad de derechos para IA.

I. La personalidad electrónica

En la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL), el Principio 59 del capítulo dedicado a los Principios generales relativos al desarrollo de la robótica y la inteligencia artificial para uso civil promueve:

«Crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente».

Aparecía así en un texto que no era normativo, pero sí procedente de una autoridad política, el concepto de personalidad electrónica, que ha ocasionado un debate considerable. Pero antes de referirme a él es precisa una aclaración previa: la distinción entre robot e inteligencia artificial; no es una cuestión fácil, porque el concepto mismo de robot no está claro. En cualquier caso, el término robot debe ser empleado para máquinas con cierta corporeidad física que interactúan con su entorno, y que a veces pueden estar dota-

dos de programas de inteligencia artificial[1]. En las páginas que siguen me ocuparé de la inteligencia artificial (IA) en general, no específicamente de los robots, aunque buena parte de las reflexiones sobre la personalidad electrónica están pensadas sobre todo para los robots; en cualquier caso, lo escrito sobre robots es aplicable a la IA.

Con independencia de la condición robótica de la máquina de la que se trate, la inicial intención de la UE hacia el reconocimiento de la personalidad electrónica no ha prosperado. La regulación de la responsabilidad civil por daños provocados por la IA no ha adoptado la solución de otorgar personalidad a la máquina, sino establecer un régimen de responsabilidad repartida entre los diferentes autores y usuarios de la tecnología. La reciente *Directiva (UE) 2024/2853 del Parlamento y del Consejo sobre responsabilidad por los daños causados por productos defectuosos* incluye los programas de IA en los productos que regula y atribuye la responsabilidad al fabricante con una serie de requisitos y circunstancias; el poder europeo rechaza, por tanto, la posibilidad de que la IA tenga algún tipo de imputación de actuaciones jurídicas y, desde luego, elimina cualquier atisbo de personalidad electrónica. Esa era ya la tendencia en la UE en los proyectos de Directivas, a pesar de la inicial manifestación del Parlamento[2]. En otros ámbitos jurídicos, como el Derecho Penal, está presente la misma tendencia: la responsabilidad sólo es atribuible a los seres humanos que diseñan y manejan estas herramientas tecnológicas[3].

No obstante, algunos juristas, como Moisés Barrio, mantienen aún la viabilidad de la personalidad electrónica para solucionar algunas de las cuestiones que genera el empleo de la IA con relevancia jurídica. La propuesta no deja de ser atendible, a pesar de los problemas que genera, porque, al fin y al cabo, la persona jurídica es una ficción y podría ser admisible situar en un programa un punto de imputación para resultados jurídicos. Barrio explica que esa atribución de personalidad, que puede implicar incluso el reconocimiento de algunos derechos, es limitada sólo a algunas manifestaciones de robots dotados de IA y limitada también al número de derechos que podrían ser reconocidos, referidos sólo a lo preciso para solventar los problemas derivados de la institución jurídica de la que se trate[4]. No obstante, las dificulta-

1 *Cfr*. Barrio Andrés, M., «Del Derecho de Internet al Derecho de los robots», en Barrio, M., *Derecho de los robots*., Wolters Kluwer, Madrid, 2018, pp. 63 y ss.

2 *Vid*. Castilla Barea, M., «¿Qué queda hoy de la Resolución del parlamento europeo de 20 de octubre de 2020 sobre un régimen de responsabilidad civil en materia civil en materia de inteligencia artificial?», en Cervilla, M.ª D. / Ballesteros, A., (Eds.), *Temas actuales de Derecho Privado II*, Cizur Menor, Aranzadi, 2023, pp. 121 y ss.

3 *Vid*. los capítulos contenidos en Barrio Andrés, M. (Dir.), *Derecho de los Robots*, cit., pp. 131 y ss.

4 *Vid*. Barrio Andrés, M., «De nuevo sobre la Persona Jurídica», *Inteligencia Artificial*, 27, 2024, pp. 65 y ss. Fernando Llano expone los diferentes aspectos del debate sobre la

des de esa conceptuación de la IA hacen preferible buscar la solución fuera de la imputación de responsabilidad a la máquina.

Más allá de esta cuestión de técnica jurídica, desde hace unos años en un plano académico ha surgido el debate acerca del reconocimiento de derechos a la IA, aunque con frecuencia, como acabo de decir más arriba, se hable de derechos de los robots. Este asunto es diferente. No se trata de resolver problemas de responsabilidad civil, sino de la cuestión más profunda de la titularidad de los derechos, de la subjetividad. En efecto, la noción de derechos, de derechos subjetivos, presenta una variedad de dimensiones jurídicas. Abarca poderes y facultades de actuar concretas, diversificadas en los diferentes problemas y materias reguladas por el derecho civil, mercantil, administrativo, etc. Pero en su núcleo más esencial, el concepto de derechos implica la existencia de un sujeto dotado de un estatus especial, la dignidad, que le hace acreedor de un respeto especial. La historia de los últimos siglos de la ciencia jurídica y la política en Occidente ha visto la lucha por perfilar la figura de la persona como ser dotado de dignidad y de derechos. Que actualmente esa personalidad, como titular de respeto especial, quiera extenderse a las máquinas de IA obliga a una reflexión detenida desde el plano de la filosofía jurídica. Y el primer paso es averiguar si la IA presenta alguna similitud con la personalidad humana que justifique esa extensión de los derechos.

II. La realidad de la IA

La búsqueda de esa similitud a la que acabo de aludir parece ya resuelta, porque precisamente el nombre de IA indicaría la presencia de una cierta racionalidad compartida con lo humano. Sin embargo, un acercamiento a la realidad de esta tecnología mostrará que estamos ante una máquina y difícilmente puede poseer un estatus que la haga merecedora de derechos. Responder a esta cuestión requiere tener claro qué es la IA.

Sintetizando mucho es posible decir que ésta es un programa que trata información. Y también debemos tener presente la diferencia establecida por los ingenieros entre IA en sentido débil y en sentido fuerte. La primera es la que existe ahora, capaz de realizar algunas tareas que imitan dimensiones de la inteligencia humana, aunque no es una inteligencia en sentido auténtico; la IA fuerte o general es la que posee esas características. Tendría autoconciencia, podría realizar actividades muy diversas, empatizaría, etc. La posibilidad de diseñarla está hoy fuera del alcance de esta tecnología[5].

personalidad electrónica, como modo de solventar problemas y daños generados por máquinas de IA imprevisibles en *Homo ex Machina*., Tirant lo Blanch, Valencia, 2024, pp. 118 y ss.

5 *Cfr*. López de Mántaras, R., y Meseguer González, P., *Inteligencia artificial*, CSIC, Madrid, 2018, pp. 38-39, 41, 72-73, 152.

Algunos visionarios hablan de superinteligencia, aquella que superará al ser humano en capacidades. Ray Kurzweil es uno de sus principales adalidades y denomina «singularidad» al momento en el que la IA rebasará la inteligencia humana, un momento que considera cercano[6]. Kurzweil mezcla en su ensayo disquisiciones sobre IA propiamente dicha con la interacción de biología humana y tecnología para crear cyborgs que supuestamente superarían las limitaciones de la corporalidad humana. En cualquier caso, los ingenieros no señalan modo alguno de que se pueda conseguir tal cosa. Y es que la IA es tecnología, no es auténtica inteligencia.

Ante todo, es preciso recordar que es tecnología diversa. Efectivamente, la IA es una etiqueta que acoge diferentes tipos de herramientas, técnicas, programas, etc.[7]. A pesar de esa multiplicidad, es posible dividirla en dos bloques básicos[8]. El primero está compuesto por la IA basada en la lógica, es decir, en programas que emplean estructuras lógicas de diversa procedencia (deductiva, derrotable, difusa, no monótona, multivalente, etc.) para obtener resultados precisos desde las premisas iniciales introducidas por los programadores; el sistema funciona de manera automática aplicando reglas lógicas ante las cuestiones nuevas introducidas en él. Ocurre que estos sistemas tienen las ventajas y limitaciones de funcionamiento de la lógica. Como es bien sabido, esta es necesariamente formal, pero no sirve para crear argumentos y premisas iniciales. Garantiza la corrección de pasos en el razonamiento, pero no el de sus contenidos. Por otra parte, aunque las lógicas no deductivas han pretendido superar los problemas del deductivismo a la hora de afrontar problemas mediados por circunstancias y variaciones, en la realidad, los sistemas de IA basados en lógica son estructuras que formalizan a posteriori razonamientos humanos concretos.

La otra variante de IA es la basada en el manejo de datos, frecuentemente conectada con el análisis de datos masivos. Estos sistemas funcionan mediante la comparación de parámetros para establecer concordancias entre los introducidos inicialmente y los que componen los nuevos casos. Particularmente relevante en este sector es el aprendizaje automático, que posibilita a los algoritmos del sistema el entrenamiento ante éxitos y errores y reforzar la vía de soluciones que lleva al éxito, en función de las líneas programadas inicialmente. Algunas vertientes del aprendizaje automático, como el profundo, alcanzan una complejidad extraordinaria, de forma que los procesos de funcionamiento del algoritmo no son ya completamente previsibles por el programador. Por otra parte, estos sistemas también emplean elementos de lógica en su funcionamiento.

---

6 *Vid*. Kurzweil, R. *La Singularidad está cerca*. Trad. C. García. Lola Books, Berlín, 2012.

7 *Cfr*. Boden, M., *Inteligencia Artificial*, trad. de I. Pérez Parra, Turner Libros 2017, p. 29.

8 *Vid*. una exposición sintética de esa diferencia en Branting, K.; «Data-centric and logic-based models for automated legal problem solving», *Artificial Intelligence and Law*, 25, 2017, per totum.

Los modelos de IA más eficaces actualmente son los basados en datos, entre los que se encuentra la popular IA generativa o de propósito general. Es importante subrayar que esa calificación de general no implica que estas IA sean de IA en sentido fuerte o general: no tienen nada que ver con la inteligencia humana. No obstante, las características de la IA basada en datos han hecho pensar en una cierta autonomía del algoritmo. En realidad, no existe tal autonomía. Como ha señalado Ryan Calo sobre los robots (aunque sería aplicable a la IA en general), en el momento actual no son ya meros autómatas que actúen completamente programados. Se organizan a sí mismos, ya que son capaces de un comportamiento emergente, es decir, son sistemas que parten de una programación y se comportan de manera compleja y no completamente anticipable; en efecto, el programador no controla necesariamente dicho comportamiento[9]. Recordemos que el algoritmo, en la base de estos sistemas, es un conjunto de instrucciones dirigido a conseguir un resultado preciso en un plazo rápido mediante pasos formalizados. Su diseño no requiere intuiciones, ni conjeturas ni interpretaciones, como las instrucciones propias de la interacción humana. En cambio, los algoritmos de aprendizaje profundo exigen menos precisión[10]. Ahora bien, dicha emergencia no implica intención ni voluntad de actuar. A pesar de la literatura amplia que existe sobre los desarrollos de una IA pensante, la pretensión de estos programas no es pensar como humanos, sino superar problemas concretos mediante la formalización del lenguaje y la representación de conceptos complejos a partir de datos sencillos, algo difícil en el conocimiento abstracto e intuitivo que está en el ámbito humano[11].

Es necesario subrayar que la realidad denominada IA abarca sofisticados sistemas para procesar datos que reproducen algunas vertientes del pensar humano, pero sin conformar un auténtico pensar. Por ese motivo tales sistemas no pueden ser considerados agentes morales. El desarrollo de las virtudes, la captación de principios, la empatía, la comprensión de las exigencias de una situación requiere de unas características, que incluyen la corporalidad (auténtica, no la estructura mecánica del robot) y la coexistencia, que son irreproducibles en la IA[12]. Cuestión distinta es que el empleo de sistemas

9 *Cfr*. Calo, R., Robots as Legal Metaphors, *Harvard Journal of Law and Technology*, 30, 2016, pp. 227 y ss.

10 *Cfr*. Desai, R., y Kroll, J.A., Trust but Verify, *Harvard Journal of Law and Technology*, pp. 23-24.

11 *Cfr*. Bosch Rue, A., Casas Roma, J., y Lozano Bagen, T., *Deep Learning. Principios y fundamentos*. UOC, Barcelona, 2019, pp. 17 y ss. López de Mántaras y Meseguer sostienen que la IA no podrá ser como la humana, porque ésta requiere interacciones con entorno, procesos de socialización y culturación además de corporalidad. *Cfr*. López de Mántaras, R. / Meseguer, P., *Inteligencia artificial*, cit., p. 152.

