

DERECHOS HUMANOS E INTELIGENCIA ARTIFICIAL

José Justo Megías Quirós*

Resumen

La Inteligencia Artificial (IA) ha traído consigo importantes avances y beneficios para la humanidad, pero sería imprudente ignorar los riesgos conexos que, para los derechos humanos, entraña su uso indiscriminado. Las primeras alertas de instituciones privadas y públicas sirvieron para concienciar a los organismos internacionales de la necesidad de proceder a su regulación, de modo significativo en la Unión Europea (UE) y con un impulso cada vez más relevante en el seno de Naciones Unidas. Estas páginas recogen los últimos pasos en las propuestas de regulación de la IA a nivel internacional con el objetivo de proteger los derechos humanos.

Palabras clave: derechos humanos, inteligencia artificial, ética, regulación legal.

HUMAN RIGHTS AND ARTIFICIAL INTELLIGENCE

Abstract

Artificial Intelligence (AI) has brought important advances and benefits to humanity, but it would be unwise to ignore the related risks to human rights that its indiscriminate use would entail. The first alerts from private and public institutions have served to make international organizations aware of the need to proceed with their regulation, significantly in the European Union (EU) and with an increasingly relevant impulse within the United Nations. These pages collect the last steps in the proposals for the regulation of AI at an international level with the aim of protecting human rights.

Key words: human rights, artificial intelligence, ethic, legal regulation

* Catedrático de Filosofía del Derecho, Universidad de Cádiz. Avda. de la Universidad 4. 11406 Jerez de la Fra., Cádiz, España. Dirección electrónica: josejusto.megias@uca.es. El artículo es resultado del Proyecto I+D PID2019-108669RB-100/AEI/10.13039/501100011033, “*Derecho e Inteligencia Artificial: nuevos horizontes jurídicos de la personalidad y la responsabilidad robótica*”, financiado por el Ministerio de Ciencia, Innovación y Universidades (España).

La insistencia de la doctrina en la necesidad de regular la IA por los riesgos que entraña para los derechos humanos¹ ha contribuido a que, en la actualidad, tengamos en curso de aprobación dos modelos internacionales de regulación, uno en el ámbito de Naciones Unidas y otro en el de la UE. En el primero ha sido la UNESCO la que ha acometido este objetivo con una propuesta de Recomendación que posiblemente sea aprobada en noviembre de 2021 y que recoge exigencias éticas que deberían ser asumidas por las legislaciones internas para asegurar el respeto de los derechos humanos². La UE, en cambio, ha planteado directamente una regulación legal de la IA. Si bien en una primera etapa apostó por un marco ético, aboga ahora por la aprobación de normas jurídicas vinculantes complementadas con exigencias éticas. Así se refleja en la propuesta en 2020 de dos Reglamentos (norma vinculante de aplicación directa) del Parlamento Europeo, uno para establecer un marco de principios éticos y obligaciones jurídicas que regularían la IA, y otro para regular la responsabilidad civil por daños. El paso definitivo en este sentido lo ha dado la Comisión Europea en 2021 al proponer la aprobación de un Reglamento más ambicioso que, además de armonizar el derecho europeo, regularía los sistemas de IA con el fin de proteger los derechos humanos, y que tendría la denominación de *Ley de Inteligencia Artificial*.

1. Derechos humanos y riesgos reales asociados a la IA

Desde que comenzara a utilizarse la IA, hemos visto los beneficios que reporta, pero también los riesgos que su uso indiscriminado entraña para los derechos

¹ Vid., por ejemplo, Llano Alonso, F.H. (ed.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*. Cizur Menor, Aranzadi, 2021; García San José, D., “Implicaciones jurídicas y bioéticas de la Inteligencia Artificial (IA). Especial consideración al marco normativo internacional”, *Cuadernos de Derecho Transnacional* vol. 13, nº 1, 2021, pp. 255-276; Robles Carillo, M., “La gobernanza de la Inteligencia Artificial: contexto y parámetros generales”, *REEI* 39, 2020, pp. 1-27; Santos González, M.J., “Regulación legal de la robótica y la Inteligencia Artificial: retos de futuro”, *Revista Jurídica de la Universidad de León* 4, 2017, pp. 25-50; De Asís, R., *Una mirada a la robótica desde los derechos humanos*. Marcial Pons, Madrid, 2015.

² La propuesta toma cuerpo a partir de un documento del Grupo de Trabajo de la UNESCO, *Estudio preliminar sobre la ética de la Inteligencia Artificial*, SHS/COMEST/EXTWG-ETHICS-AI/2019/1, París, 26 de febrero de 2019, y otro de su Conferencia General, Documento 206 EX/42, *Estudio preliminar sobre los aspectos técnicos y jurídicos relativos a la conveniencia de disponer de un Instrumento normativo sobre la Ética de la Inteligencia Artificial*. París, 28 de marzo de 2019. Aprobado por la Conferencia General de la UNESCO en Resolución 40 C/67, de 30 de julio de 2019.

humanos, tolerados como una consecuencia natural, en ocasiones inevitable, a cambio de alcanzar beneficios que ofrece la IA, como agilidad y eficacia en la relación con la Administración Pública, gestión de expedientes, detección de errores en los procedimientos, etc. Sin embargo, se advierten riesgos considerables para los derechos humanos³.

Así, en el campo de la salud, la IA permite realizar diagnósticos y previsión de riesgos, prescripción de tratamientos, cirugía robótica, asistencia médica remota, tratamiento de imágenes, mapas sanitarios, control de propagación de enfermedades, etc. Sin embargo, existen riesgos claros en lo que afecta al derecho a la integridad, al tratamiento de datos personales muy sensibles, a la autonomía del paciente, al consentimiento y control humano sobre las decisiones finales de los sistemas, etc.⁴

En el campo de la seguridad ciudadana, la IA resulta muy útil para el reconocimiento biométrico, registro de actividad, análisis de comportamiento, interceptación y análisis de comunicaciones, búsqueda de personas desaparecidas, etc. Pero, de nuevo se advierte el riesgo del tratamiento abusivo de datos de carácter personal, preocupando en especial la *inteligencia anticipativa* con fines de vigilancia policial predictiva, excesivamente dependiente de sesgos discriminatorios⁵.

En la administración de justicia, la IA facilita la sistematización y búsqueda de información jurídica útil para jueces, abogados y para la sociedad en general, y permite la predicción a partir de decisiones judiciales ya adoptadas que podrían ser replicadas, pero sin capacidad de argumentar y garantizar una decisión equitativa, justa, precisa y apropiada, lo que la hace susceptible de “inexactitudes, resultados

³ Un estudio actualizado sobre derechos humanos en Anzalone, A., *Lo humano de los derechos humanos*. Dykinson, Madrid, 2021. Sobre los riesgos de la IA para los derechos humanos vid. Cotino Hueso, L., “Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales”, *Ilemata* 24, 2017, pp. 131-150, y “Ética en el diseño para el desarrollo de una Inteligencia Artificial, robótica y big data confiables y su utilidad desde el Derecho”, *Revista Catalana de Dret Públic* 58, 2019, pp. 29-48.

⁴ Vid. Andrés Segovia, B., “El reinicio tecnológico de la Inteligencia Artificial en el servicio público de salud”, *Ius et Scientia* vol. 7, nº 1, 2021, pp. 327-356; Juan, G.R., “Inteligencia Artificial y Filosofía del Bioderecho: una tesis crítica y una propuesta ética”, *Ius et Scientia* vol. 6, nº 2, 2020, pp. 96-110; Cárcar Benito, J.E., “La Inteligencia Artificial (IA): aplicación jurídica y regulación en los servicios de salud”, *Derecho y Salud* vol. 29, nº extra 1, 2019, pp. 265-277.

