

ANÁLISIS COMPARADO DE LOS DISTINTOS ENFOQUES REGULATORIOS DE LA INTELIGENCIA ARTIFICIAL EN LA UNIÓN EUROPEA, EE. UU., CHINA E IBEROAMÉRICA

A comparative analysis of regulatory approaches to artificial intelligence: perspectives from the European Union, the United States, China, and Ibero-America

MARÍA PÉREZ-UGENA
Universidad Rey Juan Carlos
maria.perezugena@urjc.es

Cómo citar/Citation

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.

doi: <https://doi.org/10.18042/cepc/aijc.28.05>

Resumen

Este estudio lleva a cabo un examen comparativo del panorama regulatorio que rodea a la inteligencia artificial (IA) en Iberoamérica, la Unión Europea, los Estados Unidos de América y China. La investigación comienza con una introducción que describe las características definitorias de la IA e identifica los principios generales que deberían guiar su desarrollo. Estos principios sirven como ejes básicos para la creación de marcos jurídicos generales destinados a fundamentar posibles restricciones en esta tecnología con el fin de proteger los derechos humanos. El estudio subraya la necesidad de emprender acciones coordinadas a nivel internacional, en consonancia con los problemas que plantea la IA dentro de un contexto jurídico global.

Palabras clave

Inteligencia artificial; regulación; derecho comparado; derechos humanos.

Abstract

This study conducts a comparative examination of the regulatory landscape surrounding artificial intelligence (AI) across Ibero-America, the European Union, the United States of America, and China. The investigation begins with an introduction outlining the defining characteristics of AI and identifying the general principles that should guide its development. These principles serve as foundational pillars for the establishment of comprehensive legal frameworks aimed at justifying potential restrictions on this technology to safeguard human rights. The study emphasizes the necessity of undertaking coordinated international efforts, in line with the challenges posed by AI within a global legal context.

Keywords

Artificial intelligence; regulation; Comparative law; human rights.

SUMARIO

I. INTRODUCCIÓN A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL (IA). II. DE LA NECESIDAD DE UN MARCO ÉTICO GLOBAL PARA LA PROTECCIÓN DE LOS DERECHOS HUMANOS: 1. Privacidad. 2. «Explicabilidad» y transparencia. 3. Justicia y no discriminación. 4. Seguridad y protección. III. DISTINTOS MODELOS DE REGULACIÓN: 1. Europa. 2. EE. UU. 3. China. IV. ESTADO DE LA CUESTIÓN EN IBEROAMÉRICA. V. ANÁLISIS DE LAS DIFERENCIAS. VI. CONCLUSIONES. *BIBLIOGRAFÍA.*

I. INTRODUCCIÓN

La definición de inteligencia artificial (IA) resulta compleja, entre otras razones, debido a su naturaleza dinámica, sujeta a las variaciones derivadas de los avances tecnológicos. La falta de consenso en una tecnología que tiene un carácter global complica aún más la tarea de conceptualizarla¹. De entre la infinidad de definiciones de IA (Russell y Norvig, 2021)², resulta ilustrativa la que entiende que se trata de la convergencia entre «la ciencia y la ingeniería para crear máquinas inteligentes, especialmente programas informáticos inteligentes. Está relacionada con la tarea similar de utilizar ordenadores para comprender la inteligencia humana, pero la IA no tiene por qué limitarse a métodos que sean biológicamente observables» (Mccarthy, 2007)³.

¹ Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial, y se modifican determinados actos legislativos de la Unión, Bruselas, 21 de abril de 2021, COM(2021) 206 final. Base de la que se ha venido en denominar «Ley IA de la UE», señala: «Los sistemas de IA vendrían definidos por el *software* que se desarrolla, empleando ciertas técnicas o estrategias, 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».

² Diferencia cuatro posibles objetivos o definiciones de la IA que se distinguen por la racionalidad y pensamiento frente a la actuación. En un primer grupo de «enfoque humano», se ubican los «sistemas que piensan como humanos» y los «sistemas que actúan como humanos»; en un segundo grupo, que denominan «enfoque ideal», se incluyen los «sistemas que piensan racionalmente» y los «sistemas que actúan racionalmente».

³ Es habitual tratar de definir la IA sobre la base de la comparación con la inteligencia humana.

Los efectos de la inteligencia artificial (IA) en derechos humanos plantean cuestiones sobre cuáles son los aspectos que deben regularse y por qué, aunque no se trata de un problema nuevo, sino de una muy peculiar manifestación del desarrollo tecnológico en su conjunto (Sánchez Barrilao, 2016). La razón fundamental detrás de esta necesidad de regulación radica en los riesgos inherentes asociados con los sistemas de IA. Sin embargo, no todas las aplicaciones de IA afectan a los derechos de la misma manera, ni con el mismo nivel de riesgo. Por consiguiente, un enfoque clave para abordar esta cuestión es identificar y regular aquellas manifestaciones de IA que se consideren «peligrosas», en función del riesgo que representen. Esta perspectiva se refleja ampliamente en el marco regulatorio europeo y, en menor medida, en el de los EE. UU. Si bien China también está activa en el ámbito de la regulación de la IA, su enfoque varía. En el contexto de Iberoamérica, las decisiones regulatorias pueden inspirarse tanto en los modelos europeos como en los estadounidenses, lo que impacta, directamente, en la configuración de la normativa relacionada con los derechos individuales en el entorno digital.

La complejidad de la IA radica en sus diversas formas, pues la IA no es un solo fenómeno, sino que hay muchas formas de inteligencia artificial, lo que exige tratar la IA desde sus posibles tipologías, de acuerdo con los riesgos asociados. Los sistemas de IA reproducen funciones cognitivas humanas, como razonamiento y resolución de problemas, automatizando actividades vinculadas al pensamiento humano⁴. Existen diferentes niveles de IA, desde la débil o estrecha, centrada en tareas específicas, hasta la fuerte o general, que aspira a replicar la inteligencia humana en todos los aspectos. Todas comparten el uso de redes neuronales, inspiradas en el funcionamiento del cerebro humano, capaces de aprender de grandes cantidades de datos y reglas (Mayer-Schönberger y Cukier, 2019). Dentro de la IA, destacan el aprendizaje automático y el aprendizaje profundo (Mitchell, 1997). El primero extrae conocimientos mediante experiencia supervisada, mientras que el segundo permite que el sistema aprenda y piense de manera autónoma, generando redes neuronales a gran escala⁵.

⁴ Según la Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)).

⁵ Agencia Española de Protección de Datos, *Malentendidos sobre el machine learning (aprendizaje automático)*, disponible en: <https://tinyurl.com/2e9ztmd2>. Además, el uso de sistemas de aprendizaje automático plantea un problema: la incapacidad para discernir las causas subyacentes de las relaciones identificadas. La correlación no implica causalidad, requiriendo un análisis profundo y contextual para establecer vínculos significativos.

La tecnología que produce creciente preocupación en términos de protección de derechos es la conocida como IA generativa. A lo largo de los años, la inteligencia artificial (IA) ha atravesado diversos ciclos de avance, y el lanzamiento de ChatGPT de OpenAI parece ser un hito significativo. En este contexto, la atención se centra en la IA generativa, que actualmente experimenta un salto cualitativo en el procesamiento de lenguajes naturales y supone un cambio de paradigma.

La IA generativa se basa en modelos fundacionales, por lo que marca un cambio fundamental en comparación con la IA tradicional, centrada en el análisis de datos. Estos modelos no solo pueden analizar y expresar lo que observan, sino también crear algo completamente nuevo, utilizando los mismos datos. Constituyen la base de la IA generativa y no se limitan al lenguaje, sino que pueden aprender la gramática de códigos de *software*, moléculas, imágenes naturales y diversos tipos de datos. A diferencia del aprendizaje automático tradicional, que se basa en el aprendizaje supervisado, un modelo fundacional es un tipo de IA «de propósito general» que se entrena «sin supervisión». Para ello, utiliza grandes cantidades de texto o datos provenientes de Internet. Estos modelos, fundamentados en la matemática y las ciencias, tienen la capacidad de crear poesía, redactar códigos y participar en conversaciones sobre diversos temas culturales, lo que permite una capacidad única de generación y adaptación en diversos contextos⁶.

El enorme impacto y efecto transformador de la IA, consecuencia de los avances que se han producido de forma casi repentina, especialmente con la IA generativa, resulta absolutamente innovador y exige considerar la regulación de la IA desde una perspectiva global. Los sistemas de IA no conocen límites geográficos y su impacto se produce a nivel mundial. Las normativas locales pueden resultar insuficientes para abordar las necesidades que plantea la IA en un mundo interconectado, lo que hace que resulte urgente trabajar en la creación de estándares éticos y legales comunes para la protección de los derechos (Balaguer Callejón, 2023).