12 *Cfr*. Wallach, W., y Valler, S., «Moral machines. From Value Alignment to Embodied Virtue», en *Ethics of Artificial Intelligence*. OUP, 2020, pp. 386 y ss. Misselhorn, C., «Artificial Moral Agents», en *The Cambridge Handbook of Responsible AI*. Cambridge University Press,

de IA para resolver problemas con relevancia moral y jurídica requiera una atención muy delicada de los patrones y criterios utilizados en la selección de las premisas y datos con los que construir y entrenar dichos sistemas. Pero es preciso recordar que el papel ineludible de la supervisión ética y jurídica corresponde a los autores humanos del programa. Ellos son los que deciden, porque la máquina dotada de IA no puede decidir[13]. Una decisión requiere de una serie de elementos. Implica la existencia de alternativas y elecciones y de un sujeto decisorio. En aquellas situaciones en las que el curso de un acontecimiento se deriva de manera necesaria e ineludible desde unas reglas, axiomas, principios, etc., y no cabe elegir entre alternativas, no hay una decisión. Pero la vida humana (lo que incluye el derecho) está lleno de situaciones en los que es preciso decidir, porque la solución al problema planteado en esa situación no es derivable automáticamente desde una premisa. La filosofía hermenéutica, a partir de una reelaboración de la doctrina aristotélica de la prudencia, ha puesto de manifiesto esta realidad en los asuntos humanos[14]. Es cierto que a veces todo se resuelve aplicando mecánicamente la regla, pero, otras muchas, no es así. La misma realidad encontramos en el derecho. En todas esas situaciones en las que es preciso seleccionar, elegir y decidir intervienen varios elementos que contribuyen a aumentar la complejidad del proceso. Es necesario que exista un sujeto decisor, alternativas para que opte entre ellas, las circunstancias que componen la decisión, los fines perseguidos al decidir, los medios, la valoración de fines y medios, y la jerarquía

2022, pp. 33 y ss. Peter Railton sostiene la importancia de diseñar máquinas sensibles a preocupaciones éticas capaces de responder a ese tipo de cuestiones. Serían «máquinas éticas» que no podrían adquirir competencias éticas de la misma forma que los humanos, porque las máquinas no comprenden, no tienen experiencias compartidas ni sentido común, no son capaces de aplicar principios. Pero este autor considera que sí es posible entrenar a la máquina de IA mediante una programación de aprendizaje a partir de datos masivos observando comportamientos humanos. Esto, a juicio de este autor, permite a las máquinas decidir y mantener autonomía. *Cfr*. Railton, P., «Ethical Learning, Natural and Artificial», en *Ethics of Artificial Intelligence*, cit., pp. 45 y ss., 64 y ss. Lo cierto es que ese tipo de entrenamiento puede servir para que la máquina detecte similitudes entre sus datos y los parámetros del caso nuevo. Eso no significa autonomía ni decisión. Intentos de utilizar lógica deóntica para estructurar decisiones éticas en la máquina de IA sólo consiguen formalizar a posteriori decisiones humanas. Es lo que ocurre por ejemplo en la propuesta de Arkoudas, K., Bringsford, S., y Bello P., «Toward Ethical Robots via Mechanized Deontic Logic», en *AAAI Fall Symposium in Machine Ethics*, The AAAI Press, , Menlo Park, CA, USA 2005, pp. 17-23.

13 John McCarthy, uno de los principales impulsores de la IA, ha planteado la posibilidad de que los sistemas de IA tengan libre arbitrio, pero McCarthy lo ve como capacidad para conseguir fines mediante el resultado del razonamiento desarrollado por la IA. Pero no explica en qué consiste la decisión ni el razonar de tales sistemas. *Vid*. Maccarthy, J., «Free Will – Even for Robots», *Journal of Experimental and Theoretical AI*, 2000, pp. 3 y ss.

14 *Vid*. Gadamer, H.G., *Verdad y método*. Trad. de A. Agud y R. de Agapito, Sígueme, Salamanca,1977, pp. 438 y ss.

de criterios utilizados en las valoraciones[15]. La presencia de esos elementos (cuyo peso en la decisión es variable según el caso en el que nos encontremos) imposibilita que la decisión humana pueda ser algo automático. Ahora bien, todas estas características son ajenas a la IA. Esta no es un sujeto decisor, porque carece de autoconciencia y, por tanto, de la capacidad de hacerse cargo de una situación problemática y de comprender los bienes y necesidades que están en juego. En suma, no es capaz de actuar, sólo de producir un resultado mecánico, aunque este no sea completamente previsible.

Dada esta carencia, de nada remotamente semejante a la personalidad, parece que no tiene mucho sentido plantear el reconocimiento de derechos a la IA. Sin embargo, no son pocas las propuestas que han surgido en ese sentido[16]. Y la respuesta es más compleja de lo que pudiera parecer inicialmente.

Esa complejidad se debe a varias causas procedentes del curso de la ciencia jurídica europea. Las más importantes son dos: el concepto mismo de persona en el Derecho y la noción de derecho en sentido subjetivo. Evidentemente no es posible en el espacio de estas páginas siquiera resumir los problemas planteados por esos conceptos, pero sí quiero mencionar algunos aspectos que son esenciales para responder a la cuestión de los derechos de la IA.

No voy a entrar en la discusión filosófica del concepto de persona, pero sí quiero mencionar que dentro del derecho la persona no siempre se ha identificado con el individuo. En el Derecho romano, los esclavos eran personas desde el punto de vista natural (por derecho natural todo son libres), pero no lo eran dentro del *ius civile*: desde el punto de vista del orden jurídico eran cosas[17]. A partir del siglo XVII las teorías jurídicas modernas erigieron a la persona individualmente considerada como clave del orden jurídico; esa persona era el centro de un conjunto de derecho naturales emanados de la libertad individual, derechos que debían ser protegidos por el Estado. Pero una lectura atenta de los integrantes de la denominada Escuela de Derecho Natural Moderno durante los siglos XVII y XVIII muestra que los auténticos sujetos de derecho eran los varones propietarios[18]. Por otra parte, diversos ordenamientos han extendido la personalidad jurídica a entes que no son individuos

15 *Cfr*. Robles Morchón, G., *Teoría del Derecho. Fundamentos de Teoría Comunicacional del Derecho*. Vol. III, Civitas, Cizur Menor 2021, pp. 74 y ss. Aunque este autor se refiera a la decisión específicamente jurídica, la caracterización que hace del proceso decisorio es extensiva al ámbito general de la acción humana.

16 Vid. un elenco (ya en 2021) de la bibliografía y de las tendencias existentes sobre la materia en Harris, J., y Anthis, J.R., «The Moral Consideration of Artificial Entities: A Literature Review», *Science and Engineerig Ethics*, 2021, per totum.

17 Vid. Megías Quirós, J.J., «Dignidad y derecho: de la Antigüedad a la Edad Media», *Anuario de Filosofía del Derecho*, Madrid, (34), 2018, pp. 303-331.

18 Vid, Carpintero Benítez, F., *Historia breve del derecho natural*, Colex, Madrid, 2000, pp. 167 y ss.

humanos, como las sociedades en el Derecho español. Cabría decir que es decisión de cada ordenamiento conferir personalidad jurídica a entes diversos, con independencia del hecho natural de la persona individual[19].

Por otra parte, la noción misma de derechos también es controvertida. La historia del desarrollo de los derechos subjetivos, es decir, del empleo de la palabra «derechos» para definir potestades o facultades de actuar en manos de un titular es complicada. En primer lugar, parece difícil la existencia de un orden jurídico que no otorgue poderes de actuar ni establezca limitaciones a tales poderes de acuerdo con un conjunto de normas (con independencia de su origen y forma). Cuestión diferente es que la conformación de tales facultades sea objeto de teorización por una ciencia jurídica establecida. Los romanos, por ejemplo, no tenían una ciencia jurídica preocupada por los conceptos generales; aunque su orden jurídico contuviera facultades, ellos no explicitaron un concepto de derecho (*ius*) como facultad. Fueron la teología y la ciencia jurídica europea siglos después las que elaboraron una concepción teórica del derecho como facultad (que en el siglo XIX los alemanes llamaron derecho subjetivo)[20]. Pero esas facultades pueden tener una configuración muy variada; pueden depender de normas legales previas o pueden surgir desde la misma naturaleza humana, tal y como hizo la Escuela de derecho natural moderno; en este segundo caso, los derechos consisten en libertades naturales propias de todo individuo. Una concepción que está en la base de los Derechos Humanos. En consecuencia, las facultades jurídicas pueden ser poderes concretos dentro de un ordenamiento (para ejercitar acciones judiciales, por ejemplo) o una exigencia general de respeto a la libertad y a la igualdad; estas exigencias que constituyen el núcleo de los Derechos Humanos, a su vez, no tienen una delimitación conceptual sencilla[21], pero están claramente ceñidos a los seres humanos. Con independencia de esa pluralidad de vertientes en las facultades jurídicas, no ha existido duda de que sólo eran atribuibles a sujetos humanos. Es cierto que la sociedad del Antiguo Régimen a veces se perseguía penalmente a los animales, pero en general la participación en el derecho sólo era posible para los seres humanos.

La complejidad del problema aumenta si tenemos en cuenta que la filosofía desde hace décadas niega la existencia de la persona como una categoría

19 *Cfr.* Robles Morchon, G., *Teoría del Derecho*. Vol. II, Cizur Menor, Thomson, 2015, pp. 912 y ss. Este autor explica que esta dependencia de la concepción jurídica de persona respecto de una decisión del ordenamiento es independiente de la existencia de la persona como una realidad filosófica y moral vinculada al sujeto individual.

20 Un estudio de las tendencias que llevaron al desarrollo del concepto de derecho subjetivo está en la obra colectiva de Carpintero Benítez, F., de Mora, E. V., Megías, J.J., y Rodríguez Puerto, M., *El derecho subjetivo en su historia*, Servicio de Publicaciones de la Universidad de Cádiz, Cádiz, 2003.

21 *Vid.* Pérez-Luño, A., *Derechos humanos, Estado de Derecho y Constitución*., Tecnos, Madrid, 2018, cap. I de la Parte I.

ontológica. Para las filosofías calificadas de postmodernas el acercamiento a lo real no es sino una constelación de interpretaciones construidas por juegos de lenguaje compartidos; no hay sustrato objetivo tras esa pluralidad de interpretaciones. El concepto de persona también forma parte de tales juegos lingüísticos: no hay nada tras ellos.

A partir de esa disolución de objetividades un sector de la ciencia jurídica no ve imposible conferir un cierto estatus de personalidad a las máquinas de IA. De la negación de la persona como realidad ontológica pasan a la atribución de cierto estatus personal a las máquinas. Un ejemplo de esta actitud es la de Günther Teubner.

Este autor parte de afirmar que la subjetividad es un artefacto social. Él piensa que el concepto moderno de persona (volveré más adelante a esta noción de persona) es una construcción diferente del ser humano concreto de carne y hueso. Teubner piensa que todo es una construcción sistemática para reducir incertidumbre; el concepto de individuo titular de derechos contribuye a esa labor permitiendo la comunicación fluida dentro de dicho sistema. Desde este punto de vista, Teubner señala que existe comunicación social entre agentes informáticos (software) y seres humanos; aunque los primeros carezcan de las competencias psíquicas de los humanos, emerge un sistema social genuino (en opinión de Teubner) en la interacción asimétrica entre humanos y no humanos[22].

Reparemos que, para Teubner, todo lo que conforma la sociedad es una construcción dirigida a la comunicación. La persona como sujeto individual portador de libertad plasmadas en derechos también es un «artefacto semántico de comunicación». Como acabo de señalar, la persona de carne y hueso, el hombre *(Mensch)* no es auténticamente protegido por la persona como construcción semántica[23]. Es una especie de sustrato ajeno a las reducciones de complejidad en que consiste los diferentes subsistemas sociales. Es función del subsistema atribuir derechos y deberes a sus actores; dado que existe interacción entre agentes electrónicos y personas, esos agentes aparecen como híbridos y tienen un estatus jurídico. Teubner se apoya en la obra del sociólogo Bruno Latour, que defiende un concepto de actuar *(agency)* que va más allá de las actuaciones humanas. A partir de la que denomina «teoría del actor red» explica que son las redes las que influyen y conforman la subjetividad; no tiene sentido diferenciar entre sujetos y objetos en el sentido tradicional, porque el sujeto es algo artificial y trazable

22 *Cfr.* Teubner, G., «Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law», *Journal of Law & Society*, 33, 2006, p. 14.

23 *Cfr.* Teubner, G., «Die anonyme Matrix: Zu Menschenechtsverletzungen durch, private transnationale Akteure», *Der Staat*, 44, 2006, pp. Por otra parte, ese ser humano carnal ajeno a las construcciones comunicativas sociales no tiene nada que ver con un concepto ontológico de persona, como ser dotado de dignidad que reclama respeto.

según las entidades en las redes que les permiten existir. El sujeto no es algo autónomo como la persona individual de la filosofía moderna[24].

En esta línea, Teubner afirma que los agentes electrónicos obtienen ahora una nueva personificación jurídica en contextos económicos y tecnológicos, que crean nuevos centros de acción, y que pueden ocasionar daños de manera que se requiere defensa contra agentes electrónicos. El derecho se abre, en consecuencia, a la entrada de agentes electrónicos dentro de nuevas dinámicas políticas. Y no es preciso para ello que tales agentes tengan una subjetividad jurídica completa[25].