⁵ Ello ha llevado al Parlamento Europeo a aprobar la Resolución de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)), en la que propone la prohibición de este uso de la IA. No consideraremos en este artículo la seguridad militar, que ya cuenta con tratados internacionales. Vid. Corral Hernández, D., “Inteligencia Artificial en la batalla. Un guerrero preciso y letal sin alma ni corazón”, *Documento de Opinión IEEE* (Mº de Defensa) 79, 2021, pp. 1-16.

discriminatorios, sesgos implícitos o incorporados”⁶. Agilizar el acceso a la justicia no conlleva una sociedad más justa si no se garantiza que las decisiones del sistema sean correctas⁷.

El mayor riesgo en campos como la educación (personalizar contenidos, evaluar conocimientos, control de plagios, tutorización y seguimiento, etc.) o el transporte público (control de carreteras y de velocidad, semaforización inteligente, optimización del transporte público, etc.) deriva del uso que se pueda hacer de los datos personales y la injerencia en la privacidad. Esta misma amenaza se aprecia en la actividad publicitaria con la elaboración de perfiles personales sin transparencia y sin consentimiento de los afectados, lo que, además de atentar contra la privacidad, permite manipular a las personas con ofertas comerciales personalizadas⁸. En el sector de recursos humanos y financieros, la IA tiene gran aceptación para la administración de personal, selección de candidatos, etc., en teoría de un modo objetivo y preciso, pero ocultando que los algoritmos suelen incorporar (o crear) sesgos discriminatorios⁹.

Preocupan también los riesgos en el sector informativo, en particular los “problemas relacionados con el acceso a la información, la desinformación, la discriminación, la libertad de expresión, la privacidad, la alfabetización mediática y la cultura

⁶ Grupo de Trabajo UNESCO, *Estudio preliminar sobre la ética de la Inteligencia Artificial*, cit., n. 50. Vid. San Miguel Caso, C., “La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales?”, *Ius et Scientia* vol. 7, n° 1, 2021, pp. 286-303; De Asís Pulido, M., “La incidencia de las nuevas tecnologías en el derecho al debido proceso”, *Ius et Scientia* vol. 6, n° 2, 2020, pp. 186-199.

⁷ Vid. Rodríguez Puerto, M.J., “¿Puede la Inteligencia Artificial interpretar normas jurídicas? Un problema de razón práctica”, *CEFD* 44, 2021, pp. 74-96; Solar Cayón, J.I., “Inteligencia Artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Univ. de Alcalá, Alcalá de Henares, 2020, pp. 125-172. Quizás el documento más explícito es el de la Comisión Iberoamericana de Ética Judicial (CIEJ), Noveno Dictamen, *Sobre el uso por el juez de las nuevas tecnologías: ventajas y desafíos éticos*, 12 de marzo de 2020.

⁸ Vid. Álvarez Rubio, J., “Inteligencia Artificial y protección jurídica de los consumidores” y Tomillo Urbina, J., “La responsabilidad por la prestación de servicios a través de plataformas de intermediación on line”, ambos en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial...*, cit., pp. 275-331 y 333-364 respectivamente.

⁹ El riesgo es muy frecuente en el sector financiero, en el que los algoritmos seleccionan candidatos basándose en sesgos que excluyen sin razón objetiva a ciertos sectores de población. Vid. Alonso, A., Carbó, J.M., *Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation*. Banco de España. Documento de Trabajo N° 2105, Madrid, 2021.

en materia de información”¹⁰. La IA no sólo es utilizada para limitar la libertad de expresión, sino también para generar información falsa u ofensiva, para modificar audios e imágenes con gran credibilidad, etc., lo que genera desinformación y repercute negativamente sobre la limpieza del debate público y los procesos democráticos¹¹.

Por último, debemos mencionar los riesgos derivados del uso de la IA en actividades privadas, realidad que nos desborda: robótica del hogar, asistentes de búsqueda de información o contenidos, configuración de rutas, entrenamiento deportivo, asistencia médica (incluida la psíquica), etc. Todos ellos se basan en el tratamiento de datos personales con mayor o menor transparencia y el consiguiente riesgo para la privacidad, pero pueden además inducir a conductas nocivas para la integridad personal como la automedicación sin control, la modificación de conductas interpersonales, el aislamiento social, etc., que pueden poner en peligro la propia esencia de la persona, en especial la última generación de asistente emocional, que, al imitar capacidades humanas para generar confianza, es capaz de acceder a sentimientos y emociones en la interacción¹².

Así, pues, junto a los beneficios que la IA ha traído consigo para la sociedad, resultan evidentes los riesgos para la dignidad de la persona y los derechos humanos, en especial los derechos a la integridad física, la vida privada, la igualdad, la libertad de expresión y de reunión, la tutela judicial efectiva, etc., riesgos que se agudizan cuando afectan a grupos vulnerables (menores, inmigrantes, personas con discapacidad, edad avanzada, etc.). En la actualidad se intenta minimizar estos riesgos con dos

¹⁰ Grupo de Trabajo UNESCO, *Estudio preliminar sobre la ética de la Inteligencia Artificial*, cit., n. 69. “La transparencia y la credibilidad plantean problemas en los casos en los que los consumidores no saben o no pueden saber cuándo el contenido lo generan máquinas, de qué fuentes procede, y en qué medida se ha verificado o es incluso falsa la información, con debates actuales sobre los *deep fakes* o *falsedades profundas* como casos extremos”. Documento citado, n. 77.

¹¹ Vid. Pelayo González-Torres, A., “TIC, Inteligencia Artificial y crisis de la democracia” y Sánchez Martínez, M.O., “Desafíos democráticos en el ecosistema digital”, ambos en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial...*, cit., pp. 55-78 y 79-119 respectivamente; vid. también Petit, M., “Por una crítica de la razón algorítmica. Estado de la cuestión sobre la inteligencia artificial, su influencia en la política y su regulación”, *Quaderns del CAC* 44, vol. XXI, 2018, pp. 5-15.

¹² La *Emotional Chatting Machine* reconoce emociones del usuario, se retroalimenta con ellas para aprender y ofrece respuestas según las circunstancias del usuario en un intento de suplir la relación verdaderamente humana y en ocasiones le impide que distinga, incluso, si interactúa con una persona o una máquina. Vid. Piedra Alegría, J., “El despertar de los robots (sociales)”, en Bermúdez Vázquez, M., (coord.), *Lucas en el camino: filosofía y ciencias sociales en tiempos de desconcierto*. S.I., Dykinson, 2021, pp. 29-51.

modelos de regulación distintos, uno ético y otro jurídico. Trataremos de aportar un breve análisis de ellos.

2. Propuesta de la UNESCO: un marco ético

La UNESCO acordó en julio de 2021 que su Conferencia General debatiera, en su reunión de noviembre, la *Recomendación sobre la Ética de la Inteligencia Artificial*¹³. En caso de aprobación, será el primer texto de carácter universal sobre la IA basado “en el derecho internacional y en un enfoque normativo mundial y centrado en la dignidad humana y los derechos humanos, así como en la igualdad de género, la justicia social y económica, el bienestar físico y mental, la diversidad, la interconexión, la inclusión y la protección del medio ambiente”¹⁴. El modelo por el que opta es el del establecimiento de principios éticos universales que sirvan de base para la regulación legal de la IA en cada Estado. Parece adecuado para, desde unos principios éticos bien definidos, adaptar los deberes a la cambiante realidad de la IA y salvar desventajas de la regulación legal, que, por su sometimiento a requisitos formales para su modificación, adolecería de constantes lagunas y la impunidad de quienes las aprovecharan para hacer un uso abusivo de la IA¹⁵. Este marco constituiría únicamente el primer paso, instando la Recomendación a los Estados a seguirlo para elaborar un marco legal propio que ofrezca seguridad jurídica.

