Revista de las Cortes Generales. ISSN: 0213-0130. ISSNe: 2659-9678
N.º 116, Segundo semestre (2023): pp. 81-133
https://doi.org/10.33426/rcg/2023/116/1775

LA PROPUESTA DE «LEY DE INTELIGENCIA ARTIFICIAL» EUROPEA[*]

THE PROPOSAL FOR AN "ARTIFICIAL INTELLIGENCE ACT" IN THE EUROPEAN UNION (EU)

Miguel Ángel Presno Linera
Catedrático de Derecho Constitucional.
Universidad de Oviedo
https://orcid.org/0000-0002-0033-6159

A Ibán García del Blanco.

Fecha de recepción del artículo: noviembre 2023
Fecha de aceptación y versión final: diciembre 2023

RESUMEN

En este trabajo se analiza el origen y el desarrollo de la propuesta de «Ley de inteligencia artificial» europea, es decir, del proyecto de la Unión Europea de aprobar un conjunto de normas armonizadas para el desarrollo, la introducción en el mercado y la utilización de sistemas de IA de manera que esos objetivos sean compatibles con la dignidad humana y la autonomía individual, los derechos humanos y las libertades fundamentales, el funcionamiento de la democracia y el respeto del Estado de Derecho. Cuando se terminaron estas páginas (20 de noviembre de 2023) no había concluido el proceso legislativo pero es interesante conocer su desarrollo, los importantes y veloces cambios tecnológicos a los que ha debido hacer frente y el impacto que ya tiene tanto dentro como fuera de la Unión Europea.

Palabras clave: Unión Europea, derechos fundamentales, inteligencia artificial, riesgo tecnológico, modelos fundacionales, efecto Bruselas.

---

[*] Esta publicación es uno de los resultados del proyecto PID2022-136548NB-I00 *Los retos de la inteligencia artificial para el Estado social y democrático de Derecho*, financiado por el Ministerio de Ciencia e Innovación en la Convocatoria Proyectos de Generación de Conocimiento 2022.

ABSTRACT

This paper aims at analyzing the origin and development of the European proposal for an 'Artificial Intelligence Act', i.e. the European Union's project to adopt a set of harmonised rules for the development, market introduction and use of AI systems in a way that is compatible with human dignity and individual autonomy, human rights and fundamental freedoms, the functioning of democracy and respect for the rule of law. At the time this paper is being written (20 November 2023) the legislative process has not been completed. Nonetheless, it is interesting to learn about its development, the significant and quick technological changes it has had to cope with and the impact that the proposal it is already having on the European Union and abroad.

Keywords: European Union, Fundamental Rights, Artificial Intelligence, Technology Risk, Foundational Models, Brussels Effect.

I. Introducción

En el año 2022, la Fundación del Español Urgente, promovida por la Agencia EFE y la Real Academia Española, otorgó el título de «palabra del año» a la expresión compleja inteligencia artificial (IA). Esta elección se justificó, entre otras razones, porque «el análisis de datos, la ciberseguridad, las finanzas o la lingüística son algunas de las áreas que se benefician de la inteligencia artificial. Este concepto ha pasado de ser una tecnología reservada a los especialistas a acompañar a la ciudadanía en su vida cotidiana: en forma de asistente virtual (como los que incorporan los teléfonos inteligentes), de aplicaciones que pueden crear ilustraciones a partir de otras previas o de chats que son capaces de mantener una conversación casi al mismo nivel que una persona. No obstante, también ha estado muy presente por las implicaciones éticas que supone el desarrollo de la inteligencia de las máquinas»[1].

Dos años antes, en el primer párrafo del *Libro Blanco sobre la inteligencia artificial* de la Comisión Europea, de 19 de febrero de 2020, se dijo que «la 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].

Como veremos en las siguientes líneas, el impacto de la IA ya se había detectado con anterioridad y con él la consciencia de que estamos inmersos en una *infoesfera* (Floridi, 2012, p. 11), en un ambiente

[1] https://www.fundeu.es/recomendacion/inteligencia-artificial-es-la-expresion-del-2022-para-la-fundeurae/ (consulta el 20 de noviembre de 2023).

[2] https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1, (consultado el 20 de noviembre de 2023).

global compuesto por organismos informacionales interconectados y este mestizaje ontológico entre lo biológico y lo técnico, entre lo carbónico y lo silícico (Campione, 2020, p. 13), exige dar respuestas jurídicas a preguntas como las que formuló en el plano ético el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías en su *Declaración sobre Inteligencia artificial, robótica y sistemas «*autónomos*»*, de 9 de marzo de 2018: ¿cómo podemos construir un mundo con IA y dispositivos «autónomos» interconectados que sea seguro y cómo podemos estimar los riesgos involucrados? ¿Quién es responsable de resultados no deseados y en qué sentido es responsable? ¿Cómo se deben rediseñar nuestras instituciones y leyes para que estén al servicio del bienestar de las personas y la sociedad, y para hacer de la sociedad un lugar seguro ante la aplicación de estas tecnologías? ¿Cómo evitar que, a través del aprendizaje automático, los datos masivos y las ciencias del comportamiento se manipulen las arquitecturas de toma de decisiones según fines comerciales o políticos? En suma, ¿cómo se puede prevenir que estas poderosas tecnologías sean utilizadas como herramientas para socavar sistemas democráticos y como mecanismos de dominación?[3].

Pues bien, en este texto se analiza el intento de la Unión Europea, impulsado formalmente el 21 de abril de 2021, de aprobar un conjunto de normas armonizadas para el desarrollo, la introducción en el mercado y la utilización de sistemas de IA en la Unión a partir de un enfoque proporcionado basado en los riesgos al que se denominaría, de manera gráfica, «Ley de inteligencia artificial», aunque su nombre técnico, caso de que salga adelante, sería 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[4].

Al cierre de estas páginas se acaba de superar la fase última de deliberación a través de los «trílogos» entre las instituciones europeas

---

[3] https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/ethics-artificial-intelligence-statement-ege-released-2018-03-09_en (consulta el 20 de noviembre de 2023).

[4] Sobre las normas armonizadas en la Unión Europea, Álvarez García, 2020; sobre dichas normas en relación con la inteligencia artificial, Álvarez García y Tahiri Moreno, 2023.

implicadas[5], y aunque no es seguro sí es probable que culmine con éxito este proyecto normativo; en todo caso, creemos que merece la pena conocer el desarrollo de este proceso, los importantes y veloces cambios tecnológicos a los que ha debido hacer frente y el impacto que ya tiene, antes de, en su caso, aprobación y publicación tanto dentro como fuera de la UE.

II. Las iniciativas para la regulación supranacional de la inteligencia artificial

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»[6]. 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 inteligencia artificial (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.

---

[5] https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/asuntos-economicos/Paginas/2023/081123-reglamento-identidad-digital-europea.aspx (consulta el 20 de noviembre de 2023) y 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/ (consultado el 10 de diciembre de 2023).

[6] https://www.consilium.europa.eu/media/21604/19-euco-final-conclusions-es.pdf, p. 7 (consultado el 20 de noviembre de 2023).

Transcurrido poco más de un año, el 7 de diciembre de 2018, la Comisión Europea presentó al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones 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[7]. 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[8].

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*[9], 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

---

[7] https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:-52018DC0795 (consultado el 20 de noviembre de 2023).

[8] https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/es/pdf, p. 8 (consultado el 20 de noviembre de 2023).

[9] https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1, (consultado el 20 de noviembre de 2023).

finalidad del presente Libro Blanco es formular alternativas políticas para alcanzar estos objetivos…»

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[10], la responsabilidad civil[11] y los derechos de propiedad intelectual[12], a las que siguieron, ya en 2021, resoluciones sobre el uso de la IA y en los sectores educativo, cultural y audiovisual[13] y en materia penal[14].

Antes de esta última Resolución ya se había publicado el documento con el que formalmente se abrió el procedimiento normativo que nos ocupa: el 21 de abril de 2021 se conoció 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, elaborada por la Comisión (Cotino, 2021)[15]. Según se explica en la exposición de motivos, «la propuesta establece normas armonizadas para el desarrollo, la introducción en el mercado y la utilización de sistemas de IA en la Unión a partir de un enfoque proporcionado basado en los riesgos. También propone una definición única de la IA que puede resistir el paso del tiempo. Asimismo, prohíbe determinadas prácticas particularmente perjudiciales de IA por ir en contra de los valores de la Unión y propone restricciones y

---

[10] 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) (consultada el 20 de noviembre de 2023).

[11] Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un régimen de responsabilidad civil en materia de inteligencia artificial, 2020/2014(INL) (consultada el 20 de noviembre de 2023).

[12] 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) (consultada el 20 de noviembre de 2023).

[13] 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 (consultado el 20 de noviembre de 2023).

[14] 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 (consultado el 20 de noviembre de 2023).

[15] https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52021PC0206 (consultada el 20 de noviembre de 2023).

salvaguardias específicas en relación con determinados usos de los sistemas de identificación biométrica remota con fines de aplicación de la ley. La propuesta establece una sólida metodología de gestión de riesgos para definir aquellos sistemas de IA que plantean un ”alto riesgo” para la salud y la seguridad o los derechos fundamentales de las personas. Dichos sistemas de IA tendrán que cumplir una serie de requisitos horizontales obligatorios que garanticen su fiabilidad y ser sometidos a procedimientos de evaluación de la conformidad antes de poder introducirse en el mercado de la Unión. Del mismo modo, se imponen obligaciones previsibles, proporcionadas y claras a los proveedores y los usuarios de dichos sistemas, con el fin de garantizar la seguridad y el respeto de la legislación vigente protegiendo los derechos fundamentales durante todo el ciclo de vida de los sistemas de IA. En el caso de determinados sistemas de IA, solo se proponen obligaciones mínimas en materia de transparencia, en particular cuando se utilizan robots conversacionales o ultrafalsificaciones».

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 que, en el momento de escribir estas líneas, son los famosos «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[16].

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[17], donde se señala que «para garantizar que la

[16] OpenAI entrenó el chat GPT-4 mediante la utilización de 170 billones de parámetros y un conjunto de datos de entrenamiento de 45 Gigabytes, unidad que equivale a (aproximadamente) a 10 elevado a 9 (mil veinticuatro millones) de bytes, la unidad más pequeña de información. Sobre los datos que «sirven de alimento» a la inteligencia artificial, Jove Villares, 2023.

[17] https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf (consultada el 20 de noviembre de 2023).

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».

Finalmente, y por lo que se conoce al finalizar estas páginas (noviembre de 2023), tenemos las enmiendas introducidas por el Parlamento Europeo en el texto de la Comisión y aprobadas el 14 de junio de 2023[18] (Barrio Andrés, 2023), que han incorporado 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»[19]; también de lo que se entiende 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 amplían los sistemas de IA prohibidos.

---

[18] https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html (a 20 de noviembre de 2023).

[19] En el 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, se define el «Sistema de inteligencia artificial» como el «diseñado para funcionar con un cierto nivel de autonomía y que, basándose en datos de entradas proporcionadas por máquinas o por personas, infiere cómo lograr un conjunto de objetivos establecidos utilizando estrategias de aprendizaje automático o basadas en la lógica y el conocimiento, y genera información de salida, como contenidos (sistemas de inteligencia artificial generativos), predicciones, recomendaciones o decisiones, que influyan en los entornos con los que interactúa»; por lo que respecta a los modelos fundacionales, se conciben como los entrenados en una gran cantidad de datos no etiquetados a escala (generalmente mediante aprendizaje autosupervisado y/o con recopilación automática de contenido y datos a través de internet mediante programas informáticos) que da como resultado un modelo que se puede adaptar a una amplia gama de tareas posteriores.

2. En otros contextos supranacionales

En un contexto más amplio como es el del Consejo de Europa cabe destacar, en primer lugar, el trabajo de investigación sobre algoritmos y derechos humanos, según el cual la IA afectará a un gran número, si no a la práctica totalidad, de nuestros derechos fundamentales[20]; 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, cabe recordar que el Comité de Ministros ha decidido adoptar un enfoque transversal de la inteligencia artificial en los diversos sectores del 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. Al mismo tiempo, el instrumento debe favorecer la innovación.

---

[20] *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, Published by the Council of Europe, 2018, disponible en https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (a 20 de noviembre de 2023).

En tercer lugar, la Asamblea Parlamentaria del Consejo de Europa ha aprobado 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, y ha instado al Comité de Ministros a elaborar un instrumento jurídicamente vinculante que regule la IA. La Asamblea tiene un Subcomité sobre Inteligencia Artificial y Derechos Humanos.