Para entender la propuesta de Teubner es preciso tener en cuenta su teoría autopoiética de lo jurídico derivada de Niklas Luhmann. No es posible entrar ahora a analizarla, pero sí quiero destacar que el derecho aparece como un sistema autorreferencial: es jurídico lo que el propio sistema considera como tal: la subjetividad no es sino un atributo del sistema para regular determinados procesos comunicativos dentro del mismo. Es interesante señalar las relaciones entre la teoría social de Luhmann y la noción de sistema cibernético, es decir, con el tratamiento tecnificado de la información. Su Teoría de Sistemas es, en última instancia, un estudio de la manera en que unas estructuras procesan información para reducir su complejidad y poder manejarla[26]. Y, al fin y al cabo, la tecnología de la IA es una manifestación de cibernética que trabaja precisamente con esa realidad, con esa información, para ofrecer resultados.

A pesar de la complejidad de las doctrinas de Luhmann y Teubner, el resultado no es muy diferente del aportado por la Teoría Pura kelseniana.

Por otra parte, la idea de subjetividad personal está disuelta en este tipo de filosofía, de forma que resulta un tanto irrelevante adjudicar derechos a la IA. Un representante conspicuo de este tipo de planteamiento es Karl-Heinz Ladeur, quien niega la existencia de una identidad personal racional atemporal y fija en favor de una racionalidad y subjetividad descentralizadas en una pluralidad de juegos de lenguaje; en esta situación, el de sujeto, es un término signo para un conjunto de funciones heterogéneas[27]. Aunque no se refiera a robots ni máquinas con IA, la subjetividad disuelta que defiende podría aplicarse a éstas. Además, Ladeur admite la existencia de derechos fundamentales dentro de un orden jurídico a modo de espacios de expansión y juego de esas identidades fragmentadas.

---

24 *Cfr*. Latour, B., *Reassembling the Social. An Introduction to Actor-Network-Theory*., Oxford Universit Press, Oxford, 2007, pp. 116 y ss., 208 y ss.

25 *Cfr*. Teubner, G., *«Rights of Non-humans?»* P. 20.

26 *Cfr*. Pérez Luño, A.E., *Nuevas tecnologías, sociedad y derecho*, Fundesco, Madrid, 1987, pp. 30-33.

27 *Cfr*. Ladeur, K.H., *Postmoderne Rechtstheorie,* en Duncker & Humblot, Berlín, 1996, pp. 34, 37, 47 y 82 y ss., 151.

No obstante, las propuestas de atribución de derechos a las máquinas de IA no se sitúan en el plano de estas filosofías tan complejas (y complicadas), sino que buscan acercamientos más pragmáticos. No es sorprendente, porque buena parte de ellas viene del ámbito anglosajón más aficionado al pragmatismo que a las abigarradas sutilezas de la filosofía que la Europa continental produce desde hace décadas.

Es llamativo también que la defensa de derechos para la IA esté planteada como derechos para los robots. Ya aludí antes a las diferencias entre robot e IA. Claro está que estas propuestas de conceder derechos a los robots sólo se refieren a los que están dotados con programas de IA. El por qué los defensores de esta extensión hablan de robots es que la corporeidad y apariencia de éstos los hacen más reconocible por un humano. Esto es importante, porque ahí reside la clave de los derechos atribuidos a las máquinas.

No es fácil discernir el rigor de estas propuestas, porque a veces se basan en una noción de IA que no es real, al implicar unas capacidades que no posee esta tecnología, salvo en la ciencia ficción. En otras ocasiones, la propuesta es más modesta y más realista.

La tesis que sostienen autores como Mark Coeckelberg o David Gunkel es dejar de lado las consideraciones ontológicas sobre la personalidad tradicionales (como la kantiana) y la extensión de ésta a la IA. En este sentido, Coeckelberg expone un argumento de justificación social-relacional de la toma en consideración moral, basado en la apariencia surgida de una ontología ecológica y social. En vez de atender a las características ontológicas, esta propuesta se fija en las experiencias que suscita una relación concreta con el robot; la consideración moral no reside en el sujeto ni en el objeto, sino en la relación misma. Hombre y robot son entidades relacionales dentro de una ecología social[28]. La tesis es que, en la medida en que el ser humano entra en algún tipo de interacción con el robot, éste adquiere para él cierto estatus que requiere un reconocimiento.

Una vía similar adopta David Gunkel cuando plantea la extensión de derechos a los robots, en una doctrina de reconocimiento del otro basada parcialmente en la filosofía de Enmanuel Levinas. La apreciación humana de la presencia del robot, su relación física cara a cara, puede llevar a ese reconocimiento de un estatus moral del robot. Esa apreciación no lleva necesariamente al reconocimiento de derechos a esa entidad robótica, sino a una reflexión sobre la ética y el problema del otro que supere ontologías del individuo moderno y visiones antropocéntricas[29]. No obstante, posteriormente

28 *Cfr*. Coeckelberg, M., «*Robot Rights? Towards a social-relational justification moral consideration*», Ethics, Inf Technology, 12, 2010, pp. 209, 210, 215 y ss.

29 *Vid.* Gunkel, D.J., *Robot Rights*., MIT Press, Cambridge, 2018, pp. 159 y ss. *Vid.* también en ese sentido, Gellers, J.C., «*Rights for Robots. Artificial Intelligence, Animal and Environmental Law*», Routledge, Londres / Nueva York, 2021, cap. 5.

Gunkel explica que, al reconocer derechos a los robots, el Estado no pretende ocuparse de grandes cuestiones morales que afectan a la personalidad de la IA, sino que sólo pretende proporcionar un marco jurídico para integrar los robots en las prácticas jurídicas y necesidades sociales existentes[30]. Por eso, los derechos de los robots no se basan en especulaciones metafísicas, no son derechos naturales, sino en una concesión política mediante derechos legales[31]. Es, nuevamente, la idea pragmática de conceder derechos para soluciones ciertos problemas.

Reparemos en que estas propuestas sólo justificarían los derechos de una IA incorporada a una máquina con algún tipo de apariencia que resultara amable para el humano. Pero en realidad, lo que sería protegido con esos derechos robóticos es el afecto o la simpatía suscitada en un humano, no la dignidad propia del robot, como ocurre con los derechos de los humanos[32]. Y lo cierto es que algunas investigaciones empíricas de la opinión pública muestran cierta tendencia a reconocer el potencial interactivo de los robots e incluso cierta preferencia por otorgarles «derechos básicos» como el derecho a acceder a la energía y a la actualización de programas, y escasa aceptación de concederles derechos sociopolíticos como la propiedad[33]. Claro que se trata de sondeos encuestados que muestran mentalidades, pero meras aceptaciones empíricas no pueden servir de justificación para responder a la cuestión de la concesión de derechos a la IA[34].

III. La imposibilidad de derechos para IA

Es interesante observar que cuando autores como Coeckelberg o Gunkel estudian si el robot (entendamos ahora la IA) posee características ontológicas que le hagan merecedor de la categoría de personalidad dotada de derechos, mencionan filosofías occidentales de la persona individual, como el dualismo cartesiano, a las que consideran inadecuadas para afrontar esta cuestión, plantean (y rechazan) la aplicabilidad de teorías morales deontológicas y utilitaristas. Y optan finalmente por planteamientos relacionales que, supuestamente, superarían los defectos de una base ontológica para los

30 *Cfr.* Gunkel, D. J., «The Rights of Robots», en Alvarez- Nakagawa, A. / Douzinas, C. (Eds.), *Non-Human Rigths Critical Perspectives*. Edward Elgar, 2022, p. 7.

31 *Cfr.* Gunkel, D. J., «The Rights of Robots», en Alvarez-Nakagawa, A. / Douzinas, C. (Eds.), *Non-Human Rigths Critical Perspectives*. Edward Elgar, 2022, pp. 18-19.

32 Esa es la propuesta de protección especial de ciertos robots que generan un afecto particular en sus dueños aportada por Mamak, K., «Should criminal law protect love relations with robots?» *AI & Society*, 39, 2024, *per totum*.

33 *Vid.* De Graaf, M., Hindricks, F., y Hindricks, K., «Who Wants to Grant Robot Rights», *Frontiers in Robotics and AI*, 8, 2022, per totum.

34 *Cfr.* Müller, V.C., «Is it time for robot rights? Moral Status in Artificial Entities», *Ethics and Information Technologies*, 23, 2021, p. 582.

derechos. No dejan de llevar razón, porque esas filosofías han de ser matizadas para una comprensión cabal de los derechos de la persona. Y precisamente esos matices van en la vía de lo relacional, pero de una forma que imposibilita reconocer personalidad y, por tanto, derechos a la IA.

A la hora de hablar de los derechos y sus titulares es habitual pensar en la noción de persona que elaboró el pensamiento moderno, especialmente la filosofía kantiana. Esta no es sino la expresión más refinada de una corriente de pensamiento ya establecida en el siglo XVII y que identificó la persona con el individuo naturalmente libre e independiente, de forma que la única obligación legítima era aquella surgida del consentimiento. A la hora de establecer el concepto de derecho estos autores prescindieron de cualquier criterio normativo ajeno a la libertad misma; de esa libertad procedían los derechos que no eran sino plasmaciones de esa misma libertad. Hasta ese momento, la mentalidad jurídica europea estaba dominada por el *ius commune* (derecho romano reinterpretado por los juristas universitarios) que pensaban el derecho como la solución justa a determinados problemas e interacciones sociales; esas soluciones procedían de preceptos que habían de tener en cuenta circunstancias y exigencias que venían objetivizadas desde esas mismas situaciones y necesidades sociales, que muchas veces eran ajenas a la libertad individual. La propiedad privada, por ejemplo, aunque fuera concebida como una facultad para disponer de bienes, era entendida también por la mentalidad romanista como portadora de limitaciones en aras de utilidades sociales[35]. Efectivamente, aunque este orden jurídico de corte romanista reconociera facultades y poderes para actuar conforme a derecho no entendía la realidad jurídica como una constelación de derechos-poderes centrados en un individuo libre, sujeto autónomo que se situaba como ente dominante de todo lo real[36].

A fines del siglo XVIII, los teóricos del derecho que seguían la filosofía práctica kantiana habían refinado esta concepción de la persona que encontraba su plasmación en un sujeto portador de derechos, entre los que encontraba su primacía la propiedad; en función de la protección de esos derechos fue creado el Estado: su misión era proteger los derechos individuales, al modo de una sociedad aseguradora[37].

---

35 Sobre el desarrollo de la Escuela de Derecho Natural Moderno y su individualismo, *vid.* Carpintero Benítez, F., «Voluntarismo y contractualismo: una visión sucinta de la Escuela de derecho natural moderno», en *Persona y Derecho*, 13; 1985. Id., «La independencia y autonomía del individuo: los orígenes de la persona jurídica», en *Anuario de Filosofía del Derecho*, 4, 1987. Sobre las delimitaciones de la propiedad privada en cuanto a su uso, que ha de estar dirigido por el bien común, *cfr.* Megías Quirón, J.j., *Propiedad y derecho natural*, cit., pp. 146 y ss. Se refiere a Tomás de Aquino, pero esa explicación es extensible a la mentalidad premoderna sobre la propiedad.

36 René Descartes lo entendió mediante la distinción entre res cogitans y res extensa.

37 *Cfr.* Carpintero Benítez, F., *La Cabeza de Jano*, Universidad de Cádiz, Chiclana, 1989, pp. 145 y ss.

Esta filosofía fundadora de derechos tenía su sustento en una concepción de innegable relevancia: el concepto de persona basado en la dignidad y la autonomía. De esa forma, la dignidad, entendida como la capacidad de disposición y de autodeterminación, fue la base imprescindible de los derechos. La persona, para Kant, era un fin en sí mismo y la autonomía implicaba que ésta obedecía la ley moral por su propia racionalidad, no por la satisfacción de intereses o fines contingentes. Sin duda, esta noción de persona supuso históricamente una defensa de la persona frente a las desigualdades propias del Antiguo Régimen y su aportación para la defensa de la dignidad del ser humano tiene un valor indudable. Pero también es preciso reconocer que esta concepción tenía algunas insuficiencias que perjudicaban su misma pretensión liberadora.

En primer lugar, la igualdad que estaba en la base del derecho tenía sus limitaciones, porque la plasmación práctica de los derechos no alcanzaba a todos; los que realmente tenían capacidad de desenvolver sus poderes en la vida social eran los cabezas de familia propietarios[38]. No es sólo esta limitación la que me interesa destacar ahora, porque esta y otras características son relevantes para la cuestión de los derechos de la IA. En efecto, la dignidad en sentido kantiano es completamente inviable para justificar la personalidad de las máquinas, porque es evidente que carecen de esa autodeterminación moral que estaba en la base de su filosofía práctica. Conscientes de eso, los autores que buscan reconocer derechos a robots y otros artilugios rechazan el modelo de persona kantiana y buscan otras concepciones (incluso procedentes de culturas orientales o indígenas americanas) que permitan cierta dimensión personal, aunque el ente en cuestión carezca de las características del sujeto moderno occidental. Precisamente la filosofía de reconocimiento de una faz y de una actividad con la que la persona interactúa va en esa línea. Pero ese reconocimiento requiere también unas características de las que las máquinas carecen.