El texto pone al descubierto la fragilidad del marco ético al referirse a la “diversidad de orientaciones éticas y culturas en todo el mundo” y a “los valores éticos locales y regionales”¹⁶, aunque propone como objetivo “aportar un instrumento normativo aceptado mundialmente”¹⁷. Si bien descarta en principio unas exigencias éticas objetivas, conformándose con otras fruto de la negociación para salvar las dife-

¹³ El documento propuesto es el *Anteproyecto de Recomendación sobre la Ética de la Inteligencia Artificial*, elaborado por el Grupo Especial de Expertos (GEE) de la UNESCO. Documento Final SHS/BIO/AHEG-AI/2020/4 REV.2. París, 7 de septiembre de 2020.

¹⁴ *Anteproyecto de Recomendación...*, cit., Preámbulo, p. 1. Reitera de forma constante que toda regulación de la AI debe respetar la dignidad, los derechos humanos y las libertades fundamentales de todo ser humano.

¹⁵ Entre las razones por las que se opta por un marco ético y no una regulación más detallada está el carácter dinámico e innovador de la IA, con un desarrollo en ocasiones absolutamente impredecible, de modo que cualquier norma detallista corre el riesgo de quedar obsoleta en un corto espacio de tiempo. A ello se añade la dificultad que entraña la cantidad de actores implicados y la dificultad de regular sistemas con capacidad de autoaprendizaje, lo que los convierte en imprevisibles y hace casi imposible una regulación completa.

¹⁶ *Anteproyecto de Recomendación...*, cit., Preámbulo, p. 4.

¹⁷ *Anteproyecto de Recomendación...*, cit., n. 6.

rencias culturales, aclara, sin embargo, que “todos los valores y principios” propuestos “son deseables en sí mismos”¹⁸, aunque precisen ser contextualizados para su correcta aplicación, sin que ninguna disposición de la Recomendación pueda interpretarse como una autorización a realizar actividades que vulneren “los derechos humanos, las libertades fundamentales, la dignidad humana y el respeto del medio ambiente”¹⁹. Desde esta perspectiva, los objetivos de la Recomendación son: a) proporcionar un marco universal de valores, principios y acciones que oriente a los Estados en la formulación de sus leyes y políticas; b) orientar a todos los agentes para asegurar la perspectiva ética en todas las etapas de los sistemas de IA; c) promover el respeto de la dignidad humana y la igualdad de género, salvaguardar los intereses de las generaciones presentes y futuras y proteger los derechos humanos, las libertades fundamentales y el medio ambiente; d) fomentar el diálogo multidisciplinario y plural sobre cuestiones éticas de la IA; y e) promover el acceso equitativo a los avances y beneficios de la IA²⁰.

2.1. Valores universales y principios éticos para la IA

La Recomendación escalona valores, principios y exigencias particulares para el terreno práctico. Los valores vendrían a ser los ideales inspiradores de toda regulación de los sistemas de IA, mientras que los principios constituirían una primera concretización de dichos valores en exigencias éticas universales. El tercer escalón, bajo la rúbrica “Ámbitos de acción política”, está formado por otra serie de exigencias éticas más concretas agrupadas en diez ámbitos de actuación entre las que sitúa en primer lugar la obligación de los Estados de establecer un marco legal interno adecuado y promocionar el respeto del mismo²¹. Los valores propuestos son cuatro. En primer lugar, el *respeto, protección y promoción de la dignidad humana, los derechos humanos y las libertades fundamentales*. La IA no sólo debe respetar la dignidad y derechos universales de todo ser humano, sino también, en la medida de lo posible, promocionarlos sin dar cabida a discriminación alguna por razón de sexo, género, idioma, religión, opiniones políticas o de otra índole, origen nacional, étnico,

¹⁸ *Anteproyecto de Recomendación...*, cit., n. 11.

¹⁹ *Anteproyecto de Recomendación...*, cit., n. 141.

²⁰ *Anteproyecto de Recomendación...*, cit., n. 8.

²¹ *Anteproyecto de Recomendación...*, cit., n. 48.

indígena o social, orientación sexual e identidad de género, bienes, nacimiento, discapacidad, edad u otra condición²². Este primer valor se complementa con el de *diversidad e inclusión*, que se concreta en descartar el uso de la IA para *homogeneizar* a los seres humanos e impedir el desarrollo personal de acuerdo con lo que nos diferencia a unos de otros en el “estilo de vida, creencias, opiniones, expresiones o experiencias personales”²³.

Los otros dos valores son el de *prosperidad del medio ambiente y los ecosistemas*, que exige la orientación de la IA hacia el desarrollo sostenible, y el de *vivir en armonía y paz*, que exige contribuir “a la interconexión de todas las criaturas vivas entre sí y con el medio natural”²⁴ y no “segregar ni cosificar a los seres humanos ni socavar su seguridad, así como tampoco dividir y enfrentar entre sí a las personas y los grupos ni amenazar la coexistencia armoniosa entre los seres humanos, los no humanos y el medio natural”²⁵.

Junto a los valores, la Recomendación contempla principios éticos que deben orientar las regulaciones internas de la IA y, mientras tanto, servir de criterios de actuación a todos los actores. El primero de ellos es el de *proporcionalidad e inocuidad*. Desde la base de que no toda innovación tecnológica supone en sí misma un avance para la humanidad²⁶, considera relevante la finalidad y el uso que se haga de la IA. Este principio exige evitar, desde el propio diseño, cualquier daño que pueda derivar del uso de la IA, lo que implicaría no sólo que el fin perseguido sea legítimo, sino que también lo sean los medios utilizados por la IA para alcanzarlo (proporcionados al resultado perseguido, ajustados a los valores reseñados, rigurosos científicamente y siempre bajo control humano)²⁷.

El principio de *seguridad y protección* exige que todo sistema de IA esté diseñado de forma que su funcionamiento impida el riesgo de brechas en la seguridad y protección del ser humano y del entorno natural, en particular en lo referente al acceso a datos de calidad y su tratamiento²⁸. Se complementan con el principio de

²² *Anteproyecto de Recomendación...*, cit., n. 13. “Las personas nunca deberían ser cosificadas, no debería socavarse su dignidad, y sus derechos humanos nunca deberían ser objeto de violación o abusos” (n. 15).

²³ *Anteproyecto de Recomendación...*, cit., n. 20.

²⁴ *Anteproyecto de Recomendación...*, cit., n. 22.

²⁵ *Anteproyecto de Recomendación...*, cit., n. 24.

²⁶ Vid. al respecto González Arencibia, M., Martínez Cardero, D., “Dilemas éticos en el escenario de la Inteligencia Artificial”, *Economía y Sociedad* vol. 25, n° 57, 2020, pp. 1-17.

²⁷ Para su efectividad, deberán aplicarse “procedimientos de evaluación de riesgos y la adopción de medidas para impedir que se produzca ese daño”. *Anteproyecto de Recomendación...*, cit., n. 25.