Pues bien, a estas alturas conocemos el texto del proyecto de trabajo consolidado de Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho, de 7 de julio de 2023, donde se exterioriza la conocida doble vertiente que presenta la IA: por una parte, los sistemas de inteligencia artificial pueden diseñarse, desarrollarse y utilizarse para ofrecer oportunidades sin precedentes para la protección y promoción de los derechos humanos y las libertades fundamentales, la democracia y el Estado de Derecho; por otra, el diseño, el desarrollo, la utilización y el desmantelamiento de los sistemas de inteligencia artificial puedan socavar la dignidad humana y la autonomía individual, los derechos humanos y las libertades fundamentales, la democracia y el Estado de Derecho.

Se trata, como parece lógico en un contexto de integración jurídica menos intenso, de un texto mucho más escueto y no tan detallado como la propuesta de Reglamento de la Unión Europea y donde, no obstante, se establecen los principios y las obligaciones necesarios para garantizar que el diseño, el desarrollo, la utilización y el desmantelamiento de los sistemas de inteligencia artificial sean plenamente compatibles con el respeto de la dignidad humana y la autonomía individual, los derechos humanos y las libertades fundamentales, el funcionamiento de la democracia y el respeto del Estado de Derecho.

Por su parte, en un contexto de mayor globalización, los 36 países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), junto con Argentina, Brasil, Colombia,

Costa Rica, Perú y Rumanía suscribieron el 22 de mayo de 2019 los principios de la OCDE sobre la Inteligencia Artificial[21].

En síntesis, dichos principios postulan que, primero, la IA debe estar al servicio de las personas y del planeta, impulsando un crecimiento inclusivo, el desarrollo sostenible y el bienestar; en segundo lugar, los sistemas de IA deben diseñarse de manera que respeten el Estado de Derecho, los derechos humanos, los valores democráticos y la diversidad, e incorporar salvaguardias adecuadas –por ejemplo, permitiendo la intervención humana cuando sea necesario– con miras a garantizar una sociedad justa y equitativa; en tercer término, los sistemas de IA deben estar presididos por la transparencia y una divulgación responsable a fin de garantizar que las personas sepan cuándo están interactuando con ellos y puedan oponerse a los resultados de esa interacción; en cuarto lugar, estos sistemas han de funcionar con robustez, de manera fiable y segura durante toda su vida útil, y los potenciales riesgos deberán evaluarse y gestionarse en todo momento. Finalmente, las organizaciones y las personas que desarrollen, desplieguen o gestionen sistemas de IA deberán responder de su correcto funcionamiento en consonancia con los principios precedentes.

La OCDE recomienda a los Gobiernos facilitar una inversión pública y privada en investigación y desarrollo que estimule la innovación en una IA fiable; fomentar ecosistemas de IA accesibles con tecnologías e infraestructura digitales, y mecanismos para el intercambio de datos y conocimientos; desarrollar un entorno de políticas que allane el camino para el despliegue de unos sistemas de IA fiables; capacitar a las personas con competencias de IA y apoyar a los trabajadores con miras a asegurar una transición equitativa y cooperar en la puesta en común de información entre países y sectores, desarrollar estándares y asegurar una administración responsable de la IA.

Finalmente, y poco antes del cierre de estas páginas, el 1 y 2 de noviembre, Estados Unidos, China, la Unión Europea y otros 26 países llegaron a un acuerdo global para avanzar la cooperación científica y tratar de frenar los posibles peligros «catastróficos» de la

[21] https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449#mainText (consultado el 20 de noviembre de 2023).

IA, la llamada *The Bletchley Declaration*[22], donde, entre otras cosas, se acogen con satisfacción los esfuerzos internacionales pertinentes para examinar y abordar el impacto potencial de los sistemas de IA en los foros existentes y otras iniciativas pertinentes, así como el reconocimiento de que es necesario abordar la protección de los derechos humanos, la transparencia y la explicabilidad, la equidad, la rendición de cuentas, la regulación, la seguridad, la supervisión humana adecuada, la ética, la mitigación de los prejuicios, la privacidad y la protección de datos.

Con esos fines, se acordó apoyar una red internacional inclusiva de investigación científica sobre la seguridad en las fronteras de la IA que abarque y complemente la colaboración multilateral, plurilateral y bilateral existente y nueva, incluso a través de los foros internacionales existentes y otras iniciativas pertinentes, para facilitar el suministro de la mejor ciencia disponible para la formulación de políticas y el bien público. Adicionalmente, y en reconocimiento del potencial positivo transformador de la IA, y como parte de la garantía de una cooperación internacional más amplia en materia de IA, resuelven mantener un diálogo mundial inclusivo que implique a los foros internacionales existentes y otras iniciativas pertinentes y contribuya de manera abierta a debates internacionales más amplios, y continuar la investigación sobre la seguridad de la IA en las fronteras para garantizar que los beneficios de la tecnología puedan aprovecharse de manera responsable para el bien de todos.

III. Los fundamentos jurídicos de la propuesta de regulación europea de la inteligencia artificial

La propuesta de la Comisión Europea explica cuál es su fundamento jurídico: en primer lugar, el artículo 114 del Tratado de Funcionamiento de la Unión Europea (TFUE), cuyo apartado 1 dispone que «el Parlamento Europeo y el Consejo, con arreglo al procedimiento legislativo ordinario y previa consulta al Comité Económico y Social, adoptarán las medidas relativas a la aproximación de las disposiciones

[22] https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 (consultado el 20 de noviembre de 2023).

legales, reglamentarias y administrativas de los Estados miembros que tengan por objeto el establecimiento y el funcionamiento del mercado interior». Añade la propuesta que «constituye una parte fundamental de la Estrategia para el Mercado Único Digital de la UE. Su objetivo primordial es garantizar el correcto funcionamiento del mercado interior mediante el establecimiento de normas armonizadas, en particular en lo que respecta al desarrollo, la introducción en el mercado de la Unión y el uso de productos y servicios que empleen tecnologías de IA o se suministren como sistemas de IA independientes» (Álvarez García y Tahiri Moreno, 2023, pp. 7 y ss).

En segundo lugar, la propuesta invoca el principio de subsidiariedad: «la naturaleza de la IA, que a menudo depende de conjuntos de datos amplios y variados y que puede integrarse en cualquier producto o servicio que circula libremente por el mercado interior, implica que los Estados miembros no pueden alcanzar de manera efectiva los objetivos de esta propuesta por sí solos. Asimismo, está surgiendo un mosaico de normas nacionales con posibles divergencias que entorpecerá la circulación fluida en la UE de productos y servicios asociados a sistemas de IA y no garantizará de manera efectiva la seguridad y la protección de los derechos fundamentales y los valores de la Unión en los distintos Estados miembros. Las estrategias nacionales orientadas a afrontar estos problemas solo crearán inseguridad jurídica y barreras adicionales, y ralentizarán la adopción de la IA por parte del mercado.

Resultará más fácil alcanzar los objetivos de esta propuesta a escala de la Unión para evitar que el mercado único se fragmente en marcos nacionales potencialmente contradictorios que impidan la libre circulación de bienes y servicios que lleven IA incorporada. Por otro lado, el establecimiento de un marco reglamentario europeo sólido para conseguir que la IA sea fiable garantizará la igualdad de condiciones y protegerá a todas las personas, al tiempo que reforzará la competitividad y la base industrial de Europa en el ámbito de la IA. Además, la única manera de proteger la soberanía digital de la UE y de aprovechar sus herramientas y competencias reguladoras para crear normas y reglas globales es mediante la adopción de medidas comunes a escala de la Unión».

En tercer lugar, se apela al principio de proporcionalidad: «la propuesta se fundamenta en los marcos jurídicos existentes y es

proporcionada y necesaria para alcanzar sus objetivos, ya que sigue un enfoque basado en los riesgos y únicamente impone cargas normativas cuando es probable que un sistema de IA entrañe altos riesgos para los derechos fundamentales y la seguridad… Unas normas armonizadas, y los instrumentos de orientación y cumplimiento en que se apoyan, ayudarán a los proveedores y los usuarios a cumplir los requisitos establecidos en la propuesta y a reducir al mínimo sus gastos. Los costes en que incurren los operadores son proporcionales a los objetivos logrados y a los beneficios que pueden obtener gracias a esta propuesta en términos económicos y de reputación».

Finalmente, se justifica el instrumento jurídico elegido –el Reglamento– «por la necesidad de aplicar uniformemente las nuevas normas, tales como la definición de IA, la prohibición de determinadas prácticas perjudiciales que la IA permitiría y la clasificación de determinados sistemas de IA. Puesto que, de conformidad con el artículo 288 del TFUE, los Reglamentos son directamente aplicables, la elección de este instrumento reducirá la fragmentación jurídica y facilitará el desarrollo de un mercado único de sistemas de IA legales, seguros y fiables… Al mismo tiempo, las disposiciones del Reglamento no son excesivamente prescriptivas y permiten que los Estados miembros actúen a distintos niveles en relación con aquellos elementos que no socavan los objetivos de la iniciativa, en particular en lo que respecta a la organización interna del sistema de vigilancia del mercado y la adopción de medidas para promover la innovación».

IV. ¿De qué hablamos cuando hablamos de inteligencia artificial?

En la corta historia de la IA –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 propuso John McCarthy– se han proporcionado distintas definiciones que, en general, aluden al desarrollo de sistemas que imitan o reproducen el pensamiento y obrar humanos,

actuando racionalmente –en el sentido de hacer lo «correcto» en función de su conocimiento– e interactuando con el medio[23].

La IA pretende sintetizar o reproducir los procesos cognitivos humanos, tales como la percepción, la creatividad, la comprensión, el lenguaje o el aprendizaje (Russel y Norvig, 2008, pp. 1 y ss.). 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

---

[23] En esos primeros momentos cundió el optimismo sobre la IA y su impacto: Herbert Simon predijo que «en veinte años las máquinas serán capaces de hacer el trabajo de una persona» y Marvin Minsky declaró en 1970 a la revista *Life* que «dentro de tres a ocho años tendremos una máquina con la inteligencia general de un ser humano». No hay que olvidar que poco antes (1969) se había llegado a la Luna y en el cine (1968) se había estrenado *2001: una odisea del espacio*, la película de Stanley Kubrick basada en varios cuentos de Arthur C. Clarke, en la que, como es conocido, el ordenador HAL 900 desempeña un papel decisivo y en la trama argumental se cuestiona si una máquina como esa puede tener emociones y sentimientos y, en última instancia. «morir».

Pero como estas optimistas previsiones no se cumplieron, entre otras razones por la existencia de pocos datos y la escasa capacidad de la computación del momento, a principios de los años setenta se enfriaron las expectativas, que volvieron a coger auge y financiación durante los años ochenta pero que decayeron de nuevo en los noventa hasta que, en el presente siglo, el acceso a cantidades ingentes de datos –Big Data–, la disponibilidad de procesadores muy potentes a bajo coste y el desarrollo de redes neuronales profundas y complejas consolidaron definitivamente la IA (Oliver, 2020, pp. 36 y ss.) y han despejado las dudas sobre su decisiva importancia en los próximos años.

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» (González Cabanes, Díaz Díaz; 2022, pp. 58 y 64).

La dificultad de ofrecer una definición «acabada» de la IA se presenta también en el ámbito jurídico (Barrio Andrés, 2022, pp. 14-21; Ruíz Tarrías, 2023, pp. 91-119), como se puede comprobar leyendo las diferentes versiones que se han ido ofreciendo en la propuesta de Reglamento de inteligencia artificial: 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 salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa» (artículo 3) pero, tras las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023, se define 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 los considerandos previos al articulado se explica con un poco más de detalle este concepto precisando que las principales características de la inteligencia artificial son su capacidad de aprendizaje, de razonamiento o de modelización, diferenciándola así de otros sistemas de software y planteamientos de programación más sencillos. «Los sistemas de IA están diseñados para operar con distintos niveles de autonomía, lo que significa que tienen al menos cierto grado de independencia de las acciones de los controles hu-

manos y ciertas capacidades para operar sin intervención humana. El término «basado en máquinas» se refiere al hecho de que todo sistema de IA funciona con máquinas. La referencia a objetivos explícitos o implícitos subraya que los sistemas de IA pueden operar con arreglo a objetivos explícitos definidos por el ser humano o a objetivos implícitos. Los objetivos del sistema de IA pueden ser diferentes de la finalidad prevista del sistema de IA en un contexto específico. La referencia a las predicciones incluye el contenido, una forma de predicción en tanto una posible información de salida producida por un sistema de IA. A efectos del Reglamento, los entornos deben entenderse como los contextos en los que operan los sistemas de IA, mientras que la información de salida generada por el sistema de IA, es decir, las predicciones, recomendaciones o decisiones, responden a los objetivos del sistema, sobre la base de las entradas de dicho entorno. Dicha información de salida influye a su vez en el entorno, por el simple hecho de introducir nueva información en él».

