La regulación de la inteligencia artificial en Europa

Presno Linera, M.Á.; Meuwese, A.C.M.

LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EUROPA[1]

MIGUEL ÁNGEL PRESNO LINERA
Catedrático de Derecho constitucional
Universidad de Oviedo

ANNE MEUWESE
Catedrática de Derecho público y gobernanza de la inteligencia artificial
Universidad de Leiden

TRC, n.º 54, 2024, pp. 131-161
ISSN 1139-5583

Sumario

I. PRESENTACIÓN

En el mes de marzo de 2024, y con muy pocos días de diferencia, culminaron los trabajos de elaboración tanto del texto del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia

1 Este trabajo es uno de los resultados del Proyecto de investigación 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. Agradecemos las sugerencias, críticas y comentarios de quienes evaluaron anónimamente la primera versión del texto.

de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión Europea (en adelante, RIA), como del Convenio Marco del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho (en lo sucesivo, CIA). Ambas normas, sobre las que nos ocuparemos en las líneas siguientes, son el resultado presente de un proceso de debate legislativo, institucional, social, tecnológico y económico que se remonta a varios años atrás y que parte del convencimiento, como se dice en el primer párrafo del Libro Blanco sobre la inteligencia artificial de la Comisión Europea, de 19 de febrero de 2020, de que la inteligencia artificial (en lo sucesivo, IA) «se está desarrollando rápido. Cambiará nuestras vidas, pues mejorará la atención sanitaria (por ejemplo, incrementando la precisión de los diagnósticos y permitiendo una mejor prevención de las enfermedades), aumentará la eficiencia de la agricultura, contribuirá a la mitigación del cambio climático y a la correspondiente adaptación, mejorará la eficiencia de los sistemas de producción a través de un mantenimiento predictivo, aumentará la seguridad de los europeos y nos aportará otros muchos cambios que de momento solo podemos intuir. Al mismo tiempo, la IA conlleva una serie de riesgos potenciales, como la opacidad en la toma de decisiones, la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas o su uso con fines delictivos»[2].

Cabe recordar que el impacto de la IA ya se había detectado con anterioridad, así como la consciencia de que estamos inmersos en una *infoesfera*[3], en un ambiente global compuesto por organismos informacionales interconectados y este mestizaje ontológico entre lo biológico y lo técnico, entre lo carbónico y lo silícico[4], exige conocer los posibles modelos de interacción entre las personas y los sistemas inteligentes[5] y 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

2 https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1, (fecha de consulta: 14/06/2024).

3 Floridi (2012: 11).

4 Campione (2020: 13).

5 Llano Alonso (2024: 138-144).

estas poderosas tecnologías sean utilizadas como herramientas para socavar sistemas democráticos y como mecanismos de dominación?[6]

II. LAS INICIATIVAS PARA LA REGULACIÓN SUPRANACIONAL DE LA INTELIGENCIA ARTIFICIAL[7]

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»[8].

Se comienza a evidenciar así la preocupación de las instituciones de la UE a propósito, por lo que aquí interesa, de la regulación jurídica de la IA, de manera que se pueda aprovechar todo lo que supone en materia de innovación y desarrollo tecnológico y, al mismo tiempo, queden garantizados de manera adecuada los derechos fundamentales y el propio Estado social y democrático de Derecho.

Transcurrido poco más de un año, el 7 de diciembre de 2018, la Comisión Europea presentó la Comunicación titulada «Plan Coordinado sobre la Inteligencia Artificial», junto con el Plan Coordinado sobre el Desarrollo y Uso de la Inteligencia Artificial «Made in Europe»-2018, preparado por los Estados miembros (como parte del Grupo sobre la Digitalización de la Industria Europea y la Inteligencia Artificial), Noruega, Suiza y la Comisión[9]. 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

6 https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/ethics-artificial-intelligence-statement-ege-released-2018-03-09_en (fecha de consulta: 14/07/2024); sobre las «nuevas leyes de la robótica», Pasquale (2024), sobre el impacto de la IA en los derechos fundamentales, Presno Linera (2022).

7 Una panorámica más amplia en Pérez-Ugena (2024: 124-156).

8 https://www.consilium.europa.eu/media/21604/19-euco-final-conclusions-es.pdf, p. 7 (fecha de consulta: 14/07/2024).

9 https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52018DC0795 (fecha de consulta: 14/07/2024).

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[10].

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*[11], donde se afirma que «la Comisión respalda un enfoque basado en la regulación y en la inversión, que tiene el doble objetivo de promover la adopción de la inteligencia artificial y de abordar los riesgos vinculados a determinados usos de esta nueva tecnología. La finalidad del presente Libro Blanco es formular alternativas políticas para alcanzar estos objetivos…»

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[12], la responsabilidad civil[13] y los derechos de propiedad intelectual[14], a las que siguieron, ya en 2021, resoluciones sobre el uso de la IA y en los sectores educativo, cultural y audiovisual[15] y en materia penal[16].

Antes de esta última Resolución ya se había publicado el documento con el que formalmente se abrió uno de los procedimientos normativos que nos ocupan: el 21 de abril de 2021 se conoció la Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecer normas armonizadas en materia de

10 https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/es/pdf, p. 8 (fecha de consulta: 14/07/2024).

11 https://op.europa.eu/es/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1, (fecha de consulta: 14/07/2024).

12 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) (fecha de consulta: 14/07/2024).

13 Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un régimen de responsabilidad civil en materia de inteligencia artificial, 2020/2014(INL) (fecha de consulta: 14/07/2024).

14 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) (fecha de consulta: 14/07/2024).

15 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 (fecha de consulta: 14/07/2024).

16 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 (fecha de consulta: 14/07/2024).

inteligencia artificial (Ley de inteligencia artificial) y se modifican determinados actos legislativos de la Unión, elaborada por la Comisión[17].

Cabe recordar ahora, por lo que luego veremos, que la IA se definió entonces (artículo 3.1) como «el software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa». También que en esa propuesta no se incluyó referencia alguna a los luego llamados «modelos fundacionales», es decir, los sistemas de IA entrenados con una cantidad ingente de datos y que son capaces de realizar una gran variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural[18], los que el Reglamento europeo ha acabado denominando sistemas de inteligencia artificial de uso general y a los que nos referiremos más adelante.

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[19], donde señaló que «para garantizar que la definición de los sistemas de IA proporcione criterios suficientemente claros para distinguirlos de otros sistemas de software más clásicos, el texto transaccional restringe la definición del artículo 3, apartado 1, a los sistemas desarrollados a través de estrategias de aprendizaje automático y estrategias basadas en la lógica y el conocimiento», es decir, se introduce el criterio del aprendizaje automático como una de las características de los sistemas de IA, algo que no estaba previsto así en la propuesta de la Comisión; también se amplían los sistemas que se pretende prohibir y se introducen cambios en los considerados de «alto riesgo».

A continuación, cabe mencionar las enmiendas introducidas por el Parlamento Europeo en el texto de la Comisión y aprobadas el 14 de junio de 2023[20], que incorporaron una nueva definición de sistema de IA —«un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida —como predicciones, recomendaciones o decisiones— que influya en entornos reales o virtuales»; también de lo que se entendió por un modelo fundacional —«un modelo de sistema de IA entrenado con un gran volumen de datos, diseñado para producir información de salida de carácter general y capaz de adaptarse a una

17 https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A52021PC0206 (fecha de consulta: 14/07/2024).

18 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.

19 https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/es/pdf (fecha de consulta: 14/07/2024).

20 https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.html (fecha de consulta: 14/07/2024).

amplia variedad de tareas diferentes—», al tiempo que, entre otras cosas. se ampliaron los sistemas de IA prohibidos.

En el mes de diciembre de 2023 tuvieron lugar los *trílogos* entre las instituciones europeas implicadas[21] (Parlamento, Comisión, Consejo) para limar las diferencias en cuestiones tan relevantes como la prohibición, o no, del uso de los sistemas de identificación biométrica remota en tiempo real en espacios de acceso público[22] o el alcance de la regulación de los entonces denominados «modelos fundacionales», que pasaron a llamarse «modelos de IA de uso general».