Jesús Ballesteros ha destacado la insuficiencia del concepto moderno de persona y ha alertado de que el modelo individualista en el que está basado resulta al final incompatible con la protección universal de los derechos. Porque el individuo diseñado por la modernidad está pensado para la consecución del propio interés, desde un punto de vista técnico y economicista, un centro de poder que busca expandirse de manera ilimitada[39]. En esa misma

38 *Vid.* Böhme, H.-J., *Politische Rechte des Einzelnen in der Naturrechtslehre des 18. Jahrhunderts und in der Staatstheorie des Frühkonstitutionalismus*, Duncker & Humblot, Berlín, 1993, pp. 14 y ss. Este autor sostiene que esa teoría cambió un poco a final de la Escuela, abriendo el arco de sujetos participantes, pero la lectura de Kant y sus discípulos muestra que para ellos tanto mujeres, como niños o no propietarios están excluidos del ámbito público.

39 *Cfr.* Ballesteros, J., «El individualismo como obstáculo a la universalidad de los derechos humanos», en *Persona y Derecho*, 41, 1999, p. 17. Id., *Postmodernidad: Decadencia o resistencia*. Tecnos, Madrid, 1989, pp. 26 y ss.

línea, Andrés Ollero ha señalado que para la modernidad los derechos se han conformado como «cotos exclusivos» del individuo protectores de la arbitrariedad según el modelo de la propiedad privada[40].

Ya he señalado que la personalidad moderna, sobre todo desde el punto de vista kantiano, está basada en la autonomía. Pero esa autonomía, recuerda Ballesteros, era en realidad una fuente de poder, que confunde autonomía con autarquía y autosuficiencia. Esto tiene unas consecuencias particularmente relevantes, porque limita la titularidad de los derechos a quienes tienen esa capacidad de manifestar esa autonomía mediante la capacidad de elegir de hecho, adultos autosuficientes[41]. Estas consideraciones sobre la subjetividad, identificada como una fuente de poder, resultan inadecuadas, a juicio de Ballesteros, para un reconocimiento verdaderamente universal de los derechos, porque excluyen a quienes no pueden elegir, a los que han perdido capacidad de autonomía, como los menores o las personas con discapacidad; la persona ve reducido así su ámbito[42].

Desde el punto de vista de la doctrina de los derechos humanos, es preciso superar esa unión típicamente moderna de individualismo y voluntarismo; frente a estas características, es necesario partir de otra consideración sobre la persona, que no la ve como ser aislado e insuficiente, sino precisamente como un ser que se abre a lo real a la coexistencia con los otros[43]. A primera vista, este planteamiento parece tener similitudes con las propuestas de quienes atribuyen derechos a la IA, que también hablan de reconocimiento de otros, de abandono del concepto de individuo moderno. Pero precisamente esa exigencia de apertura a la otredad es la que falta a la IA.

En efecto, el reconocimiento implica la coexistencia. Sergio Cotta ha convertido esa realidad en clave para entender lo humamo. Según Cotta, el ser humano participa de una serie de «formas coexistenciales», entre las que aparece el derecho. El yo, explica este filósofo, sólo se comprende a sí mismo en una «relación intramundana de pertenencia-extraneidad relativa a otros entes»[44]. De ahí se deriva una cierta copertenencia, porque el ser humano es consciente de su propia indigencia y finitud, y eso le lleva a la consciencia de que no puede ser causa de sí mismo: se desenvuelve en una dependen-

40 *Cfr*. Ollero Tassara, A., *Derechos humanos y metodología jurídica*. Centro de Estudios Connstitucionales, Madrid, 1989, pp. 152-153.

41 *Cfr*. Ballesteros, J. «El individualismo…», p. 17. *Cfr*, en ese sentido Fernández Ruiz-Gálvez, E., «Del universalismo abstracto a la universalidad concreta», *Persona y Derecho*, 41, 1999, pp. 62 y ss.

42 *Cfr*. Cabrera, L., «Autonomía y dignidad: la titularidad de los derechos», *Anuario de Derechos Humanos*, 3, 2003. pp. 9 y ss. Id., «Personas y seres humanos, ¿distinción o identidad?» *Ius et Scientia*, 3, 2017,

43 *Cfr*. Ballesteros, J. *Postmodernidad*…, cit., p. 147.

44 *Cfr*. Cotta. S., *El derecho en la existencia humana*. Trad. de I. Peidró. EUNSA, Pamplona,1987, p. 57.

cia recíproca que no es simplemente uso egoísta del otro, sino reciprocidad auténtica. Tampoco una disolución en la relación: el yo es un «ente en relación», que sólo se constituye en receptividad[45]. En el derecho, esta receptividad se manifiesta en reconocer al otro como igual en su «esencia ontológica»; de ahí surge el auténtico vínculo jurídico[46]. Como señala Andrés Ollero, el derecho consiste, en última instancia, en un ajustamiento de coexistencias y mutuas acogidas, no en el mero establecimiento de individualidades cerradas. De ahí que la determinación del contenido esencial de los derechos deba tener en cuenta como punto de partida esa coexistencialidad[47].

En la línea de Cotta, Ballesteros explicaba que el derecho ha de estar fundado en el respeto universal al otro y en la igualdad mínima[48]. Dicho de otra forma, el derecho ha de reconocer la igualdad de todos y, en consecuencia, ha de admitir que todos tienen derechos. Desde este punto de vista, no es el poder de expansión desde un centro individual, tal y como proclaman los modernos, la clave de los derechos, sino la vulnerabilidad e indigencia ontológicas propias de lo humano. Nuevamente es Ballesteros quien señala un camino de acercamiento a los derechos humanos que tengan en cuenta un sujeto interdependiente, integrado en contextos y ocupado en el cuidado, de otros y del entorno, en vez de en el cumplimiento de deseos egoístas[49]. Es en esa mutua aceptación en la que se cimenta la apreciación de la dignidad, que consiste justamente en la apreciación del respeto universal de todos los seres humamos, a causa de ese desvalimiento ontológico[50].

La filosofía hermenéutica que elaboró Hans-Georg Gadamer tiene algunas coincidencias con estas reflexiones sobre la coexistencialidad a las que acabo de aludir. Porque el sujeto hermenéutico tampoco es el individuo moderno descrito más arriba, centro de poder expansivo sobre una realidad inerte presta a la apropiación. En esa línea, Francesco Viola, jurista seguidor de esa línea hermenéutica, sostiene que la persona está caracterizada por la elección de bienes, y la privación de ese acceso vulnera su identidad. Y parte esencial de esos bienes es la exigencia de participar en una ética del discurso público; en ella se desenvuelve la autonomía personal mediante la relación dialógica según las razones propias del horizonte común de significado[51]. Y

45 *Cfr*. Cotta. S., *op. cit*., p. 71-74.

46 *Cfr*. Cotta. S., *op. cit*., p. 112.

47 *Cfr*. Ollero Tassara, A. *Derechos humanos y metodología jurídica*, cit., pp. 151 y ss.

48 *Vid*. *Cfr.* Ballesteros, J., «El derecho como no-discriminación y no-violencia», en Anuario *de Filosofía del Derecho*, 17, 1974, pp. 162 y ss.

49 *Cfr*. Ballesteros, J. «El individuo...», cit., pp. 19 y ss.

50 *Vid*. al respecto. Megías Quiros, J.J., «Dignidad, universalidad y derechos humanos», en *Anuario de Filosofía del Derecho*, 22, 2005,

51 *Cfr*. Viola, F., *De la naturaleza a los derecho*s., Trad. de V. Bellver, Comares, Granada, 1998, pp. 33, 345, 349.

ese horizonte común no es homogéneo, sino que supone una pluralidad de situaciones y condiciones existenciales que llevan a la expansión y florecimiento de una «vida buena» coordinada con otros[52].

Estas reflexiones sobre la coexistencialidad como base para los derechos sirven para mostrar la imposibilidad de otorgar derechos a la IA. Efectivamente, los intentos de extender esa figura a máquinas chocan con dificultades de orden ontológico. Y las posibles justificaciones buscadas en el reconocimiento o afectividad que pueda generar el contacto con la máquina sólo encubre la ausencia de un auténtico reconocimiento mutuo, que es la base del derecho. Resulta difícil imaginar cómo máquinas diseñadas según procedimientos de computación puedan desarrollar esos horizontes de interacción de los que habla Francesco Viola. Precisamente Vincent Müller destaca que los intentos de dotar de derechos a la IA fallan, porque parten de un concepto de agente individualizado, cuando la auténtica actuación requiere una contextualización e interacción[53]. Este autor considera que esa sería la vía para avanzar en los derechos de las máquinas. El problema reside en que esa contextualización no es posible con las máquinas, porque un programa computacional, por muy sofisticado que sea, no se desenvuelve mediante coexistencias surgidas de la comprensión mutua[54].

No obstante, los términos del debate son más amplios que el reconocer derechos a ciertas máquinas, como los robots. Creo que tiene un alcance que puede resultar peligroso, porque se inserta en una corriente actual desvirtuadora de los derechos. La conciencia extendida acerca del carácter valioso de ciertas realidades ha llevado a un afán quizá algo desmedido por el uso de la categoría de los derechos.

Dentro de esta tendencia, la Manga del Mar Menor en España tiene personalidad jurídica, aunque no esté del todo claro qué significa tal cosa ni la efectividad de tal concesión para la mejor protección de ese ecosistema[55]. Por supuesto, no se trata de negar el valor del medio ambiente y la necesidad de protección jurídica, sino de plantear la pertinencia desde el punto de vista de la ciencia jurídica, del empleo del concepto de derecho subjetivo para amparar los daños medioambientales, sobre todo teniendo en cuenta que pueden existir deberes (frente a los ecosistemas) sin que exista la contrapartida de un derecho subjetivo con un sujeto titular. Y en cualquier caso esa

52 *Cfr*. Viola, F. *op. cit*., p. 358.

53 *Cfr*. Müller, V., *op. cit*., p. 585.

54 *Vid*. Carabantes López, M., *Inteligencia artificial. Una perspectiva filosófica*., Escolar y Mayo, Madrid, 2016, per totum.

55 *Vid*. al respecto, Gutiérrez Castillo, V.L., «Estudio del primer ecosistema con personalidad jurídica en Europa: análisis a la luz de la obligación internacional de proteger y del derecho comparado», *Cuadernos de Derecho Transnacional*, 15, 2023, pp. 578-601.

realidad no es una mera cosa pasiva dispuesta para la explotación arbitraria por parte del ser humano[56].

Los defensores de los derechos de las máquinas relacionan esa cuestión con los derechos de la naturaleza y de los animales, pero es necesario destacar que se trata de problemas diferentes, especialmente en el caso de los animales sentientes, que manifiestan cierta inteligencia y capacidad de sentir. Porque es preciso subrayar que la sedicente IA no es sino tratamiento de información de manera altamente sofisticada. Atribuirle derechos implica minusvalorar la importancia que estos tienen para el respeto al humano.

En efecto, el curso histórico para el reconocimiento de los derechos a todos los seres humanos ha sido arduo; lo es aún, en realidad. No me refiero sólo a las facultades para actuar dentro del orden jurídico, sino a la protección de los bienes más básicos ínsitos en la persona. Pensar en trasladar esas exigencias a flujos automatizados de información no parece razonable.

Claro que tal vez lo que algunos pretenden es justamente eso: la minusvaloración de lo verdaderamente humano. Ese es el trasfondo del denominado transhumanismo, que es campo diferente de la IA, pero utiliza a veces a esta para la supuesta superación tecnológica de la humanidad. En este sentido, Jesús Ballesteros relaciona el desarrollo de una IA entendida como intento de superar el cuerpo y sus limitaciones con el rechazo de lo humano, entendido como vulnerabilidad y reconocimiento del otro[57]. Las pretensiones de la IA a la que se refiere Ballesteros es la que busca la superinteligencia, como la propugnada por Ray Kurzweil, basada en cábalas un tanto futuristas. Su finalidad es más la transformación tecnológica de lo humano, que otorgar derechos a las máquinas.

Si las propuestas de Kurzweil y quienes siguen esa tendencia ponen en peligro la real dignidad del humano, las proposiciones más modestas que otorgan algunos derechos a los robots también pueden resultar inquietantes. Como ha señalado Silvia Tamayo, los robots no son entidades vivas, no son conscientes, no pueden, por tanto, ser sujetos de derechos; si se les dan derechos se rompe la frontera con el ser humano y se pone en peligro los fundamentos humanísticos de Europa[58].

---

56 *Cfr*. Megías Quirón, J.J., «Derechos Humanos y medio ambiente». *Humana Iura*, 10, 2000, pp. 232 y ss. Este mismo autor señala, acerca de la protección de los animales, que su adecuada protección jurídica frente al sufrimiento y daño no implica concederles derechos. Cfr. Megías Quirón, J.J., «Ser humano y animales: estatuto ontológico», *Cuadernos de Bioética*, 31, 2020, p. 67.