Como ya se ha dicho, en el 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, se define el «Sistema de inteligencia artificial» como el «diseñado para funcionar con un cierto nivel de autonomía y que, basándose en datos de entradas proporcionadas por máquinas o por personas, infiere cómo lograr un conjunto de objetivos establecidos utilizando estrategias de aprendizaje automático o basadas en la lógica y el conocimiento, y genera información de salida, como contenidos (sistemas de inteligencia artificial generativos), predicciones, recomendaciones o decisiones, que influyan en los entornos con los que interactúa».

Previamente, en la Resolución del Parlamento Europeo, de 3 de mayo de 2022, sobre la inteligencia artificial en la era digital se recordaba que hay una diferencia significativa entre la IA simbólica, que constituye el principal enfoque de la IA entre los años cincuenta y los años noventa, y la IA basada en datos y aprendizaje automático, que domina desde el año 2000: durante la primera oleada, la IA se desarrolló codificando los conocimientos y la experiencia de los expertos en un conjunto de reglas que luego ejecutaba una máquina;

en la segunda oleada, los procesos de aprendizaje automatizados de algoritmos basados en el procesamiento de grandes cantidades de datos, la capacidad de reunir datos procedentes de múltiples fuentes diferentes y de elaborar representaciones complejas de un entorno dado, y la determinación de patrones convirtieron a los sistemas de IA en sistemas más complejos, autónomos y opacos, lo que puede hacer que los resultados sean menos explicables; en consecuencia, la IA actual puede clasificarse en muchos subcampos y técnicas diferentes[24].

Por lo que respecta al concepto de IA asumido en otros contextos internacionales, cabe mencionar que en el texto del proyecto de trabajo consolidado de Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho, se define, de manera bastante diferente a la propuesta de Reglamento de la UE, como «cualquier sistema algorítmico o combinación de tales sistemas que utilice métodos computacionales derivados de la estadística u otras técnicas matemáticas y que genere texto, sonido, imágenes u otros contenidos o que ayude o sustituya la toma de decisiones humana. La Conferencia de las Partes podrá, en su caso, decidir interpretar esta definición de forma coherente con los avances tecnológicos pertinentes» (artículo 3); por su parte, la OCDE, en la línea del concepto de IA de la propuesta de Reglamento de la UE, lo entiende como un sistema operado por una máquina capaz de influir en su entorno produciendo resultados (como predicciones, recomendaciones o decisiones) para cumplir un conjunto determinado de objetivos. Utiliza datos y entradas generados por la máquina y/o introducidos por el ser humano para (i) percibir entornos reales y/o virtuales; (ii) producir una representación abstracta de estas percepciones en forma de modelos derivados de análisis automatizados (por ejemplo, aprendizaje automático) o manuales; y (iii) utilizar las inferencias del modelo para formular diferentes opciones de resultados. Los sistemas de IA están diseñados para funcionar de forma más o menos autónoma.

---

[24] https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_ES.html (consultada el 20 de noviembre de 2023).

V. El ámbito de aplicación de la propuesta de ley de inteligencia artificial europea

Tras las enmiendas introducidas por el Parlamento Europeo se ha ampliado el ámbito de futura aplicación de la «Ley» europea de IA, que abarcaría, en primer lugar, a los proveedores que introduzcan en el mercado o pongan en servicio sistemas de IA en la Unión, con independencia de si dichos proveedores están establecidos en la Unión o en un tercer país; de esta manera, si un proveedor no establecido en el territorio comunitario quiere poner en servicio un sistema de IA en la Unión Europea estará sometido a las exigencias y prohibiciones previstas en el Reglamento.

En segundo lugar, se aplicarán las previsiones del Reglamento a los implementadores de sistemas de IA que estén establecidos o se encuentren en la Unión; también, a los proveedores e implementadores de sistemas de IA que estén establecidos o se encuentren en un tercer país, cuando se aplique la legislación de un Estado miembro en virtud de un acto de Derecho internacional público o cuando esté previsto que la información de salida generada por el sistema se utilice en la Unión y a los proveedores que introduzcan en el mercado o pongan en servicio fuera de la Unión un sistema de IA contemplado en el artículo 5 [los prohibidos por el Reglamento] cuando el proveedor o distribuidor de dicho sistema esté situado dentro de la Unión; una nueva forma de «exportar» los efectos de la normativa europea más allá de las fronteras de la UE.

En tercer lugar, se aplicará el Reglamento a las personas afectadas, tal como se definen en el artículo 3, apartado 8 bis, que se encuentren en la Unión y cuya salud, seguridad o derechos fundamentales se vean perjudicados por el uso de un sistema de IA comercializado o puesto en servicio en la Unión.

Esta normativa no se aplicará a las autoridades públicas de terceros países ni a las organizaciones internacionales que entren dentro del ámbito de aplicación de este Reglamento cuando dichas autoridades u organizaciones utilicen sistemas de IA en el marco de la cooperación internacional o de acuerdos internacionales con fines de aplicación de la ley y cooperación judicial con la Unión o con uno o varios Estados miembros y sean objeto de una decisión de la Comisión adoptada de conformidad con el artículo 36 de la Directiva (UE)

2016/680 o el artículo 45 del Reglamento (UE) 2016/679 (decisión de adecuación) o formen parte de un acuerdo internacional celebrado entre la Unión y el tercer país o la organización internacional de que se trate con arreglo al artículo 218 del TFUE que ofrezca garantías adecuadas con respecto a la protección de la intimidad y los derechos y libertades fundamentales de las personas.

Tampoco se aplicará el Reglamento a las actividades de investigación, pruebas y desarrollo en relación a un sistema de IA antes de su introducción en el mercado o puesta en servicio, siempre que dichas actividades se lleven a cabo respetando los derechos fundamentales y la legislación aplicable de la Unión. Las pruebas en condiciones reales no estarán cubiertas por esta exención. La Comisión estará facultada para adoptar actos delegados con el fin de especificar esta exención y evitar que se abuse o pueda abusarse de ella.

Finalmente, el Reglamento no se aplicará a los componentes de IA proporcionados en el marco de licencias libres y de código abierto, salvo en la medida en que sean comercializados o puestos en servicio por un proveedor como parte de un sistema de IA de alto riesgo o de un sistema de IA incluido en el ámbito de aplicación de los títulos II o IV. Esta exención no se aplicará a los modelos fundacionales.

Por lo que respecta al uso de los sistemas de IA con fines militares, desaparece del articulado la previsión incluida en el texto presentado por la Comisión conforme a la cual no se aplicaría a los sistemas de IA desarrollados o utilizados exclusivamente con fines militares. No obstante, se mantiene en la exposición de motivo que los sistemas de IA desarrollados o utilizados exclusivamente con fines militares deben quedar excluidos del ámbito de aplicación del presente Reglamento cuando su uso sea competencia exclusiva de la política exterior y de seguridad común regulada en el título V del Tratado de la Unión Europea (TUE).

Por otra parte, y de acuerdo con el instrumento jurídico adoptado para la aprobación de esta normativa, el Reglamento no impedirá que los Estados miembros o la Unión mantengan o introduzcan disposiciones legales, reglamentarias o administrativas que sean más favorables a los trabajadores en lo que atañe a la protección de sus derechos respecto al uso de sistemas de IA por parte de los emplea-

dores o fomenten o permitan la aplicación de convenios colectivos que sean más favorables a los trabajadores.

En el texto del proyecto de trabajo consolidado de Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho se prevé su aplicación al diseño, desarrollo, uso y desmantelamiento de sistemas de inteligencia artificial con potencial para interferir en el respeto de los derechos humanos y las libertades fundamentales, el funcionamiento de la democracia y el respeto del Estado de Derecho. No se aplicará a las actividades de investigación y desarrollo relativas a los sistemas de inteligencia artificial, a menos que los sistemas se prueben o utilicen de un modo que pueda interferir con los derechos humanos y las libertades fundamentales, la democracia y el Estado de Derecho (artículo 4).

VI. Los principios generales aplicables a todos los sistemas de inteligencia artificial

Merced a las enmiendas introducidas por el Parlamento Europeo se incorporaron una serie de principios generales aplicables a todos los sistemas de IA y a los modelos fundacionales con el objetivo de promover un enfoque europeo coherente centrado en el ser humano con respecto a una inteligencia artificial ética y fiable, que esté plenamente en consonancia con la Carta, así como con los valores en los que se fundamenta la Unión. Dichos principios, por otra parte, bastante obvios, son los siguientes:

a) «Intervención y vigilancia humanas»: los sistemas de IA se desarrollarán y utilizarán como una herramienta al servicio de las personas, que respete la dignidad humana y la autonomía personal, y que funcione de manera que pueda ser controlada y vigilada adecuadamente por seres humanos.
b) «Solidez y seguridad técnicas»: los sistemas de IA se desarrollarán y utilizarán de manera que se minimicen los daños imprevistos e inesperados, así como para que sean sólidos en caso de problemas imprevistos y resistentes a los intentos de modificar el uso o el rendimiento del sistema de IA para permitir una utilización ilícita por parte de terceros malintencionados.

c) «Privacidad y gobernanza de datos»: los sistemas de IA se desarrollarán y utilizarán de conformidad con las normas vigentes en materia de privacidad y protección de datos, y tratarán datos que cumplan normas estrictas en términos de calidad e integridad.
d) «Transparencia»: los sistemas de IA se desarrollarán y utilizarán facilitando una trazabilidad y explicabilidad adecuadas, haciendo que las personas sean conscientes de que se comunican o interactúan con un sistema de IA, informando debidamente a los usuarios sobre las capacidades y limitaciones de dicho sistema de IA e informando a las personas afectadas de sus derechos.
e) «Diversidad, no discriminación y equidad»: los sistemas de IA se desarrollarán y utilizarán incluyendo a diversos agentes y promoviendo la igualdad de acceso, la igualdad de género y la diversidad cultural, evitando al mismo tiempo los efectos discriminatorios y los sesgos injustos prohibidos por el Derecho nacional o de la Unión.
f) «Bienestar social y medioambiental»: los sistemas de IA se desarrollarán y utilizarán 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.

En el caso de los sistemas de IA de alto riesgo, los principios generales serán aplicados y cumplidos por los proveedores o implementadores mediante los requisitos establecidos en el Reglamento. En el caso de los modelos fundacionales, los principios generales serán aplicados y cumplidos por los proveedores o implementadores.

La Comisión y la Oficina de IA incorporarán estos principios rectores en las peticiones de normalización, así como en las recomendaciones consistentes en orientaciones técnicas destinadas a prestar asistencia a proveedores e implementadores en cuanto al modo de desarrollar y utilizar los sistemas de IA. Las organizaciones europeas de normalización tendrán en cuenta los principios generales como objetivos basados en los resultados cuando elaboren las correspondientes normas armonizadas para los sistemas de IA de alto riesgo.

Por su parte, y en una línea muy similar, en el Capítulo III del proyecto de trabajo consolidado de Convenio Marco sobre inteligen-

cia artificial, derechos humanos, democracia y Estado de Derecho se incluyen los principios de diseño, desarrollo, utilización y desmantelamiento de sistemas de inteligencia artificial que se imponen a los Estados parte: transparencia y control, rendición de cuentas y responsabilidad, igualdad y no discriminación, intimidad y protección de datos; seguridad, protección y solidez e innovación segura.

VII. El enfoque basado en los riesgos

La regulación de la IA, tal y como se concibe en la UE aunque no solo en este espacio jurídico, requiere la aplicación de un enfoque basado en los riesgos que esté claramente definido, 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 en cuestión. Cuando el nivel de riesgo alcance determinada intensidad será necesario prohibir determinadas prácticas de inteligencia artificial que se consideren inaceptables (considerando 6 de la exposición de motivos de la propuesta) y es que , dependiendo de las circunstancias de su aplicación y utilización concretas, así como del nivel de desarrollo tecnológico, la inteligencia artificial puede generar riesgos y menoscabar los intereses públicos o privados y los derechos fundamentales de las personas físicas que protege el Derecho de la Unión, de manera tangible o intangible, incluidos daños físicos, psíquicos, sociales y económicos (considerando 4).