²⁸ *Anteproyecto de Recomendación...*, cit., n. 27. Vid. Téllez Carbajal, E., “Derechos humanos, ética y transparencia algorítmica”, *Ius et Scientia* vol. 7, n° 1, 2021, pp. 370-386.

privacidad, uno de los más importantes dada la relevancia de los datos de carácter personal para la tecnología IA, lo que exige que el marco jurídico interno proteja eficazmente frente a la recopilación, tratamiento y uso ilícito de los datos, así como que garantice los derechos conexos que permiten el control de los datos por su titular²⁹. Sitúa al mismo nivel la exigencia ética que afecta a diseñadores, creadores y entrenadores de los sistemas de IA, que deberán evaluar el impacto de los algoritmos sobre la privacidad y adoptar cautelas para impedir intromisiones ilegítimas. En relación al medio ambiente (incluido el ser humano), el principio se complementa con el de *sostenibilidad*, que exige una “evaluación continua de los efectos sociales, culturales, económicos y ambientales de las tecnologías de la IA” siempre ajustados a los Objetivos de Desarrollo Sostenible de NU³⁰.

Otro principio de gran importancia es el de *equidad y no discriminación*. Conlleva la obligación de hacer partícipe a toda la humanidad de los beneficios de la IA según las necesidades singulares de poblaciones desfavorecidas, marginadas y vulnerables³¹ y la obligación de impedir todo sesgo discriminatorio por razón de edad, cultura, usos lingüísticos, discapacidad, etc., en los que puedan incurrir los algoritmos de los sistemas de IA por reproducción de los existentes en la realidad social o creación de uno nuevo.

Del principio de *supervisión y decisión humanas* deriva la obligación de las autoridades de velar de forma eficaz por el diseño, desarrollo y uso ético de los sistemas de IA, asegurar que toda decisión final del sistema esté bajo control humano y que siempre sea una persona, física o jurídica, la obligada a rendir cuentas y asumir la responsabilidad de las consecuencias³². Guarda estrecha relación con el principio de *responsabilidad y rendición de cuentas*, que se concreta en la obligación de todos

²⁹ “Estos mecanismos (...) se refieren (...) al ejercicio de derechos por parte de los interesados, así como al derecho de las personas a que se eliminen sus datos personales, garantizando así un objetivo legítimo y una base jurídica válida para el tratamiento (...), la transparencia, las salvaguardias adecuadas para los datos sensibles y una supervisión independiente eficaz”. *Anteproyecto de Recomendación...*, cit., n. 33. También el último principio propuesto, el principio de *gobernanza y colaboración adaptativas y de múltiples partes interesadas*, concreta la exigencia de respetar no sólo la regulación establecida por el derecho internacional para la recopilación, tratamiento y uso de datos personales, sino también la regulación interna de cada Estado para los datos generados en sus territorios. *Anteproyecto de Recomendación...*, cit., n. 46.

³⁰ *Anteproyecto de Recomendación...*, cit., n. 31.

³¹ *Anteproyecto de Recomendación...*, cit., n. 28. Se exige a los países más desarrollados tecnológicamente que compartan sus innovaciones y beneficios con los menos avanzados y “contribuyan a un orden mundial más equitativo en lo que respecta a la información, la comunicación, la cultura, la educación, la investigación y la estabilidad socioeconómica y política” (*ibidem*).

³² *Anteproyecto de Recomendación...*, cit., nn. 35-36.

los actores de la IA de actuar con responsabilidad en todas sus etapas (desde el diseño hasta su uso final)³³ y rendir cuentas de los procesos y resultados (decisiones y acciones)³⁴.

El principio de *transparencia y explicabilidad*, supuesto el derecho “a saber cuándo se toma una decisión sobre la base de algoritmos de IA”, se concreta en la exigencia de ofrecer información precisa sobre el uso, funcionamiento, fines y resultados del sistema de IA para poder ejercer un control eficaz. La explicabilidad, por su parte, exige hacer comprensibles el funcionamiento de los algoritmos, los procedimientos seguidos por el sistema y los resultados finales, permitiendo así llevar a cabo la trazabilidad de todo el proceso³⁵. En especial, resulta obligado informar de forma comprensible a los ciudadanos sobre el impacto que los sistemas de IA tienen sobre sus derechos fundamentales.

2.2. Ámbitos de actuación

El último bloque de la Recomendación está dedicado a concretar exigencias éticas que las futuras legislaciones internas deberían recoger como obligaciones legales en ciertos ámbitos de actuación. Así, en relación con la *evaluación del impacto ético* de los sistemas de IA, toda regulación debería establecer un modelo de evaluación para determinar los riesgos, beneficios e impacto sobre los derechos humanos y el Estado de Derecho, así como medidas de prevención, atenuación y seguimiento de los riesgos. Entre los estándares fijados por el modelo de evaluación deben figurar: a) mecanismos de supervisión para determinar, prevenir y mitigar riesgos y rendir cuentas³⁶; b) mecanismos para evaluar los efectos socioeconómicos y minimizar

³³ Vid. Terrones Rodríguez, A.L., “Inteligencia Artificial y ética de la responsabilidad”, *Cuestiones de Filosofía* vol. 4, n° 22, 2018, pp. 141-170.

³⁴ *Anteproyecto de Recomendación...*, cit., n. 43. Exigiría el establecimiento de mecanismos de auditoría y trazabilidad que permitan comprobar si el sistema IA atenta o no contra los derechos humanos y determinar, en su caso, la responsabilidad.

³⁵ *Anteproyecto de Recomendación...*, cit., n. 38. El cumplimiento de estas exigencias permitirá evaluar la fiabilidad de la IA y el control de su incidencia sobre los derechos humanos. Se complementa con el principio de *sensibilización y alfabetización*, que impone a todos los actores, públicos y privados, la exigencia “de garantizar una participación pública efectiva, de modo que todos los miembros de la sociedad puedan adoptar decisiones informadas sobre la utilización de los sistemas de IA y estén protegidos de influencias indebidas”. *Anteproyecto de Recomendación...*, cit., n. 44.

³⁶ “Mecanismos de diligencia debida y supervisión para determinar, prevenir y mitigar los riesgos y rendir cuentas sobre la forma en que abordan el impacto de los sistemas de IA en los derechos humanos, el estado de derecho y las sociedades inclusivas”. *Anteproyecto de Recomendación...*, cit., n. 51.

las brechas sociales³⁷; c) medidas de vigilancia de los sistemas y sus algoritmos en todo el ciclo de vida³⁸; d) procedimientos para evaluar “el impacto ético de los sistemas de IA”³⁹; e) mecanismos de auditabilidad, trazabilidad, explicabilidad y supervisión de los sistemas desplegados⁴⁰; y f) mecanismos para seguir y evaluar “las iniciativas y políticas relacionadas con la ética de la IA”⁴¹.

El segundo ámbito de actuación está referido a la *gobernanza y administración éticas*. Además de exigir la participación de todos los actores en la elaboración de los instrumentos de gobernanza, éstos deben ser transparentes e incluir aspectos de previsión, protección y reparación⁴². Ello se concreta en una serie de características y elementos insoslayables: a) “ajustarse al derecho internacional de los derechos humanos y promover los derechos humanos y las libertades fundamentales”⁴³; b) reforzar “la capacidad del poder judicial para adoptar decisiones relacionadas con los sistemas de IA en el marco del estado de derecho y de conformidad con las normas internacionales, en particular en lo que respecta a la utilización de los sistemas de IA en sus deliberaciones”⁴⁴; c) mecanismos de auditabilidad y trazabilidad de los sistemas que permitan investigar y reparar los daños causados a los usuarios; d) mecanismo de certificación, contemplando “diferentes niveles de auditoría de los sistemas, los datos y el cumplimiento de las directrices éticas”⁴⁵; e) creación de una au-

³⁷ *Anteproyecto de Recomendación...*, cit., n. 51.