57 *Cfr*. Ballesteros, J., «Biotecnología, Biopolítica y Posthumanismo», en Ballesteros, J. / Fernández, E. (Eds.), *Biotecnología y Posthumanismo*. Thomson / Aranzadi, 2007, pp. 28 y ss.

58 *Cfr*. Tamayo Haya, S., «Los robots como entes jurídicos», Solar, J.I. (Ed.), *Dimensiones éticas y jurídicas de las IA en el marco del Estado de Derecho*. Alcalá, Universidad de Alcalá, 2020, pp. 212 y ss.

La apariencia de vida y sentimiento que pueden proporcionar algunas máquinas de IA sin duda generan efectos en los seres humanos. Pero es sólo apariencia. Claro que nuestra cultura presta atención especial a la apariencia. La difusión de filosofías denominadas postmodernas a las que aludí más arriba afirman la disolución de lo real, del sujeto, de los bienes específicamente humanos... Todo parece ser un juego de espejos que no reflejan nada real y en esa situación no resulta chocante que la IA tenga derechos. Pero la realidad se impone: los daños a la dignidad de la persona aparecen una y otra vez y entonces resuena de nuevo la lucha por los derechos humanos, por la exigencia de ámbitos de actuación que liberen de opresiones asfixiantes o, en un plano, más modesto, permitan que la persona se desenvuelva con cierta autonomía en el tráfico jurídico. En ese momento, resulta clara la importancia de los derechos y su vinculación con lo propiamente humano; el plano de las máquinas, aunque sean valiosas y merecedoras de protección, sin duda, no puede equipararse.

Bibliografía

Arkoudas, K., Bringsford, S., y Bello P., «Toward Ethical Robots via Mechanized Deontic Logic», en *AAAI Fall Symposium in Machine Ethics*, The AAAI Press, Menlo Park, CA, USA 2005.

Ballesteros, J., «Biotecnología, Biopolítica y Posthumanismo», en Ballesteros, J. / Fernández, E. (Eds.), *Biotecnología y Posthumanismo*. Thomson / Aranzadi, 2007.

Ballesteros, J., «El individualismo como obstáculo a la universalidad de los derechos humanos», en *Persona y Derecho*, 41, 1999.

Ballesteros, J. *Postmodernidad: Decadencia o resistencia*. Tecnos, Madrid, 1989.

Ballesteros, J., «El derecho como no-discriminación y no-violencia», en *Anuario de Filosofía del Derecho*, 17, 1974.

Barrio Andrés, M., «De nuevo sobre la Persona Jurídica», *Inteligencia Artificial*, 27, 2024.

Barrio Andrés, M., «Del Derecho de Internet al Derecho de los robots», en Barrio, M., *Derecho de los robots*., Wolters Kluwer, Madrid, 2018.

Boden, M., *Inteligencia Artificial*, trad. de I. Pérez Parra, Turner Libros 2017.

Böhme, H.-J., *Politische Rechte des Einzelnen in der Naturrechtslehre des 18. Jahrhunderts und in der Staatstheorie des Frühkonstitutionalismus*, Duncker & Humblot, Berlín, 1993.

Bosch Rue, A., Casas Roma, J. y Lozano Bagen, T., *Deep Learning. Principios y fundamentos*. UOC, Barcelona, 2019.

Branting, K.; «Data-centric and logic-based models for automated legal problem solving», *Artificial Intelligence and Law*, 25, 2017.

Cabrera, L., «Personas y seres humanos, ¿distinción o identidad?», *en Ius et Scientia,* 3. 2017.

Cabrera, L., «Autonomía y dignidad: la titularidad de los derechos», en *Anuario de Derechos Humanos,* 3, 2002.

Calo, R., «Robots as Legal Metaphors», *Harvard Journal of Law and Technology*, 30, 2016.

Carabantes López, M., *Inteligencia artificial. Una perspectiva filosófica*., Escolar y Mayo, Madrid, 2016.

Carpintero Benítez, F., de Mora, E. V., Megías, J.J., y Rodríguez Puerto, M., *El derecho subjetivo en su historia*, Servicio de Publicaciones de la Universidad de Cádiz, Cádiz, 2003.

Carpintero Benítez, F., *Historia breve del derecho natural*, Colex, Madrid, 2000.

Carpintero Benítez, F., «Voluntarismo y contractualismo: una visión sucinta de la Escuela de derecho natural moderno», en *Persona y Derecho*, 13; 1998.

Carpintero Benítez, F., *La Cabeza de Jano*, Universidad de Cádiz, Chiclana, 1989.

Carpintero Benítez, F., «La independencia y autonomía del individuo: los orígenes de la persona jurídica», en *Anuario de Filosofía del Derecho*, 1987.

Castilla Barea, M., «¿Qué queda hoy de la Resolución del parlamento europeo de 20 de octubre de 2020 sobre un régimen de responsabilidad civil en materia civil en materia de inteligencia artificial?», en Cervilla, M.ª D. / Ballesteros, A., (Eds.), *Temas actuales de Derecho Privado II*, Cizur Menor, Aranzadi, 2023.

Coeckelberg, M., *«Robot Rights? Towards a social-relational justification moral consideration»*, Ethics, Inf Technology, 12, 2010.

Cotta. S., *El derecho en la existencia humana*. Trad. de I. Peidró. EUNSA, Pamplona,1987.

De Graaf, M., Hindricks, F., y **Hindricks, K.,** «Who Wants to Grant Robot Rights», *Frontiers in Robotics and AI*, 8, 2022.

Desai, R. y Kroll, J.A., Trust but Verify, *Harvard Journal of Law and Technology.*

Fernández Ruiz-Gálvez, E., «Del universalismo abstracto a la universalidad concreta», en *Persona y Derecho*, 41, 1999.

Gadamer, H.G., *Verdad y método*. Trad. de A. Agud y R. de Agapito, Sígueme, Salamanca,1977.

Gellers, J.C., «*Rights for Robots. Artificial Intelligence, Animal and Environmental Law*», Routledge, Londres / Nueva York, 2021.

Gunkel, D. J., «The Rights of Robots», en Alvarez- Nakagawa, A. / Douzinas, C. (Eds.), *Non-Human Rigths Critical Perspectives,* Edward Elgar, 2022.

Gunkel, D.J., *Robot Rights*, MIT Press, Cambridge, 2018.

Gutiérrez Castillo, V.L., «Estudio del primer ecosistema con personalidad jurídica en Europa: análisis a la luz de la obligación internacional de proteger y del derecho comparado», en *Cuadernos de Derecho Transnacional*, 15, 2023.

Harris, J., y **Anthis, J.R**., «The Moral Consideration of Artificial Entities: A Literature Review», en *Science and Engineerig Ethics*, 2021.

Kurzweil, R. *La Singularidad está cerca*. Trad. C. García. Lola Books, Berlín, 2012.

Ladeur, K.H., *Postmoderne Rechtstheorie*, en Duncker & Humblot, Berlín, 1996.

Latour, B., *Reassembling the Social. An Introduction to Actor-Network-Theory*., Oxford Universit Press, Oxford, 2007.

Llano Alonso, F.H., *Homo ex Machina*., Tirant lo Blanch, Valencia, 2024.

López de Mántaras, R. y **Meseguer González, P**., *Inteligencia artificial*, CSIC, Madrid, 2018.

Maccarthy, J., «Free Will – Even for Robots», *Journal of Experimental and Theoretical AI*, 2000.

Mamak, K., «Should criminal law protect love relations with robots?» *AI & Society*, 39, 2024.

Megías Quirón, J.J., «Ser humano y animales: estatuto ontológico», en *Cuadernos de Bioética*, 31, 2020.

Megías Quirós, J.J., «Dignidad y derecho: de la Antigüedad a la Edad Media», en *Anuario de Filosofía del Derecho*, Madrid, 2018.

Megías Quirós, J.J., «Dignidad, universalidad y derechos humanos», en *Anuario de Filosofía del Derecho*, 22, 2005.

Megías Quirón, J.J., «Derechos Humanos y medio ambiente», en *Humana Iura*, 10, 2000.

Misselhorn, C., «Artificial Moral Agents», en *The Cambridge Handbook of Responsible AI*. Cambridge University Press, 2022.

Müller, V.C., «Is it time for robot rights? Moral Status in Artificial Entities», *Ethics and Information Technologies*, 23, 2021.

Ollero Tassara, A., *Derechos humanos y metodología jurídica*. Centro de Estudios Connstitucionales, Madrid, 1989.

Pérez-Luño, A., *Derechos humanos, Estado de Derecho y Constitución*., Tecnos, Madrid, 2018.

Pérez Luño, A.E., *Nuevas tecnologías, sociedad y derecho*, Fundesco, Madrid, 1987.

Railton, P., «Ethical Learning, Natural and Artificial», en *Ethics of Artificial Intelligence*

Robles Morchón, G., *Teoría del Derecho. Fundamentos de Teoría Comunicacional del Derecho*. Vol. III, Civitas, Cizur Menor 2021.

Robles Morchon, G., *Teoría del Derecho*. Vol. II, Cizur Menor, Thomson, 2015.

Tamayo Haya, S., «Los robots como entes jurídicos», en Solar, J.I. (Ed.), *Dimensiones éticas y jurídicas de las IA en el marco del Estado de Derecho*. Alcalá, Universidad de Alcalá, 2020.

Teubner, G., «Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law», *Journal of Law & Society*, 33, 2006.

Teubner, G., «Die anonyme Matrix: Zu Menschenechtsverletzungen durch, private transnationale Akteure», *Der Staat*, 44, 2006.

Viola, F., *De la naturaleza a los derechos*., Trad. de V. Bellver, Comares, Granada, 1998.

Wallach, W. y Valler, S., «Moral machines. From Value Alignment to Embodied Virtue», en *Ethics of Artificial Intelligence*. OUP, 2020.

16.

INTELIGENCIA ARTIFICIAL: REGULACIÓN Y ÉTICA PARA LA PROTECCIÓN DE LOS DERECHOS FUNDAMENTALES

Idoia Salazar García

Fundadora y Presidenta de la Fundación OdiseIA

SUMARIO: I. Inteligencia artificial, ventajas y oportunidades. II. Riesgos y retos reales de la inteligencia artificial. 2.1. La Explicabilidad. 2.2. Responsabilidad. 2.3. El futuro del trabajo y de las prestaciones sociales. 2.4. Relaciones personas-IA. 2.5. Concentración de poder y riqueza en unas pocas empresas muy grandes. 2.6. El uso malicioso de la inteligencia artificial. III. La ética en la IA. IV. Principios éticos de la inteligencia artificial. V. Normativa más allá de la ética.

I. Inteligencia artificial, ventajas y oportunidades

El término sistema de inteligencia artificial no es simple de definir. Según la propuesta de regulación de la Comisión Europea, se define como: un sistema diseñado para funcionar con un cierto nivel de autonomía y que, basándose en datos proporcionados por máquinas y/o personas, infiere cómo lograr un conjunto determinado de objetivos utilizando enfoques basados en el aprendizaje de máquinas y/o en la lógica y el conocimiento. Así, produce resultados generados por el sistema, como contenidos, predicciones, recomendaciones o decisiones, que influyen en los entornos con los que interactúa.

Otros se refieren al concepto general de Inteligencia Artificial como: el estudio de la informática que se centra en el desarrollo de estrategias y tácticas por el cual sistemas y máquinas manifiestan una inteligencia humana[1].

1 López Briega, R.E., *Introducción a la Inteligencia Artificial. Matemáticas, Análisis de Datos y Python,* 2017. https://relopezbriega.github.io/blog/2017/06/05/introduccion-a-la-inteligencia-artificial/

En definitiva, son muchos los que, a lo largo de los siglos XX y XXI, han intentado abordar el intento de dotar de significado a estas dos palabras: Inteligencia Artificial.

Fue a mediados del siglo XX, cuando empezó a evolucionar de manera notable y realista el pensamiento científico en este campo, durante esta época, fue abrumador. El rápido desarrollo de grandes inventos tecnológicos permitió comenzar a pensar en la posibilidad real de dotar a las «máquinas», a los «autómatas», de algo hasta entonces intrínseco únicamente en el ser humano: la inteligencia, la capacidad para pensar y razonar siguiendo una lógica. Alan Turing, considerado por muchos el padre de la Inteligencia Artificial, aludía a este concepto en su histórico artículo, publicado en 1950, en el que planteaba la pregunta: ¿Pueden las máquinas pensar? Para concretarlo propuso su «Juego de imitación»[2], posteriormente conocido como «El test de Turing». Para este juego era necesario un interrogador (persona), ubicado en una habitación aislada, y en otra estancia una persona y una computadora. Ambos deberían responder a la pregunta que les realizase el interrogador de forma aleatoria. La máquina tenía que ser capaz de hacerse pasar por un ser humano; si el interrogador no podía distinguir entre el individuo y el ordenador, se consideraba que la máquina había alcanzado un determinado nivel de inteligencia. Para Turing, la inteligencia artificial comenzaba a existir cuando los humanos éramos incapaces de distinguir, en una conversación a ciegas, si nuestro interlocutor era también humano o máquina.