Este aproximación basada en el riesgo también está presente en el proyecto de trabajo consolidado de Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho, donde se dice (artículo 2) que cada parte mantendrá y adoptará, en su ordenamiento jurídico interno, las medidas graduadas y diferenciadas que sean necesarias y apropiadas teniendo en cuenta la gravedad y la probabilidad de las repercusiones negativas sobre los derechos humanos y las libertades fundamentales, la democracia y el Estado de Derecho del diseño, el desarrollo, la utilización y el desmantelamiento de los sistemas de inteligencia artificial.

En definitiva, estamos ante una concreción del bien conocido «principio de precaución», que ya está presente en el artículo 18.4 de la Constitución española, pues, como se recordará, se mandata a la ley para que limite el uso de la informática a fin de «garantizar el

honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos», y es igualmente un principio que, como también es sabido, guía la actuación de la Unión Europea[25] (San Martín Segura, 2023, pp. 231 y ss); así, y por citar únicamente dos ejemplos en 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,

[25] 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 CE 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 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://web.archive.org/web/20071222063317/http://europa.eu/scadplus/leg/es/lvb/l32042.htm (consulta el 20 de noviembre de 2023).

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», y 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»[26]

VIII. Los sistemas de inteligencia artificial objeto de prohibición

En la propuesta de la Comisión las prohibiciones eran cuatro y englobaban las prácticas que tienen un gran potencial para manipular a las personas mediante técnicas subliminales que trasciendan su consciencia o que aprovechan las vulnerabilidades de grupos de personas concretos, como los menores o las personas con discapacidad, para alterar de manera sustancial su comportamiento de un modo que es probable que les provoque perjuicios físicos o psicológicos a ellos o a otras personas. La propuesta prohibía igualmente que las autoridades públicas realizasen calificación social basada en IA con fines generales. Por último, también se prohibía, salvo excepciones limitadas, el uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público con fines de aplicación de la ley.

Pues bien, tras las enmiendas introducidas por el Parlamento Europeo, se ha ampliado el abanico de prácticas prohibidas por entender que suponen un riesgo inaceptable y se eleva el número a nueve: se mantiene, en primer lugar, la prohibición de los sistemas de IA que se sirvan de técnicas subliminales 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

[26] Sobre el tratamiento de las situaciones de riesgo tecnológico y científico por parte del Derecho, Esteve Pardo, 1999 y 2009; sobre el concepto de «riesgo algorítmico», San Martín Segura, 2023, pp. 272 y ss.; sobre los riesgos en el ámbito de la inteligencia artificial, Barrio Andrés, 2021, pp. 186 y ss.

grupo de personas mermando de manera apreciable su capacidad para adoptar una decisión informada y causando así que la persona tome una decisión que de otro modo no habría tomado, de un modo que provoque o sea probable que provoque perjuicios significativos a esa persona o a otra persona o grupo de personas; se conserva igualmente la prohibición de los sistemas de IA que aprovechen alguna de las vulnerabilidades de una persona o un grupo específico de personas con el objetivo o el efecto de alterar de manera sustancial su comportamiento de un modo que provoque o sea probable que les provoque perjuicios significativos a esa persona o a otra.

En tercer lugar, se mantiene, aunque con modificaciones, la prohibición de sistemas de IA con el fin de evaluar o clasificar a las personas físicas o grupos de personas físicas a efectos de su calificación social durante un período determinado de tiempo 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 una o varias de las situaciones siguientes: un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos enteros en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente; la prohibición iba dirigida inicialmente a las autoridades públicas o a quien ejerciera su representación y con la enmienda del Parlamento Europeo se incluye a los sujetos privados, físicos y jurídicos (sobre estos sistemas, García Sánchez, 2022, pp. 779-793).

Finalmente, también subsiste la prohibición del uso de sistemas de identificación biométrica remota «en tiempo real» en espacios de acceso público, pero se han eliminado todas las excepciones que proponía la Comisión[27], algo que, en mi opinión, está por ver que se mantenga en el texto si resulta finalmente aprobado.

---

[27] i) la búsqueda selectiva de posibles víctimas concretas de un delito, incluidos menores desaparecidos;

ii) la prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas o de un atentado terrorista;

iii) la detección, la localización, la identificación o el enjuiciamiento de la persona que ha cometido o se sospecha que ha cometido alguno de los delitos mencionados en el artículo 2, apartado 2, de la Decisión Marco 2002/584/JAI del Consejo, para el que la normativa en vigor en el Estado miembro implicado imponga una pena o una medida de seguridad

Además, se han incorporado cinco nuevas prohibiciones: 1) la introducción en el mercado, la puesta en servicio o la utilización de sistemas de categorización biométrica que clasifiquen a personas físicas con arreglo a atributos o características sensibles o protegidos, o sobre la base de una inferencia de dichos atributos o características; 2) la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales a partir de internet o de imágenes de circuito cerrado de televisión; 3) la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA para inferir las emociones de una persona física en los ámbitos de la aplicación de la ley y la gestión de fronteras, en lugares de trabajo y en centros educativos; 4) la puesta en servicio o la utilización de sistemas de IA para el análisis de imágenes de vídeo grabadas de espacios de acceso público que empleen sistemas de identificación biométrica remota «en diferido», salvo que estén sujetos a una autorización judicial previa de conformidad con el Derecho de la Unión y sean estrictamente necesarios para una búsqueda selectiva destinada a fines de aplicación de la ley y relacionada con un delito grave (según la definición del artículo 83, apartado 1, del TFUE) concreto que ya se haya cometido; 5) la introducción en el mercado, la puesta en servicio o la utilización de sistemas de IA para llevar a cabo evaluaciones de riesgo de personas físicas o grupos de personas físicas con el objetivo de determinar el riesgo de que estas personas cometan delitos o infracciones o reincidan en su comisión, o para predecir la comisión o reiteración de un delito o infracción administrativa reales o potenciales, mediante la elaboración del perfil de personas físicas o la evaluación de rasgos de personalidad y características, en particular la ubicación de la persona o las conductas delictivas pasadas de personas físicas o grupos de personas físicas.

Esta última prohibición también es posible que se elimine o que los sistemas prohibidos por el Parlamento pasen finalmente a ser calificados como de «alto riesgo» pues es discutible que los Estados acepten renunciar a todas estas herramientas de «inteligencia artificial policial» o de «policía predictiva». Y es que, como señala Miró

---

privativas de libertad cuya duración máxima sea al menos de tres años, según determine el Derecho de dicho Estado miembro.

Llinares (2019, p. 100), «hoy, y en parte gracias a las expectativas que parece dar la IA, la sociedad no espera solo que la policía reaccione a los accidentes de tráfico, a los hurtos en los lugares turísticos o a los altercados y agresiones violentas relacionadas con manifestaciones deportivas o políticas, sino que no sucedan, que se intervenga incluso antes de que acontezcan (…)».

A este respecto, en España es bien conocida la herramienta predictiva *VioGén* (González Álvarez *et alii*, 2018; Presno Linera, 2023; San Martín Segura, 2023), que desde su entrada del sistema en funcionamiento y hasta el 30 de septiembre de 2023 ha permitido la evaluación de 770.944 casos de violencia de género. Del total de los casos registrados había, en esa fecha, 81.319 activos, es decir, con seguimiento policial, y 675.698 inactivos. De los activos, 31.685 eran sin riesgo apreciado, 35.715 con riesgo bajo, 13.028 con riesgo medio, 868 con riesgo alto y 23 con riesgo extremo (Estadísticas Ministerio del Interior, 2023). En conjunto estamos hablando del mayor sistema del mundo en ese ámbito.

Esta herramienta predictiva no es, en rigor, IA, pues no usa algoritmos que «aprenden» a partir del procesamiento de datos; es un sistema actuarial que utiliza modelos estadísticos para inferir el riesgo que puede correr una víctima (tanto de agresión como de homicidio) así como su evolución en base a un conjunto de indicadores que han sido determinados y posteriormente evaluados por un grupo de expertos (Fundación Éticas, p. 10). No obstante, algunas informaciones del Ministerio del Interior parecen mostrar que no está descartada la incorporación de un algoritmo de autoaprendizaje[28].

Si llevamos a cabo una «evaluación de su impacto algorítmico» (Simón Castellanos, 2023, pp. 67 y ss. y 101 y ss..) sobre los derechos del presunto agresor para valorar si las eventuales limitaciones se corresponden con las conocidas exigencias del principio de proporcionalidad cabe concluir, en primer lugar, *VioGén* parece una herramienta idónea, pues las medidas que adoptar y que, en su caso, afectarán al presunto agresor, están vinculadas al nivel de riesgo detectado, es decir, se orientan a proteger la vida y la integridad física de la mujer en tanto que sean necesarias para ello. En segun-

[28] https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2020/151220-inteligencia.aspx. (a 20 de noviembre de 2023].

do término, *VioGén* es, en principio, una herramienta necesaria en términos relativos, pues no parecen existir sistemas menos gravosos para el presunto agresor que al tiempo garanticen de forma similar los derechos fundamentales de la mujer denunciante; así, si el riesgo detectado es alto se hará un control aleatorio de sus movimientos y contactos esporádicos con personas que frecuente o de su entorno; si el riesgo se califica como extremo se hará un control intensivo de sus movimientos hasta que este deje de ser una amenaza inminente para la seguridad de la víctima. Finalmente, es un sistema que aporta proporcionalidad en un sentido estricto entre los derechos que pueden verse limitados como resultado de su aplicación y los derechos que pretenden garantizarse, pues se trata de predicciones hechas para tener validez durante un período de tiempo no muy largo que, en su caso, implicarán medidas limitativas que se correspondan con el pronóstico concreto de peligrosidad alcanzado, y todo ello bajo la supervisión de personas especializadas en el tratamiento y seguimiento de la violencia de género.

No obstante, y a nuestro juicio, para llevar a cabo con más rigor una evaluación del impacto algorítmico y, en suma, de la proporcionalidad de *VioGén*, es imprescindible más información sobre su metodología que la que actualmente ofrece el Ministerio del Interior; en particular, sobre el valor relativo de los diferentes tipos de datos con los que trabaja y sobre cómo se combinan para el resultado final (más extensamente, Martínez Garay et al., 2023, p. 18). Faltan, en suma, mayor rendición de cuentas, transparencia y explicabilidad (Presno Linera, 2022, pp. 25 y ss; 2023, p. 7) pero, dados sus resultados, nos parece un ejemplo de policía predictiva que, difícilmente, será objeto de una prohibición total si llega a convertirse en un sistema de IA.

IX. Los requisitos para los sistemas de inteligencia artificial de alto riesgo

Siguiendo con el criterio de los riesgos, tanto en la propuesta de la Comisión como en el texto de las enmiendas aprobadas por el Parlamento Europeo se conviene en establecer normas comunes para todos los sistemas de IA considerados de alto riesgo al objeto de garantizar un nivel elevado y coherente de protección de los intereses públicos en lo que respecta a la salud, la seguridad y los derechos

fundamentales (Soriano Arnanz, 2021; Añón Roig, 2022, pp. 17-49; de Hoyos Sancho, 2022, pp. 403-422; San Martín Segura, 2023, p. 288). El Parlamento Europeo ha añadido como bienes a proteger la democracia, el Estado de Derecho y el medio ambiente. Dichas normas deben ser coherentes con, entre disposiciones, la Carta de los Derechos Fundamentales de la Unión Europea, no deben ser discriminatorias y deben estar en consonancia con los compromisos de la Unión en materia de comercio internacional.

Entre los sistemas de IA se considerarán de alto riesgo, entre otros, los que se incluyen en un Anexo que acompañará al Reglamento si presentan un riesgo significativo de causar perjuicios para la salud, la seguridad o los derechos fundamentales de las personas físicas o, en su caso, si presentan un riesgo significativo de causar perjuicios medioambientales[29].