Finalmente, se aprobó la Resolución legislativa del Parlamento Europeo, de 13 de marzo de 2024, sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión y se publicó en el Diario Oficial de la UE, el 12 de julio de 2024, el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

2. **En el Consejo de Europa**

En un contexto más amplio como es el de los 46 Estados que componen el Consejo de Europa, incluidos los 27 de la Unión Europea, cabe destacar, en primer lugar, el trabajo de investigación sobre algoritmos y derechos humanos, según el cual la IA afectará a un gran número, sino a la práctica totalidad, de nuestros derechos fundamentales[23]; 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

21 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/ (fecha de consulta: 14/07/2024)

22 Sobre estos sistemas, Crawford (2023: 150 y ss.); sobre su regulación, Cotino Hueso (2023: 347-402) y Ridaura Martínez (2024: en prensa).

23 Consejo de Europa, *Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, 2018, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (fecha de consulta: 14/07/2024).

sindical, derecho de huelga); en cuarto lugar, y a su vez vinculado a muchos otros derechos, al de no sufrir discriminación por raza, género, edad, orientación sexual…; en quinto lugar, a los derechos dependientes del acceso a los servicios públicos (educación, sanidad…) y, en general, a los derechos sociales (prestaciones por desempleo, enfermedad, jubilación…); finalmente, y por no extendernos mucho más, al derecho a intervenir en procesos participativos de índole política (elecciones, referendos, iniciativas legislativas populares…) y en, general, a las libertades en el ámbito ideológico (de pensamiento, conciencia y religión).

En segundo lugar, la Asamblea Parlamentaria del Consejo de Europa aprobó, el 22 de octubre de 2020, un conjunto de principios éticos básicos que deberían respetarse al elaborar y establecer aplicaciones de IA, incluida la transparencia, la justicia y la equidad, la responsabilidad humana de la toma de decisiones, la seguridad, la privacidad y la protección de datos. Ha identificado la necesidad de crear un marco normativo transversal para la IA, con principios específicos basados en la protección de los derechos humanos, la democracia y el Estado de derecho e instó al Comité de Ministros a elaborar un instrumento jurídicamente vinculante que regule la IA.

En tercer lugar, el Comité de Ministros adoptó, el 20 de mayo de 2022, un enfoque transversal de la inteligencia artificial en los diversos sectores del Consejo de Europa, estableciendo el Comité sobre Inteligencia Artificial (CAI) y encomendándole la elaboración de un Convenio [marco] jurídicamente vinculante sobre el desarrollo, diseño y aplicación de sistemas de IA, basado en las normas del Consejo de Europa en materia de derechos humanos, democracia y estado de derecho, sobre la base de estos principios fundamentales.

El Comité de Ministros también decidió permitir la inclusión en las negociaciones de la Unión Europea y de los Estados no europeos interesados que compartan los valores y objetivos del Consejo de Europa; a saber, Argentina, Australia, Canadá, Costa Rica, la Santa Sede, Israel, Japón, México, Perú, los Estados Unidos de América y Uruguay. El Consejo de Europa también hizo partícipes a actores no estatales en las negociaciones: un total de 68 representantes de la sociedad civil y de la industria participaron en calidad de observadores, interviniendo junto con Estados y representantes de otras organizaciones internacionales, como la Organización para la Seguridad y la Cooperación en Europa (OSCE), la Organización para la Cooperación y el Desarrollo Económicos (OCDE), la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) y los órganos y comités pertinentes del Consejo de Europa. La Unión Europea también participó en las negociaciones representada por la Comisión Europea, incluyendo en su delegación también a representantes de la Agencia de los Derechos Fundamentales de la Unión Europea (FRA) y del Supervisor Europeo de Protección de Datos (SEPD). Aunque se formularon acusaciones de que se excluía a las ONG de las negociaciones, supuestamente porque los Estados Unidos se negaban a compartir cierta información con participantes no estatales,

desde el Consejo de Europa se justificó el proceso argumentando que la manera de desarrollar el borrador (discutirlo primero a puerta cerrada entre los participantes estatales y solo publicarlo y discutirlo en el plenario después) era conforme con las normas y prácticas vigentes[24].

El resultado final ha sido el citado Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho[25], que se abrió a la firma durante la conferencia de Ministros de Justicia del Consejo de Europa celebrada en Vilna el 5 de septiembre de 2024 y ha sido firmado por Andorra, Georgia, Islandia, Noruega, Moldavia, San Marino, el Reino Unido, Israel, Estados Unidos y la Unión Europea.

3. **Otras iniciativas y acuerdos internacionales sobre la regulación de la inteligencia artificial**

En un contexto de mayor globalización, los 36 países miembros de la 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[26] y asumieron, en la línea de lo acordado en Europa, que un sistema de IA es el que, para objetivos explícitos o implícitos, deduce, a partir de las entradas recibidas, cómo generar resultados de salida como previsiones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los distintos sistemas de IA tienen diversos grados de autonomía y adaptabilidad una vez desplegados.

Y los mencionados 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

24 Hendrickx, Victoria & Wannes Ooms, 'An interview with KCDS-CiTiP Fellow Jan Kleijssen on the AI Convention of the Council of Europe, 23 May 2023, https://lirias.kuleuven.be/retrieve/714817.

25 Pueden verse los estudios de Cotino Hueso y Gascón Marcén sobre el Convenio Marco.

26 https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0449#mainText (fecha de consulta: 14/07/2024).

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, el 1 y 2 de noviembre de 2023, 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 IA, la llamada *The Bletchley Declaration*[27], 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. La declaración ha tenido una acogida desigual, principalmente por el carácter voluntario de las medidas[28].

27 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 (fecha de consulta: 14/07/2024).

28 Leslie et al. (2024).

III. ¿DE QUÉ SE HABLA CUANDO SE HABLA DE INTELIGENCIA ARTIFICIAL EN LA NORMATIVA EUROPEA?

En la corta historia de la IA[29] 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. La IA pretende sintetizar o reproducir los procesos cognitivos humanos, tales como la percepción, la creatividad, la comprensión, el lenguaje o el aprendizaje[30]. Para ello, la IA utiliza todas las herramientas a su alcance, entre las que destacan las proporcionadas por la computación, incluidos los algoritmos. No obstante, los sistemas de IA no usan cualquier algoritmo sino, esencialmente, los que «aprenden» a base del procesamiento de datos.

Por otro lado, en ocasiones se habla de IA cuando en realidad estamos hablando de un subcampo, el aprendizaje automático (o *machine learning* en inglés, AA en lo sucesivo). El AA trata de encontrar patrones en datos para construir sistemas predictivos o explicativos; por tanto, puede considerarse una rama de la IA ya que a partir de la experiencia (los datos) toma decisiones o detecta patrones significativos y eso es una característica fundamental de la inteligencia humana. Es importante resaltar que para que un sistema de AA tenga éxito es tan necesario utilizar los algoritmos adecuados como realizar una correcta gestión y tratamiento de los datos utilizados para desarrollar el sistema.

Profundizando un poco más, nos encontramos con las redes neuronales, también llamadas redes neuronales artificiales, que son un modelo computacional de aprendizaje automático que procesa la información a través de un conjunto de unidades llamadas neuronas, o neuronas artificiales, que están conectadas entre sí y organizadas por capas, formando una red. Los datos de entrada atraviesan la red neuronal, donde son procesados mediante operaciones matemáticas, generando una salida. Por su parte, el concepto de aprendizaje profundo (*Deep learning*) hace referencia a las redes neuronales de un gran número de capas. No existe un criterio claro en cuanto a partir de qué número de capas ocultas podemos considerar una red neuronal como profunda y, por tanto, aprendizaje profundo, pero hay una opinión cada vez más extendida entre los expertos que afirma que cualquier red con más de 2 capas ocultas puede considerarse «profunda»[31].

29 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; pueden verse al respecto el libro de Mitchell (2024: 27 y ss.) y el discurso de ingreso de la profesora Asunción Gómez-Pérez en la Real Academia Española con el título Inteligencia artificial y lengua española, https://www.rae.es/sites/default/files/2023-05/Discurso%20de%20ingreso%20de%20Asuncion%20Gomez-Perez.pdf (fecha de consulta: 14/07/2024).

30 Russel y Norvig (2008: 1 y ss.).

31 González Cabanes y Díaz Díaz (2023: 58 y 64).

La dificultad de ofrecer una definición «acabada» de la IA se presenta también en el ámbito jurídico[32], como se puede comprobar leyendo las diferentes versiones que se han ido ofreciendo durante el proceso de aprobación del RIA: así, en el texto que presentó la Comisión el 21 de abril de 2021 se entendía como «el software que se desarrolla empleando una o varias de técnicas y estrategias que figuran en el Anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa» (artículo 3).