A continuación, se incluyen algunos de los principios que deben necesariamente respetarse al objeto de crear un marco global de protección de derechos humanos. Tras la exposición de tales principios, junto con las referencias a algunos de los estudios más destacados sobre la cuestión, se incluye un análisis comparativo de los principales modelos de regulación, al objeto de obtener una comprensión general del estado de la cuestión.

⁶ Véase la web de IBM, explicativa de conceptos sobre IA: <https://tinyurl.com/mufcudd4>.

II. DE LA NECESIDAD DE UN MARCO ÉTICO GLOBAL PARA LA PROTECCIÓN DE LOS DERECHOS HUMANOS.

Pese a que sea necesario, la regulación global de la IA resulta compleja debido, además de a las razones aludidas, a la imprevisibilidad de los avances tecnológicos. La anticipación es difícil, y las normas jurídicas surgen una vez que se aplican las tecnologías y los daños a los derechos ya están ocurriendo (Reche Tello *et al.*, 2022)⁷. El principio de neutralidad tecnológica debe guiar a las autoridades reguladoras, asegurando que la regulación sea resistente al avance del tiempo, considerando la rápida evolución tecnológica. Enfrentarse a los problemas implica priorizar el contenido, el fondo y los derechos afectados, más allá de la plataforma o técnica utilizada (García Herrera, 2021: 265-300). El principio de precaución, aplicable como guía ética y regulatoria, exige la toma de medidas preventivas aun cuando no haya evidencia concluyente sobre los riesgos, lo que se traduce en que los desarrolladores, reguladores y usuarios estén obligados a considerar cuidadosamente los posibles efectos negativos de la IA y tomar medidas para mitigarlos.

LA IA precisa de marcos generales reguladores que aseguren un desarrollo y un uso de la IA ético y responsable, puesto que la complejidad de normas aplicables, de distinto orden, relativas a los aspectos tecnológicos y a los que afectan al contenido que desarrolla, que actualmente regulan de forma indirecta algunos usos de la IA dificulta en extremo enfrentarse con éxito a los riesgos que la IA conlleva para la protección de los derechos.

La investigación que aborda la cuestión ética, como base para la determinación de los principios generales de carácter jurídico que deben regir el desarrollo tecnológico de la IA, ha propuesto varios métodos (Hagendorf, 2020). Se plantean reflexiones relativas al modo en que los principios de la ética pueden ser implementados en los sistemas autónomos (Anderson y Anderson, 2014: 1-10). En otros casos, se trata de metaestudios sobre la ética de la IA, por ejemplo, sobre la base de «palabras clave» (Vakkuri y Abrahamsson, 2018: 1-6)⁸. Además, hay una infinidad de propuestas sobre cuáles son los principios que deben presidir el desarrollo de IA⁹, como ejes fundamentales en

⁷ Véase la «Recomendación sobre la ética de la inteligencia artificial», adoptada el 23 de noviembre de 2021 por la UNESCO. Se trata del primer marco normativo universal sobre ética de la IA. Disponible en: <https://tinyurl.com/44k82hfm>.

⁸ Véase el estudio de mapeo sistemático (SMS) basado en palabras clave utilizadas en la IA y la ética para ayudar a identificar, desafiar y comparar los conceptos principales utilizados en el discurso actual sobre la ética de la IA.

⁹ En la UE destacan, de entre los muchos documentos relativos a la ética en IA, la Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre un marco de

torno a los cuales deben crearse los marcos jurídicos generales, que justifiquen posibles limitaciones a las nuevas tecnologías en la protección de los derechos (Jeld *et al.*, 2020).

Son las normas jurídicas las que, finalmente, deben establecer limitaciones y prohibiciones específicas de manifestaciones que sean contrarias a principios y bienes de la máxima relevancia constitucional (Sánchez Barrilao, 2016), sobre la base de disposiciones jurídicas, lo que incluye el necesario sometimiento a un control jurisdiccional para la salvaguarda de los derechos (Presno Linera, 2022).

A continuación, de acuerdo con los trabajos enunciados, se destacan los siguientes principios básicos, que se suman a otros a los que se ha hecho referencia, como el principio de neutralidad o el principio de precaución, que deben estar presentes en el desarrollo de un sistema IA, lo que no implica que sean siempre suficientes. La exposición se estructura en los siguientes bloques: privacidad; explicabilidad y transparencia; justicia y no discriminación; seguridad y protección.

los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas, 2020/2012(INL). Comisión Europea, *Generar confianza en la inteligencia artificial centrada en el ser humano* (COM[2019] 168); Grupo de expertos de alto nivel sobre inteligencia artificial, *Directrices éticas para una IA fiable*, 2019; Grupo de expertos de alto nivel sobre inteligencia artificial, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment (Lista de evaluación para una inteligencia artificial fiable con fines de autoevaluación)*, 2020. La Alianza de la IA es un foro de múltiples interesados que se creó en junio de 2018. Para más información, véase el siguiente enlace: <https://tinyurl.com/45cxcj96>. La OECD ha señalado cuáles deben ser los principios éticos de la IA: la IA debe estar al servicio de las personas y del planeta, impulsando un crecimiento inclusivo, el desarrollo sostenible y el bienestar; 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; 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; los sistemas de IA 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; las organizaciones y las personas que desarrollen, desplieguen o gestionen sistemas de IA deberán responder de su correcto funcionamiento en consonancia con los principios precedentes. En mayo de 2019, cuarenta y dos países adoptan los Principios de la OCDE sobre Inteligencia Artificial. Disponible en: <https://tinyurl.com/z88y96rp>.

1. PRIVACIDAD

Este eje incluye el control humano de la tecnología, más concretamente, también el de la posible revisión humana de decisiones automáticas y el control sobre el uso de los datos, el *habeas data*. Los sistemas de IA deben desarrollarse de forma que protejan la privacidad y que permitan el ejercicio de derechos relacionado con la protección de datos.

En el sistema jurídico europeo, el Reglamento (UE) 2016/679, Reglamento General de Protección de Datos (RGPD)¹⁰, prevé una especial limitación aplicable a los sistemas de IA y referida a la toma de decisiones automatizadas. Así, se reconoce a cualquier persona el derecho a no estar sujeta a decisiones automatizadas que, basadas únicamente en el procesamiento de datos, puedan tener un impacto legal significativo en ellas o afectarles de manera similar (Rebollo Delgado, 2023)¹¹. La garantía se refiere a decisiones importantes o significativas que afecten a derechos y se basen, únicamente, en procesos automatizados, mediante algoritmos o sistemas informáticos, sin intervención humana directa (Palma Ortigosa, 2022).

El tratamiento automatizado de datos también abarca la elaboración de perfiles¹², que es la evaluación de datos personales para analizar o predecir aspectos relacionados con el desempeño laboral, la situación económica, la salud, preferencias personales, fiabilidad, comportamiento, ubicación o movimientos de una persona, cuando estas evaluaciones pueden tener consecuencias legales significativas o un impacto similar¹³.

La creación de perfiles tiene el potencial de perpetuar los estigmas preexistentes y fomentar la fragmentación social. De manera paralela, puede

¹⁰ Reglamento del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.

¹¹ Art. 22 del RGPD.

¹² Debe ser una forma automatizada de tratamiento, incluyen todos aquellos tratamientos que tienen participación parcialmente humana; debe llevarse a cabo respecto a datos personales; y el objetivo de la elaboración de perfiles debe ser evaluar aspectos personales sobre una persona física. Véase el informe de la agencia de Protección de Datos *Adecuación al Reglamento General de Protección de Datos de tratamientos que incorporan inteligencia artificial. Una introducción*. Disponible en: <https://tinyurl.com/muzc4s7b>.

¹³ Los campos del sector bancario y financiero, la asistencia sanitaria, la fiscalidad, los seguros, la mercadotecnia y la publicidad son ejemplos típicos en los cuales se lleva a cabo de manera recurrente la elaboración de perfiles para enriquecer el proceso de toma de decisiones. Está recogido en el art. 4.4 del RGPD.

encajonar a una persona en una etiqueta específica, restringiendo sus opciones a las sugerencias preestablecidas. Este fenómeno puede minar la libertad de elección, limitando las opciones y decisiones sobre productos y servicios. En ocasiones, la creación de perfiles puede conllevar predicciones erróneas, mientras que en otros casos puede dar lugar a la negación de servicios y bienes, perpetrando así una discriminación injustificada.