Y para mantener el control del riesgo se establecerá, implantará, documentará y mantendrá un sistema de gestión durante todo el ciclo de vida del sistema, lo que requerirá revisiones y actualizaciones periódicas. Como parece lógico, el control de los sistemas de IA de alto riesgo se realizará antes de su introducción en el mercado o puesta en servicio y los ensayos se realizarán a partir de parámetros y um-

[29] En el 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 se entiende por «Sistema de inteligencia artificial de alto riesgo» el que cumpla alguno de los siguientes supuestos: a) Un sistema de inteligencia artificial que constituya un producto regulado por la legislación de armonización de la Unión especificada en el anexo VII del presente real decreto se considerará de alto riesgo si debe someterse a una evaluación de la conformidad por un tercero con vistas a la introducción en el mercado o puesta en servicio de dicho producto con arreglo a la legislación mencionada. b) Un sistema de inteligencia artificial que vaya a ser utilizado como un componente que cumple una función de seguridad y cuyo fallo o defecto de funcionamiento pone en peligro la salud y la seguridad de las personas o los bienes en un producto regulado por una norma armonizada de la Unión Europea, si debe someterse a una evaluación de conformidad por parte de un tercero con vistas a la introducción en el mercado o puesta en servicio de dicho producto con arreglo a la legislación de armonización aplicable. Este supuesto será aplicable, aunque el sistema de inteligencia artificial se comercialice o se ponga en servicio independientemente del producto. c) Sistemas de inteligencia artificial mencionados en el anexo II, siempre que la respuesta del sistema sea relevante respecto a la acción o decisión a tomar, y pueda, por tanto, provocar un riesgo significativo para la salud, los derechos de las personas trabajadoras en el ámbito laboral o la seguridad o los derechos fundamentales (artículo 3.4).

brales de probabilidades previamente definidos que sean adecuados para la finalidad prevista o el uso indebido razonablemente previsible del sistema de IA de alto riesgo de que se trate.

Los sistemas de alto riesgo estarán sometidos a una serie de exigencias orientadas a su control y que cabe resumir como de registro de actividades, de medición de su impacto ambiental, de transparencia, de inteligibilidad, de supervisión humana, de seguridad en el diseño, de resistencia a los errores, de subsanación de eventuales sesgos y de resistencia frente a los intentos de usos no autorizados.

Con más detalle, estos sistemas se diseñarán y desarrollarán con capacidades que permitan registrar automáticamente eventos («archivos de registro») mientras están en funcionamiento.

En segundo lugar, deberán registrar el consumo de energía, la medición o el cálculo del uso de los recursos y el impacto ambiental del sistema de IA de alto riesgo durante todas las fases de su ciclo de vida.

En tercer lugar, estos sistemas se diseñarán y desarrollarán de un modo que garantice que funcionan con un nivel de transparencia suficiente para que los proveedores y usuarios entiendan razonablemente el funcionamiento del sistema. El usuario estará capacitado para comprender y utilizar adecuadamente el sistema de IA sabiendo, en general, cómo funciona y qué datos trata, lo que le permitirá explicar las decisiones adoptadas por el sistema a la persona afectada.

En cuarto lugar, los sistemas de IA de alto riesgo irán acompañados de las instrucciones de uso inteligibles correspondientes en un formato digital adecuado o puestas a disposición de otra forma en un medio durable, las cuales incluirán información concisa, correcta, clara y, en la medida de lo posible, completa que proporcione asistencia en el funcionamiento y el mantenimiento del sistema de IA, que contribuya a fundamentar la toma de decisiones informada por parte de los usuarios y sea razonablemente pertinente, accesible y comprensible para los usuarios.

En quinto término, estos sistemas se diseñarán y desarrollarán de modo que sean vigilados de manera efectiva por personas físicas, lo que incluye dotarlos de una herramienta de interfaz humano-máquina adecuada, entre otras cosas, de forma proporcionada a los riesgos

asociados a dichos sistemas. Las personas físicas encargadas de garantizar la vigilancia humana tendrán un nivel suficiente de alfabetización en materia de IA y contarán con el apoyo y la autoridad necesarios para ejercer esa función durante el período en que los sistemas de IA estén en uso y para permitir una investigación exhaustiva tras un incidente.

En la propuesta enmendada por el Parlamento se explica que el objetivo de la vigilancia humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad, los derechos fundamentales o el medio ambiente que pueden surgir cuando un sistema de IA de alto riesgo se utiliza conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persisten a pesar de aplicar otros requisitos establecidos y cuando las decisiones basadas únicamente en el procesamiento automatizado por parte de sistemas de IA producen efectos jurídicos o significativos de otro tipo para las personas o grupos de personas con los que se deba utilizar el sistema.

En sexto lugar, los sistemas de IA de alto riesgo se diseñarán y desarrollarán siguiendo el principio de seguridad desde el diseño y por defecto. Teniendo en cuenta su finalidad prevista, deben alcanzar un nivel adecuado de precisión, solidez, seguridad y ciberseguridad y funcionar de manera consistente en esos sentidos durante todo su ciclo de vida.

En séptimo lugar, se deberán adoptar medidas técnicas y organizativas para garantizar que estos sistemas sean lo más resistentes posible a los errores, fallos e incoherencias que pueden surgir en los propios sistemas o en el entorno donde operan, en particular a causa de su interacción con personas físicas u otros sistemas.

En octavo lugar, los sistemas de IA de alto riesgo que continúan aprendiendo tras su introducción en el mercado o puesta en servicio se desarrollarán de tal modo que los posibles sesgos en la información de salida que influyan en los datos de entrada en futuras operaciones («bucle de retroalimentación») y la manipulación maliciosa de los datos de entrada utilizados para el aprendizaje durante el funcionamiento se subsanen debidamente con las medidas de mitigación oportunas.

Finalmente, estos sistemas serán resistentes a los intentos de terceros no autorizados de alterar su uso, comportamiento, información de salida o funcionamiento aprovechando las vulnerabilidades del sistema.

Una exigencia previa para que el control de los riesgos sea eficaz es la llamada «alfabetización» en materia de IA, cuya promoción se configura como una obligación tanto para la Unión Europea como para los Estados miembros y que se extiende a los proveedores e implementadores de sistemas de IA, garantizando, en todo caso, un equilibrio adecuado en materia de género y de edad, con vistas a permitir un control democrático de los sistemas de IA. En particular, dichas medidas de alfabetización consistirán en la enseñanza de nociones y capacidades básicas sobre sistemas de IA y su funcionamiento, incluidos los distintos tipos de productos y usos, sus riesgos y sus beneficios.

X. Los modelos fundacionales

Si hubiera que destacar una novedad incorporada en las fases finales de elaboración del Reglamento y que no estaba prevista ni en la propuesta de la Comisión ni en la orientación general del Consejo es, sin duda, la de los llamados modelos fundacionales, que en las enmiendas parlamentarias se han definido como un avance reciente en el que se desarrollan modelos de IA a partir de algoritmos diseñados para optimizar la generalidad y versatilidad de la información de salida. A menudo, estos modelos se entrenan con un amplio abanico de fuentes de datos y grandes volúmenes de datos a fin de llevar a cabo una extensa gama de tareas posteriores, incluidas algunas para las que no han sido desarrollados y entrenados específicamente[30].

[30] En el 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 se entiende por «modelo fundacional» un modelo de inteligencia artificial entrenado en una gran cantidad de datos no etiquetados a escala (generalmente mediante aprendizaje autosupervisado y/o con recopilación automática de contenido y datos a través de internet mediante programas informáticos) que da como resultado un modelo que se puede adaptar a una amplia gama de tareas posteriores (artículo 3.6).

El modelo fundacional puede entrenarse con diferentes métodos, como el aprendizaje supervisado[31] o el aprendizaje reforzado[32]. Los sistemas de IA con una finalidad prevista específica o los sistemas de IA de uso general pueden constituir aplicaciones concretas de un modelo fundacional, lo que significa que cada modelo fundacional puede reutilizarse en innumerables sistemas de IA de etapas posteriores o sistemas de IA de uso general. Estos modelos tienen una importancia cada vez mayor para numerosas aplicaciones y sistemas de etapas posteriores".

Se trata de modelos que han alcanzado una gran popularidad en los últimos tiempos, como el ChatGPT[33] o BARD[34], en especial por su extraordinaria capacidad para generar textos, códigos e imágenes, algo que ha generado, en palabras introducidas por el Parlamento Europeo en la propuesta de Reglamento, «una incertidumbre significativa sobre el modo en que evolucionarán los modelos fundacionales, tanto en lo que se refiere a la tipología de los modelos como a su autogobernanza».

---

[31] Es el aprendizaje en el cual los datos de entrenamiento que se aportan al algoritmo incluyen la solución deseada para que pueda aprender. Se dice que están «etiquetados». Para el ejemplo del filtro de spam en el correo electrónico, entrenando el sistema a partir de un conjunto de emails etiquetados con *spam* y no *spam*, el sistema podría predecir qué tipo correo sería uno recién recibido (González Cabanes y Díaz Díaz, 2022, p. 49).

[32] Un tipo de aprendizaje en el que el sistema es un simulador o «agente» y aprende en base a ensayo-error. Tras cada ensayo llega una recompensa o una penalización, y el sistema aprende generando una estrategia o «política» que refuerza las acciones que le han llevado a la recompensa, definiendo qué acciones debe escoger el agente en una situación dada. Este aprendizaje se potencia con supercomputadoras que aceleran el aprendizaje con varias simulaciones en paralelo, de forma que en poco tiempo el sistema adquiere el aprendizaje de plazos mucho más largos. Veamos algunos casos de uso: • Ampliando el caso del filtro de *spam* para el correo, una vez entrenado el sistema como supervisado, en la bandeja de *spam* nos pregunta si los correos son realmente *spam* o no lo son, para generar esa política que potencie las decisiones que ha tomado de forma correcta. • Tiene muchas aplicaciones a la robótica, por ejemplo, para el control de calidad en líneas de producción. Los descartes, piezas defectuosas, son reforzados mediante validación del defecto por parte de operarios y el sistema aprende de esta forma potenciando los criterios que le llevaron a considerarlo como tal, defectuoso… (González Cabanes y Díaz Díaz, 2022, pp. 49 y 50).

[33] Chat Generative Pre-Trained Transformer, desarrollado en 2022 por la empresa OpenAI: página web: https://chat.openai.com/auth/login

[34] Sistema conversacional de inteligencia artificial desarrollado por Google en 2023; página web: https://bard.google.com/

Y se añade que, «teniendo en cuenta la complejidad de dichos modelos y su impacto imprevisible, así como la falta de control del proveedor de IA de etapas posteriores sobre el desarrollo del modelo fundacional y el consiguiente desequilibrio de poder, y con el fin de garantizar un reparto equitativo de las responsabilidades a lo largo de la cadena de valor de la IA… deben evaluar y mitigar los posibles riesgos y perjuicios mediante un diseño, unas pruebas y un análisis adecuados, aplicar medidas de gobernanza de datos –en particular, una evaluación de los sesgos– y cumplir requisitos de diseño técnico que garanticen niveles adecuados de rendimiento, previsibilidad, interpretabilidad, corregibilidad, seguridad y ciberseguridad, así como cumplir las normas medioambientales… Los modelos fundacionales generativos deben garantizar la transparencia sobre el hecho de que el contenido ha sido generado por un sistema de IA y no por seres humanos…»

Estos modelos estarán sujetos a las obligaciones generales impuestas a los sistemas de IA, y ya mencionadas, y los proveedores de modelos fundacionales destinados específicamente a generar, con distintos niveles de autonomía, contenidos como texto, imágenes, audio o vídeo complejos («IA generativa») formarán y, en su caso, diseñarán y desarrollarán el modelo fundacional de manera que se garanticen salvaguardias adecuadas contra la generación de contenidos que infrinjan el Derecho de la Unión, en consonancia con el estado de la técnica generalmente reconocido y sin perjuicio de los derechos fundamentales, incluida la libertad de expresión; además, documentarán y pondrán a disposición del público un resumen suficientemente detallado del uso de los datos de formación protegidos por la legislación sobre derechos de autor.

Adicionalmente, y por su vinculación con la generación de contenidos por los modelos fundacionales, se ha introducido en la propuesta de Reglamento la definición de «ultrafalsificación» como un contenido de sonido, imagen o vídeo manipulado o sintético que puede inducir erróneamente a pensar que es auténtico o verídico, y que muestra representaciones de personas que parecen decir o hacer cosas que no han dicho ni hecho, producido utilizando técnicas de IA, incluido el aprendizaje automático y el aprendizaje profundo.

XI. Las autoridades de supervisión de la inteligencia artificial

Como es conocido, existen garantías orgánicas especializadas en materia de protección de datos, tanto en el ámbito nacional como en el europeo, con organismos como la Agencia Española de Protección de Datos y el Comité Europeo de Protección de Datos; la primera es una autoridad administrativa independiente de ámbito estatal, de las previstas en la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, con personalidad jurídica y plena capacidad pública y privada, que actúa con plena independencia de los poderes públicos en el ejercicio de sus funciones; el segundo es un organismo europeo independiente que tiene como objetivo garantizar la aplicación coherente del Reglamento General de Protección de Datos y la directiva europea sobre protección de datos en el ámbito policial. Parece obvio que estas entidades también están llamadas a jugar un papel importante en todo aquello que vincula a la IA con la protección de datos personales.