Tras las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023, se definió como «un sistema basado en máquinas diseñado para funcionar con diversos niveles de autonomía y capaz, para objetivos explícitos o implícitos, de generar información de salida —como predicciones, recomendaciones o decisiones— que influya en entornos reales o virtuales».

En el texto definitivo se ha conceptualizado como un sistema basado en máquinas diseñado para funcionar con distintos niveles de autonomía, que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de las entradas que recibe, salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales (artículo 3).

En el Preámbulo del RIA se explica que una característica clave de los sistemas de IA es su capacidad para inferir. Esta inferencia se refiere al proceso de obtención de resultados, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos y virtuales, y a una capacidad de los sistemas de IA para derivar modelos y/o algoritmos a partir de entradas/datos. Las técnicas que permiten la inferencia al construir un sistema de IA incluyen enfoques de aprendizaje automático que aprenden, a partir de los datos, cómo alcanzar determinados objetivos; y enfoques basados en la lógica y el conocimiento que infieren a partir del conocimiento codificado o la representación simbólica de la tarea que debe resolverse. La capacidad de un sistema de IA para inferir va más allá del procesamiento básico de datos, permitiendo el aprendizaje, el razonamiento o el modelado. El término «basado en máquinas» se refiere al hecho de que los sistemas de IA funcionan en máquinas. A efectos del presente Reglamento, los entornos deben entenderse como los contextos en los que operan los sistemas de IA, mientras que los resultados generados por el sistema de IA reflejan diferentes funciones realizadas por los sistemas de IA e incluyen predicciones, contenidos, recomendaciones o decisiones. Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que tienen cierto grado de independencia de las acciones de la intervención humana y capacidades para funcionar sin intervención humana. La capacidad de adaptación que

32 Barrio Andrés (2022: 14-21), Presno Linera (2023: 95-99) y Bustos Gisbert (2024: 153-158).

puede mostrar un sistema de IA tras su despliegue se refiere a las capacidades de autoaprendizaje, que permiten al sistema cambiar mientras se utiliza.

Esta definición coincide con la que asume el CIA, en cuyo artículo 2 se dice que, a los efectos del presente Convenio, se entenderá por «sistema de inteligencia artificial» un sistema basado en máquinas que, con objetivos explícitos o implícitos, infiere, a partir de los datos que recibe, cómo generar resultados, como predicciones, contenidos, recomendaciones o decisiones que puedan influir en entornos físicos o virtuales. Los diferentes sistemas de inteligencia artificial varían en sus niveles de autonomía y adaptabilidad después de la implementación».

Como se expone en el documento explicativo sobre el Convenio, esta definición de sistema de inteligencia artificial se extrae de la última definición revisada adoptada por la OCDE el 8 de noviembre de 2023. La elección de los redactores de utilizar este texto en particular es significativa no sólo por la alta calidad del trabajo realizado por la OCDE y sus expertos, sino también por la necesidad de mejorar la cooperación internacional en el tema de la inteligencia artificial y facilitar los esfuerzos destinados a armonizar la gobernanza de la inteligencia artificial a nivel mundial, en particular mediante la armonización de la terminología pertinente. La definición refleja una amplia comprensión de lo que son los sistemas de inteligencia artificial, específicamente en comparación con otros tipos de sistemas de software tradicionales más simples basados en las reglas definidas únicamente por personas físicas para ejecutar operaciones automáticamente. Su objetivo es garantizar la precisión y la seguridad jurídicas, sin dejar de ser lo suficientemente abstracto y flexible como para seguir siendo válido a pesar de los futuros avances tecnológicos. La definición se redactó a los efectos de la Convención Marco y no tiene por objeto dar un significado universal al término pertinente. Los redactores tomaron nota de la exposición de motivos que acompaña a la definición actualizada de sistema de inteligencia artificial que figura en la Recomendación de la OCDE sobre Inteligencia Artificial (OCDE/LEGAL/0449, 2019, modificada en 2023) para una explicación más detallada de los distintos elementos de la definición. Si bien esta definición proporciona un entendimiento común entre las Partes en cuanto a lo que son los sistemas de inteligencia artificial, las Partes pueden especificarlo en sus ordenamientos jurídicos nacionales para mayor seguridad jurídica y precisión, sin limitar su alcance.

IV. LOS PRINCIPIOS QUE INSPIRAN LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL EN EUROPA

El Grupo independiente de expertos de alto nivel sobre IA creado por la Comisión Europea hizo públicos en abril de 2019 una serie de principios con el objeto de contribuir a garantizar la fiabilidad y el fundamento ético de la IA: acción y supervisión humanas; solidez técnica y seguridad; gestión de la

privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas[33].

Y, como se recuerda en el considerando 27 del RIA, por acción y supervisión humanas se entiende que los sistemas de IA se desarrollan y utilizan como una herramienta al servicio de las personas, que respeta la dignidad humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos; por solidez técnica y seguridad se entiende que los sistemas de IA se desarrollan y utilizan de manera que sean sólidos en caso de problemas y resilientes frente a los intentos de alterar el uso o el funcionamiento del sistema de IA para permitir su uso ilícito por terceros y reducir al mínimo los daños no deseados; por gestión de la privacidad y de los datos se entiende que los sistemas de IA se desarrollan y utilizan de conformidad con normas en materia de protección de la intimidad y de los datos, al tiempo que tratan datos que cumplen normas estrictas en términos de calidad e integridad; por transparencia se entiende que los sistemas de IA se desarrollan y utilizan de un modo que permita una trazabilidad y explicabilidad adecuadas, y que, al mismo tiempo, haga que las personas sean conscientes de que se comunican o interactúan con un sistema de IA e informe debidamente a los responsables del despliegue acerca de las capacidades y limitaciones de dicho sistema de IA y a las personas afectadas acerca de sus derechos; por diversidad, no discriminación y equidad se entiende que los sistemas de IA se desarrollan y utilizan de un modo que incluya a diversos agentes y promueve la igualdad de acceso, la igualdad de género y la diversidad cultural, al tiempo que se evitan los efectos discriminatorios y los sesgos injustos prohibidos por el Derecho nacional o de la Unión; por bienestar social y ambiental se entiende que los sistemas de IA se desarrollan y utilizan de manera sostenible y respetuosa con el medio ambiente, así como en beneficio de todos los seres humanos, al tiempo que se supervisan y evalúan los efectos a largo plazo en las personas, la sociedad y la democracia.

Estos principios se concretan a lo largo del articulado del Reglamento y, en especial, cuando se regulan las condiciones de funcionamiento de los sistemas de alto riesgo; así, por ejemplo, estos sistemas (artículo 14) se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas. El objetivo de la supervisión humana será prevenir o reducir al mínimo los riesgos para la salud, la seguridad o los derechos fundamentales que pueden surgir cuando se utiliza un sistema de IA de alto riesgo conforme a su finalidad prevista o cuando se le da un uso indebido razonablemente previsible, en particular cuando dichos riesgos persistan a pesar de la aplicación de otros requisitos. Las medidas de supervisión serán proporcionales a

33 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (fecha de consulta: 14/07/2024).

los riesgos, al nivel de autonomía y al contexto de uso del sistema de IA de alto riesgo. Entre otras, las personas responsables de la supervisión podrán decidir, en cualquier situación concreta, no utilizar el sistema de IA de alto riesgo o descartar, invalidar o revertir los resultados de salida que este genere; también, intervenir en el funcionamiento del sistema o interrumpirlo pulsando un botón de parada o mediante un procedimiento similar que permita que se detenga de forma segura.

En segundo lugar, a estos sistemas de alto riesgo se les exige (artículo 15) que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida, algo a lo que pueden contribuir soluciones de redundancia técnica, tales como copias de seguridad o planes de prevención contra fallos. Entre las soluciones técnicas destinadas a subsanar vulnerabilidades específicas de la IA figurarán medidas para prevenir, detectar, combatir, resolver y controlar los ataques que traten de manipular el conjunto de datos de entrenamiento («envenenamiento de datos»), o los componentes entrenados previamente utilizados en el entrenamiento («envenenamiento de modelos»), la información de entrada diseñada para hacer que el modelo de IA cometa un error («ejemplos adversarios» o «evasión de modelos»), los ataques a la confidencialidad o los defectos en el modelo.