El RGPD establece disposiciones con el propósito de garantizar que la práctica de elaboración de perfiles y la toma de decisiones automatizadas, ya sea con o sin la elaboración de perfiles, se realicen de manera que no conlleven un impacto injustificado en los derechos. Entre estas disposiciones se incluyen: la imposición de requisitos específicos en cuanto a transparencia y equidad; la asignación de mayores responsabilidades en términos de proactividad; la definición de bases legales específicas para el procesamiento de datos; el otorgamiento de derechos individuales para oponerse a la elaboración de perfiles, en particular con fines de *marketing*; la exigencia de llevar a cabo una evaluación de impacto en la protección de datos cuando se cumplan ciertas condiciones.

2. EXPLICABILIDAD Y TRANSPARENCIA

La explicabilidad y la transparencia son principios obligados en el ámbito de la IA. Tienen que ver con códigos abiertos de datos y algoritmos, notificación cuando se interactúa con IA, notificación cuando la IA toma una decisión acerca de una persona, requerimientos regulares de información y derecho a ser informado, y contratación abierta referida a gobiernos (Jeld *et al.*, 2020).

Se trata de cuestiones cruciales en el uso de la inteligencia artificial (IA) para garantizar el respeto de los derechos humanos. La transparencia se refiere a la capacidad de comprender y comunicar cómo funcionan los sistemas de IA, incluyendo los datos y algoritmos utilizados, las decisiones tomadas y su impacto en la ciudadanía. Por otro lado, la explicabilidad implica hacer inteligibles los resultados de los sistemas de IA, proporcionando información sobre su funcionamiento y contribución a los resultados. Ambos principios son esenciales para garantizar, también, el derecho a la tutela judicial efectiva contenido en el art. 24 de la Constitución (Cotino Hueso y Castellanos Claramunt, 2022). No solo se refiere a que las personas tienen derecho a acceder a la justicia, a obtener una resolución fundada en derecho, a ser oídas y a defender sus intereses legítimos, sino a que este derecho implica que las personas puedan estar plenamente informadas cuando una decisión se basa en algoritmos de IA o se toma a partir de ellos, en particular cuando afecta a su seguridad o a sus derechos. Dado su carácter instrumental, la transparencia no

es un concepto estático y su contenido evoluciona según la finalidad y el contexto en el que se aplique (Cotino Hueso y Castellanos Claramunt, 2022).

La transparencia y la explicabilidad exigen saber no solo qué datos procesa el sistema, sino también cómo los valora (Presno Linera, 2022). Ambas están estrechamente relacionadas con la responsabilidad y la rendición de cuentas. La responsabilidad implica que los actores de la IA asuman las consecuencias de sus acciones y decisiones, y puedan ser sancionados o reparados si causan daño o vulneran derechos. La rendición de cuentas implica que los actores de la IA puedan demostrar y justificar que sus acciones y decisiones son conformes con las normas y principios aplicables, y estén sometidos a controles (Cotino Hueso y Castellanos Claramunt, 2022).

En el ámbito práctico, la falta de transparencia plantea importantes cuestiones sobre la confianza, la verificación y la impugnación de decisiones tomadas por sistemas de IA. La verificación permite comprobar y validar que los sistemas de IA cumplen con los requisitos y estándares establecidos. La impugnación permite cuestionar y oponerse a decisiones tomadas por sistemas de IA si se consideran injustas o perjudiciales.

En definitiva, transparencia y la explicabilidad son elementos fundamentales para garantizar el respeto de los derechos en el uso de la IA (Pauner Chulvi, 2023), además de constituirse en derechos en sí mismos (Arellano Toledo, 2019). Estos principios no solo son importantes en el ámbito teórico y jurídico, sino también en la práctica, donde contribuyen a la confianza, la responsabilidad y la rendición de cuentas en el desarrollo y aplicación de sistemas de IA.

3. JUSTICIA Y NO DISCRIMINACIÓN

Este eje alude al tratamiento equitativo de personas, o grupos de personas, por parte de un sistema de IA. Se plantea la necesidad de controlar los sistemas para que la IA ayude a la toma de decisiones más justas, a contrarrestar los sesgos humanos y a promover la inclusión (Jeld *et al.*, 2020).

El uso de algoritmos conlleva un riesgo relacionado con la generación de sesgos, los cuales pueden reforzar prejuicios existentes o históricos, sin ser siempre evidentes en las aplicaciones (Belloso Martín, 2022). Estos sesgos pueden surgir por factores de distinta índole y tienen como efecto el perjuicio a minorías o grupos poco representados en los datos utilizados para el aprendizaje computacional (Mehrabi *et al.*, 2022). El resultado es la reproducción de situaciones que los sistemas IA van a perpetuar (Presno Linera, 2022). La incapacidad de los algoritmos para contrarrestar conscientemente los sesgos incorporados por sus desarrolladores presenta un riesgo de discriminación, contrario al principio de igualdad reconocido en el art. 14 de la Constitución.

La discriminación puede manifestarse de formas distintas, afectando a personas o grupos de manera diferenciada y desfavorable (Serra Cristóbal, 2020). Los sesgos no son meros errores técnicos, sino que tienen implicaciones significativas al perpetuar dinámicas de dominación y discriminación, distorsionando análisis y decisiones. Estos sesgos pueden originarse en la selección de datos que no representan fielmente la realidad, como sesgo de muestreo, o al reflejar prejuicios existentes en los datos de entrenamiento del algoritmo (Mehrabi *et al.*, 2022).

La subrepresentación de ciertos grupos en conjuntos de datos evidencia un desequilibrio y distorsiona las conclusiones basadas en ellos. Resulta necesario tener en cuenta esos desequilibrios para evitar sesgos discriminatorios y garantizar una representación adecuada de diversos grupos. La inclusión de diversidad geográfica en conjuntos de datos ampliamente utilizados, como ImageNet y Open Images, puede mitigar sesgos de representación. En el ámbito del procesamiento del lenguaje natural (NLP, por sus siglas en inglés) se identifican sesgos en bases de conocimiento utilizadas en diversas aplicaciones (Mehrabi *et al.*, 2022).

Se han planteado varios casos que ilustran esta situación. Así, el programa SyRI, un sistema implementado en los Países Bajos bajo la Ley de Organización de Implementación y Estructura de Ingresos (SUWI), diseñado para evaluar la probabilidad de utilización ilegal de fondos gubernamentales de seguridad social, dio lugar a la sentencia del Tribunal de Distrito de La Haya el 5 de febrero de 2020. Esta puso de manifiesto ciertas deficiencias en el funcionamiento de SyRI. Se hizo hincapié en la falta de proporcionalidad y transparencia en su aplicación, así como en la vulneración del derecho a la privacidad consagrado en el Convenio Europeo de Derechos Humanos. Se evidenciaron los problemas que conllevan la dependencia de algoritmos para el análisis de datos y la evaluación del riesgo, sin la divulgación adecuada de su funcionamiento interno (Belloso Martín, 2022). En un sentido similar, en Italia, el Tribunal Ordinario de Bolonia emitió una sentencia el 31 de diciembre de 2020, relacionada con la empresa Deliveroo. En este caso, el algoritmo utilizado para gestionar la asignación de pedidos a los trabajadores, denominado «Frank», reconoció la existencia de discriminación al no tener en cuenta determinadas circunstancias relacionadas con las ausencias justificadas de las personas empleadas¹⁴.

¹⁴ En España la Ley 12/2021, de 28 de septiembre, conocida como la ley de *riders*, en el art. 64.4, señala que el comité de empresa, con la periodicidad que proceda en cada caso, tendrá derecho a «ser informado por la empresa de los parámetros, reglas e instrucciones en los que se basan los algoritmos o sistemas de inteligencia artificial que

En definitiva, la presencia de sesgos en algoritmos representa un riesgo evidente para los derechos humanos, ya que pueden perpetuar y amplificar desigualdades existentes, contraviniendo principios fundamentales de igualdad y justicia. Además, los efectos pueden agravarse y afectar a mayor número de personas por el uso de sistemas de IA (Presno Linera, 2022). Es esencial tener presente la posibilidad de que existan estos sesgos para garantizar que los sistemas de IA sean justos y equitativos en su funcionamiento y aplicación en cada una de las fases de desarrollo del sistema (Belloso Martín, 2022).

Finalmente, cabe destacar otro aspecto crítico relacionado con la inteligencia artificial: el sesgo en la interpretación de sus resultados. Este fenómeno se manifiesta al aceptar pasivamente los resultados generados por los sistemas de IA, sin someterlos a un análisis crítico sobre su validez, de forma que se otorga una autoridad indiscutible a la solución propuesta por el algoritmo, basada únicamente en la confianza depositada en el sistema. En otras palabras, este sesgo implica una aceptación acrítica de los resultados de la inteligencia artificial, lo que puede llevar a decisiones erróneas o injustas al estar influenciadas por prejuicios o suposiciones no fundamentadas. Reconocer la posibilidad de sesgos en la interpretación de los resultados de la inteligencia artificial implica la necesidad de adoptar un enfoque crítico y reflexivo, con el objetivo de mitigar los posibles impactos negativos y asegurar una aplicación justa y equitativa del derecho en el contexto de la inteligencia artificial¹⁵.