Muchos han sido los que, desde entonces, han intentado que sus programas informáticos superaran el «Test de Turing». En 1990 se creó el concurso Premio Loebner en el que intentaban superarlo, lo cual no fue posible hasta 2010 cuando el robot Suzette, de Bruce Wilcox, lo consiguió. Suzette, tenía 16.000 reglas de conversación y era capaz de mantener 40 horas de conversación ininterrumpida. Tenía una personalidad coherente y respondía emocionalmente. Hoy día, las teorías de Turing siguen aplicándose a los fundamentos de la robótica y la inteligencia artificial.

El estadounidense Marvin Minsky también revolucionó este campo durante el siglo XX y sus descubrimientos y pensamientos supusieron un antes y un después. En 1956 acuñó el acuñó el término «Inteligencia Artificial» —ya ideado por Turing— y estableció sus fundamentos guiados por el sueño visionario de dotar a los ordenadores con la capacidad para razonar. Él la definía de esta manera: «Es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hubiera hecho un humano». Dos años más tarde ingresó como profesor en el Massachusetts Institute of Technology (MIT), en Boston, donde trabajó más de cuatro décadas y donde creó su Laboratorio de Inteligencia Artificial, incubadora de muchos inventos robóticos actua-

2 Turing, A.M., «Computing machinery and intelligence», en *Mind*, Volume LIX, Issue 236, 1 October 1950, pags. 433-460, https://doi.org/10.1093/mind/LIX.236.433

les. Creó el primer prototipo de una máquina capaz de aprender de manera autónoma «Snarc», diseñó las primeras manos con sensores táctiles y formó parte del equipo que diseñó ARPAnet, el precursor de lo que hoy conocemos como Internet.

A pesar de estos avances durante el siglo XXI, la verdadera ola actual de progreso y entusiasmo por la IA comenzó alrededor de 2010, impulsada por cuatro principales factores, íntimamente ligados el uno al otro[3]:

- La disponibilidad del Big Data, gratis o a precio reducido, para el comercio electrónico, empresas, redes sociales, medios de comunicación, ciencia y gobierno.
- Esta materia prima (Big Data), de la que se nutren la IA, mejora considerablemente el llamado machine learning y los algoritmos.
- El procesamiento de datos mejora exponencialmente gracias a las altas capacidades de los ordenadores.
- La recuperación económica y el aumento significativo, por parte de las empresas tecnológicas, en inversión en IA. La confianza en las oportunidades de esta nueva tecnología es extremadamente alta. Además, tiene aplicaciones transversales en prácticamente cualquier área.

Desde entonces los progresos en IA han sido sorprendentes. Entre los avances más destacados está el reconocimiento de imágenes, con casi un 100 % de fiabilidad o las mejoras en el lenguaje natural en el que son capaces de expresarse y comprender los «robots». Sin embargo, aún nos encontramos en una fase muy temprana de la IA, la conocida como IA débil, consistente en desarrollos muy específicos como juegos estratégicos, traducción de idiomas, reconocimiento de imágenes, o los primeros pasos en la autonomía de los vehículos. Aun así, en el momento actual, ya ofrece grandes ventajas, por ejemplo, en los diagnósticos médicos, sistemas de recomendación y orientación de anuncios o planificadores de viajes.

La llamada IA general o IA fuerte se refiere a la Inteligencia Artificial del futuro, en el que el comportamiento inteligente de la máquina estará mucho más desarrollado en tantas competencias como podría estarlo en una persona. Este es uno de los grandes miedos que despierta esta tecnología en la actualidad: el hecho de que los robots lleguen a ser algún día más inteligente que los humanos. Frente esto hay opiniones de expertos que auguran que estos sistemas inteligentes del futuro podrían crear otros algoritmos aún más inteligentes que ellos, hasta llegar a la llamada «Singularidad Tecnológica», el momento en que los robots dotados de IA superarían en todos los aspectos a la raza humana. Tomarían sus propias decisiones independientemente del

3 Salazar García, I., *La Revolución de los Robots. Cómo la Inteligencia Artificial y la robótica afectan a nuestro futuro*, TREA, 2019. ISSN: 1 978-8417767-34-1

criterio humano y, en el caso de descontrolarse, el mundo se sumiría en un caos, con pocas opciones para las personas.

Existe una visión más positiva sostenida por muchos investigadores que ve en el desarrollo de la IA como un compañero de viaje de los humanos. Una ayuda. Un asistente capaz de operar de manera segura y con ética.

En cualquier caso, es muy necesario que los distintos gobiernos y organismos internacionales conozcan muy bien estas tecnologías., sus ventajas y sus posibles riesgos, y ayuden a minimizar el posible impacto negativo con suficiente antelación.

Mientras avanzamos en una dirección un otra, la evolución exponencial de la IA continúa, con retos y riesgos, que serán necesarios controlar si queremos aprovechar también todas las ventajas que trae esta tecnología.

II. Riesgos y retos reales de la inteligencia artificial

Existen muchos prejuicios en torno a la Inteligencia Artificial alimentados, en buena parte, por la ciencia ficción. Estas cuestiones pueden suponer, a medio plazo, la ralentización de su implantación a nivel internacional, motivado por la falta de confianza. Algunas carecen de fundamento. Otras no. Los riesgos actuales más relevantes de estas tecnologías de IA empezarían por los «efectos secundarios» no deseados, los denominados FATE *(Fair, Accountable, Transparent and Explainable)*. En estas cuestiones, la posible discriminación es uno de los puntos más importantes por su impacto social. Aunque el aprendizaje automático *(machine learning)* es capaz de resolver tareas complejas con un alto rendimiento, puede utilizar información que, desde el punto de vista de la sociedad o de la humanidad, no es deseable. Por ejemplo, conceder un préstamo a personas en función de la raza o la religión. Aunque es posible eliminar esos atributos «no deseados» de los conjuntos de datos, hay otros atributos menos obvios que están altamente correlacionados con esos atributos «no deseados» cuya eliminación es menos sencilla. El aprendizaje automático es objetivo y encuentra cualquier relación que haya en los datos, independientemente de las normas y valores específicos.

Otro de los grandes riesgos es que el conjunto de los datos utilizados para entrenar al sistema de IA esté sesgado de manera consciente o inconsciente. El aprendizaje automático basa sus conclusiones en los datos. Si no se usan muestras lo suficientemente representativas de todos los sectores implicados la conclusión no será válida. Aparte del sesgo en los datos de entrenamiento, el sesgo también puede provenir del algoritmo. Un algoritmo de *Machine Learning* trata de ser lo más preciso posible al ajustar el modelo a los datos de entrenamiento. Si no se elige bien el modelo puede provocar «falsos positivos» y /o «falsos negativos».

La explicabilidad es otro de los grandes problemas a los que se enfrentan los sistemas de IA de *Deep Learning* en la actualidad. Para algunas aplicacio-

nes, la explicación de las decisiones es una parte esencial de la propia decisión, y su falta hace que la decisión sea inaceptable. Por ejemplo, un sistema de IA que ayuda a un juez en su toma de decisiones, extrayendo conclusiones sobre un litigio entre un cliente y una aseguradora de salud es inaceptable sin la explicación de la decisión. Esto se conoce como el problema de la «interpretabilidad»[4].

Otras de las cuestiones de mayor relevancia en cuanto a los efectos secundarios en el uso de la IA son: Privacidad, transparencia y control de los datos. Todos los datos y sistemas de IA explotan datos, y muchas veces estos son datos personales. Utilizarlos tiene como efecto secundario que la privacidad puede verse comprometida, aunque sea de forma involuntaria. El escándalo de Cambridge Analytica/Facebook[5] muestra que este es un problema mayor de lo que podríamos haber pensado. En este caso, ocurrido en 2018, el investigador de la Universidad de Cambridge Aleksandr Kogan, recopiló datos de decenas de miles de usuarios de Facebook y sus amigos, a través de una aplicación que cumplía con las condiciones de Facebook estipuladas para una investigación concreta. A través de la app consiguió información sobre más de 70 millones de usuarios de esta red social. El problema empezó cuando este investigador transfirió, de manera ilegal (es decir, en contra de la política de datos de Facebook), estos datos a la empresa Cambridge Analytica, especializada en el marketing político online, sobre todo en campañas electorales, usando tecnologías de Big Data. Esta empresa prestó servicios a la campaña de Trump para influir en las elecciones americanas de 2016 y también a la campaña pro Brexit en el referéndum del Reino Unido.

2.1. La explicabilidad

Otro efecto secundario importante, se presenta con la explicabilidad. En la actualidad, los sistemas de IA dependen íntimamente de su programador inicial. Es esta persona la que decide qué sistemas usar y con qué datos entrenar el modelo. Por lo tanto, en el tema de la «explicabilidad» de las conclusiones a que llegan estos sistemas, el primer responsable es el ingeniero.

Los resultados de los algoritmos más potentes de la IA, los de *deep learning* son difíciles de entender, ya que el proceso que desarrolla es opaco. Se les conoce como Black Box, caja negra. Sin embargo, hay muchos algoritmos de aprendizaje automático que son perfectamente inteligibles (cajas blancas) para saber cómo ha llegado a una determinada conclusión. Por ejemplo, los llamados «árboles de decisión» son perfectamente comprensibles.

4 O´Neil, C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.Random House LCC US, 2016.

5 Wylie, C., *Mindf*ck: Cambridge Analytica. La trama para desestabilizar el mundo*, Roca Editorial, 2020.

Aunque, hasta la fecha, la mayoría de los ingenieros prefieren usar deep learning (una caja negra) por sus prestaciones superiores, cada vez son más las empresas que consideran usar un algoritmo interpretable (caja blanca) si su prestación es suficientemente buena para resolver el problema de negocio, aunque sea su rendimiento un poco peor que un algoritmo de caja negra. La razón es la explicabilidad, que, en ciertas aplicaciones, debe tener un peso importante, como en el sector médico. De todas formas, hoy en día hay mucha investigación trabajando en mejorar la explicabilidad de los sistemas de *deep learning*[6].

2.2. Responsabilidad

Por otro lado, cuando los sistemas se vuelven autónomos y autodidactas, la responsabilidad del comportamiento y las acciones de esos sistemas se hace menos evidente. En el mundo anterior a la IA, el uso incorrecto de un dispositivo es responsabilidad del usuario, mientras que el fallo del dispositivo es responsabilidad del fabricante. Cuando los sistemas se vuelven autónomos y aprenden con el tiempo, algunos comportamientos podrían no ser previstos por el fabricante. Por lo tanto, no está claro quién sería responsable en caso de que algo vaya mal. Un ejemplo de ello son los coches sin conductor. Según la Comisión Legislativa del Reino Unido la culpa pasaría a ser del desarrollador o fabricante del coche. Este argumento se relaciona con la fabricación del propio producto, el creador hace verificaciones, y distintas pruebas con el fin de que cuando sea lanzado al mercado, sea un producto o servicio totalmente seguro. Sin embargo, se tiene que analizar en cada producto o servicio el grado de inteligencia artificial que tiene para poder responder a la pregunta.

Por otro lado, presenta una amenaza a la democracia, y crea constantemente las denominadas cámaras de eco online[7]. Consiste en sistemas que se basan en los comportamientos previos de una persona a nivel digital, y luego archiva esos datos con el fin de mostrarles contenido híper-personalizado y un entorno único en el que lo que la persona observa es lo que desea. Por lo cual, el entorno habitual en el que los seres humanos nos movemos va desapareciendo. El debate, el público pluralista, la diversidad de contenidos son cosas que poco a poco se vuelven menos frecuentes. Todo esto conlleva a la separación y polarización a nivel público, puesto que lo único que hace es perfilar a los seres humanos, enseñándoles únicamente contenido de su interés, creencias, gustos, y acostumbrándolos aquello. Se puede evidenciar a su vez puesto que el 62,7 % de la encuesta realizada, afirma que le gustaría que le muestren únicamente contenido personalizado en base a sus intereses.

6 Darlington, K., «Sistemas de IA explicables: comprender las decisiones de las máquinas», en *OpenMind,* BBVA, 2017.

7 CDE: *Inteligencia Artificial: amenazas y oportunidades*. Centro de Documentación Europea de la Universidad de Almería, 2020. https://www.cde.ual.es/inteligencia-artificial-amenazas-y-oportunidades/

2.3. El futuro del trabajo y de las prestaciones sociales

Otro de los grandes riesgos es el relativo al futuro del trabajo y las prestaciones sociales. La IA puede asumir determinadas tareas repetitivas o peligrosas. Pero si esto ocurre a escala masiva, tal vez desaparezcan muchos puestos de trabajo y el desempleo se dispare.