En tercer lugar, los sistemas de alto riesgo (artículo 13) se diseñarán y desarrollarán de un modo que se garantice que funcionan con un nivel de transparencia suficiente para que los responsables del despliegue interpreten y usen correctamente sus resultados de salida. Irán acompañados de las instrucciones de uso correspondientes en un formato digital o de otro tipo adecuado, las cuales incluirán información concisa, completa, correcta y clara que sea pertinente, accesible y comprensible para los responsables del despliegue. Además (artículo 50), los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA, excepto cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización. Por su parte, los proveedores de sistemas de IA, entre los que se incluyen los sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto, velarán por que los resultados de salida del sistema estén marcados en un formato legible por máquina y que sea posible detectar que han sido generados o manipulados de manera artificial.

En una línea similar, aunque de manera más genérica, el Convenio del Consejo de Europa prevé, entre las obligaciones generales para los Estados parte en relación con los sistemas de IA, las de supervisión[34], igualdad y no

34 Artículo 8 — Transparencia y supervisión. Cada Parte adoptará o mantendrá medidas para garantizar que existan requisitos adecuados de transparencia y supervisión adaptados a los contextos y riesgos

discriminación[35], privacidad y protección de datos personales[36]; fiabilidad[37] y transparencia[38].

V. UN ENFOQUE DE LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL BASADO EN LOS RIESGOS

La regulación de la IA, tal y como se concibe en Europa, aunque no solo en este espacio jurídico, requiere la aplicación de un enfoque basado en los riesgos, que adapte el tipo de las normas y su contenido a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA[39]. Se trata de una de las concreciones del bien conocido «principio de precaución», que guía la actuación de la Unión Europea[40]; así, y por citar únicamente dos ejemplos en el ámbito que nos

específicos con respecto a las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, incluida la identificación de los contenidos generados por los sistemas de inteligencia artificial.

35 Artículo 10 — Igualdad y no discriminación. 1. Cada Parte adoptará o mantendrá medidas con vistas a garantizar que las actividades relacionadas con el ciclo de vida de los sistemas de inteligencia artificial respeten la igualdad, incluida la igualdad de género, y la prohibición de la discriminación, según lo dispuesto en la legislación internacional y nacional aplicable. 2.Cada Parte se compromete a adoptar o mantener medidas destinadas a superar las desigualdades para lograr resultados justos, equitativos y equitativos, en consonancia con sus obligaciones nacionales e internacionales aplicables en materia de derechos humanos, en relación con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial.

36 Artículo 11 — Privacidad y protección de datos personales. Cada Parte adoptará o mantendrá medidas para garantizar que, en lo que respecta a las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial: a. los derechos de privacidad de las personas y los datos personales están protegidos, incluso a través de leyes, normas y marcos nacionales e internacionales aplicables; b. Se han establecido garantías y salvaguardias efectivas para las personas, de conformidad con las obligaciones jurídicas nacionales e internacionales aplicables.

37 Artículo 12 — Fiabilidad. Cada Parte adoptará, según proceda, medidas para promover la fiabilidad de los sistemas de inteligencia artificial y la confianza en sus resultados, que podrían incluir requisitos relacionados con la calidad y la seguridad adecuadas a lo largo de todo el ciclo de vida de los sistemas de inteligencia artificial.

38 Artículo 15 — Garantías procesales. 2. Cada Parte procurará garantizar que, según corresponda al contexto, se notifique a las personas que interactúen con sistemas de inteligencia artificial que están interactuando con dichos sistemas y no con un ser humano.

39 Soriano Arnanz (2021).

40 San Martín Segura (2023: 231 y ss.); como se explica en la comunicación de la Comisión Europea, de 2 de febrero de 2000, sobre el recurso al principio de precaución, el Tratado de la Comunidad Europea solo contiene una referencia explícita al principio de precaución, a saber, en el título dedicado a la protección del medio ambiente. No obstante, en la práctica, su ámbito de aplicación es mucho más amplio y se extiende asimismo a la política de los consumidores y a la salud humana, animal o vegetal. A falta de una definición del principio de precaución en el Tratado o en otros textos comunitarios, el Consejo solicitó a la Comisión, en su Resolución de 13 de abril de 1999, que elaborase líneas directrices claras y eficaces con vistas a la aplicación de este principio. La comunicación de la Comisión es una respuesta a esta solicitud. El establecimiento de líneas directrices comunes acerca de la aplicación del principio de precaución tendrá asimismo repercusiones positivas a escala internacional.

La Comisión subraya que el principio de precaución solo puede invocarse en la hipótesis 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

ocupa, en la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica, se dice que las actividades de investigación en el ámbito de la robótica deben llevarse a cabo de conformidad con el principio de precaución, anticipándose a los posibles impactos de sus resultados sobre la seguridad y adoptando las precauciones debidas, en función del nivel de protección, al tiempo que se fomenta el progreso en beneficio de la sociedad y del medio ambiente; por su parte, en la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas se recuerda que tal enfoque debe estar en consonancia con el principio de precaución que guía la legislación de la Unión y debe ocupar un lugar central en cualquier marco regulador para la inteligencia artificial.

El RIA define el riesgo como la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio (artículo 3.2) y, como consecuencia, en algunos casos el riesgo resultante será inaceptable y eso conducirá a la prohibición de los sistemas de IA que lo generen. Otros sistemas se caracterizan de alto riesgo porque son capaces de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas y a ellos se les aplicarán las cautelas que ya hemos visto en el apartado anterior y se les someterá (artículo 9) a un proceso iterativo continuo, planificado y ejecutado durante todo el ciclo de vida, que requerirá revisiones y actualizaciones sistemáticas periódicas. Constará de las siguientes etapas: a) la determinación y el análisis de los riesgos conocidos y previsibles que el sistema de IA de alto riesgo pueda plantear para la salud, la seguridad o los derechos fundamentales cuando el sistema de IA de alto riesgo se utilice de conformidad con su finalidad prevista; b) la estimación y la evaluación de los riesgos que podrían surgir cuando el sistema de IA de alto riesgo se utilice

efectos potencialmente negativos, evaluación de los datos científicos disponibles y determinación del grado de incertidumbre científica.

Medidas que se derivan del recurso al principio de precaución. El recurso al principio de precaución debe guiarse por tres principios específicos: 1) la aplicación del principio debe basarse en una evaluación científica lo más completa posible; en cada etapa esta evaluación debe determinar, en la medida de lo posible, el grado de incertidumbre científica; 2) toda decisión de actuar o de no actuar en virtud del principio de precaución debe ir precedida de una determinación del riesgo y de las consecuencias potenciales de la inacción; 3) tan pronto como se disponga de los resultados de la evaluación científica o de la determinación del riesgo, todas las partes interesadas deben tener la posibilidad de participar, con la máxima transparencia, en el estudio de las diferentes acciones que pueden preverse.

Aparte de estos principios específicos, siguen siendo aplicables los principios generales de una buena gestión de los riesgos cuando se invoca el principio de precaución. Se trata de los cinco principios siguientes: la proporcionalidad entre las medidas adoptadas y el nivel de protección elegido; la no discriminación en la aplicación de las medidas; la coherencia de las medidas con las ya adoptadas en situaciones similares o utilizando planteamientos similares; el análisis de las ventajas y los inconvenientes que se derivan de la acción o de la inacción y la revisión de las medidas a la luz de la evolución científica.

https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2000:0001:FIN:es:PDF (fecha de consulta: 14/07/2024).

de conformidad con su finalidad prevista y cuando se le dé un uso indebido razonablemente previsible; c) la evaluación de otros riesgos que podrían surgir, a partir del análisis de los datos recogidos con el sistema de vigilancia post-comercialización; d) la adopción de medidas adecuadas y específicas de gestión de riesgos diseñadas para hacer frente a los riesgos detectados.

Con vistas a eliminar o reducir los riesgos asociados a la utilización del sistema de IA de alto riesgo, se tendrán debidamente en cuenta los conocimientos técnicos, la experiencia, la educación y la formación que se espera que posea el responsable del despliegue, así como el contexto en el que está previsto que se utilice el sistema.