4. SEGURIDAD Y PROTECCIÓN

La seguridad y la protección en el ámbito de la inteligencia artificial (IA) se relacionan con la confianza, predictibilidad y seguridad del diseño (Jeld *et al.*, 2020). Se recogen entre los principios generales de carácter jurídico, por su carácter instrumental para la protección de los derechos frente al desarrollo tecnológico. Al igual que la explicabilidad y transparencia, la seguridad y protección buscan asegurar la fiabilidad y equidad en el diseño y funcionamiento de los sistemas de IA.

Estos principios instrumentales se complementan entre sí para garantizar una toma de decisiones objetiva y equitativa, reduciendo al mínimo los riesgos de discriminación, sesgo o falta de rendición de cuentas en el contexto de la

afectan a la toma de decisiones que pueden incidir en las condiciones de trabajo, el acceso y mantenimiento del empleo, incluida la elaboración de perfiles».

¹⁵ Véase el documento de la Agencia Española de Protección de Datos. Disponible en: <https://tinyurl.com/muzc4s7b>.

aplicación de la ley. En el ámbito judicial, es obligado un control humano adecuado sobre la IA, que garantice el respeto a los principios y garantías procesales, así como a los derechos y libertades de las personas (Barona Villar, 2021), ya que contribuye a reducir al mínimo el riesgo de adoptar decisiones asistidas por IA erróneas o sesgadas de aplicación de la ley por el poder judicial. Asimismo, la seguridad de los sistemas de inteligencia artificial (IA) es un principio fundamental para mitigar los riesgos asociados a la censura algorítmica y proteger los derechos humanos, referidos al ámbito de las libertades de comunicación. La censura algorítmica, fenómeno por el cual los sistemas de IA restringen o manipulan la información en el entorno digital, plantea serios riesgos en este sentido, con causas y consecuencias diversas, intrínsecamente ligadas al diseño, funcionamiento y uso de la IA (Rebollo, 2023). En este contexto, el derecho de sufragio debe ser objeto de especial preocupación ante la IA (González de la Garza, 2019). La preservación de este derecho frente a las amenazas digitales, como ataques informáticos y desinformación, es esencial para garantizar la integridad de las elecciones y la confianza en los representantes elegidos. Por ello, resulta obligado implementar medidas efectivas para contrarrestar estos riesgos, así como una mayor confianza y aceptación de estas tecnologías en la sociedad.

En definitiva, la interconexión entre los principios de explicabilidad y transparencia en el uso de la inteligencia artificial se entrelaza estrechamente con la necesidad de garantizar la seguridad y protección de los derechos en el ámbito digital. Al promover la transparencia y la explicabilidad en el desarrollo y aplicación de los sistemas de IA, se fortalece la protección de los derechos, la privacidad, la igualdad, la tutela judicial efectiva, así como los derechos de carácter político, referidos a las libertades de comunicación o el derecho al sufragio.

La referencia a los principios expuestos pretende servir como guía básica, que debería formar parte de cualquier modelo de regulación de la IA. A continuación, se recogen los aspectos más destacados de los distintos modelos de regulación, desde una perspectiva comparada. Se pretende, con ello, ofrecer una visión general de la forma en que se hace frente a la posible vulneración de derechos como consecuencia del desarrollo tecnológico que ha supuesto la IA.

III. DISTINTOS MODELOS DE REGULACIÓN

Los enfoques de regulación de la IA en Estados Unidos, China y Europa difieren en sus prioridades y métodos, aunque todos persiguen el objetivo común de encontrar un equilibrio entre la promoción de la innovación tecnológica y la protección de los derechos humanos y los valores fundamentales.

La forma en que se plantea ese equilibrio, que se traduce en cómo se lleva a cabo la regulación de la IA, no solo moldea el desarrollo y la aplicación de la tecnología, sino que también refleja los valores y las prioridades de la sociedad de que se trate en su conjunto¹⁶.

1. LA REGULACIÓN EN EUROPA

En abril de 2021, la Comisión Europea presentó una propuesta para establecer un marco normativo de la Unión Europea sobre inteligencia artificial (IA). Esta iniciativa, conocida como el proyecto de ley de IA, tiene como objetivo principal regular de manera horizontal el campo de la IA. Resulta de un proceso de consulta iniciado en el año 2018, que ha venido trabajando con diversos actores para establecer un marco normativo sobre la base de los principios éticos que se han señalado. El régimen legal propuesto se centra en la utilización específica de los sistemas de IA y los riesgos asociados con el objetivo de lograr una actuación coordinada respecto de las implicaciones éticas y humanas de la IA, desde un enfoque humanista y centrado en la persona, especialmente en lo que respecta a la protección de los derechos. Este enfoque es fundamental frente a los riesgos de que se produzcan arbitrariedades, sesgos, discriminaciones y otras vulneraciones de los derechos (Cotino Hueso, 2021).

La posición del regulador europeo se basa en los riesgos, lo que resulta acertado (Cotino Hueso, 2021), y se hace depender del uso de la IA, de forma que tiene en cuenta la probabilidad de daños derivados, especialmente, de la inteligencia artificial generativa (Barrio Andrés, 2022). La UE pretende, con esta regulación, situarse como un líder mundial en el desarrollo de sistemas IA seguros y éticos¹⁷. En diciembre de 2021, el Consejo de la UE acordó la posición general de los Estados miembros sobre la propuesta. Posteriormente, el 14 de junio de 2023, el Parlamento Europeo emitió su posición al respecto. El 8 de diciembre del mismo año se firmó un acuerdo entre el Parlamento Europeo y el Consejo sobre la mal llamada ley de inteligencia artificial, que ha marcado un hito significativo en la regulación de esta tecnología a nivel mundial. Este acuerdo conlleva posibles implicaciones globales. Es probable que otras regiones consideren estas normativas como referencia al desarrollar

¹⁶ A efectos de conocer los distintos proyectos que se debaten en materia de IA, véase el informe de la Universidad de Stanford sobre el tema. Disponible en: <https://aiindex.stanford.edu/report/>.

¹⁷ Comisión Europea (2021). «Generar confianza mediante el primer marco jurídico sobre la IA». En *Excelencia y confianza en la inteligencia artificial*, <https://tinyurl.com/4kz8kyw9>.

sus propias leyes. Lo que podría contribuir a la formación de estándares globales para el desarrollo y la implementación de la inteligencia artificial¹⁸.

En general, el acuerdo representa un paso significativo hacia la creación de un marco regulatorio sólido para la inteligencia artificial y refleja las preocupaciones y dificultades asociadas con su rápida evolución¹⁹. Ciertas prácticas de IA que amenazan la seguridad, los medios de vida y los derechos de las personas implican un nivel de riesgo inaceptable y están prohibidas. El registro, la evaluación de riesgos, las sanciones, la supervisión humana y otras obligaciones se imponen a las denominadas IA de alto riesgo, mientras que las IA de riesgo limitado estarían sujetas a las obligaciones de transparencia propuestas.

El acuerdo se estructura en cuatro categorías de riesgo: inaceptable, elevado, limitado y mínimo. En concreto, se acordó prohibir: a) los «sistemas de categorización biométrica que utilicen características sensibles» (por ejemplo, creencias políticas, orientación sexual o raza); b) la «extracción no dirigida de imágenes faciales» de Internet o grabaciones de circuito cerrado de televisión (CCTV) para crear bases de datos de reconocimiento facial; c) los «sistemas de reconocimiento de emociones» en el lugar de trabajo y en instituciones educativas; d) los «sistemas de crédito social» basados en el comportamiento social o características personales; y e) los «sistemas que manipulen el comportamiento humano» para quebrantar su libre albedrío o que exploten las vulnerabilidades de las personas (por su edad, discapacidad o situación socioeconómica)²⁰.

¹⁸ Sobre su contenido, véase: <https://tinyurl.com/45xsfe4s>; véanse los trabajos del Comité de Ministros del Consejo de Europa (2021). *Recommendation CM/Rec(2021)8 of the Committee of Ministers to member States on the protection of individuals with regard to automatic processing of personal data in the context of profiling* (Adopted by the Committee of Ministers on 3 November 2021 at the 1416 meeting of the Ministers' Deputies). Disponible en: rm.coe.int/0900001680a46147; Comité sobre Inteligencia Artificial del Consejo de Europa (2023). *Consolidated working draft of the framework convention on artificial intelligence, human rights, democracy and the rule of law*. Disponible en: <https://tinyurl.com/42ft39jw>.