³⁸ “Medidas adecuadas para vigilar todas las etapas del ciclo de vida de los sistemas de IA, incluidos el comportamiento de los algoritmos utilizados para la adopción de decisiones, los datos y los actores de la IA que participan en el proceso”. *Anteproyecto de Recomendación...*, cit., n. 52.

³⁹ Las evaluaciones deben ser “transparentes y abiertas al público (...), multidisciplinarias, multiculturales, pluralistas e inclusivas” y prever “las repercusiones, atenuar los riesgos, evitar las consecuencias perjudiciales, facilitar la participación de los ciudadanos”. *Anteproyecto de Recomendación...*, cit., n. 53.

⁴⁰ *Anteproyecto de Recomendación...*, cit., n. 53. Propuesta reiterada en n. 58.

⁴¹ *Anteproyecto de Recomendación...*, cit., n. 54.

⁴² *Anteproyecto de Recomendación...*, cit., n. 55. Se alienta a utilizar “los prototipos de políticas y los entornos de pruebas reguladores para acelerar la formulación de leyes, reglamentos y políticas acordes con el rápido desarrollo de las nuevas tecnologías y garantizar que las leyes y los reglamentos se puedan poner a prueba en un entorno seguro antes de su aprobación oficial” (n. 70).

⁴³ *Anteproyecto de Recomendación...*, cit., n. 62. Lo que obligará a los gobiernos a “aplicar políticas destinadas a garantizar que las acciones de los actores de la IA se ajusten al derecho, las normas y los principios internacionales de derechos humanos durante todo el ciclo de vida de los sistemas de IA” (n. 65).

⁴⁴ *Anteproyecto de Recomendación...*, cit., n. 69.

⁴⁵ *Anteproyecto de Recomendación...*, cit., n. 57.

toridad independiente para la evaluación de impacto ético y seguimiento de los sistemas⁴⁶; f) mecanismos: para supervisar la adquisición de sistemas de IA de “utilización delicada en materia de derechos humanos”⁴⁷ y “la seguridad y la protección de los sistemas de IA”⁴⁸, para combatir “todo tipo de estereotipos (...), la difusión de información no fiable o la diseminación de ideas antidemocráticas”⁴⁹, y para “lograr la rendición de cuentas y la responsabilidad por el contenido y los resultados de los sistemas”⁵⁰; y g) requisitos “de transparencia y explicabilidad de los sistemas de IA”⁵¹.

El tercer ámbito de actuación, referido a la *política de datos*, tiene como objetivo el equilibrio entre el tratamiento legítimo de datos y la protección de la privacidad. El marco regulador deberá garantizar la calidad, seguridad y protección de los datos, así como prever mecanismos de retroinformación para que el sistema aprenda de los errores⁵². Deberá asegurar al más alto nivel la transparencia y tratamiento seguro de datos delicados, establecer mecanismos de rendición de cuentas eficaces y velar por “el pleno disfrute de los derechos de los interesados, en particular el derecho de acceso y el derecho a la eliminación de sus datos personales en los sistemas de IA, un nivel adecuado de protección mientras se utilicen los datos con fines comerciales (...) o sean transferidos al extranjero y una supervisión eficaz e independiente”⁵³.

En los dos siguientes ámbitos de actuación, *desarrollo y cooperación internacional y medio ambiente*, se exige la colaboración de los Estados, empresas transnacionales y organismos internacionales en la investigación, innovación y análisis del impacto ético de los sistemas de IA en esferas de desarrollo como “la atención de la salud, la agricultura, el suministro de alimentos, la educación, los medios de comunicación, la cultura, el medio ambiente, la ordenación de los recursos hídricos, la

⁴⁶ *Anteproyecto de Recomendación...*, cit., nn. 59 y 63.

⁴⁷ *Anteproyecto de Recomendación...*, cit., n. 63.

⁴⁸ *Anteproyecto de Recomendación...*, cit., n. 64.

⁴⁹ *Anteproyecto de Recomendación...*, cit., n. 66.

⁵⁰ *Anteproyecto de Recomendación...*, cit., n. 68.

⁵¹ *Anteproyecto de Recomendación...*, cit., n. 71.

⁵² *Anteproyecto de Recomendación...*, cit., n. 72. “La privacidad ha de ser respetada, protegida y promovida a lo largo del ciclo de vida de los sistemas de IA” (n. 73).

⁵³ *Anteproyecto de Recomendación...*, cit., n. 74. Especialmente, “garantizar una mayor seguridad de los datos personales y los datos confidenciales que, de ser divulgados, puedan causar daños, lesiones o dificultades excepcionales a una persona. Cabe citar como ejemplos los datos relativos a infracciones, procesos penales y condenas, así como a las medidas de seguridad conexas; los datos biométricos y genéticos; y los datos personales relativos al origen étnico o social, las opiniones políticas, la pertenencia a sindicatos, las creencias religiosas y de otro tipo, la salud y la vida sexual” (n. 75).

gestión de infraestructuras y la planificación y el crecimiento económico”⁵⁴, de acuerdo con los valores y principios establecidos y las normas de desarrollo sostenible⁵⁵.

Los cuatro ámbitos siguientes se centran en un uso de la IA que fomente la igualdad universal. Así, se insta a que las evaluaciones de impacto ético incluyan perspectiva de género para evitar la discriminación de la mujer⁵⁶. En el ámbito de la *cultura*, se insta a los Estados a promover que la IA colabore en la extensión de la cultura para todos⁵⁷, y en el de la *educación e investigación*, a instrumentar la colaboración internacional para que toda la humanidad se pueda formar en materia de IA, y promocionar la investigación ética de la IA y su uso en la educación sin discriminación⁵⁸. Por último, en el ámbito de la *economía y trabajo* se insta a los Estados a promover, en colaboración con todos los actores sociales, una formación para trabajar en entornos más digitalizado y preparar a los trabajadores para la transición a los nuevos entornos⁵⁹.

⁵⁴ *Anteproyecto de Recomendación...*, cit., n. 80. Insiste en esta colaboración en los nn. 81-85.

⁵⁵ *Anteproyecto de Recomendación...*, cit., n. 86. También deberán incentivar el uso de la IA para “la vigilancia, protección y regeneración del medio ambiente y los ecosistemas”, promocionando medidas para: “a) apoyar la protección, el seguimiento y la gestión de los recursos naturales; b) apoyar la prevención, el control y la gestión de los problemas relacionados con el clima; c) apoyar un ecosistema alimentario más eficiente y sostenible; d) contribuir a acelerar el acceso a la energía sostenible y su adopción a gran escala; e) facilitar y promover la incorporación de infraestructuras sostenibles, modelos empresariales sostenibles y financiación sostenible al servicio del desarrollo sostenible; f) detectar los contaminantes o prever los niveles de contaminación y, de ese modo, ayudar a las partes interesadas pertinentes a definir, planificar y poner en marcha intervenciones específicas para prevenir y reducir la contaminación” (n. 87).