Por su parte, la ya citada resolución del Parlamento Europeo a propósito de normas de Derecho civil sobre robótica prevé la creación de una agencia europea para la robótica y la inteligencia artificial que «proporcione los conocimientos técnicos, éticos y normativos necesarios para apoyar la labor de los actores públicos pertinentes, tanto a nivel de la Unión como a nivel de los Estados miembros, en su labor de garantizar una respuesta rápida, ética y fundada ante las nuevas oportunidades y retos –sobre todo los de carácter transfronterizo– que plantea el desarrollo tecnológico de la robótica, por ejemplo en el sector del transporte» y «considera justificado, en vista del potencial de la robótica, de los problemas que suscita y de la actual dinámica de inversiones, que esa agencia europea esté dotada de un presupuesto adecuado y de un personal compuesto por reguladores y por expertos externos en cuestiones técnicas y deontológicas dedicados a controlar, desde un punto de vista intersectorial y pluridisciplinar, las aplicaciones basadas en la robótica, a determinar las normas en materia de mejores prácticas y, en su caso, a recomendar medidas reguladoras, a definir nuevos principios y a hacer frente a posibles problemas de protección de los consumidores y desafíos sistémicos; pide a la Comisión (y a la agencia europea, en el caso de que se cree)

que informen anualmente al Parlamento sobre los últimos avances de la robótica, así como sobre las medidas que resulten necesarias».

Ya centrada en la IA y no tanto en la robótica, la propuesta de Reglamento que estamos analizando prevé (artículo 59, enmendado por el Parlamento Europeo) que «cada Estado miembro designará una autoridad nacional de supervisión, que se organizará de manera que se preserve la objetividad e imparcialidad de sus actividades y funciones» a más tardar … [tres meses después de la fecha de entrada en vigor del presente Reglamento]; esta autoridad garantizará la aplicación y la ejecución del Reglamento y deberá actuar de manera independiente, imparcial y objetiva. Se dispone que cada autoridad nacional de supervisión ejercerá sus funciones en el territorio de su Estado miembro y, de darse un caso que afecte a dos o más autoridades nacionales de supervisión, la del Estado miembro en el que haya tenido lugar la infracción será considerada la autoridad de supervisión principal.

A efectos de que puedan cumplir estos objetivos, se prescribe que tal autoridad debe disponer de recursos técnicos, financieros y humanos adecuados, así como de las infraestructuras para el desempeño eficaz de sus funciones; en concreto, dispondrá permanentemente de suficiente personal cuyas competencias y conocimientos técnicos incluirán un conocimiento profundo de las tecnologías de inteligencia artificial, datos y computación de datos, la protección de datos personales, la ciberseguridad, el Derecho en materia de competencia, los riesgos para los derechos fundamentales, la salud y la seguridad, y conocimientos acerca de las normas y requisitos legales vigentes.

Pues bien, España, adelantándose no solo al plazo previsto en el Reglamento sino a la propia aprobación del mismo, ya cuenta con su autoridad nacional de supervisión: la Ley 22/2021, de 28 de diciembre, de Presupuestos Generales del Estado para el año 2022, recogió, en su disposición adicional centésimo trigésima, la «creación de la Agencia Española de Supervisión de Inteligencia Artificial», autorizando al Gobierno a impulsar una ley para la creación de la Agencia Española de Supervisión de Inteligencia Artificial, configurada como una Agencia Estatal dotada de personalidad jurídica pública, patrimonio propio y autonomía en su gestión, con potestad administrativa. Por su parte, la Ley 28/2022, de 21 de diciembre,

de fomento del ecosistema de las empresas emergentes, prevé la «creación de la Agencia Española de Supervisión de Inteligencia Artificial», cumpliendo con ello la exigencia prevista en el artículo 91 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público y eso es lo que se ha hecho a través del Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial (Barrio Andrés, 2023).

Con arreglo al artículo 4 del real decreto, le corresponde a la Agencia llevar a cabo tareas de supervisión, el asesoramiento, la concienciación y la formación dirigidas a entidades de derecho público y privado para la adecuada implementación de toda la normativa nacional y europea en torno al adecuado uso y desarrollo de los sistemas de inteligencia artificial, más concretamente, de los algoritmos. Además, la Agencia tendrá la función de inspección, comprobación, sanción y demás que le atribuya la normativa europea que le resulte de aplicación y, en especial, en materia de inteligencia artificial. En el ámbito de la competencia estatal, ejercerá las funciones de autoridad responsable de la supervisión, y en su caso sanción, de los sistemas de inteligencia artificial con el objeto de eliminar o reducir los riesgos para la integridad, la intimidad, la igualdad de trato y la no discriminación, en particular entre mujeres y hombres, y demás derechos fundamentales que pueden verse afectados por el mal uso de los sistemas[35].

Una de las funciones de la Agencia es la promoción de entornos de prueba que permitan una correcta adaptación de sistemas

[35] De acuerdo con el artículo 8, La Agencia observará los principios de interés general por los que debe regirse la actuación. En el ejercicio de sus funciones específicas se regirá, además, por los siguientes principios básicos:

a) Autonomía, entendida como la capacidad de la Agencia de gestionar, en los términos previstos en su Estatuto, los medios puestos a su disposición para alcanzar los objetivos comprometidos.

b) Independencia técnica, basada en la capacitación, especialización, profesionalidad y responsabilidad individual del personal al servicio de la Agencia que deberá observar los valores de competencia, ética profesional y responsabilidad pública que son de aplicación. En el desempeño de sus funciones y en el ejercicio de sus competencias, la Agencia actuará con plena autonomía.

c) Transparencia en todas las actividades administrativas y cumplimiento de las obligaciones de buen gobierno por parte de los responsables públicos de la Agencia, así como

innovadores de inteligencia artificial a los marcos jurídicos en vigor (artículo 10.1.a) y, a este respecto, también se ha aprobado el anteriormente mencionado 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, que tiene por objeto establecer un entorno controlado de pruebas para ensayar el cumplimiento de ciertos requisitos por parte de algunos sistemas de inteligencia artificial que puedan suponer riesgos para la seguridad, la salud y los derechos fundamentales de las personas. Asimismo, regula el procedimiento de selección de los sistemas y entidades que participarán en dicho entorno[36].

Este decreto, a la hora de excluir de los entornos de pruebas a los sistemas de IA cuya prohibición se pretende en el texto europeo, incluye literalmente los que en su día incorporó la Comisión y avaló el Consejo[37], no la lista más extensa y restrictiva aprobada por el Par-

---

la rendición de cuentas y compromisos para presentar la información precisa y completa sobre todos los resultados y procedimientos utilizados en la gestión.

d) Eficacia en su actuación, utilizando todos los medios disponibles para el logro de los fines definidos en su Estatuto.

e) Eficiencia en la asignación y utilización de recursos públicos y evaluación continuada de la calidad de los procesos de gestión y de los procedimientos de actuación, que se efectuará atendiendo a los criterios de legalidad, celeridad, simplificación y accesibilidad electrónica.

f) Cooperación interinstitucional, entendido como la búsqueda de sinergias en la colaboración con otras Administraciones Públicas, agentes e instituciones, públicas o privadas, nacionales e internacionales para el fomento del conocimiento en todos sus ámbitos.

g) Integración del principio de igualdad de trato entre mujeres y hombres, promoviendo la perspectiva de género y una composición equilibrada de mujeres y hombres en sus órganos, consejos y comités y actividades…

[36] A los efectos de esa norma, se entiende por entorno controlado de pruebas o experiencia el entorno o experiencia, con una duración determinada, que proporciona un contexto estructurado para el desarrollo de las actuaciones necesarias que posibiliten a proveedores y usuarios de los sistemas de inteligencia artificial de alto riesgo, sistemas de propósito general y modelos fundacionales, que realicen las pruebas necesarias para la implementación de los requisitos establecidos en este real decreto, bajo la supervisión del órgano competente (artículo 3.2).

[37] Los sistemas de inteligencia artificial propuestos por los proveedores IA solicitantes no podrán estar incluidos en los siguientes supuestos:

a) Sistemas de inteligencia artificial comercializados o puestos en servicio para actividades militares, de defensa o seguridad nacional, cualquiera que sea la entidad que desarrolle esas actividades.

b) Sistemas de inteligencia artificial que se sirvan de técnicas subliminales que trasciendan la consciencia de una persona con el objetivo o el efecto de alterar efectivamente su

lamento y a la que nos referimos con anterioridad. Tras la valoración de las solicitudes se emitirá una resolución motivada[38].

Pero, además de estas autoridades nacionales, se contempla también (artículo 56), la constitución de la Oficina Europea de Inteligencia Artificial, como un organismo independiente de la Unión que tendrá personalidad jurídica propia y estará situada en Bruselas.

Entre sus múltiples funciones cabe destacar las siguientes: La Oficina de IA desempeñará las siguientes funciones: apoyar, asesorar y cooperar con los Estados miembros, las autoridades nacionales de supervisión, la Comisión y otras instituciones, órganos y organismos de la Unión en lo que respecta a la aplicación del presente Reglamento; llevar un seguimiento de la aplicación efectiva

---

comportamiento de un modo que provoque o pueda provocar, con probabilidad razonable, perjuicios físicos o psicológicos a esa persona o a otra.

c) Sistemas de inteligencia artificial que aprovechen alguna de las vulnerabilidades de un grupo específico de personas debido a su edad o discapacidad o una situación social o económica específica con el objetivo o el efecto de alterar efectivamente el comportamiento de una persona de ese grupo de un modo que provoque o pueda provocar, con probabilidad razonable, perjuicios físicos o psicológicos a esa persona o a otra.

d) Sistemas de inteligencia artificial que tengan el fin de evaluar o clasificar personas físicas durante un período determinado de tiempo atendiendo a su conducta social o a características personales o de su personalidad conocidas o predichas, de forma que la clasificación social resultante provoque una o varias de las situaciones siguientes: 1.º Un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos en contextos sociales que no guarden relación con los contextos donde se generaron o recabaron los datos originalmente. 2.º Un trato perjudicial o desfavorable hacia determinadas personas físicas o colectivos que sea injustificado o desproporcionado con respecto a su comportamiento social o la gravedad de este.

e) Sistemas de identificación biométrica remota «en tiempo real» para su uso en espacios de acceso público con fines de aplicación de la ley, salvo y en la medida en que dicho uso sea estrictamente necesario para alcanzar uno o varios de los objetivos siguientes: 1.º La búsqueda selectiva de posibles víctimas concretas de un delito, incluido personas menores desaparecidas. 2.º La prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas físicas, para infraestructuras críticas, o un atentado terrorista. 3.º La detención, la localización, la identificación o el enjuiciamiento de la persona que ha cometido o se sospecha que ha cometido alguno de los delitos mencionados en el artículo 2, apartado 2, de la Decisión marco 584/2002/JAI del Consejo, para el que la normativa en vigor en el Estado miembro implicado imponga una pena o una medida de seguridad privativas de libertad cuya duración máxima sea al menos tres años, según determine el Derecho de dicho Estado miembro (artículo 5.6).

[38] El órgano competente para la instrucción del procedimiento es la Subdirección General de Inteligencia Artificial y Tecnologías Habilitadoras Digitales con el apoyo de la Oficina del Dato dependiente de la Secretaría de Estado de Digitalización e Inteligencia

y coherente del Reglamento y garantizarla; contribuir a la coordinación entre las autoridades nacionales de supervisión; actuar como mediador en los debates sobre desacuerdos graves que puedan surgir entre las autoridades competentes en relación con la aplicación del Reglamento; coordinar investigaciones conjuntas; contribuir a la cooperación efectiva con autoridades competentes de terceros países y con organizaciones internacionales; examinar, por propia iniciativa o a petición de su Consejo de Administración o de la Comisión, las cuestiones relativas a la aplicación del Reglamento y emitir dictámenes, recomendaciones o contribuciones escritas; asistir a las autoridades en el establecimiento y el desarrollo de espacios controlados de pruebas

---

Artificial. Les corresponderá la evaluación de las solicitudes presentadas para la participación en el entorno.