¹⁹ «The EU AI Act Newsletter #42: Provisional Agreement Reached», 29/11/23-11/12/23, <https://tinyurl.com/3chv7a73>.

²⁰ De forma excepcional, se permite, en cambio, solo en determinados supuestos el uso por las fuerzas de seguridad de sistemas biométricos en lugares públicos, siempre que se cuente con autorización judicial. Los supuestos, no obstante, que permiten el uso de este tipo de tecnología son los que tienen por finalidad localizar a la víctima de un delito muy grave, prevenir delitos de terrorismo o perseguir crímenes también muy graves. Este es uno de los temas más complicados, que ha exigido necesariamente la toma de la decisión

Junto con los sistemas prohibidos, se relacionan otros considerados de riesgo mínimo y los de alto riesgo. Los primeros únicamente quedan sometidos a obligaciones básicas de transparencia, en particular, la necesidad de garantizar la accesibilidad a su documentación técnica. Con carácter voluntario, eso sí, las empresas podrán someterse a códigos de conducta adicionales para estos sistemas de IA²¹.

Los sistemas de IA clasificados como de alto riesgo forman parte de este grupo, porque conllevan un posible daño a la salud, la seguridad, los derechos fundamentales, el medio ambiente, la democracia y el Estado de derecho. Se incluyen también los que se usen para influir en resultados electorales y la conducta de los votantes, que quedan sujetos, adicionalmente, a una serie de medidas. En primer lugar, la evaluación y mitigación de riesgos, entre las cuales exige una evaluación de impacto sobre derechos fundamentales. Además, exige las garantías de alta calidad de los datos empleados; una medida que permite el control son los registros de actividad. Exige, igualmente, medidas apropiadas de supervisión humana. Y, por último, la información y control ciudadano, en forma de explicaciones sobre las decisiones basadas en este tipo de sistemas que afecten a sus derechos, y derecho a presentar quejas sobre este tipo de sistemas²².

Respecto a la IA generativa, estos sistemas y los modelos sobre los que se basan (por ejemplo, los GPT de OpenAI) son asimismo objeto de regulación. Todos ellos deberán cumplir requisitos de transparencia que incluyen la elaboración de documentación técnica y la difusión de resúmenes detallados sobre el contenido digital utilizado en su entrenamiento, el cual, a su vez, habrá de respetar la normativa europea sobre derechos de autor²³.

Finalmente, para aquellos clasificados como «de alto impacto» por su potencial riesgo sistémico (en la misma línea de los citados para los sistemas de alto riesgo), y siempre que reúnan ciertos criterios, se disponen obligaciones más estrictas. Así, además de la obligación común con los sistemas de riesgo alto de evaluar y mitigar tales riesgos sistémicos, deberán realizar pruebas contradictorias (el conocido como *red teaming*), están obligados a informar a la Comisión sobre incidentes graves, así como sobre su eficiencia energética, y deben garantizar la ciberseguridad²⁴.

mediante consenso. *The EU AI Act Newsletter* #42: «Provisional Agreement Reached», 29/11/23-11/12/23, disponible en: <https://tinyurl.com/3chv7a73>.

²¹ *Ibidem*.

²² *Ibidem*.

²³ *Ibidem*.

²⁴ *Ibidem*.

El acuerdo explica con claridad el estado actual del proceso legislativo y los plazos previstos para la entrada en vigor de la ley, que deberá someterse a aprobación final por el Parlamento Europeo y por el Consejo. La totalidad del futuro reglamento no será aplicable hasta dos años después, si bien algunas de sus disposiciones, como las relativas a las prohibiciones y a la IA generativa, serán aplicables al cabo de 6 y 12 meses desde la entrada en vigor, respectivamente. Se da a la Comisión Europea un papel clave para la implementación de la ley, ya que le permite adoptar actos delegados y de ejecución para especificar los requisitos técnicos y metodológicos de los sistemas de IA, así como para actualizar la lista de usos prohibidos o de alto riesgo. Esto implica que la Comisión tendrá que estar al día de los avances científicos y tecnológicos en el campo de la IA y consultar a las partes interesadas, como los proveedores, los usuarios, las autoridades nacionales y los expertos independientes²⁵. El acuerdo también implica que los Estados miembros tendrán que adaptar sus legislaciones nacionales para garantizar la aplicación efectiva y coherente de la ley, especialmente en lo que se refiere a las infracciones y sanciones²⁶.

2. LA REGULACIÓN EN EE. UU.

La referencia actual en EE. UU. a esta cuestión es de octubre de 2023, mediante una orden ejecutiva del presidente americano²⁷. La orden establece una serie de principios y prácticas para el uso responsable y confiable de la IA por parte de las agencias federales, como la transparencia, la equidad, la rendición de cuentas, la seguridad y la privacidad. La orden también crea un comité asesor de expertos del sector, la academia y la sociedad civil para supervisar la implementación de la regulación y proponer recomendaciones para mejorarla. La orden no se aplica a las empresas privadas ni a los usos militares de la IA.

Además de la acción de las propias empresas tecnológicas —como el plan de Microsoft de mayo de 2023, «Governing AI: A Blueprint for the Future»²⁸, en el que se plantean cuestiones relacionadas, entre otras, con la seguridad y transparencia—, destaca, como antecedente, la Ley de Iniciativa Nacional de Inteligencia Artificial de 2020, que coordina un plan federal de investigación y ejecución de la IA. Su meta es asegurar la continua primacía de EE. UU. en la investigación y creación de la IA²⁹.

²⁵ *Ibidem*.

²⁶ Véase la nota «Alerta: Ley europea de Inteligencia Artificial», disponible en: <https://tinyurl.com/yu6nbsfm>.

²⁷ Consultar hoja informativa en: <https://tinyurl.com/2wsfey69>.

²⁸ Véase contenido del plan, disponible en: <https://tinyurl.com/4ku7xr49>.

²⁹ Sobre el contenido de la ley, véase: <https://tinyurl.com/mvz7arzd>.

Por último, el pasado 3 de febrero de 2022 se presentó en la Cámara de Representantes de Estados Unidos el proyecto de Ley de Responsabilidad Algorítmica de 2022³⁰, centrado en la transparencia, la rendición de cuentas y la equidad de las decisiones automatizadas. El proyecto exige realizar evaluaciones de impacto en cuanto a sesgo, eficacia y otros factores.

En Estados Unidos, la regulación de la IA se ha caracterizado por un enfoque más flexible y descentralizado, ya que la mención a ciertos valores no va acompañada de una regulación clara. Las pautas y directrices de las agencias gubernamentales suelen tener un carácter voluntario, sin que puedan considerarse regulaciones en sentido estricto, lo que refleja la valoración de la innovación y la libertad empresarial por encima de la protección de derechos. El resultado de esa flexibilidad es la inexistencia de una regulación uniforme de estándares éticos y de seguridad en todo el país (Hernández Peña, 2022)³¹.

3. LA REGULACIÓN EN CHINA

China ha optado por un enfoque más centralizado y proactivo en la regulación de la IA. El Gobierno chino emite políticas y directrices específicas para supervisar el desarrollo y la implementación de la IA. Prioriza, a estos efectos, la seguridad y la estabilidad social. Esta centralización permite un control más estricto sobre el desarrollo tecnológico y su impacto en la sociedad, pero también conlleva posibles riesgos sobre la libertad y la privacidad³².

La primera regulación específica de IA de la República Popular China fue el Reglamento de Recomendación de Algoritmos, que entró en vigor el 1 de marzo de 2022. Regula el uso de tecnologías de recomendación de algoritmos para proporcionar servicios en línea. En junio de 2023 publicó una serie de normas mediante el denominado Reglamento de Síntesis Profunda³³, que se aplica a los servicios disponibles para el público en general en China y exige que los proveedores de servicios de IA realicen revisiones de seguridad y registren sus algoritmos ante el Gobierno si sus servicios pueden influir en la

³⁰ Véase las bases del proyecto, disponibles en: <https://tinyurl.com/5n7rxuzp>.

³¹ Véase «El presidente Biden emite una orden ejecutiva para desarrollar en los EE. UU. una inteligencia artificial segura y confiable», *Diario La Ley*, disponible en: <https://tinyurl.com/2y53z2dh>.

³² La fuente de información sobre la actual situación de la IA en China: <https://tinyurl.com/29fbu9tt>.