También el Convenio Marco contempla un marco jurídico de gestión de riesgos e impactos (artículo 16), imponiendo a los Estados parte medidas para la identificación, evaluación, prevención y mitigación de los riesgos planteados por los sistemas de inteligencia artificial, teniendo en cuenta los impactos reales y potenciales en los derechos humanos, la democracia y el Estado de Derecho. Dichas medidas se graduarán y diferenciarán, según proceda, y deberán: a). tener debidamente en cuenta el contexto y el uso previsto de los sistemas de inteligencia artificial, en particular en lo que respecta a los riesgos para los derechos humanos, la democracia y el Estado de Derecho; b). tener debidamente en cuenta la gravedad y la probabilidad de los posibles impactos; c). considerar, cuando proceda, las perspectivas de las partes interesadas pertinentes, en particular de las personas cuyos derechos puedan verse afectados; d). aplicar de forma iterativa a lo largo de las actividades dentro del ciclo de vida del sistema de inteligencia artificial; e). incluir el monitoreo de los riesgos e impactos adversos para los derechos humanos, la democracia y el estado de derecho; f). incluir documentación de los riesgos, los impactos reales y potenciales, y el enfoque de gestión de riesgos; g). exigir, cuando proceda, que se sometan a ensayo los sistemas de inteligencia artificial antes de ponerlos a disposición para su primer uso y cuando se modifiquen significativamente. Finalmente, cada Parte evaluará la necesidad de una moratoria o prohibición u otras medidas adecuadas con respecto a determinados usos de los sistemas de inteligencia artificial cuando considere que dichos usos son incompatibles con el respeto de los derechos humanos, el funcionamiento de la democracia o el Estado de Derecho.

VI. LAS DIFERENCIAS ESTRUCTURALES ENTRE EL REGLAMENTO DE LA UNIÓN EUROPEA Y EL CONVENIO MARCO DEL CONSEJO DE EUROPA

Ya se ha visto que el Reglamento y el Convenio Marco se terminaron de elaborar en fechas muy próximas y siendo conscientes en cada institución de lo que estaba haciendo la otra; de hecho, el Comité de Ministros del Consejo de Europa decidió permitir la inclusión en las negociaciones de la Unión Europea, que

participó representada por la Comisión Europea, incluyendo en su delegación también a representantes de la Agencia de los Derechos Fundamentales de la Unión Europea y del Supervisor Europeo de Protección de Datos. Incluso el artículo 27.2 del Convenio Marco prevé que las Partes que sean miembros de la Unión Europea aplicarán, en sus relaciones mutuas, las propias normas de la Unión Europea que regulen las materias comprendidas en el ámbito de aplicación del Convenio, sin perjuicio de su objeto y fin y de su plena aplicación con las demás Partes.

Entrando en las diferencias estructurales más destacadas entre ambas normas cabe destacar, en primer lugar, que el ámbito de aplicación del Reglamento es regional, el territorio de la Unión Europea[41], al margen de que pueda generar, como veremos más adelante, un «efecto Bruselas», mientras que el Convenio Marco ya nace con una cierta vocación de globalidad; así, en su Preámbulo se invoca la necesidad de establecer, con carácter prioritario, un marco jurídico aplicable a escala mundial que establezca principios generales y normas comunes que rijan las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, preservando eficazmente los valores compartidos y aprovechando los beneficios de la inteligencia artificial para la promoción de esos valores de manera que favorezca la innovación responsable, y el artículo 30.1 prevé que el Convenio estará abierto a la firma de los Estados miembros del Consejo de Europa, de los Estados no miembros que hayan participado en su elaboración y de la Unión Europea.

En segundo lugar, el Reglamento, de conformidad con lo dispuesto en el artículo 288 del Tratado de Funcionamiento de la Unión Europea, «será obligatorio en todos sus elementos y directamente aplicable en cada Estado miembro»; el Convenio Marco será obligatorio en la medida en los Estados se incorporen a él. De acuerdo con esta característica, los destinatarios del CIA son los Estados, mientras que el RIA se dirige directamente a los proveedores y los responsables del despliegue.

En tercer término, el Reglamento incluye una normativa extensa (113 artículos y XIII Anexos) y muy prolija, con preceptos muy detallados, mientras que el Convenio Marco se compone de un articulado mucho más reducido (36 artículos) y, desde luego, menos detallado. El propio Convenio destaca su «carácter marco…, que podrá complementarse con otros instrumentos para abordar cuestiones específicas relacionadas con las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial» (número 11 del Preámbulo).

41 De acuerdo con el artículo 1, «el objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA (en lo sucesivo, «sistemas de IA») en la Unión…»

En cuarto lugar, el Reglamento, en términos generales, contiene más reglas, es decir, incluye comportamientos precisos de lo que puede, o no, hacerse en materia de IA; el Convenio Marco, por su parte, adopta una configuración mucho más principialista, esto es, contiene mandatos de optimización, caracterizados por el hecho de que pueden ser cumplidos en diferente grado. Así, por ejemplo, el Reglamento establece una serie de prácticas de IA que estarán prohibidas (artículo 5)[42] e impone unas obligaciones que deben cumplir los sistemas de alto riesgo[43] y, entre otros, los proveedores de modelos de IA de uso general (artículo 51). Por su parte, el Convenio Marco dispone (artículo 4) que «cada Parte adoptará o mantendrá medidas para garantizar que las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial sean compatibles con las obligaciones de protección de los derechos humanos, consagradas en el derecho internacional aplicable y en su legislación nacional».

En quinto lugar, y también con carácter general pero no absoluto y en la línea de lo dicho en el punto anterior, el Reglamento impone obligaciones de medios y de resultado mientras que el Convenio Marco incluye, esencialmente, obligaciones de resultado, dejando a los Estados la concreción de las medidas adecuadas para alcanzarlos. Así, por ejemplo, el Reglamento prevé que los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas (artículo 14.1); más adelante, dispone que los proveedores de sistemas de IA de alto riesgo que consideren o tengan motivos para considerar que un sistema de IA de alto riesgo que han introducido en el mercado o puesto en servicio no es conforme con el presente Reglamento adoptarán inmediatamente las medidas correctoras necesarias para que sea conforme, para retirarlo del mercado, desactivarlo o recuperarlo, según proceda (artículo 20.1).

Por su parte, el Convenio Marco establece (artículo 1.1 y 1.2) que sus disposiciones «tienen por objeto garantizar que las actividades realizadas durante el ciclo de vida de los sistemas de inteligencia artificial sean plenamente compatibles con

42 Así, por citar algunas, las que se sirvan de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas; las que exploten alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoquen, o sea razonablemente probable que provoquen, perjuicios considerables a esa persona o a otra; las que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión…

43 Un sistema de gestión de riesgos (artículo 9), la gobernanza de los datos (artículo 10), la documentación técnica (artículo 11), supervisión humana (artículo 14)…

los derechos humanos, la democracia y el Estado de Derecho. Cada Parte adoptará o mantendrá medidas legislativas, administrativas o de otra índole apropiadas para dar efecto a las disposiciones establecidas en el presente Convenio. Estas medidas se graduarán y diferenciarán según sea necesario en función de la gravedad y la probabilidad de que se produzcan efectos adversos para los derechos humanos, la democracia y el Estado de Derecho a lo largo del ciclo de vida de los sistemas de inteligencia artificial»; además, y por mencionar otro supuesto, conforme al artículo 5, «1. Cada Parte adoptará o mantendrá medidas que tengan por objeto garantizar que los sistemas de inteligencia artificial no se utilicen para socavar la integridad, independencia y eficacia de las instituciones y procesos democráticos, incluido el principio de separación de poderes, el respeto de la independencia judicial y el acceso a la justicia. 2. Cada Parte adoptará o mantendrá medidas que tengan por objeto proteger sus procesos democráticos en el contexto de las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial, incluido el acceso equitativo de las personas al debate público y su participación en él, así como su capacidad para formarse opiniones libremente).

Finalmente, el Reglamento contiene un sistema de sanciones: «los Estados miembros establecerán el régimen de sanciones y otras medidas de ejecución, como advertencias o medidas no pecuniarias, aplicable a las infracciones del presente Reglamento que cometan los operadores y adoptarán todas las medidas necesarias para garantizar que se aplican de forma adecuada y efectiva… Tales sanciones serán efectivas, proporcionadas y disuasorias (artículo 99.1). A continuación, el artículo especifica el importe de las multas administrativas que se deben imponer, que son significativas[44].