⁵⁶ Los Estados “deberían velar por que los estereotipos de género y sesgos discriminatorios no se trasladen a los sistemas de IA”. *Anteproyecto de Recomendación...*, cit., n. 92. Cita aspectos proclives a la discriminación como el salario, la representación profesional, puestos directivos superiores, juntas directivas, equipos de investigación, educación, acceso y utilización de la de la IA, así como en la distribución del trabajo no remunerado y de las responsabilidades de cuidado (n. 91).

⁵⁷ *Anteproyecto de Recomendación...*, cit., nn. 95-96. Insta también a promover investigaciones sobre la propiedad intelectual, por ejemplo, para determinar la titularidad “de los derechos de las obras creadas mediante tecnologías de la IA” (n. 100).

⁵⁸ Los Estados “deberían velar por que los investigadores en IA reciban formación en ética de la investigación y exigirles que incluyan consideraciones éticas en sus concepciones, productos y publicaciones”. *Anteproyecto de Recomendación...*, cit., n. 109.

⁵⁹ Deberá analizarse “el impacto de los sistemas de IA en el entorno laboral local con miras a anticipar las tendencias y los desafíos futuros. Estos estudios deberían investigar el impacto

El último ámbito de actuación al que hace referencia la Recomendación es el de la *salud y bienestar social*. Para este campo se pide a los Estados el esfuerzo “por emplear sistemas eficaces de IA para mejorar la salud humana y proteger el derecho a la vida” en el que prime “la solidaridad internacional para hacer frente a los riesgos e incertidumbres de la salud en el plano mundial (...) conforme al derecho internacional y al derecho, las normas y los principios internacionales de derechos humanos”⁶⁰. La relevancia de la IA en este ámbito lleva a la Recomendación a exigir una reglamentación rigurosa de la predicción, detección y tratamiento médicos de las aplicaciones de IA, estableciendo: a) supervisión para minimizar sesgos discriminatorios; b) atención a la privacidad, con vigilancia constante; c) garantizar el conocimiento y consentimiento informado para el tratamiento de datos; y d) garantizar que el cuidado humano y la decisión final sobre el diagnóstico y el tratamiento correspondan en última instancia a un ser humano. Se trata, en definitiva, de garantizar un trato humano y digno a las personas⁶¹.

3. La regulación europea de la IA: del marco ético a la norma vinculante

La UE ha sometido el impacto de la IA sobre los derechos humanos a un profundo debate durante los últimos años, plasmando el resultado en numerosos documentos del Grupo de Expertos de Alto Nivel sobre IA y en los Informes de la Comi-

de los sistemas de IA en los sectores económico, social y geográfico, así como en las interacciones entre seres humanos y robots y entre los propios seres humanos, a fin de asesorar sobre las mejores prácticas de reconversión y reasignación profesional”. *Anteproyecto de Recomendación...*, cit., n. 117.

⁶⁰ *Anteproyecto de Recomendación...*, cit., n. 119. Se encarece el cuidado “de las relaciones del paciente con su familia y con el personal sanitario”, así como la participación de todos los implicados en el desarrollo de los sistemas (n. 120).

⁶¹ *Anteproyecto de Recomendación...*, cit., n. 121. En esta línea, se exigen estudios rigurosos sobre los efectos de la IA en materia de salud mental, elaborar directrices sobre las interacciones entre seres humanos y robots y sus repercusiones en las relaciones entre seres humanos, velar por que las interacciones entre seres humanos y robots se ajusten a los mismos valores y principios que se aplican a cualquier otro sistema de IA, proteger el derecho de los usuarios a determinar fácilmente si interactúan con un ser vivo o con un sistema de IA que imita las características humanas o animales, aplicar políticas de sensibilización sobre la antropomorfización de las tecnologías de la IA, promover la investigación colaborativa sobre los efectos de la interacción a largo plazo de personas (especialmente los niños) con sistemas de IA, etc. (nn. 122-129).

sión de Libertades Civiles, Justicia y Asuntos de Interior y de la Comisión de Asuntos Jurídicos, base para las propuestas del Parlamento y de la Comisión⁶². En este camino se pueden distinguir dos etapas, una primera en la que la opción prioritaria era establecer un marco ético que complementara la legislación ya vigente, y una segunda en la que se opta claramente por un marco legal innovador concebido explícitamente para la IA.

3.1. Primera apuesta: el marco ético

Los primeros textos aprobados por la UE siguieron un camino muy similar al de la UNESCO con principios éticos objetivos que complementarían la escasa regulación de la IA. La diferencia es que sus exigencias, además de estar dirigidas a los Estados para que adaptaran su legislación nacional, obligaban también a los propios órganos de la UE a modificar la legislación europea ya vigente.

El primer texto del Parlamento Europeo con contenido ético significativo fue su propuesta de *Carta sobre robótica* de 2017⁶³. Establecía que la IA debería regirse por los principios éticos de beneficencia, no perjuicio o maleficencia, autonomía del usuario, justicia, respeto de derechos fundamentales, precaución, participación,

⁶² Vid., por ejemplo, Comunicación, de 25 de abril de 2018, Inteligencia artificial para Europa (COM(2018)0237); Comunicación, de 7 de diciembre de 2018, Plan coordinado sobre la inteligencia artificial (COM(2018)0795); Comunicación, de 8 de abril de 2019, Generar confianza en la inteligencia artificial centrada en el ser humano (COM(2019)0168); Comunicación, de 19 de febrero de 2020, Una Estrategia Europea de Datos (COM(2020)0066); Comunicación, de 19 de febrero de 2020, Configurar el futuro digital de Europa (COM(2020)0067); y Comunicación, de 21 de abril de 2021, Fostering a European approach to Artificial Intelligence (COM(2021)) 205 final.

⁶³ Resolución, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica.

transparencia, rendición de cuentas, seguridad, reversibilidad de la acción, privacidad, maximización de beneficios y minimización de daños⁶⁴. Le siguió su Resolución de 2019⁶⁵, en la que, además de instar a la Comisión a revisar y adaptar la legislación europea a la nueva realidad desde una perspectiva ética, concretaba nuevos principios para complementar la legislación e insistía en la aprobación “de una carta ética de buenas prácticas para la IA”⁶⁶. Dos de sus resoluciones de 2020 reconocieron la insuficiencia del marco ético para la protección eficaz de los derechos fundamentales frente a la IA, en especial la privacidad, la autonomía y el normal desarrollo de una sociedad democrática en el “ecosistema digital”⁶⁷, para los que proponía aprobar un reglamento con obligaciones legales uniformes para todos los Estados⁶⁸. En

⁶⁴ Consciente de que las normas jurídicas no eran suficiente, se afirma que es “preciso un marco ético claro, estricto y eficiente que oriente el desarrollo, diseño, producción, uso y modificación de los robots” (Resolución citada, n. 11), marco que “debe basarse en los principios de beneficencia, no maleficencia, autonomía y justicia, así como en los principios consagrados en la Carta de los Derechos Fundamentales de la Unión Europea, como la dignidad humana, la igualdad, la justicia y la equidad, la no discriminación, el consentimiento informado, la vida privada y familiar y la protección de datos, así como en otros principios y valores inherentes al Derecho de la Unión, como la no estigmatización, la transparencia, la autonomía, la responsabilidad individual, y la responsabilidad social (n. 13).

⁶⁵ Resolución, de 12 de febrero de 2019, *sobre una política industrial global europea en materia de inteligencia artificial y robótica* (2018/2088(INI)). Incluye recomendaciones para la modificación del marco jurídico en los nn. 114-137.