2. Se procederá a la evaluación de las solicitudes para la participación en el entorno, evaluándose para cada uno de los sistemas de inteligencia artificial recibidos, lo siguiente:

a) Grado de innovación o complejidad tecnológica del producto o servicio.

b) Grado de impacto social, empresarial o de interés público que presenta el sistema de inteligencia artificial propuesto.

c) Grado de explicabilidad y transparencia del algoritmo incluido en el sistema de inteligencia artificial presentado.

d) Alineamiento de la entidad y el sistema de inteligencia artificial con la Carta de Derechos Digitales del Gobierno de España.

e) Tipología de alto riesgo del sistema de inteligencia artificial, buscando una representación variada de tipologías en la selección.

f) Cuando se trate de sistemas de inteligencia artificial de propósito general, se evaluará también su potencial de ser transformados en un sistema de inteligencia artificial de alto riesgo.

g) Cuando se trate de modelos fundacionales de inteligencia artificial se evaluará la capacidad de despliegue y utilización, así como el impacto relativo o absoluto en la economía y sociedad.

h) El grado de madurez del sistema de inteligencia artificial, considerando que ha de estar lo suficientemente avanzado como para ser puesto en servicio o en el mercado en el marco temporal del entorno controlado de pruebas o a su finalización. Se buscará una representación variada de madurez de los sistemas de inteligencia artificial.

i) La calidad de la memoria técnica.

j) El tamaño o tipología del proveedor IA solicitante, según número de trabajadores o volumen de negocios anual, valorándose positivamente la condición de empresa emergente, pequeña o mediana empresa para garantizar una mayor diversidad de tipologías de empresas participantes. Se buscará una representación variada de tamaño y tipología de proveedor IA en la selección.

k) Y en su caso, la evaluación de la declaración responsable que acredite el cumplimiento de la norma relativa a la protección de datos personales. De igual forma se podrá solicitar documentación acreditativa adicional según recoge el anexo V del presente real decreto (artículo 8).

y facilitar la cooperación entre los espacios controlados de pruebas; promover la sensibilización del público y su comprensión de las ventajas, los riesgos, las salvaguardias y los derechos y obligaciones en relación con la utilización de sistemas de IA; realizar un seguimiento de los modelos fundacionales y organizar un diálogo periódico con los desarrolladores de modelos fundacionales en lo que respecta a su cumplimiento, así como a los sistemas de IA que utilizan dichos modelos de IA; promover la alfabetización en materia de inteligencia artificial.

XII. Las reglas en materia de sanciones

La propuesta de Reglamento incluye un sistema sancionador como forma de asegurar que se apliquen sus disposiciones y corresponderá a los Estados miembros determinar el régimen aplicable a las infracciones cometidas por cualquier operador. Las sanciones deberán ser efectivas, proporcionadas y disuasorias y tendrán particularmente en cuenta los intereses de las pymes y las empresas emergentes, así como su viabilidad económica.

La cuantía variará, como es lógico, de acuerdo con la gravedad de la infracción: a) el incumplimiento de la prohibición de las prácticas de inteligencia artificial estará sujeto a multas administrativas de hasta 40.000.000 de euros o, si el infractor es una empresa, de hasta el 7 % del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior; b) el incumplimiento de los requisitos relativos a los sistemas de alto riesgo estará sujeto a multas administrativas de hasta 20.000.000 de euros o, si el infractor es una empresa, de hasta el 4 % del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior; c) el incumplimiento por parte del sistema de IA o del modelo fundacional de cualquiera de los requisitos u obligaciones establecidos en el Reglamento distintos de los previstos en las letras a) y b) estará sujeto a multas administrativas de hasta 10.000.000 de euros o, si el infractor es una empresa, de hasta el 2 % del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior; d) la presentación de información inexacta, incompleta o engañosa a organismos notificados y a las autoridades nacionales competentes en respuesta a una solicitud estará sujeta a

multas administrativas de hasta 5.000.000 de euros o, si el infractor es una empresa, de hasta el 1 % del volumen de negocio total anual mundial del ejercicio financiero anterior, si esta cuantía fuese superior.

Podrán imponerse multas adicionales a las medidas no monetarias como órdenes o advertencias, en lugar de esas. Al decidir la cuantía de la multa administrativa en cada caso concreto se tomarán en consideración todas las circunstancias pertinentes de la situación correspondiente y se tendrá debidamente en cuenta lo siguiente: la naturaleza, la gravedad y la duración de la infracción y de sus consecuencias, teniendo en cuenta el propósito del sistema de IA, así como, cuando proceda, el número de particulares afectados y el nivel de los daños que hayan sufrido; si otras autoridades nacionales de supervisión de uno o varios Estados miembros han impuesto ya multas administrativas al mismo operador por la misma infracción; el tamaño y el volumen de negocio anual del operador que comete la infracción; las acciones emprendidas por el operador para mitigar los perjuicios o los daños sufridos por las personas afectadas; la intencionalidad o negligencia en la infracción; el grado de cooperación con las autoridades nacionales competentes con el fin de poner remedio a la infracción y mitigar sus posibles efectos adversos; el grado de responsabilidad del operador, teniendo en cuenta las medidas técnicas y organizativas que aplique; la forma en que las autoridades nacionales competentes tuvieron conocimiento de la infracción, en particular si el operador notificó la infracción y, en tal caso, en qué medida; la adhesión a códigos de conducta o a mecanismos de certificación aprobados; cualquier infracción previa pertinente del operador y cualquier otro factor agravante o atenuante aplicable a las circunstancias del caso.

Estas sanciones, así como los costes de litigio asociados y las reclamaciones de indemnización, no podrán ser objeto de cláusulas contractuales ni otras formas de acuerdo de reparto de cargas entre los proveedores y distribuidores, importadores, implementadores o cualquier otro tercero.

Se prevé también que el supervisor europeo de protección de datos podrá imponer multas administrativas a las instituciones, las agencias y los organismos de la Unión comprendidos en el ámbito de aplicación del Reglamento.

XIII. El acuerdo de 8 de diciembre de 2023 entre el Consejo y el Parlamento

Tras prolongados e intensos debates, en la noche del pasado 8 de diciembre se llegó a un acuerdo provisional en el seno de la Unión Europea sobre la propuesta de Ley de inteligencia artificial. A fecha 20 de diciembre no se conoce el texto definitivo que tendrá el Reglamento tras el acuerdo citado pero, como parecía evidente, en estas negociaciones finales no se ha ido más allá de las enmiendas introducidas por el Parlamento (nueva definición de la IA, regulación de los modelos fundacionales, prohibición de numerosas prácticas de IA…); más bien cabía pensar, por los diferentes intereses en presencia, que algunos de los cambios del Parlamento iban a ser, a su vez, modificados a la baja y se intuía que el uso de sistemas de reconocimiento facial en tiempo real –prohibidos por el Parlamento– y de instrumentos de policía predictiva iban a ser objeto de profunda discusión en orden a permitir, con cautelas, su uso.

Tampoco ha trascendido que en estas últimas negociaciones se haya cambiado la definición de lo que se entenderá por IA a efectos del Reglamento aunque nos atrevemos a intuir que no habrá modificaciones esenciales respecto al concepto aprobado por el Parlamento y que va en la línea del defendido por la OCDE.

Queda, igualmente, por ver qué ocurrirá con la prohibición parlamentaria de sistemas de IA para llevar a cabo evaluaciones de riesgo de personas físicas o grupos de personas con el objetivo de determinar el riesgo de que cometan delitos o infracciones o reincidan en su comisión, o para predecir la comisión o reiteración de un delito o infracción administrativa reales o potenciales. No me parece probable que los Estados acepten renunciar a todas estas herramientas de «inteligencia artificial policial» o de «policía predictiva».

Lo que sí se ha anticipado es que habrá diferentes períodos de *vacatio legis* del Reglamento, que podrán ir de unos pocos meses a, parece, que dos años.

Por lo que respecta a las instituciones de gobierno y control de la IA, se ha anunciado que se creará una Oficina de IA dentro de la Comisión encargada de supervisar los modelos de IA más avanzados, de contribuir a fomentar las normas y las prácticas de ensayo, y de hacer cumplir las normas comunes en todos los Estados miembros.

Habrá, además, un Consejo de Inteligencia Artificial, compuesto por representantes de los Estados miembros como plataforma de coordinación y órgano consultivo de la Comisión. Por último, se creará un foro consultivo para las partes interesadas, como los representantes de la industria, las pymes, las empresas emergentes, la sociedad civil y el mundo académico, con el fin de proporcionar conocimientos técnicos al Consejo de IA.

XIV. ¿Generará la regulación europea de la inteligencia artifical un «efecto Bruselas»?

En un conocido artículo publicado en 2012, que adoptó formato de libro en 2020[39], Anu Bradford explicó cómo y por qué las normas y Reglamentos «de Bruselas» han penetrado en muchos aspectos de la vida económica dentro y fuera de Europa a través del proceso de «globalización normativa unilateral», algo que se produce cuando un Estado o una organización supranacional es capaz de externalizar sus leyes y Reglamentos fuera de sus fronteras a través de mecanismos de mercado, dando lugar a la globalización de las normas. La globalización normativa unilateral es un fenómeno en el que una ley de una jurisdicción migra a otra sin que la primera la imponga activamente o la segunda la adopte voluntariamente (pp. 3 y 4)[40].

La potencia del mercado interior de la UE, unido a unas instituciones reguladoras con buena reputación, obliga a las empresas extranjeras que quieran participar en ese mercado a adaptar su conducta o su producción a las normas de la UE, que a menudo son las más estrictas; la alternativa es la renuncia a ese mercado, lo que no parece una opción razonable. Explica Bradford que las empresas multinacionales suelen tener un incentivo para estandarizar su producción a escala mundial y adherirse a una única norma. Esto convierte a la norma de la UE en una norma mundial: es el «efecto Bruselas de facto». Y, una vez que estas empresas orientadas a la exportación hayan ajustado sus prácticas empresariales para cumplir las estrictas normas

---

[39] *The Brussels Effect: How the European Union Rules the World*, Oxford University Press, 2020.

[40] "The Brussels Effect", *Northwestern University Law Review*, 1, 2012, disponible (a 20 de noviembre de 2023) en https://scholarship.law.columbia.edu/faculty_scholarship/271

de la UE, a menudo tienen el incentivo de presionar a sus gobiernos para que adopten esas mismas normas en un esfuerzo por igualar las condiciones frente a las empresas nacionales no exportadoras: el «efecto Bruselas de *iure*» (p. 7).

Y añade que la preferencia de los responsables políticos de la UE por una regulación estricta refleja su aversión al riesgo y su compromiso con una economía social de mercado. Además, y como ya hemos visto con anterioridad, la UE sigue el principio de precaución, que apuesta por la acción reguladora precautoria, incluso en ausencia de una certeza absoluta y cuantificable del riesgo, siempre que haya motivos razonables para temer que los efectos potencialmente peligrosos puedan ser incompatibles con el nivel de protección elegido (pp. 15 y 16).

Pues bien, cabría pensar que la regulación europea de la IA podría generar, en la línea de lo que ha ocurrido en ámbitos como la vida privada y la protección de datos[41], una exportación del contenido de esa nueva normativa a otros Estados, un «efecto Bruselas» sobre la regulación de la IA (De la Sierra, 2023, pp. 15 y ss). Sin embargo, la propia Anu Bradford se ha mostrado escéptica al respecto en su último trabajo –*Digital Empires: The Global Battle to Global Battle to Regulate Technology*–, de 2023, recordando que Estados Unidos sigue siendo un modelo basado en el mercado abierto, China un modelo de centralismo estatal y la Unión Europea sigue apostando por la regulación.

Ahora bien, Estados Unidos también ha optado por aprobar normas que regulen la IA, aunque no sea con la misma intensidad que en la Unión Europea; así, el 30 de octubre de 2023 el presidente Biden emitió *la Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*[42], donde se proclama que el Gobierno Federal tratará de promover principios y acciones responsables de seguridad

---

[41] El Reglamento General de Protección de Datos (RGPD) de la UE, del 14 de abril de 2016, ha tenido un efecto global: así, en 2017 Japón creó una agencia independiente para gestionar las quejas sobre vida privada con el fin de ajustarse al nuevo reglamento de la UE y gigantes tecnológicos como Facebook y Microsoft anunciaron en 2018 que se acogerían RGPD.