Por su parte, el Convenio Marco se limita a disponer que «cada Parte establecerá o designará uno o más mecanismos eficaces para supervisar el cumplimiento de las obligaciones establecidas en el Convenio» (artículo 26.1). Sin embargo, al tratarse de un tratado, en función de las normas del sistema jurídico nacional con respecto a la aplicación del derecho internacional, los tribunales nacionales podrán declarar las violaciones de determinados artículos. Para muchos

44 Así, el no respeto de la prohibición de las prácticas de IA a que se refiere el artículo 5 estará sujeto a multas administrativas de hasta 35.000.000 de euros o, si el infractor es una empresa, de hasta el 7 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior. El incumplimiento de cualquiera de las disposiciones que figuran a continuación en relación con los operadores o los organismos notificados, distintas de los mencionados en el artículo 5, estará sujeto a multas administrativas de hasta 15.000.000 de euros o, si el infractor es una empresa, de hasta el 3 % de su volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior: a) las obligaciones de los proveedores con arreglo al artículo 16; b) las obligaciones de los representantes autorizados con arreglo al artículo 22; c) las obligaciones de los importadores con arreglo al artículo 23; las obligaciones de los distribuidores con arreglo al artículo 24; e) las obligaciones de los responsables del despliegue con arreglo al artículo 26… La presentación de información inexacta, incompleta o engañosa a organismos notificados o a las autoridades nacionales competentes en respuesta a una solicitud estará sujeta a multas administrativas de hasta 7.500.000 de euros o, si el infractor es una empresa, de hasta el 1 % del volumen de negocios mundial total correspondiente al ejercicio financiero anterior, si esta cuantía fuese superior (artículo 99.3, 4 y 5).

signatarios, esto significará que los tribunales u otras autoridades tendrán que determinar si las normas de la CIA son suficientemente precisas para ser consideradas autoejecutables[45]. De todas maneras, el CIA probablemente funcionara como documento interpretativo para el Tribunal Europeo de Derechos Humanos (TEDH), que no está mencionado en el texto del CIA, que prevé una «Conferencia de las partes» como mecanismo de resolución de disputas (artículo 23).

VII. LOS MODELOS DE INTELIGENCIA ARTIFICIAL DE USO GENERAL EN EL REGLAMENTO DE LA UNIÓN EUROPEA

Si hubiera que destacar una novedad incorporada en las fases finales de elaboración del RIA 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 de IA de uso general, que fueron introducidos en el texto durante su paso por el Parlamento Europeo, si bien con el nombre entonces de «modelos fundacionales». Como se explica en el Preámbulo, estos modelos suelen entrenarse usando grandes volúmenes de datos y a través de diversos métodos, como el aprendizaje auto-supervisado[46], no supervisado o por refuerzo[47]. Los modelos de IA de uso general pueden introducirse en el mercado de diversas maneras, por ejemplo, a través de bibliotecas, interfaces de programación de aplicaciones (API), como descarga directa o como copia física. Estos modelos pueden modificarse o perfeccionarse y transformarse en nuevos modelos. Aunque los modelos de IA son componentes esenciales de los sistemas de IA, no constituyen por sí mismos sistemas de IA. Los modelos de IA requieren que se les añadan otros componentes, como, por ejemplo, una interfaz de usuario, para convertirse en sistemas de IA. Los modelos de IA suelen estar integrados en los sistemas de IA y formar parte de dichos sistemas… Los grandes modelos de IA generativa son un ejemplo típico de un modelo de IA de

45 Ziller (2024).

46 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: 49).

47 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: 49 y 50).

uso general, ya que permiten la generación flexible de contenidos, por ejemplo, en formato de texto, audio, imágenes o vídeo, que pueden adaptarse fácilmente a una amplia gama de tareas diferenciadas (considerandos 97 y 99).

Estos modelos, se insiste en el Preámbulo, presentan unas oportunidades de innovación únicas, pero también representan un desafío para los artistas, autores y demás creadores y para la manera en que se crea, distribuye, utiliza y consume su contenido creativo. El desarrollo y el entrenamiento de estos modelos requiere acceder a grandes cantidades de texto, imágenes, vídeos y otros datos. Las técnicas de prospección de textos y datos pueden utilizarse ampliamente en este contexto para la recuperación y el análisis de tales contenidos, que pueden estar protegidos por derechos de autor y derechos afines. Todo uso de contenidos protegidos por derechos de autor requiere la autorización del titular de los derechos de que se trate, salvo que se apliquen las excepciones y limitaciones pertinentes en materia de derechos de autor (considerando 105).

Además, estos modelos de IA de uso general pueden plantear riesgos sistémicos; por ejemplo, cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios. Debe entenderse que los riesgos sistémicos aumentan con las capacidades y el alcance de los modelos, pueden surgir durante todo el ciclo de vida del modelo y se ven influidos por las condiciones de uso indebido, la fiabilidad, equidad y la seguridad del modelo, su nivel de autonomía, su acceso a herramientas, modalidades novedosas o combinadas, las estrategias de divulgación y distribución, la posibilidad de eliminar las salvaguardias y otros factores.

Entre otras medidas que sería muy prolijo mencionar aquí, los proveedores de modelos de IA de uso general con riesgos sistémicos deben evaluarlos y tratar de mitigarlos pero si, con todo, provocan un incidente grave, el proveedor debe, sin demora indebida, hacer un seguimiento del incidente y comunicar toda la información pertinente y las posibles medidas correctoras a la Comisión Europea y a las autoridades nacionales competentes. Además, los proveedores deben garantizar que el modelo y su infraestructura física, si procede, tengan un nivel adecuado de ciberseguridad durante todo el ciclo de vida del modelo. Esta protección debe tener en cuenta las fugas accidentales de modelos, las divulgaciones no autorizadas, la elusión de las medidas de seguridad y la defensa contra los ciberataques, el acceso no autorizado o el robo de modelos. Esa protección podría facilitarse asegurando los pesos, los algoritmos, los servidores y los conjuntos de datos del modelo, por ejemplo, mediante medidas específicas ciberseguridad, soluciones técnicas adecuadas y controles de acceso cibernéticos y físicos, en función de las circunstancias pertinentes y los riesgos existentes.

Además, quienes utilicen un sistema de IA para generar o manipular un contenido de imagen, audio o vídeo generado o manipulado por una IA que se

asemeje notablemente a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos (ultrasuplantaciones) deben hacer público, de manera clara y distinguible, que este contenido ha sido creado o manipulado de manera artificial etiquetando los resultados de salida generados por la IA en consecuencia e indicando su origen artificial. El cumplimiento de esta obligación de transparencia no debe interpretarse como un indicador de que la utilización del sistema de IA o de sus resultados de salida obstaculiza el derecho a la libertad de expresión y el derecho a la libertad de las artes y de las ciencias, en particular cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, con sujeción a unas garantías adecuadas para los derechos y libertades de terceros. En tales casos, la obligación de transparencia se limita a revelar la existencia de tales contenidos generados o manipulados de una manera adecuada que no obstaculice la presentación y el disfrute de la obra, así como su explotación y uso normales, al tiempo que se conservan la utilidad y calidad de la citada obra (considerando 134).

VIII. LOS POSIBLES «EFECTO BRUSELAS» Y «EFECTO ESTRASBURGO» DE LA REGULACIÓN EUROPEA DE LA INTELIGENCIA ARTIFICIAL

En un conocido artículo publicado en 2012, que adoptó formato de libro en 2020[48], la profesora Anu Bradford explicó cómo y por qué las normas y reglamentos «de Bruselas», ciudad sede de buena parte de las instituciones de la Unión Europea, 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)[49].

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

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

49 «The Brussels Effect», en *Northwestern University Law Review*, n.º 1, 2012, en https://scholarship.law.columbia.edu/faculty_scholarship/271 (consulta: 14/07/2024).

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 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 (p. 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[50], una exportación del contenido de esa nueva normativa a otros Estados, un «efecto Bruselas» sobre la regulación de la IA. 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.

Marco Almada y Anca Radu también hacen un análisis crítico centrado en una versión todavía provisional del Reglamento: dado que esta norma sigue la legislación de la UE sobre seguridad de los productos, sus disposiciones ofrecen una protección limitada a algunos de los valores que la política de la UE pretende garantizar, como la protección de los derechos fundamentales. Estas deficiencias se ven agravadas por los esfuerzos activos de la UE para dar forma a instrumentos alternativos, como el convenio sobre IA propuesto por el Consejo de Europa, que sigue las líneas de la Ley de IA. En consecuencia, la difusión de la Ley de IA como norma mundial tendrá consecuencias para la agenda política de la UE en materia de IA y para la conceptualización del efecto Bruselas[51].

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)[52],

50 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.

51 (2024: 1-18); véase también Barrio Andrés (2024) y Meuwese y Wolswinkel (2022).

52 https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/ (consulta: 14/07/2024).

no parece en absoluto, por lo que está ocurriendo en otros espacios jurídicos, que esta propuesta vaya a tener repercusiones únicamente hacia dentro de la Unión.