⁶⁶ Resolución citada, n. 139. El marco ético “debe basarse en los principios de beneficencia, no maleficencia, autonomía y justicia, así como en los principios consagrados (...) como la dignidad humana, la igualdad, la justicia y la equidad, la no discriminación, el consentimiento informado, la vida privada y familiar y la protección de datos, así como en otros principios y valores inherentes al Derecho de la Unión, como la no estigmatización, la transparencia, la autonomía, la responsabilidad individual, y la responsabilidad social, sin olvidar las actuales prácticas y códigos éticos” (n. 147).

⁶⁷ Resolución de 20 de octubre de 2020, sobre la Ley de servicios digitales y las cuestiones relacionadas con los derechos fundamentales (2020/2022(INI)). Estima necesario “garantizar un ecosistema digital justo en el que se respeten los derechos fundamentales consagrados (...), en particular la libertad de expresión y de información, la no discriminación, la libertad y el pluralismo de los medios de comunicación y la protección de la intimidad y de los datos personales” (n. 1).

⁶⁸ Es necesario “un marco regulador operativo y plenamente armonizado en el ámbito de las tecnologías de IA; sugiere que dicho marco adopte la forma de un reglamento y no de una directiva”. Resolución 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)), n. 3. En su Considerando E afirma la necesidad de “adoptar un enfoque de la IA centrado en el ser humano que respete los principios éticos y los derechos humanos”.

las últimas Resoluciones, sin abandonar la idea del marco ético, el Parlamento promueve claramente una regulación legal de la realidad emergente proponiendo los textos de dos Reglamentos⁶⁹.

Por su parte, la Comisión Europea siguió derroteros muy similares, apostando primero por un marco ético con la aprobación del *Libro Blanco sobre la Inteligencia Artificial*⁷⁰ en el que, admitiendo los beneficios que reporta la IA en todos los sectores, reconocía también riesgos para los derechos humanos como “la discriminación de género o de otro tipo, la intromisión en nuestras vidas privadas o su uso con fines delictivos”⁷¹, de ahí la necesidad de asentar las políticas sobre valores indiscutidos y promover el uso ético de la IA mediante la exigencia de requisitos esenciales, incorporando a los ya contemplados en el marco legal europeo (solidez técnica y seguridad, gestión de la privacidad y de datos, no discriminación y equidad, bienestar social y medioambiental y rendición de cuentas) otros requisitos esenciales como la transparencia, el seguimiento y la supervisión humana para proteger “la libertad de expresión, la libertad de reunión, la dignidad humana, la ausencia de discriminación por razón de sexo, raza u origen étnico, religión o credo, discapacidad, edad u orientación sexual, y, en su aplicación en determinados ámbitos, la protección de los datos personales y de la vida privada, el derecho a una tutela judicial efectiva y a un juicio justo, o la protección de los consumidores”⁷². En realidad, el Libro Blanco propone

⁶⁹ Esta combinación de marcos ético y jurídico se aprecia especialmente en su Resolución de 20 de enero de 2021, sobre inteligencia artificial: cuestiones de interpretación y de aplicación del Derecho internacional en la medida en que la UE se ve afectada en los ámbitos de los usos civil y militar, así como de la autoridad del Estado fuera del ámbito de la justicia penal (2020/2013(INI)). Tanto para el ámbito civil como militar, exige el respeto de los principios consagrados por la Carta de los Derechos Fundamentales y la Carta de Naciones Unidas, “así como por el Derecho internacional, por los principios de los derechos humanos y el respeto de la dignidad humana y por un entendimiento común de los valores universales y de los derechos inviolables e inalienables de la persona, la libertad, la democracia, la igualdad y el Estado de Derecho” (n. 2). Es la Resolución más exhaustiva en lo que se refiere a los principios éticos que deben inspirar la IA.

⁷⁰ Comunicación, de 19 de febrero de 2020, *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado a la excelencia y la confianza* (COM(2020)0065).

⁷¹ *Libro Blanco sobre la IA*, cit., p. 1.

⁷² *Libro Blanco sobre la IA*, cit., p. 13. También propone, en aras de la seguridad jurídica, establecer la obligación de “supervisión humana como garantía, desde la fase de diseño y a lo largo de todo el ciclo de vida” de la IA, “obligaciones explícitas para los productores con relación a los riesgos para la salud mental de los usuarios”, obligaciones relacionadas con “los riesgos derivados de los datos incorrectos en la fase de diseño, así como mecanismos para garantizar que la calidad de los datos se mantenga mientras se usen los productos y sistemas de IA”, requisitos de transparencia para evitar la “opacidad de los sistemas basados

un nuevo marco regulador desde la perspectiva del riesgo que someta los sistemas de IA a requisitos legales obligatorios para su aprobación y uso cuando entrañen un riesgo elevado para los derechos humanos⁷³.

3.2. *El cambio: propuestas legislativas en marcha*

El Parlamento Europeo no se conformó en 2020 con apuntar los riesgos más preocupantes de la IA e instar a respetar los principios éticos, sino que también propuso a la Comisión la aprobación de dos textos legislativos (reglamentos) de aplicación directa en los Estados de la UE y la adaptación de otros ya vigentes.

El primer Reglamento⁷⁴ propone un marco regulador de la IA con la conversión de principios éticos en obligaciones jurídicas, porque “los principios éticos comunes solo son eficaces cuando están también asentados en Derecho”⁷⁵ y porque las orientaciones éticas son un buen punto de partida, pero no garantizan que los desarrolladores, desplegados y usuarios actúen de manera justa ni aseguran la protección eficaz de las personas⁷⁶. Con dicho marco legal debería lograrse el respeto de la dignidad humana, la autodeterminación de la persona, la prevención de daños, la equidad, la inclusión y transparencia, la eliminación de discriminación y sesgos, así como el servicio de la tecnología a la persona⁷⁷ y convertiría en obligación legal la seguridad, transparencia y rendición de cuentas de los sistemas, así como la evaluación previa de riesgo los derechos de la persona, en conformidad con “todos los regímenes jurídicos aplicables, en particular el Derecho internacional humanitario y

en algoritmos”, obligaciones relacionadas con “programas autónomos comercializados separadamente o descargados en un producto” tras su comercialización y obligaciones de “colaboración entre los agentes económicos de la cadena [de suministro] y los usuarios”. *Libro Blanco sobre la IA*, cit., p. 19.

⁷³ Para el resto de sistemas de IA de riesgo no elevado, las normas deberían establecer un sistema de control y etiquetado voluntario con información para los usuarios sobre la seguridad y los posibles riesgos de los productos. Cf. *Libro Blanco sobre la IA*, cit., p. 29.

⁷⁴ Resolución 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 (2020/2012(INL)). Lo citaremos en adelante como Propuesta de Reglamento del marco regulador.

⁷⁵ Propuesta de Reglamento del marco regulador, Considerando Y.

⁷⁶ Propuesta de Reglamento del marco regulador, Considerando Z.

⁷⁷ Cf. Propuesta de Reglamento del marco regulador, n. 2. Insistirá en que la IA debe “adaptarse a las necesidades humanas, en consonancia con el principio según el cual su desarrollo, despliegue y uso deben estar siempre al servicio del ser humano y nunca al revés y deben tener por objeto aumentar el bienestar y la libertad individual, así como preservar la paz, prevenir los conflictos y reforzar la seguridad” (n. 10).

el Derecho internacional de los derechos humanos, y sea conforme a la legislación, los principios y los valores de la Unión”⁷⁸.