³³ El término «tecnologías de síntesis», tal como se define en el Reglamento de Síntesis Profunda, se refiere a «tecnologías que utilizan algoritmos generativos y sintéticos, como el aprendizaje profundo y la realidad virtual, para generar texto, imagen, audio, vídeo, escenas virtuales y otra información de Internet».

opinión pública o pueden «movilizar» al público. Además, las normas obligan a revelar los contenidos generados artificialmente, así como los datos utilizados para entrenar cualquier modelo lingüístico de gran tamaño³⁴. El Reglamento de Síntesis Profunda³⁵ impone obligaciones más amplias que el código de la UE³⁶ y se aplica a todos los participantes involucrados en servicios de «síntesis profunda», incluidos los «proveedores de servicios», «técnicos de soporte», «usuarios» y «plataformas en línea». Proporciona un marco integral de obligaciones desde varias perspectivas, incluidas «la ciberseguridad, la gestión de datos y la información personal, protección, auditoría y archivo de algoritmos, verificación de nombre real y moderación de contenido».

También en 2023 se ha promulgado el Proyecto de Revisión Ética Medida, que se aplicaría a cualquier actividad científica y tecnológica que involucre a seres humanos, animales de laboratorio o que, de otro, modo implique riesgos éticos, e impone obligaciones de revisión ética a cualquier «universidad», «institución de investigación científica», «institución médica y de salud» y «empresa» que participe en las actividades científicas y tecnológicas relacionadas³⁷.

Finalmente, es preciso aludir también a las Disposiciones Provisionales sobre Gestión de Servicios de Inteligencia Artificial Generativa por las que se aprueba el Reglamento de IA Generativa, publicado el 13 de julio de 2023, que entró en vigor el 15 de agosto de 2023. Tiene por objeto regular «modelos y tecnologías con capacidad de generar textos, imágenes, sonidos, vídeos y otros contenidos», por lo que el término a primera vista es más amplio que «tecnologías de síntesis», aunque reduce su ámbito subjetivo de obligaciones, ya que obliga principalmente a los «proveedores de servicios».

Los aspectos de mayor interés relativos al régimen de derechos en China afectan a los usuarios de sistemas IA. En concreto, al poder de controlar su perfil. Lo que va a requerir que los desarrolladores estén obligados a crear una interfaz donde puedan ver sus perfiles y seleccionar y eliminar activamente las

³⁴ Véase «China aprueba una regulación de la inteligencia artificial y de la inteligencia artificial generativa», *Diario La Ley*, disponible en: <https://tinyurl.com/y24mnmew>.

³⁵ Sobre el contenido de la regulación del reglamento, véase: <https://tinyurl.com/4zrrxjyn>.

³⁶ El código de la UE se centra en regular la información errónea y la desinformación, con otras áreas que se contemplan en el Reglamento General de Protección de Datos —protección de la información personal—, la Ley de Servicios Digitales modificada —transparencia, obligaciones y ciertas protecciones de usuario aplicables a servicios y plataformas en línea— y la ley de IA de la UE —regulación integral de los sistemas de IA—. Disponible en: <https://tinyurl.com/mr4btuej>.

³⁷ Publicada el 14 de abril de 2023, para consulta pública que cerró el 3 de mayo de 2023.

palabras clave utilizadas para la recomendación. Respecto de las obligaciones de los servicios que utilicen algoritmos, estos deben informar a los usuarios de las circunstancias en las que utilizan el algoritmo de recomendación y publicar los principios básicos, intenciones y mecanismos operativos del servicio de recomendación de algoritmo. Además, se incluyen requisitos de moderación de contenidos y se responsabiliza a las plataformas de cualquier contenido ilegal o indeseable que se recomiende a los usuarios mediante algoritmos. Se trata de una extensión de estipulaciones similares de la ley de ciberseguridad de China que responsabiliza a las plataformas por albergar contenido ilegal o indeseable.

Esta regulación dispuesta específicamente para la IA tiene que complementarse con las otras leyes existentes en la República Popular China, como las de protección de datos, ciberseguridad, competencia desleal y comercio electrónico, dado que todas ellas pueden aplicarse al uso de la IA. En el futuro, es previsible que el Reglamento de Recomendación de Algoritmos, el Reglamento de Síntesis Profunda, el Proyecto de Medida de Revisión Ética, una vez que entre en vigor, y el Reglamento de IA Generativa, todos ellos conocidos como «reglamentos de IA» —a la espera de la formulación de una ley integral de IA—, sean las principales leyes que regulen los servicios y productos relacionados con la IA en la República Popular China, incluida la IA generativa y el contenido generado por IA.

IV. ESTADO DE LA CUESTIÓN EN IBEROAMÉRICA

La Red Iberoamericana de Protección de Datos (RIPD), foro regional constituido por dieciséis autoridades de protección de datos de doce países de la región, elaboró un documento en 2019 con el objetivo de establecer unos principios y buenas prácticas para garantizar el respeto a los derechos humanos en el desarrollo y el uso de la inteligencia artificial. Su aproximación a la ética de la IA es, básicamente, desde la protección de los datos, para lo que el documento se basa en los estándares internacionales de protección de datos, así como en las normativas nacionales de los países miembros de la red³⁸.

El documento hace referencia, también, a aspectos como la transparencia, la explicabilidad, la seguridad, la privacidad, la equidad, la inclusión, la sostenibilidad, la diversidad, la participación, la rendición de cuentas y el bien común. Además, las recomendaciones específicas se refieren a cuestiones relacionadas con el régimen de autodeterminación informativa, e incluyen el

³⁸ Véanse recomendaciones de la RIPD. Disponible en: <https://tinyurl.com/mrysw2d9>.

consentimiento, la minimización de datos, la calidad de datos, el derecho de acceso, el derecho de rectificación, el derecho de supresión, el derecho de oposición, el derecho a la portabilidad, el derecho a la limitación del tratamiento, el derecho a no ser objeto de decisiones automatizadas, el derecho a la información y el derecho a la tutela judicial efectiva.

La red se ha preocupado también por el efecto que ChatGPT puede tener en los derechos de protección de datos. Se hace constar que esta tecnología puede conllevar riesgos para los derechos y libertades de los usuarios relativos al tratamiento de sus datos personales, que abarcan aspectos tales como los fundamentos legales para dichos tratamientos, la información que sobre los tratamientos se brinda al usuario, el ejercicio de los derechos reconocidos en las normativas de protección de datos, las posibles transferencias de datos personales a terceros sin contar con el consentimiento de los titulares, el no contar con medidas de control de edad para impedir que menores accedan a su tecnología, así como no saber si cuenta con adecuadas medidas de seguridad para la protección y confidencialidad de los datos personales recabados³⁹.

Iberoamérica presenta una diversidad de iniciativas legislativas y regulatorias sobre la IA que reflejan las diferentes realidades y prioridades de cada país. Hay una serie de países que se han posicionado como líderes regionales en el ámbito de la IA ética. Destacan Argentina, Brasil, Chile, Colombia, México y Uruguay.

En Argentina, el Plan Nacional de Inteligencia se lanzó en 2020 bajo el nombre de ARGENIA⁴⁰ con el propósito de catalizar el desarrollo y la aplicación de esta tecnología en el país. Se sigue un sistema colaborativo con la participación de actores comprometidos en el campo de la inteligencia artificial. Se estructura en cinco ejes estratégicos, que incluyen investigación y desarrollo, talento humano, transformación productiva, gobierno abierto y ética y derechos humanos. Este último eje propone la integración de la ética y los derechos humanos desde las fases iniciales del diseño de sistemas inteligentes. El plan se enfocó en establecer prácticas sólidas y mecanismos de control para asegurar la fiabilidad, transparencia, inclusividad y responsabi-

³⁹ Sobre las referencias de la RIPD a IA, véase: <https://tinyurl.com/mrjd775v>.

⁴⁰ El plan formó parte integral de la Agenda Digital Argentina 2030 y la Estrategia Nacional de Ciencia, Tecnología e Innovación Argentina Innovadora 2030, buscando fortalecer la innovación y la competitividad digital del país. La coordinación de ARGENIA estuvo a cargo de la Secretaría de Ciencia, Tecnología e Innovación Productiva, con la participación activa de diversos ministerios y organismos públicos y privados.

lidad de la inteligencia artificial. Incluye un enfoque ético que resulta innovador y diferencial en el panorama de la inteligencia artificial en Iberoamérica (García Benítez y Ruvalcaba-Gómez, 2021). Argentina ha estado considerando la regulación de la IA de la forma descrita, aunque, hasta el momento, no se ha implementado una legislación específica y las discusiones sobre este tema siguen en curso.