Pero, además, de este «efecto Bruselas» también se podría hablar de un posible «efecto Estrasburgo», sede de las instituciones del Consejo de Europa, pero en este caso de una manera voluntaria, pues, como ya hemos visto, en el proceso de elaboración del Convenio Marco se incluyó en las negociaciones a la Unión Europea y a Estados no europeos (Argentina, Australia, Canadá, Costa Rica, la Santa Sede, Israel, Japón, México, Perú, los Estados Unidos de América y Uruguay), además de a representantes de otras organizaciones internacionales, como la Organización para la Seguridad y la Cooperación en Europa (OSCE), la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO).

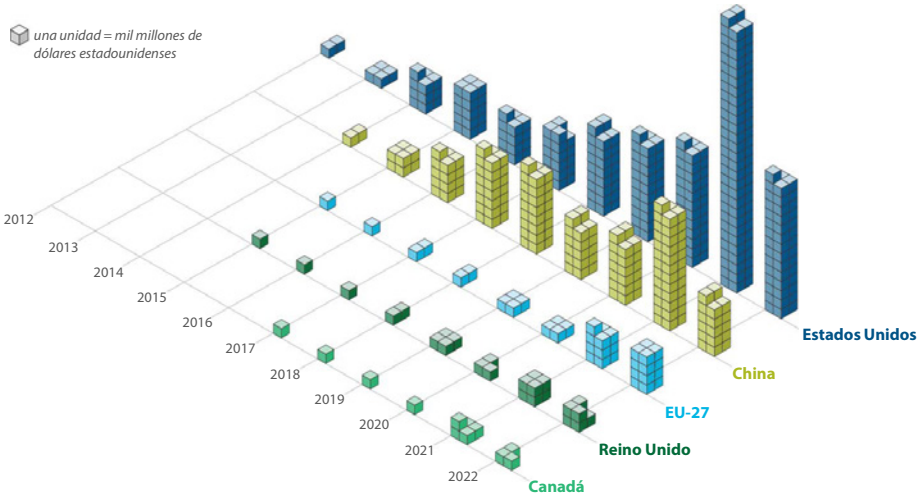
Debido a las ambiciones globales del CIA, se necesita garantizar que el sector privado no quede completamente fuera de su ámbito. El compromiso alcanzado en las negociaciones es que los Estados signatarios pueden elegir entre aplicar las normas de la CIA a los actores privados directamente o, más bien, de otra manera pero teniendo en cuento los objetivos del Convenio (Artículo 3 — 1b).

Además, en el artículo 25.1 del Convenio se alienta a las Partes a que, según proceda, presten asistencia a los Estados que no sean Partes en la Convención para que actúen de conformidad con sus disposiciones y pasen a ser Partes en ella. Y en el artículo 31.1 se prevé que, después de la entrada en vigor del Convenio, el Comité de Ministros del Consejo de Europa podrá, previa consulta a las Partes y obteniendo su consentimiento unánime, invitar a cualquier Estado no miembro del Consejo de Europa que no haya participado en la elaboración del Convenio a adherirse al mismo mediante una decisión adoptada por la mayoría prevista en el artículo 20.d del Estatuto del Consejo de Europa. Europa, y por unanimidad de los representantes de las Partes con derecho a formar parte del Comité de Ministros. En suma, el Convenio Marco del Consejo de Europa se abre a la posibilidad de ser una primera norma global.

IX. ALGUNAS CONCLUSIONES.

En primer lugar, hay que recordar que Europa va muy por detrás de Estados Unidos y de China en materia de investigación, desarrollo e innovación de la inteligencia artificial; en palabras del Tribunal de Cuentas Europeo, «la UE cuenta con una sólida comunidad de investigación pública en materia de IA (el mayor número de publicaciones científicas revisadas por pares sobre IA en el mundo en 20223), pero afronta una serie de escollos en la carrera mundial por la inversión en IA. La inversión privada en IA ha sido menor que en otras regiones

mundiales que lideran este campo (Estados Unidos y China) desde 2015»[53]. Al respecto, es muy ilustrativa la siguiente imagen sobre las inversiones de capital riesgo en el sector de la IA y de los datos en miles de millones de dólares (Fuente: OCDE, noviembre de 2023).



Por lo que respecta a la UE, y en coherencia con sus principios estructurales, la innovación y el desarrollo tecnológicos, así como el funcionamiento del mercado interior, no son objetivos aislados sino que deben garantizar «al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA» (artículo 1 RIA).

Así pues, el Reglamento es la respuesta jurídica de la UE al tremendo desafío que supone promover la innovación y la competitividad en IA de una manera compatible con la garantía de los derechos fundamentales y del Estado social y democrático de Derecho, algo que no parece preocupe en la misma medida en Estados Unidos ni, mucho menos, en China: el 30 de octubre de 2023 el presidente Biden emitió *la Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*[54], donde se proclama que el Gobierno Federal tratará de promover principios y acciones responsables de seguridad y protección de la IA con otras naciones, «incluidos nuestros competidores, al tiempo que lidera conversaciones y

53 «Informe Especial 08/2024: Ambición de UE en materia de inteligencia artificial — Una gobernanza más sólida y una inversión mayor y mejor orientada son fundamentales de cara al futuro», https://www.eca.europa.eu/es/publications/SR-2024-08 (consulta: 16/09/2024).

54 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/ (consulta: 14/07/2024).

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». China, por su parte, aprobó, en agosto de 2023, una Ley general reguladora de la Inteligencia Artificial y, en paralelo, otra regulación específica de la IA generativa donde se apela a la seguridad y robustez; a la apertura, la transparencia y la explicabilidad y a la protección del medio ambiente[55], pero, al tiempo ha apostado por la IA como herramienta de férreo control de la disidencia y por sistemas como el «crédito social»[56], que estarán prohibidos en Europa.

En todo caso, y dado el evidente interés de las empresas estadounidenses y chinas en comercializar sus sistemas de IA en la UE, habrá que esperar para ver en qué medida las normas europeas, especialmente el RIA generan, aunque sea de manera mitigada, efectos «Bruselas» y/o «Estrasburgo» en las legislaciones de esos y otros países.

En segundo lugar, no hay que olvidar que el concepto de «sistema de IA» asumido por el RIA, el CIA y, en general, por las diferentes organizaciones internacionales deja fuera del ámbito de aplicación de las citadas disposiciones a sistemas que están organizados a partir de las reglas definidas con carácter previo por personas físicas para ejecutar operaciones de manera automática, a los que, en nuestra opinión, habría que dar una respuesta jurídica adecuada por parte de las autoridades nacionales, pues no siempre cuentan con ella.

En tercer término, y con buen criterio, tanto el RIA como el CIA han asumido un enfoque regulatorio basado en los riesgos que puedan generar los sistemas de IA y ello en coherencia con el objetivo, antes apuntado, de garantizar un elevado nivel de protección de la salud, la seguridad y los derechos

55 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 (consulta: 14/07/2024).

56 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; Schaefer (2020).

fundamentales. A este mismo fin responden los que se podrían considerar como principios de actuación básicos en la materia: supervisión, igualdad y no discriminación, privacidad y protección de datos personales, fiabilidad y transparencia.

En definitiva, nos encontramos en un momento crucial para la regulación de la IA y la actuación que viene realizando desde hace tiempo la UE en este sentido, especialmente a través de la nueva norma que hemos analizado aquí, junto con el impacto de tecnologías como ChatGPT han acelerado la regulación a nivel global, a la que contribuirá el proceso de adopción del Convenio Marco de IA del Consejo de Europa. Se trata de un gran intento de armonización y construcción de un marco estándar y mínimo para Europa, con alcance mundial; un esfuerzo centrado en los derechos humanos, la democracia y el Estado de derecho y no tanto en la economía y el mercado.

Europa quiere posicionarse como un faro brillante, como una piedra angular en la regulación de la IA no solo para todo el continente sino también a nivel mundial. Para los Estados no miembros de la UE, el Convenio Marco de IA puede desempeñar, obviamente, un papel normativo y legal importante. Y para la UE y los Estados miembros que lo ratifiquen el Convenio Marco tiene el potencial de complementar e, incluso, mejorar y fortalecer la protección de los derechos. Se ha destacado a este respecto que el Convenio Marco no se limita a los sistemas de IA de alto riesgo. Además, integra «principios» en el ámbito legal y político, que van más allá de los principios éticos de la IA. Estos principios legales tienen un gran potencial en manos de los operadores jurídicos para destilar reglas y obligaciones concretas, como ya ha ocurrido durante décadas con los principios relativos a la protección de datos. Además, el Convenio de IA otorga a los interesados afectados por la IA ciertos derechos y, en particular, garantías y mecanismos para la defensa de sus derechos ante autoridades independientes. Es particularmente relevante el enfoque basado en el riesgo y, en particular, el artículo 16, que prescribe una evaluación y mitigación continuas de los riesgos y los impactos adversos de cualquier tipo de sistema de IA.