Si cada vez trabaja menos gente, el gobierno recibirá menos impuestos sobre la renta, mientras que los costes de las prestaciones sociales aumentarán debido al incremento del desempleo. Este tipo de cuestiones ha desembocado en sugerencias referentes a un necesario ingreso básico universal (UBI) para todos.

2.4. Relaciones personas-IA

Por otro lado, empiezan a surgir otra serie de cuestiones relacionadas con la interacción «hombre-máquina», que deben tratarse para que no suponga un riesgo para la persona. Por ejemplo: ¿cuál debería ser la relación (permitida) entre la IA y las personas? ¿Podría un jefe ser un sistema de IA? En Asia, los robots con IA (y sin ella) ya cuidan de las personas mayores, acompañándolas en su soledad. Y, ¿podrían las personas casarse con una IA, como ya ha ocurrido en algunos lugares de Japón?

Algunas de las primeras interacciones serán las siguientes[8]:

a) Interacciones a largo plazo: en las que los robots cohabitan con los humanos en sus hogares, y lugares de trabajo, por ejemplo.

b) Robots en la educación, la terapia, la rehabilitación y el apoyo a las personas mayores. La robótica asistencial, con y sin inteligencia artificial, es un ámbito de aplicación creciente para los robots de servicio. Implica cuestiones críticas de seguridad y ética, por ejemplo, cuando los robots asumen un papel de asistencia a personas vulnerables o con necesidades especiales. Es un hecho que hoy en día, aproximadamente el 10 % de la población mundial tiene más de 60 años; en 2050 esta proporción se habrá duplicado con creces. Necesitamos incorporar técnicas de inteligencia artificial para apoyar a los adultos mayores y ayudarles a afrontar los cambios del envejecimiento, en particular el deterioro cognitivo.

c) Interacciones multimodales, expresividad y habilidades conversacionales en las interacciones: La investigación destinada a dotar a los robots de características y cualidades similares a las de los humanos

8 Salazar García, I., y Benjamins, R., «Towards a framework for understanding societal and ethical implications of Artificial Intelligence», en Suárez Álvarez, R., Fernández-Martínez, L.M., y Blanco Alfonso, I. (Coords), *Vulnerabilidad y Cultura Digital,* Dkynson, Madrid, 2020. También, en Cornell University, 2020. arXiv:2001.09750v1

se está ampliando. Se intenta crear robots con una apariencia similar a la nuestra. Y no sólo esto. También estamos intentando dotar a «sus caras» de expresiones naturales que normalmente sólo corresponderían a los humanos.

d) Cooperación y colaboración en equipos humano-robot (IA): La IA y los humanos no sólo convivirán, sino que deberán trabajar codo con codo, cada uno desarrollando tareas específicas, pero ayudándose mutuamente. Esta interacción, sobre todo al principio, podría ser bastante difícil teniendo en cuenta el concepto de «máquina», y todas sus implicaciones, que tiene la humanidad hoy en día.

e) Detección y comprensión de la actividad humana: Esto también será un problema en el futuro próximo para esta interacción. Muchas veces, los humanos hacen cosas siguiendo diferentes razones que podrían estar fuera de la lógica para una máquina.

2.5. Concentración de poder y riqueza en unas pocas empresas muy grandes

La cuestión de la concentración de poder y riqueza en las mayores empresas que generan datos es un claro riesgo. Entre ellas, las GAFAM (Google, Amazon, Facebook, Microsoft, Apple) y algunas grandes empresas chinas, como Alibaba. Esto podría conducir a un oligopolio. Aparte de la falta de competencia, existe el peligro de que esas empresas mantengan la IA como conocimiento propio, sin compartir nada con la sociedad en general más que por el precio más alto posible. Otra preocupación de esta concentración es que esas empresas pueden ofrecer IA de alta calidad como un servicio, basado en sus datos y algoritmos propios (caja negra).

2.6. El uso malicioso de la inteligencia artificial

Las consecuencias negativas citadas hasta ahora son no intencionadas. Sin embargo, pueden existir también las intencionadas. Por ejemplo: los ciberataques, usando IA. En julio de 2021 más de 200 empresas estadounidenses fueron atacadas con el mismo método. Esta vez el ataque no fue directamente dirigido a empresas que dan servicios finales, sino a un suministrador de software. Al ser infectado este software, todas sus empresas clientes que usan este software fueron afectadas. Otro ejemplo, son los ataques físicos. Por ejemplo, la inteligencia artificial puede dar a los drones la capacidad autónoma para volar solos y actuar de manera autónoma como por ejemplo grabar o incluso disparar. En el ámbito militar, los sistemas de armas autónomas letales son otro ejemplo de crear daños físicos. En el futuro también se pueden esperar ataques que tienen tanto un componente digital como un

componente físico, por ejemplo, un ciberataque a vehículos autónomos para hacer daño físico con estos vehículos «secuestrados». También es posible usar IA para manipular a la opinión pública, de manera maliciosa.

En cualquier caso, es importante estar preparados para poder asumir todos estos riesgos/retos, porque la IA no es el futuro. Es el presente. Cada vez son más las empresas y organizaciones que se suman al uso de la IA como herramienta para mejorar la eficiencia de sus diferentes procesos de negocio. Es importante actuar rápidamente en la ética y normativa para abordar esta tecnología de manera conveniente y conseguir impacto positivo. Veamos la situación internacional de la IA, en este sentido.

III. La ética en la IA

Son muchos los expertos, científicos y autoridades en la materia que abogan por una regulación en la Inteligencia Artificial. Nuevas leyes que «limiten» el uso y aplicaciones de esta revolucionaria tecnología. Pero es difícil hacerlo cuando el cambio es continuo y aún se desconocen muchas de las implicaciones reales. A cambio, se han creado multitud de códigos éticos, muestra de nuestro interés y preocupación. Sin embargo, no son de obligado cumplimiento, muestra de nuestra conciencia y reflexión. Si se incumplen, no ocurre nada. No hay castigo real. Es lo que algunos autores denominan un «lavado (de conciencias) ético»[9]. Y el tremendo potencial económico que promete la Inteligencia Artificial no es una baza a favor de la responsabilidad, el cuidado, la ética y la cautela, cuando tratamos con este tipo de tecnologías. De todas formas, cuando hablamos de ética, muchas veces damos por hecho, de manera incorrecta, que la ética es algo universal, mientras que es algo que puede depender de la cultura.

La importancia del contexto cultural en la toma de decisiones

Es un hecho que no tiene la misma visión del mundo, ni las mismas perspectivas sobre las cosas, un japonés, un alemán o un español. En este sentido, tener en cuenta la cultura, costumbres históricas, del lugar donde se ejecute el algoritmo, es muy importante si se pretende que esta tecnología gane en aceptación social. Uno de los experimentos más conocidos que demuestran, en sus conclusiones, este punto es el llamado «The Moral Machine», desarrollado por el Massachusetts Institute of Technology (MIT).

En cualquier caso, hoy por hoy no parece plausible enseñar ética a una máquina. Por lo que serán a las empresas y sus desarrolladores a quienes hay que exigir que el desarrollo y el uso de la IA sean realizados de una manera

9 Sebio Martín, M. B., *Inteligencia Artificial y ética*, Universidad Pontificia de Comillas, 2020. https://repositorio.comillas.edu/xmlui/bitstream/handle/11531/37008/IA%20y%20etica_Sebio%20Martin,%20Margarita.pdf?sequence=-1

ética, en concordancia con los derechos humanos internacionales. Y esto es el camino dominante que hoy en día están empujando los organismos internacionales. Trabajan en estos códigos éticos en Inteligencia Artificial para dirigir las actuaciones de las empresas que implementan estas tecnologías en productos y servicios.

En este sentido, en abril de 2019, la Comisión Europea presentó las líneas maestras para el desarrollo de una IA fiable y segura en la Unión Europea, resaltando los aspectos principales a tener en cuenta para evitar posibles errores y consecuencias negativas. Son las siguientes[10].

1. Supervisión humana: Los sistemas de IA deben permitir sociedades equitativas y proteger los derechos fundamentales de los humanos. No deberán disminuir o limitar la autonomía humana.
2. Robustez y seguridad: La Inteligencia Artificial requiere que los algoritmos sean lo suficientemente seguros, confiables y sólidos como para enfrentar errores o inconsistencias durante sus procesos de toma de decisiones.
3. Privacidad y gobernanza de los datos: Los ciudadanos deben de tener un control total sobre sus propios datos. Además, los datos que los conciernen no se utilizarán para perjudicarlos ni discriminarlos.
4. Transparencia: Debe garantizarse la transparencia en el proceso de toma de decisiones del sistema de Inteligencia Artificial (Explainable AI).
5. Diversidad, no discriminación y equidad: Los sistemas de IA deben considerar toda la gama de habilidades y requisitos humanos, y garantizar la accesibilidad.
6. Bienestar social y ambiental: Los sistemas de Inteligencia Artificial deben utilizarse para mejorar el cambio social positivo y mejorar la sostenibilidad y la responsabilidad ecológica.
7. Rendición de cuentas: Deben establecerse mecanismos para garantizar la responsabilidad y la rendición de cuentas de los sistemas de AI y sus resultados.

Sin embargo, aunque la Comisión Europea es una referencia internacional con sus pautas éticas, muchas otras organizaciones, de distinta índole, han publicado principios éticos para la Inteligencia Artificial. En marzo de 2019, la Universidad de Harvard estudió los principios de 32 organizaciones en todo el mundo que en ese momento habían publicado principios de IA. Los analizaron en nueve dimensiones e hicieron una puntuación de cada organización para cada dimensión: derechos humanos, valores humanos, responsabilidad profesional, control humano de la tecnología, justicia y no discriminación, transparencia y explicabilidad, seguridad, rendición de cuentas y privacidad.

10 European Comission: Regulation of the European Parliament and of the Council, Artificial Intelligence act and amending certain Union legislative acts. (21.4.2021). Bruselas, p.79.

Analizaron organizaciones de distintos tipos como organizaciones civiles, gobiernos, organizaciones intergubernamentales y compañías privadas.

También la ETH Universidad de Suiza examinó los principios de la IA de 84 organizaciones de todo el mundo y los clasificó con las siguientes dimensiones: transparencia (usado por 73 organizaciones), justicia 68), no maleficencia (60), responsabilidad (60), privacidad (47), beneficencia (41), libertad y autonomía (34), confianza (28), sostenibilidad (14), dignidad (13) y solidaridad (6).

Además, la organización sin ánimo de lucro Algorithm Watch man-tiene un inventario global (Algorithm Watch, 2018) transparencia con 83 organizaciones que han publicado principios éticos de IA entre 2018 y 2019.

De estos estudios se deriva que la mayoría de los principios incluyen temas como transparencia, igualdad, no discriminación, rendición de cuentas y seguridad; algunos mencionan los derechos humanos y beneficencia. De su reflexión, se desprende que aún son pocas las compañías con estos importantes principios.

Por otro lado, en su Libro blanco de la Inteligencia Artificial publicado el 19 de febrero de 2020, la Comisión Europea identificó sectores y aplicaciones de alto riesgo, como, por ejemplo, salud o transporte, y aplicaciones con impacto significativo en las vidas de las personas (jurídico, vida o muerte, etc.). La regulación sería necesaria si se cumplen los dos criterios (sector y al mismo tiempo uso de alto riesgo). Propuso también definir dos listados exhaustivos (sector y uso) de alto riesgo, que se actualizarán periódicamente.

Con todas estas iniciativas de instituciones internacionales, organizaciones civiles, gobiernos nacionales e internacionales y empresas, se intenta regular o autorregular el uso de la Inteligencia Artificial para que las máquinas tomen las decisiones dentro de un ámbito adecuado, que estas sean interpretables cuando impactan en la vida de las personas y que sus resultados no discriminen de una manera no deseada. A continuación, se exponen, más concretamente los principios éticos en IA.

IV. Principios éticos de la inteligencia artificial

Los principios éticos de la Inteligencia Artificial son una serie de criterios fundamentales que los miembros de una comunidad científica o profesional deben considerar para tomar decisiones en base a lo que se considera correcto e incorrecto sobre la utilización y el funcionamiento de esta tecnología en desarrollo[11].

1. **La Explicabilidad** permite a las personas afectadas por el resultado de un sistema de IA acceder y entender el funcionamiento de esta tecnología con mayor facilidad. Esto implica proporcionarles infor-

11 Salazar García, I; y Benjamins, R., *El mito del algoritmo: cuentos y cuentas de la Inteligencia Artificial,* Anaya, 2020. ISBN-10: 8441542805

mación comprensible sobre los factores y lógica que condujeron a la obtención de los resultados y se puede lograr de diferentes maneras. Por ejemplo, según el contexto, los factores principales en la toma de decisiones, los factores determinantes, los datos, la lógica o el algoritmo detrás del resultado específico. Sin embargo, requerir de este principio también tiene desventajas ya que puede afectar negativamente la precisión, rendimiento, privacidad, costes, complejidad y seguridad de algunos sistemas de inteligencia artificial, porque implica la reducción de las variables de solución a un conjunto más pequeño para que los humanos puedan comprender y cuestionar los resultados respetando las obligaciones de protección de datos.