Brasil diseñó su estrategia de inteligencia artificial en varias etapas, en las que se ha incluido consulta pública en 2019 sobre el marco legal de inteligencia artificial, con la participación de la sociedad civil y expertos en la formulación de políticas relacionadas con la IA. El objetivo principal del plan es fomentar la investigación, la innovación y el desarrollo ético en inteligencia artificial, promoviendo la participación de diversos sectores sociales. Se estructura en nueve ejes temáticos, en cada uno de los cuales se proponen acciones estratégicas específicas para su desarrollo. Uno de ellos está dirigido a promover la regulación y el uso ético de la inteligencia artificial. Su propósito es contribuir al establecimiento de principios éticos, proteger los datos personales, prevenir la discriminación, respetar los derechos humanos y asegurar la responsabilidad de los actores involucrados. El plan enfatiza la importancia de la transparencia, la diversidad y la cooperación en el ámbito de la inteligencia artificial. Estos elementos son considerados esenciales para el éxito y la responsabilidad en el desarrollo y aplicación de esta tecnología en el contexto brasileño (García Benítez y Ruvalcaba-Gómez, 2021)⁴¹.

Chile desarrolló su política de inteligencia artificial a lo largo de tres fases, también con la colaboración activa de la ciudadanía y del Ministerio de Ciencia, Tecnología, Conocimiento e Innovación⁴². El objetivo central de esta política es potenciar la participación de las personas en el uso y desarrollo de herramientas de inteligencia artificial, al mismo tiempo que estimula debates en torno a sus implicaciones legales, éticas, sociales y económicas. La política consta de varios ejes en torno a cuatro principios transversales referidos a la IA centrada en el bienestar de las personas, IA para el desarrollo sostenible, IA inclusiva e IA globalizada; la política aborda de manera integral los aspectos clave de la implementación ética de la inteligencia artificial. Uno de los ejes específicos se enfoca en la ética y los derechos humanos, con el propósito de

⁴¹ La Estrategia Brasileña de Inteligencia Artificial (EBIA) se construyó en un proceso de tres etapas durante los años 2019 y 2020.

⁴² Ver texto del Plan Nacional, disponible en: <https://tinyurl.com/bdvkar2f>, así como el informe *Política nacional de inteligencia artificial* del Ministerio de Ciencia y Tecnología, disponible en: <https://tinyurl.com/bdf5fzey>. También el informe de la UNESCO *Artificial Intelligence Readiness Assessment Report*, disponible en: <https://tinyurl.com/y8kc8v5y>.

desarrollar una agenda completa que abarque investigación, participación ciudadana, transparencia y regulación en estos aspectos (García Benítez y Ruvalcaba-Gómez, 2021). Hay un proyecto de ley chileno que sigue el enfoque de riesgos de los sistemas de IA del reglamento europeo, basado, igualmente, en el sistema de riesgo, al tiempo que crea una Comisión Nacional de Inteligencia Artificial y establece obligaciones para desarrolladores, proveedores y usuarios de sistemas de IA⁴³.

Colombia se ha destacado como un líder regional en la transformación digital e IA (Sánchez Acevedo, 2022). Ha sido uno de los países pioneros en implementar las Recomendaciones sobre la Ética de la Inteligencia Artificial de la UNESCO⁴⁴. La aproximación a la regulación de la IA se desarrolla con un enfoque transversal y multidisciplinario que contó con la participación de diversos actores sociales e institucionales⁴⁵. Tiene como objetivo establecer un marco ético que oriente el diseño, desarrollo, implementación y evaluación de sistemas de IA, asegurando el respeto de los derechos (García Benítez y Ruvalcaba-Gómez, 2021)⁴⁶. El Proyecto de Ley de Inteligencia Artificial de Colombia, con influencia europea, que se está debatiendo actualmente, pretende establecer los principios, derechos, deberes y garantías para el uso de la IA sobre la base de búsqueda de privacidad, seguridad, transparencia, control de la discriminación y responsabilidad, según un sistema basado en el riesgo, de manera parecida a cómo lo hace el reglamento europeo⁴⁷.

México lideró la región como el primer país de América Latina en iniciar la elaboración de una estrategia de inteligencia artificial (IA) en 2018, pero, a pesar de este impulso inicial, con el cambio de Gobierno en ese mismo año no se dio seguimiento a la creación de una estrategia nacional de IA, dejándola inconclusa⁴⁸. Aun así, se han discutido propuestas para abordar los problemas éticos y legales relacionados con la IA. Con la creación de una agenda nacional que tiene como objetivo capitalizar la IA para reducir la desigualdad y abordar

⁴³ Sobre el contenido del proyecto de ley, véase: <https://tinyurl.com/39myeryp>.

⁴⁴ Véase Recomendaciones sobre la Ética de IA de la UNESCO. Disponible en: <https://tinyurl.com/44k82hfm>.

⁴⁵ Véase el informe sobre marco ético de IA en Colombia. Disponible en: <https://tinyurl.com/5n73t8bu>.

⁴⁶ A finales de 2019, a través del Consejo Nacional de Política Económica y Social (CONPES), se aprobó la Política Nacional para la Transformación Digital e Inteligencia Artificial, elaborada por el Ministerio de Tecnologías de la Información y las Comunicaciones.

⁴⁷ Véase el proyecto, disponible en: <https://tinyurl.com/2eex32sc>.

⁴⁸ Véase el informe *Hacia una estrategia de IA en México: Aprovechando la revolución de la IA*, 2018, disponible en: <https://tinyurl.com/26eb7han>.

problemas sociales y ambientales. Propone un marco de gobernanza y enfatiza un enfoque ético y de derechos humanos para el desarrollo y uso de la IA, proporcionando un contrapeso significativo ante la falta de continuidad en la estrategia nacional. Además, la Agenda Nacional Mexicana de Inteligencia Artificial contempla el análisis de los derechos de libertad de expresión, privacidad, igualdad y no discriminación en búsqueda de utilizar el poder de la IA de una manera positiva (García Benítez y Ruvalcaba-Gómez, 2021).

Uruguay emprendió el camino de la adopción de inteligencia artificial (IA) en el sector público en 2019, con el objetivo de implementar servicios digitales basados en tecnologías emergentes⁴⁹. Tras la consulta ciudadana, en 2020, se dio a conocer la Estrategia de Inteligencia Artificial para el Gobierno Digital, la cual tiene como propósito fomentar y fortalecer el uso responsable de la IA en la Administración pública. Esta estrategia se fundamenta en nueve principios generales y se apoya en cuatro pilares, centrados en las acciones para el logro de una IA ética y eficaz (García Benítez y Ruvalcaba-Gómez, 2021).

En definitiva, a pesar de los esfuerzos y las estrategias diseñadas en países como Argentina, Brasil, Chile, Colombia, México y Uruguay en el ámbito de la inteligencia artificial (IA), existe una notable brecha entre la formulación de políticas y su implementación efectiva. En los distintos países son características comunes el hecho de que se hayan puesto en marcha planes de desarrollo de la IA, tras las correspondientes consultas públicas, con el objetivo de que se desarrollen en fases y sobre ejes que pretenden un uso ético y responsable de la IA. El caso de México, donde la elaboración de una estrategia nacional de IA quedó inconclusa tras un cambio de gobierno en 2018, evidencia la fragilidad de la puesta en marcha de estas propuestas, que se enfrentan con claros obstáculos que limitan su impacto en la protección de derechos.

En conclusión, se puede observar que los distintos países de Iberoamérica están avanzando en la regulación de la IA, aunque con diferentes ritmos y enfoques, lo que plantea la necesidad de una mayor coordinación y armonización a nivel global.

V. CONCLUSIONES DEL ANÁLISIS COMPARADO DE LAS REGULACIONES ACTUALES

De las diferencias que se pueden observar, lo más relevante es la tendencia de la UE a regular, con el objetivo de proteger los derechos, de una forma más

⁴⁹ Véase el documento de estrategia IA en Uruguay, disponible en: <https://tinyurl.com/4kjej4ru>.

claramente intervencionista, frente al sistema de EE. UU., que, de acuerdo con la regulación de la inteligencia artificial de octubre de 2023, mediante la referida orden ejecutiva del presidente americano, tiene un enfoque más dirigido a lograr el desarrollo tecnológico. El conjunto de normas aprobadas en China muestra una evidente inspiración en los modelos regulatorios occidentales, en particular el Reglamento Europeo de IA y la propuesta de convenio internacional sobre IA de la OCDE. Pero su sujeción, expresamente apuntada, al desarrollo del modelo político chino, conlleva implicaciones que evidencian las diferentes concepciones. Iberoamérica cuenta con planes parciales en distintos países, con características comunes, que están en proceso de desarrollo con el objetivo de lograr un uso ético y responsable de la IA. La UE ha avanzado comparativamente en la formulación de normativas específicas para la IA. A través del reglamento sobre IA, establece reglas detalladas para el desarrollo y uso de sistemas de IA. En contraste, algunos países de Iberoamérica están en diferentes etapas de consideración y desarrollo normativo, en las que se observa una influencia de las acciones de la UE, especialmente el planteamiento sobre riesgos del reglamento. En la UE, la coordinación y la armonización son básicos para garantizar la coherencia en la regulación de la IA en todos los Estados miembros. En Iberoamérica, la coordinación puede variar, y algunos países pueden estar más avanzados en la formulación de normativas que otros.