Al mismo tiempo este marco nuevo necesitará una concretización que justamente es lo que ofrece la Ley de IA. De hecho, en la Exposición de Motivos de la Decisión del Consejo de la UE relativa a la firma del Convenio Marco está claro que el Convenio «debe aplicarse exclusivamente en la Unión a través del Reglamento de Inteligencia Artificial, que armoniza plenamente las normas para la comercialización, la puesta en servicio y la utilización de sistemas de IA, así como el acervo vigente de la Unión»[57]. Aunque el Reglamento de IA no aspira a ser una

57 Comisión Europea, Propuesta de Decisión del Consejo relativa a la firma, en nombre de la Unión Europea, del Convenio Marco del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho, Bruselas, 26.6.2024, COM(2024) 264 final, eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52024PC0264.

regulación completa y exhaustiva, todos los esfuerzos se dirigen actualmente a presentarlo como una guía para reguladores de todo el mundo.

BIBLIOGRAFÍA CITADA

Almada, M. y Radu, A. (2024). The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy. *German Law Journal*, FirstView, 1-18.

Arnal, J. y Jorge Ricart, R. E. (2023). Inteligencia artificial (I): menor «efecto Bruselas», las posibles consecuencias desglobalizadoras de un enfoque regulatorio divergente y la importancia de políticas públicas para el empleo. *Real Instituto Elcano*, 88.

Barrio Andrés, M. (2022). Inteligencia artificial: origen, concepto, mito y realidad. *El Cronista del Estado Social y Democrático de Derecho (Ejemplar dedicado a Inteligencia artificial y derecho)*, 100, 14-21.

Barrio Andrés, M. (2024). Consideraciones sobre el ámbito extraterritorial del Reglamento Europeo de Inteligencia Artificial. *Derecho Digital e Innovación*, 20.

Bradford, A. (2021). The Brussels Effect. *Northwestern University Law Review*, 1.

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

Bradford, A. (2023). *Digital Empires. The Global Battle to Regulate Technology.* Oxford: Oxford University Press.

Bustos Gisbert, R. (2024). El constitucionalista europeo ante la inteligencia artificial: reflexiones metodológicas de un recién llegado. *Revista Española de Derecho Constitucional*, 131, 149-178.

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

Cotino Hueso, L. (2023). Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos. En Balaguer Callejón, F. y Cotino Hueso, L. (coords.), *Derecho público de la inteligencia artificial* (347-402). Zaragoza: Fundación Manuel Giménez Abad.

Cotino Hueso, L. (2024). El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa. *Revista Administración & Cidadanía*, EGAP.

Crawford, K. (2023). *Atlas de IA*. Ulzama: Ned Ediciones.

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

Gascón Marcén, A. (2024). Nuevo Convenio Marco sobre Inteligencia Artificial del Consejo de Europa. *Derecho Digital e Innovación. Digital Law and Innovation Review*, 20.

Gómez-Pérez, A. (2023). *Inteligencia artificial y lengua española*. Madrid: Real Academia Española.

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

Leslie, D., Ashurst, C., González, N. M., Griffiths, F., Jayadeva, S., Jorgensen, M., Katell, M., Krishna, S., Kwiatkowski, D., Martins, C. I., Mahomed, S., Mougan, C., Pandit, S., Richey, M., Sakshaug, J. W., Vallor, S., y Vilain, L. (2024). Frontier

AI,' Power, and the Public Interest: Who Benefits, Who Decides? *Harvard Data Science Review*, (Special Issue 5). https://doi.org/10.1162/99608f92.4a42495c.

Llano Alonso, F. H. (2024). *Homo ex machina. Ética de la inteligencia artificial y Derecho digital ante el horizonte de la singularidad tecnológica*. Valencia: Tirant lo Blanch.

Meuwese A.C.M. y Wolswinkel C.J. (2022). Een Wet op de Artificiële Intelligentie? De Europese wetgever haalt de nationale in. *Nederlands Juristenblad*, 97(2), 92-100.

Mitchell, M. (2024). *Inteligencia artificial. Guía para seres pensantes*. Madrid: Capitán Swing.

Pasquale, F. (2024). *Las nuevas leyes de la robótica. Defender la experiencia humana en la era de la IA*. Barcelona: Galaxia Gutenberg.

Pérez-Ugena, M. (2024). Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU., China e Iberoamérica. *Anuario Iberoamericano de Justicia Constitucional*, 28(1), 129-156.

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

Presno Linera, M. Á. (2023). La propuesta de «Ley de Inteligencia Artificial» europea. *Revista de las Cortes Generales*, 116, 81-133.

Ridaura Martínez, M. J. (2024). El reconocimiento facial para preservar la seguridad: nuevo marco del Reglamento de Inteligencia Artificial de la Unión Europea. En Caamaño Domínguez, F. y Jove Villares, D. (coords.), *Tecnologías abusivas y Derecho*. Valencia: Tirant lo Blanch (en prensa).

Russel, S. y Norvig, P. (2008). *Inteligencia Artificial: un enfoque moderno*. Madrid: Pearson Education.

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

Schaefer, K. (2020). *China's social credit system: context, competition, technology and geopolitics*. Trivium China.

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.

Ziller, J. (2024). The Council of Europe Framework Convention on Artificial Intelligence vs. the EU Regulation: two quite different legal instruments. *CERIDAP*, 2/2024, 202-227.

***

TITLE: *The regulation of artificial intelligence in Europe*

ABSTRACT: *A few weeks before finalizing this article the final texts of both the Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (the so-called AI Act) and the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law became known. Both regulatory initiatives are the outcome of a process of legislative, institutional, social, technological and economic debate that dates back several years and that is based on the conviction, as stated in the European Commission's White Paper on Artificial Intelligence, that artificial intelligence «will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine. At the same time, Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other*

kinds of discrimination, intrusion in our private lives or being used for criminal purposes». In the following pages we will analyse the legal responses that have been given in Europe to the challenge of regulating AI in a manner that is not apocalyptic but not completely integrated either: how these rules define artificial intelligence, which principles underpin them, the choice for a risk-based regulatory approach, the structural differences between the EU Regulation and the Council of Europe Framework Convention, the treatment of general-purpose AI models, as well as the possible 'Brussels effect' and 'Strasburg effect' these rules may trigger.

RESUMEN: *Pocas semanas antes de finalizar la redacción de estas páginas se conocieron tanto el texto del Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (la llamada Ley de Inteligencia Artificial) como el del Convenio Marco del Consejo de Europa sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho. Ambas normas son el resultado presente de un proceso de debate legislativo, institucional, social, tecnológico y económico que se remonta a varios años atrás y que parte del convencimiento, como se dice en el Libro Blanco sobre la inteligencia artificial de la Comisión Europea, de que la inteligencia artificial «cambiará nuestras vidas, pues mejorará la atención sanitaria, 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». En las siguientes páginas analizaremos las respuestas jurídicas que se han dado en Europa al reto de regular la IA de forma no apocalíptica pero tampoco totalmente integrada: cómo definen estas normas la inteligencia artificial, qué principios generales las han inspirado, la opción por un enfoque regulatorio basado en los riesgos, las diferencias estructurales entre el Reglamento de la Unión Europea y el Convenio Marco del Consejo de Europa, el tratamiento de los modelos de inteligencia artificial de uso general y, finalmente, los posibles «efecto Bruselas» y «efecto Estrasburgo» que pueden provocar estas normas.*

KEY WORDS: *European Union, Council of Europe, fundamental rights, artificial intelligence, artificial intelligence regulation, Brussels effect*.

PALABRAS CLAVE: *Unión Europea, Consejo de Europa, derechos fundamentales, inteligencia artificial, regulación de la inteligencia artificial, efecto Bruselas*.

FECHA DE RECEPCIÓN: *15.07.2024* FECHA DE ACEPTACIÓN: *19.09.2024*

CÓMO CITAR/ CITATION: Presno Linera, M. Á. y Meuwese, A.C.M. (2024). La regulación de la inteligencia artificial en Europa. *Teoría y Realidad Constitucional*, 54, 131-161.