2. **La Transparencia** está sumamente relacionada con la explicabilidad y se basa en revelar cuándo se está utilizando la Inteligencia Artificial en diferentes acciones, como por ejemplo en las predicciones, toma de decisiones, recomendaciones, bot conversacionales, etc. Posibilita que las personas entiendan cómo se desarrolla, opera e implementar un sistema de inteligencia artificial en el dominio de una determinada aplicación, con capacidad de explicar por qué y qué información relevante proporciona esta tecnología en la actualidad. Además, permite que los usuarios puedan tomar decisiones más informadas, incrementando la aceptación y confianza de los resultados obtenidos.
3. **El ser humano como centro:** La Inteligencia Artificial debe desarrollarse de acuerdo con los valores centrados en el ser humano, como la igualdad, libertad, equidad de estado de derecho, justicia social, protección de datos y privacidad, derechos del consumidor, equidad comercial, etc. Ya que, algunas aplicaciones o usos de los sistemas de IA pueden implicar riesgos de que los derechos humanos y valores centrados en el ser humano puedan ser infringidos de forma intencionada o accidental. Se debe promover la alineación de valores en el diseño de esta tecnología junto a la intervención y supervisión humana de su funcionamiento, para garantizar que los comportamientos de los sistemas de inteligencia artificial protejan y prioricen al ser humano frente a los avances de esta tecnología. Todo eso fortalecerá la confianza de las personas y reducirá la discriminación u otros resultados sesgados.
4. **La toma de decisiones** es un principio fundamental que se rige por implicaciones legales como la ética y moral para influenciar las acciones de los algoritmos programados por los seres humanos. Hoy en día, la Unión Europea exige a las entidades proporcionar explicación a las decisiones que fueron tomadas a través de la Inteligencia Artificial, ya que estás en cierta medida generan impacto e influyen en la sociedad. La máquina no asume ninguna responsabilidad porque esta no tiene derechos ni principios éticos. Por lo tanto, las personas

son responsables de aplicar los marcos normativos en el desarrollo y diseño de los sistemas de esta tecnología para garantizar un correcto funcionamiento de la IA y así garantizar las razones por las que fueron tomadas determinadas decisiones.

5. **No maleficencia en el uso de inteligencia artificia**l. Los sistemas de IA tienen la obligación de respetar y proteger a las personas a través de los derechos digitales. Estos consisten en el derecho a la privacidad en entornos digitales, protección a la integridad personal, grooming, ciberseguridad, derecho a la propia imagen, acceso a información veraz, etc. Es fundamental que los humanos tengan consentimiento sobre la utilización correcta de los datos y capacidad de tomar decisiones. Es decir, tener conocimiento de la causa y finalidad de los resultados obtenidos por los sistemas de inteligencia artificial.
6. **Sesgos y discriminación.** Evitar que los algoritmos tengan problemas de discriminación y buscar la manera de que los resultados obtenidos sean explicables. Los sistemas de inteligencia artificial nos ayudan en la selección de información y en la elaboración de nuevos artículos a partir del procesamiento de los datos, sin embargo, si hay sesgo del programador de la IA se pueden generar problemas de desconfianza en los resultados obtenidos. Es fundamental generar conciencia sobre el impacto que tienen las personas que trabajan con algoritmos para así evitar, detectar y mitigar sesgos de discriminación de género, raciales, etc., porque estos generan injusticias, prejuicios y concepciones de la realidad con capacidad de influir en la toma de decisiones de los seres humanos.

A pesar de estos principios éticos en IA compartidos por la mayor parte de usuarios y desarrolladores a nivel nacional e internacional no son de obligado cumplimiento. Los casos de uso que ya han tenido lugar pronostican la clara necesidad de una regulación de obligado cumplimiento que ayude a impulsar la IA, pero manteniendo la protección sobre los derechos humanos.

V. Normativa más allá de la ética

Está claro que, a la hora de combatir estos desafíos, la IA no es diferente a otras tecnologías de gran impacto como lo fue el automóvil, en su día, o Internet. Éstas también necesitaron regularse para prevenir consecuencias negativas. En este caso, y teniendo en cuenta su rápida evolución, sin duda se requiere una regulación dinámica y adaptativa. En cualquier caso, en la mayoría de los países, la regulación en materia de IA se encuentra en etapas incipientes. Por ejemplo, en Estados Unidos, la regulación ha sido más sectorial y depende en gran medida de los estados individuales, aunque existen ciertos marcos federales en áreas específicas, como la privacidad o la discriminación.

A nivel global, el reto radica en equilibrar la innovación con la protección del ciudadano. Una regulación excesiva podría sofocar la innovación, mientras que una falta de regulación podría dejar a las personas desprotegidas. En este sentido, la UE ha aprobado, en agosto de 2024, el llamado **AI ACT** en un esfuerzo común entre los organismos reguladores europeos, empresas, expertos en IA y sociedad civil. Su objetivo: proteger los derechos fundamentales de las personas, garantizar la transparencia en la toma de decisiones de los sistemas de IA y establecer mecanismos de rendición de cuentas y supervisión humana adecuados, entre otras cuestiones básicas. No pretenden regular la tecnología en sí misma, ya que esto supondría un problema para su implementación y desarrollo en la industria de la UE, sino casos de uso específicos que pueden suponer un riesgo: usos/desarrollos prohibidos, de alto riesgo (aquellos cuya implementación puede afectar a los derechos fundamentales de la persona), riesgo medio (obligaciones de transparencia. Ej: Deep Fakes) y riesgo bajo (automatismos sin riesgo).

A continuación, se describen los cuatro niveles de riesgo:

Riesgo inaceptable: los usos de IA que se encuentren en este nivel de la jerarquía quedarán prohibidos. Se refiere a las aplicaciones que suponen una amenaza para la seguridad, la vida o los derechos fundamentales. Un ejemplo de este nivel son los sistemas que incitan a la violencia o al odio, y los sistemas de puntuación social que podrían utilizar los Gobiernos para diferenciar y clasificar a los ciudadanos.

Riesgo alto: el segundo nivel se reserva, principalmente, para el ámbito de la salud y de la educación. Aunque también incluye servicios públicos y privados, control de fronteras y administración de justicia. Las aplicaciones que quieran desarrollar prestaciones en estos campos deberán contar de forma obligatoria con: un estudio de evaluación y control de riesgos, una calidad alta en las bases de datos que alimentan el sistema, un registro de la actividad para mostrar en qué se basan los resultados obtenidos, una oferta de información clara y detallada a los usuarios y una tecnología sólida, segura y precisa.

En este nivel se encuentran los sistemas biométricos de reconocimiento facial, que quedan prohibidos en espacios públicos con las siguientes excepciones desaparición de menores, alerta de ataque terrorista, localización de autores de delitos graves. Las excepciones deben contar, para poder implementarse, con una autorización judicial en la que además se delimite el tiempo y la zona geográfica en la que se aplicará.

Riesgo limitado: en este nivel se encuentran los *chatbots*. La regulación establece que los sitios web y las aplicaciones que utilicen estos servicios de comunicación con el cliente deberán notificárselo. En los artículos dedicados a este nivel se hace hincapié en la transparencia, que además da lugar al nombre del título IV dónde se encuentran.

Riesgo mínimo: para el nivel más bajo de la jerarquía no se contempla ninguna medida reguladora en la propuesta de ley. Esto es porque la Comisión

no considera que las herramientas que pertenecen a este nivel supongan un riesgo o potencial amenaza para la sociedad. Entre ellas se encuentra la aplicación de la IA a los videojuegos o los filtros para ordenar el correo electrónico.

La UE espera conseguir con esta regulación impulsar la IA, basando este impulso en la confianza que generará la regulación, al ofrecer una protección adicional a los derechos fundamentales. Quiere impulsar el desarrollo de esta tecnología sin limitar la competitividad de pymes, start ups y emprendedores, en general.

En definitiva, la IA presenta un potencial transformador inmenso. Sin embargo, como cualquier herramienta poderosa, conlleva tanto oportunidades como riesgos. Aunque existen esfuerzos en curso para desarrollar una regulación sólida, todavía hay un largo camino por recorrer para garantizar un uso seguro y ético de la IA en nuestra sociedad. En cualquier caso, debemos recordar que el peligro no es la IA, en sí misma (al menos por ahora), sino el uso que nosotros, los humanos, hacemos de ella. La clave está en aumentar la consciencia popular y profesional sobre los riesgos, y usar herramientas, desde el diseño, para prevenirlos. La cuestión es que, hoy día, no es una opción plantearse el uso de la IA, pero se debe incentivar su uso/desarrollo ético y responsable desde el diseño.

Bibliografía

Darlington, K., «Sistemas de IA explicables: comprender las decisiones de las máquinas», en *OpenMind,* BBVA, 2017.

López Briega, R.E., *Introducción a la Inteligencia Artificial. Matemáticas, Análisis de Datos y Python*, 2017.

O´Neil, C., *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Random House LCC US, 2016.

Salazar García, I; y Benjamins, R., *El mito del algoritmo: cuentos y cuentas de la Inteligencia Artificial,* Anaya, 2020

Salazar García, I., y Benjamins, R., «Towards a framework for understanding societal and ethical implications of Artificial Intelligence», en Suárez Álvarez, R., Fernández-Martínez, L.M., y Blanco Alfonso, I. (Coords), *Vulnerabilidad y Cultura Digital,* Dkynson, Madrid, 2020.

Salazar García, I., *La Revolución de los Robots. Cómo la Inteligencia Artificial y la robótica afectan a nuestro futuro*, TREA, 2019.

Sebio Martín, M. B., *Inteligencia Artificial y ética*, Universidad Pontificia de Comillas, 2020.

Turing, A.M., «Computing machinery and intelligence», en *Mind*, Volume LIX, Issue 236, 1 October 1950.

Wylie, C., *Mindf*ck: Cambridge Analytica. La trama para desestabilizar el mundo*, Roca Editorial, 2020.

COLECCIÓN

PENSAMIENTO JURÍDICO Y POLÍTICO CONTEMPORÁNEO

COLEX

TÍTULOS PUBLICADOS

1. La libertad de expresión. Avances, límites y desafíos futuros
2. Reforma del Código de Derecho foral de Aragón en materia de discapacidad
3. Inteligencia artificial, sociedad y Derecho
4. Marceliano Isábal y su «Diccionario del Derecho aragonés»
5. El Derecho ante el reto de las tecnologías disruptivas. Actualidad legislativa y jurisprudencial

DESCUBRA MÁS OBRAS EN:

www.colex.es

Editorial Colex SL Tel.: 910 600 164 info@colex.es

EL DERECHO ANTE EL RETO DE LAS TECNOLOGÍAS DISRUPTIVAS

ACTUALIDAD LEGISLATIVA Y JURISPRUDENCIAL

Nunca ha habido una etapa en la historia de la humanidad en la que se haya acumulado un volumen de información como el que disponemos en la actualidad. Pero sobre todo en ningún otro periodo ha sido ésta tan accesible. En la actualidad, la mayoría de los ciudadanos del mundo tienen a su alcance fuentes de información que superan ampliamente el total de lo que se llegó a reunir en la mítica Biblioteca de Alejandría. Sin embargo, igual que entonces, el problema es como ordenar y jerarquizar ese conocimiento. Algo que puede agudizarse en los próximos años en un Internet que nada olvida y que no deja de expandirse. En ese sentido, este libro colectivo pretende servir de guía ante algunos de los grandes debates jurídicos que se están suscitando en el marco del paradigma tecnológico actual. Para ello se cuenta con la opinión de personas altamente cualificadas y de reconocido prestigio. En total 16 autores y 16 capítulos en los que se tratan asuntos como el metaverso, la inteligencia artificial, la transparencia y el ejercicio y la protección de los derechos en el entorno digital, la ciberseguridad, las plataformas digitales, o los neuroderechos. Sin resolver de manera definitiva estas cuestiones, la obra proporciona claves importantes para navegar por su dimensión jurídica, de tal manera que los lectores sean capaces de formarse su propio criterio con rigor y de manera informada.

COORDINACIÓN

Gonzalo Castro Marquina

AUTORES

Francisco Balaguer Callejón, Moisés Barrio Andrés, José Juan Castelló Pastor, Gonzalo Castro Marquina, Lorenzo Cotino Hueso, Iban García del Blanco, Ana Gascón Marcén, M.ª José González Ordovás, Miguel L. Lacruz Mantecón, Joaquín Meseguer Yebra, Elisa Moreu Carbonell, Miguel Ángel Presno Linera, M.ª Ángeles Rueda Martín, Margarita Robles Carrillo, Manuel Jesús Rodríguez Puerto y Idoia Salazar García

COLECCIÓN

PENSAMIENTO JURÍDICO Y POLÍTICO CONTEMPORÁNEO

Director: Prof. Dr. Dr. Guillermo Vicente y Guerrero

ISBN: 979-13-7011-409-1