Como se ha podido observar a lo largo de este trabajo, las normas generales de China y EE. UU. sobre la IA en 2023, así como los planes de regulación de algunos países iberoamericanos, tienen algunas similitudes con el reglamento europeo, como la preocupación por la seguridad, la transparencia y la rendición de cuentas de los sistemas de IA, o la necesidad, en otros, de revelar los contenidos generados artificialmente. Sin embargo, también hay diferencias evidentes, como el ámbito de aplicación de las normas, el grado de intervención y control del Gobierno, o el enfoque basado en los derechos humanos y los valores democráticos que caracteriza a la regulación europea de inteligencia artificial. Tanto en Iberoamérica como en la UE, la dirección parece apuntar hacia la protección de los aspectos éticos y responsables en el desarrollo y uso de la IA. Todos los planes en los que se trabaja en Iberoamérica buscan garantizar que la implementación de la tecnología respete los derechos humanos y cumpla con ciertos estándares. También resulta coincidente en ambas regiones el que se trabaje desde un enfoque inclusivo, buscando la participación activa de la sociedad civil y expertos a través de consultas públicas para formular políticas y regulaciones relacionadas con la IA.

Al igual que su contraparte en EE. UU., el Gobierno chino promueve el desarrollo y la consolidación de sus empresas al facilitar su crecimiento inicial y su posición en el mercado. Esta estrategia se lleva a cabo mediante la implemen-

tación de legislación que se centra exclusivamente en la reducción de riesgos inaceptables o en la protección significativa de sus ciudadanos. No obstante, a diferencia de otras aproximaciones, el Gobierno chino elige no emplear la regulación como un instrumento directivo que pueda condicionar o restringir los avances en el campo de la inteligencia artificial (Hernández Peña, 2022).

Bibliografía

- Anderson, M. y Anderson, S. (2014). Toward ensuring ethical behavior from autonomous systems: A case-supported principle-based paradigm. *Industrial Robot: An International Journal*, 42 (4), 324-331. Disponible en: <https://doi.org/10.1108/IR-12-2014-0434>.
- Arellano Toledo, W. (2019). El derecho a la transparencia algorítmica en *big data*. Derecho Público, derechos y transparencia ante el uso de algoritmos, inteligencia artificial y *big data*. *Revista General de Derecho Administrativo*. Disponible en: <https://tinyurl.com/48f6kbah>.
- Balaguer Callejón, F. (2023). *La Constitución del algoritmo*. Zaragoza: Fundación Giménez Abad de Estudios Parlamentarios y del Estado Autonómico.
- Barona Villar, S. (2021). *Algoritmización del derecho y de la justicia. De la inteligencia artificial a la Smart Justice*. Valencia: Tirant lo Blanch.
- Barrio, M. (2022). *Manual de derecho digital*. Valencia: Tirant lo Blanch.
- Belloso Martín, M. (2022). La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un derecho a la protección contra los sesgos? En F. Llano (dir.). *Inteligencia artificial y filosofía del derecho* (pp. 45-78). Murcia: Ediciones Laborum. Disponible en: <https://tinyurl.com/mw4m8a7r>.
- Bahatti, B. (2023). Essential guide to foundation models and large language models. *Medium*, 06-02-2023. Disponible en: <https://tinyurl.com/yc3erjt2>.
- Cotino Hueso, L. (2021). Un análisis crítico constructivo de la Propuesta de Reglamento de la Unión Europea por el que se establecen normas armonizadas sobre la Inteligencia Artificial (Artificial Intelligence Act). *Diario La Ley*, 2-05-2021. Disponible en: <https://tinyurl.com/mrxfhnsu>.
- Cotino Hueso, L. y Castellanos Claramunt, J. (2022). *Transparencia y explicabilidad de la Inteligencia Artificial*. Valencia: Tirant Lo Blanch. Disponible en: <https://tinyurl.com/bdd8pbhm>.
- García Benítez, V. H. y Ruvalcaba-Gómez, E. A. (2021). Análisis de las estrategias nacionales de inteligencia artificial en América Latina: estudio de los enfoques de ética y de derechos humanos. *Revista de Gestión Pública*, 10 (1), 5-32. Disponible en: <https://doi.org/10.22370/rgp.2021.10.1.3151>.
- García Herrera, V. (2021). Panorama legislativo de la inteligencia artificial en la Unión Europea. En F. Lledó Yagüe, I. Benítez Ortúzar, O. Monje Balmaseda (dirs.). *La robótica y la inteligencia artificial en la nueva era de la revolución industrial 4.0 (los desafíos jurídicos, éticos y tecnológicos de los robots inteligentes)* (pp. 265-360). Madrid: Dykinson.

- González de la Garza, M. (2019). El *Big Data* como la gran herramienta de pronóstico de tendencias en los sistemas electorales: del salto del paradigma demoscópico a los datos de comportamiento real. *Revista de Gestión Pública*. Disponible en: <https://tinyurl.com/3zsafacw>.
- Hagendorf, T. (2020). The ethics of AI ethics. An evaluation of guidelines. *Minds & Machines*, 30, 99-120. Disponible en: <https://doi.org/10.1007/s11023-020-09517-8>.
- Hernández Peña, J. C. (2022). *El marco jurídico de la inteligencia artificial. Principios, procedimientos y estructuras de gobernanza*. Cizur Menor: Aranzadi. Disponible en: <https://tinyurl.com/54dv77nc>.
- Jeld, J., Nele A., Hilligoss, H., Nagy, A. y Madhulika, S. (2020). Principled Artificial Intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center for Internet & Society*. Disponible en: <https://tinyurl.com/ycyu5yz7>.
- Mayer-Schönberger, V. y Cukier, K. (2019). *Big Data. La revolución de los datos masivos*. Madrid: Noema. Disponible en: <https://tinyurl.com/3nbsed53>.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. y Galstyan, A. (2022). A survey on bias and fairness in machine learning. *arXiv.org*. Disponible en: <https://doi.org/10.48550/arXiv.1908.09635>.
- Mccarthy, J. (2007). *What is artificial intelligence?* Stanford: Computer Science Department, Stanford University. Disponible en: <https://tinyurl.com/bdz9m9ct>.
- Mitchell, T. M. (1997) *Machine Learning*. New York: McGraw-Hill.
- Palma Ortigosa, A. (2022). *Decisiones automatizadas y protección de datos personales. Especial atención a los sistemas de inteligencia artificial*. Madrid: Dykinson.
- Pauner Chulvi, C. (2023). Transparencia algorítmica en los medios de comunicación y las plataformas digitales. *Revista Española de la Transparencia*, 17, 5.
- Presno Linera, M. A. (2022). *Derechos fundamentales e inteligencia artificial*. Madrid: Marcial Pons.
- Rebollo Delgado, L. (2023). *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson.
- Reche Tello, N., Tur Ausina, R., Ortega Gimenez, C. y Sanjúan Andrés, F. J. (2022). *La teoría constitucional frente a la transformación digital y las nuevas tecnologías*. Cizur Menor: Aranzadi.
- Russell, S. J. y Norvig, P. (2021). *Artificial Intelligence. A modern approach*. Madrid: Pearson.
- Sánchez Acevedo, M. E. (2022). La inteligencia artificial en el sector público y su límite respecto de los derechos fundamentales. *Estudios Constitucionales*, 2 (20), 277-284. Disponible en: <http://dx.doi.org/10.4067/S0718-52002022000200257>.
- Sánchez Barrilao, J. F. (2016). El derecho constitucional ante la era de ultrón: la informática y la inteligencia artificial como objeto constitucional. *Estudios de Deusto*, 64 (2), 225-258. Disponible en: [https://doi.org/10.18543/ed-64\(2\)-2016pp225-258](https://doi.org/10.18543/ed-64(2)-2016pp225-258).

- Serra Cristóbal, R. (2020). El reconocimiento de la discriminación múltiple por los tribunales. *Teoría y Derecho*, 27, 137-158. Disponible en: <https://doi.org/10.36151/td.2020.008>.
- Vakkuri, V. y Abrahamsson, P. (2018). The key concepts of ethics of Artificial Intelligence. En *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)* (pp. 1-6). Disponible en: <https://doi.org/10.1109/ICE.2018.8436265>