[42] Disponible, a 20 de noviembre de 2023, en 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/

y protección de la IA con otras naciones, «incluidos nuestros competidores», al tiempo que lidera conversaciones y colaboraciones globales clave para garantizar que la IA beneficie a todo el mundo, en lugar de exacerbar las desigualdades, amenazar los derechos humanos y causar otros daños. Además, y en la línea de la UE, en esa orden se define la IA como un sistema basado en máquinas que puede, para un conjunto dado de objetivos definidos por el ser humano, hacer predicciones, recomendaciones o tomar decisiones que influyan en entornos reales o virtuales. Y se enuncian los ocho principios que deben guiar el desarrollo de la IA: la seguridad de los sistemas, la innovación responsable, el compromiso con los trabajadores, avance en igualdad y derechos, protección de los consumidores, protección de la intimidad, gestión de los riesgos y uso responsable de la IA, búsqueda del liderazgo social, económico y tecnológico.

China, por su parte, y aun apostando por la IA como herramienta de férreo control de la disidencia y por sistemas como el «crédito social»[43], que estarán prohibidos en Europa, aprobó en agosto

[43] El sistema de crédito social tiene dos características principales: la primera es la recopilación de datos a escala nacional procedentes de un amplio abanico de organismos reguladores, Gobiernos centrales y locales, el Poder Judicial y plataformas privadas. Cuando esté plenamente operativo, el sistema recopilará dos tipos básicos de información: la crediticia pública, generada por las interacciones de una empresa con órganos gubernamentales y agencias reguladoras (multas, sentencias, licencias comerciales…), y la información crediticia de mercado, generada por las interacciones de una empresa con otros agentes del mercado (reclamaciones de consumidores, datos generados por agencias de calificación de créditos…). Los datos se utilizarán en sistemas de puntuación gestionados por las administraciones locales, la mayoría de los cuales están en fase de construcción.

El segundo elemento principal es un régimen de recompensas y castigos (en forma de «listas rojas» y «listas negras») mantenido por organismos gubernamentales. Algunas listas tienen un amplio alcance, como el incumplimiento de sentencias judiciales, mientras que otras se aplican a sectores específicos de la economía, como la alimentación o la medicina.

La inclusión en una lista roja o negra es pública; en el primer caso puede implicar diversos beneficios, que van desde la ampliación del acceso a los préstamos hasta una reducción de la frecuencia de las inspecciones o el aumento de las oportunidades en los procesos de contratación pública y acceso a la financiación, sobre todo para las pequeñas y medianas entidades. La inclusión en una lista negra origina barreras de mercado, como restricciones para obtener autorizaciones gubernamentales, mayor frecuencia de inspecciones y prohibiciones para obtener financiación. Cuando una entidad es incluida en una lista negra, su representante legal y las personas directamente responsables de la infracción también se incluirán en la lista: Kendra Schaefer, 2020, disponible, a 20 de noviembre de 2023, en https://www.uscc.gov/sites/default/files/2020-12/Chinas_Corporate_Social_Credit_System.pdf.; también Yu-Hsin Lin y Milhaupt, 2023.

de 2023 una ley general reguladora de la Inteligencia Artificial y, en paralelo, otra regulación específica de la IA generativa. En la primera de ellas se vincula la IA a los sistemas automatizados que funcionan con cierto grado de autonomía, sirven a determinados objetivos y son capaces de influir en el entorno físico o virtual mediante la predicción, la recomendación o la toma de decisiones, es decir, en manera similar a lo que ocurre en Europa y Estados Unidos. También incluye una serie de principios aplicables al desarrollo de la IA: seguridad y robustez; apertura, transparencia y explicabilidad; responsabilidad proactiva y equidad e igualdad. Igualmente se fomentará el uso de energías eficientes, para la protección del medio ambiente, en el desarrollo de estas tecnologías[44].

Y, por poner otro ejemplo, Brasil también ha iniciado el procedimiento para regular la IA: el 1 de marzo de 2023 se presentó el breve Proyecto de Ley 759/2023 en la Cámara de Diputados[45] y el 3 de mayo el más exhaustivo Proyecto de Ley 2338/2023[46]; este último tiene como objetivos establecer normas nacionales generales para el desarrollo, la implementación y el uso responsable de sistemas de inteligencia artificial en Brasil para proteger los derechos fundamentales y garantizar la implementación de sistemas seguros y fiables en beneficio de la persona, el régimen democrático y el desarrollo científico y tecnológico. Se trata de una propuesta basada en los riesgos de la IA, prohibiendo los que implican «riesgos excesivos», delimitando los de «alto riesgo» y con un enfoque basado en los derechos. Incluye, además, una definición de la IA similar a las que ya hemos visto en otros ámbitos jurídicos: es un sistema informático, con diversos diferentes grados de autonomía, diseñado para inferir cómo lograr un conjunto dado de objetivos, utilizando enfoques basados en el aprendizaje automático y/o y la representación del conocimiento, utilizando datos de entrada procedentes de máquinas o de seres humanos, con el fin de producir datos de entrada procedentes

---

[44] Más información en https://diariolaley.laleynext.es/dll/2023/09/01/china-aprueba-una-regulacion-de-la-inteligencia-artificial-y-de-la-inteligencia-artificial-generativa (a 20 de noviembre de 2023).

[45] https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2349685 (a 20 de noviembre de 2023).

[46] https://www25.senado.leg.br/web/atividade/materias/-/materia/157233 (a 20 de noviembre de 2023).

de máquinas o seres humanos, con el fin de producir predicciones, recomendaciones o decisiones que puedan influir en el entorno.

No parece, por tanto, casual que en su informe sobre este proyecto elaborado por la Autoridad Nacional de Protección de Datos, hecho público el 6 de julio de 2023, se hagan varias referencias a la propuesta que se está tramitando en la Unión Europea y se diga de manera expresa que el proyecto presentado en el Senado es «semejante» a esta última[47].

Y, por lo que respecta a España y a la influencia hacia dentro del «efecto Bruselas», hemos visto que, incluso con bastante anterioridad a la aprobación y, en su caso, entrada en vigor del Reglamento, se ha creado una autoridad nacional –la Agencia Española de Supervisión de Inteligencia Artificial–, se ha asumido una definición de IA en la línea de la propuesta europea y se ha regulado el 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».

En definitiva, y aunque en el caso de la regulación de la IA el impacto del «efecto Bruselas» pueda ser menor que en otros ámbitos (Arnal y Jorge, 2023)[48], no parece en absoluto, por lo que está ocurriendo en otros Estados y espacios jurídicos, que esta propuesta vaya a tener repercusiones únicamente hacia dentro de la Unión.

Bibliografía

Álvarez García, V. (2020). *Las normas técnicas armonizadas (Una peculiar fuente del Derecho europeo)*, Iustel.

Álvarez García, V. y Tahiri Moreno, J. (2023). La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque, *Revista General de Derecho Administrativo*, (2023).

Añón Roig, M. J. (2022). Desigualdades algorítmicas: conductas de alto riesgo para los derechos humanos. *Derechos y libertades*, 47, pp. 17-49.

[47] https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial (a 20 de noviembre de 2023).

[48] https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/ (a 20 de noviembre de 2023).

Arnal, J. y Jorge Ricart, R. (2023). *Inteligencia artificial (i): el menor «efecto Bruselas», las posibles consecuencias desglobalizadoras de un enfoque regulatorio divergente y la importancia de políticas públicas para el empleo*, Real Instituto Elcano.

Barrio Andrés, M. (2021). *Introducción al Derecho de las Nuevas Tecnologías*, Wolters Kluwers.

Barrio Andrés, M. (2022). Inteligencia artificial: origen, concepto, mito y realidad, *El Cronista del Estado social y democrático de Derecho (monográfico sobre inteligencia artificial y Derecho*), 100, pp. 14-21.

Barrio Andrés, M. (2023a). Novedades en la tramitación del próximo Reglamento europeo de inteligencia artificial, *ARI Real Instituto Elcano*, 67.

Barrio Andrés, M. (2023b). Sobre la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), *Diario LA LEY*, nº 10349, Sección Tribuna.

Bradford, A. (2012). The Brussels Effect, *Northwestern University Law Review*, 1.

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

Bradford, A. (2013). *Digital Empires. The Global Battle to Regulate Technology*, Oxford University Press.

Campione, R. (2020). *La plausibilidad del Derecho en la era de la inteligencia artificial. Filosofía carbónica y filosofía jurídica del Derecho*, Dykinson.

Cotino, L. (2021). Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act), *Diario La Ley*, 2 de julio.

De Hoyos Sancho, M. (2022). El proyecto de Reglamento de la Unión Europea sobre inteligencia artificial, los sistemas de alto riesgo y la creación de un ecosistema de confianza. En S. BARONA VILAR, (ed.) *Justicia poliédrica en periodo de mudanza: Nuevos conceptos, nuevos sujetos, nuevos instrumentos y nueva intensidad*, Tirant lo Blanch, pp. 403-422.

De la Sierra, S. (2023). European integration, digitalisation and the Brussels effect, preprint, nº 3/2023.

Esteve Pardo, J. (1999). *Técnica, riesgo y Derecho (Tratamiento del riesgo tecnológico en el Derecho ambiental)*, Ariel.

Esteve Pardo, J. (2009). *El desconcierto del Leviatán (Política y Derecho ante las incertidumbres de la Ciencia)*, Marcial Pons.

Floridi, L. (2012). *La rivoluzione dell'informazione*, Codice edizioni.

García Sánchez, M. D. (2022). Propuesta de reglamento europeo sobre inteligencia artificial: especial referencia a la identificación biométrica remota y los sistemas de puntuación social. En González Pulido y Bueno de Mata *Fodertics 10.0: estudios sobre derecho digital*, Comares, pp. 779-793

Gascón Marcet, A. (2022). La propuesta de Reglamento de la Comisión Europea por el que se establecen normas armonizadas en materia de inteligencia artificial. *Retos de la sociedad digital. Regulación y fiscalidad en un contexto internacional*, Editorial Reus, pp. 15-40.

González Álvarez, J. L; López Ossorio, J. J.; Muñoz Rivas, M. (2018). *La valoración policial del riesgo de violencia contra la mujer pareja en España – Sistema VioGén*, Ministerio del Interior.

González Cabanes, F. y Díaz Díaz, N. (2023). ¿Qué es la Inteligencia Artificial? En E. Gamero Casado, y F.L. Pérez Guerrero, *Inteligencia artificial y sector público: retos, límites y medios*, Tirant lo Blanch, pp. 37-72.

Grande Sanz, M. (2022): La propuesta de reglamento sobre inteligencia artificial: presente y futuro, En González Pulido y Bueno de Mata *Fodertics 10.0: estudios sobre derecho digital*, Comares, pp. 795-807

Jove Villares, D. (2023). La protección de lo sensible, o cuando la naturaleza del dato no lo es todo, Tirant lo Blanch.

Martínez Garay, L. *et al*. (2023). *Three predictive policing approaches in Spain: VioGén, Riscanvi and Veripol*. Universitat de València (pendiente de publicación).

Miró Llinares, F. (2019). El modelo policial que viene: Mitos y realidades del impacto de la inteligencia artificial y la ciencia de datos en la prevención policial del crimen. En: J. Martínez Espasa, (coord.). *Libro blanco de la prevención y seguridad local valenciana: Conclusiones y propuestas del Congreso Valenciano de Seguridad Local: la prevención del siglo XXI*, pp. 98-113.

Presno Linera, M. A. (2022). *Derechos fundamentales e inteligencia artificial*, Marcial Pons.

Presno Linera, M. A. (2023). Policía predictiva y prevención de la violencia de género: el sistema *VioGén*, *Revista de Internet, Derecho y Política, monográfico sobre Digitalización y algoritmización de la justicia*, nº 39.

Ruíz Tarrías, S. (2023). La búsqueda de un modelo regulatorio de la IA en la Unión Europea, *Anales de la Cátedra Francisco Suárez* (ejemplar dedicado a: Inteligencia Artificial y Derecho), 57, pp. 91-119.

Russell, S. y Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno*, Pearson Education.

San Martín Segura, D. (2023). *La intrusión jurídica del riesgo*, CEPC.

Schaefer, K. (2020). China's social credit system: context, competition, technology and geopolitics, *Trivium China*, https://www.uscc.gov/sites/default/files/2020-12/Chinas_Corporate_Social_Credit_System.pdf

Simón Castellanos, P. (2023). *La evaluación de impacto algorítmico en los derechos fundamentales*, Aranzadi.

Soriano Arnanz, A. (2021). La propuesta de Reglamento de Inteligencia Artificial de la Unión Europea y los sistemas de alto riesgo, *Revista General de Derecho de los Sectores Regulados*, 8.

Yu-Hsin Lin, L. y Milhaupt, C. (2023). China's Corporate Social Credit System: The Dawn of Surveillance State Capitalism? *The China Quarterly* , pp. 1–19.