Una aportación muy importante del Reglamento es la concreción en un Anexo de sectores, usos y fines de *alto riesgo* para los derechos humanos, entre los que figuran los sectores del empleo, educación, asistencia sanitaria, transporte, energía, seguridad, defensa, finanzas y sector público (asilo, migración, control fronterizo, sistema judicial y servicios de seguridad social), recogiendo como usos y fines de alto riesgo la contratación, la clasificación y evaluación de personas, la asignación de fondos públicos, la concesión de préstamos, el comercio, los tratamientos y procedimientos médicos, los procesos electorales y campañas políticas, decisiones del sector público con impacto significativo y directo en los derechos, la conducción automatizada, la gestión del tráfico, los sistemas militares autónomos, la producción y distribución de energía, la gestión de residuos y el control de emisiones.

El segundo Reglamento propone el establecimiento de un régimen de responsabilidad civil, objetiva (arts. 4 a 7 del texto propuesto) y subjetiva (arts. 8 y 9), para que pueda ser reclamado cualquier daño, material e inmaterial, derivado del uso de la IA⁷⁹, incluyendo los derivados de “las violaciones de derechos importantes, jurídicamente protegidos, a la vida, la salud, la integridad física y la propiedad”⁸⁰. También prevé un anexo que deberá recoger los sistemas de alto riesgo y, por ello, sometidos a un régimen de responsabilidad objetiva, mientras que el resto seguiría sujeto a responsabilidad subjetiva⁸¹.

Sin embargo, el paso definitivo lo ha dado la Comisión Europea en abril de 2021 con la propuesta de la *Ley de Inteligencia Artificial*⁸², que establecería el marco

⁷⁸ Propuesta de Reglamento del marco regulador, n. 91. Ello implica la eliminación de sesgo y discriminación en los algoritmos, la garantía de la intimidad y protección de datos, la integridad física y psíquica, la libertad y autodeterminación personal, la libertad de conciencia y expresión, los derechos sociales y laborales, el derecho al medio ambiente, etc. Contempla la creación de una Autoridad europea y de autoridades nacionales encargadas de los controles y de la emisión de certificados de conformidad ética cuando los sistemas de IA cumplan con todas las exigencias.

⁷⁹ Resolución 20 de octubre de 2020, *con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial* (2020/2014(INL)). Lo citaremos en adelante como Propuesta de Reglamento de responsabilidad civil.

⁸⁰ Propuesta de Reglamento de responsabilidad civil, n. 19. Sobre esta materia, vid. Ramón Fernández, F., “Robótica, inteligencia artificial y seguridad: ¿cómo encajar la responsabilidad civil?”, *Diario La Ley* n° 9365, Sección Doctrina, 2019, pp. 1-13.

⁸¹ Propuesta de Reglamento de responsabilidad civil, Considerando 17 del texto de Reglamento propuesto.

⁸² Comisión Europea, Propuesta de Reglamento del Parlamento europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados Actos Legislativos de la Unión. COM(2021)

jurídico para garantizar una IA fiable, la seguridad de las personas y el respeto de sus derechos fundamentales. La propuesta parte de los riesgos evidentes que la IA supone para los derechos reconocidos en la Carta de Derechos Fundamentales de la UE, tales como el derecho a la dignidad humana (art. 1), el respeto de la vida privada y familiar y la protección de datos de carácter personal (arts. 7 y 8), la no discriminación (art. 21) y la igualdad entre hombres y mujeres (art. 23), la libertad de expresión (art. 11) y de reunión (art. 12), el derecho a la tutela judicial efectiva y a un juez imparcial, la presunción de inocencia y los derechos de la defensa (arts. 47 y 48), así como asegurar la promoción de derechos de determinados grupos, por ejemplo, las condiciones justas y equitativas del trabajador (art. 31), la protección del consumidor (art. 28), los derechos del niño (art. 24) y la integración de las personas discapacitadas (art. 26)⁸³.

Siguiendo la línea del Parlamento, la Comisión clasifica los sistemas de IA en atención al riesgo que generan, distinguiendo entre sistemas de riesgo inaceptable, de alto riesgo y de riesgo bajo o mínimo. Para los primeros establece su prohibición absoluta⁸⁴, recogiendo en esta categoría las prácticas con alto potencial para manipular a la persona con técnicas subliminales o que aprovechen las debilidades de grupos vulnerables (menores, personas con discapacidad, etc.) para alterar de manera sustancial su comportamiento, las que elaboren perfiles de personas físicas sin informarlas y sin obtener consentimiento previo, otras prácticas de manipulación, explotación o calificación social con fines generales y, salvo excepciones, el uso de sistemas de identificación biométrica remota *en tiempo real* en espacios de acceso público con fines de aplicación de la ley por las autoridades públicas⁸⁵.

206 final. Bruselas, 21.4.2021 (la citaremos en estas páginas como Propuesta de Ley de Inteligencia Artificial UE)

⁸³ Cf. Propuesta de Ley de Inteligencia Artificial UE, Considerando 28.

⁸⁴ Arts. 1.b) y 5 de la Propuesta de Ley de Inteligencia Artificial UE. Son contrarios a los valores de la Unión de respeto de la dignidad humana, libertad, igualdad, democracia y Estado de Derecho y de los derechos fundamentales que reconoce la UE, como el derecho a la no discriminación, la protección de datos y la privacidad, y los derechos del niño”. Propuesta de Ley de Inteligencia Artificial UE, Considerando 15.

⁸⁵ Propuesta de Ley de Inteligencia Artificial UE, Considerandos 16 a 23. La prohíbe porque las manipulaciones subliminales pueden originar perjuicios físicos y psíquicos, las calificaciones sociales pueden dar origen a un trato discriminatorio y a la exclusión social y la identificación biométrica en tiempo real causaría la sensación de vigilancia constante que podría disuadir a las personas de ejercer su libertad de reunión y otros derechos fundamentales. Para la prohibición de identificación biométrica en tiempo real contempla tres excepciones, la búsqueda de personas desaparecidas y víctimas de un delito, determinadas amenazas para la vida o seguridad de los ciudadanos o amenazas terroristas y la detección, localización, iden-

Para los sistemas de alto riesgo⁸⁶ para la salud, la seguridad o los derechos fundamentales ya citados, bien por sus componentes o por su finalidad, establece requisitos obligatorios en materia de datos, gobernanza, documentación, registro, transparencia, información a los usuarios, vigilancia humana, solidez, precisión y seguridad, cuyo cumplimiento deberá ser evaluado antes de su introducción y utilización⁸⁷. La propia propuesta normativa incluye los Anexos II y III en los que se especifican componentes o sistemas considerados de alto riesgo, a los que se podrán añadir otros en el futuro. Entre ellos figuran, además de los relacionados con la salud y los de identificación biométrica en general, los sistemas de gestión y el funcionamiento del tráfico rodado y el suministro de agua, gas, calefacción y electricidad, cuyos posibles fallos o defectos de funcionamiento pueden suponer un peligro a gran escala para la vida y la salud de las personas⁸⁸. También se incluyen los sistemas de IA que puedan perpetuar discriminaciones en la educación, la formación profesional y el entorno laboral u originar sesgos en el acceso a instituciones educativas, en la evaluación de las personas, en la gestión de trabajadores y acceso al empleo, en las decisiones de promoción o rescisión de contratos, en la asignación de tareas y seguimiento de personas, etc.⁸⁹ Por idénticas razones se incluyen los sistemas para evaluar la calificación crediticia o solvencia y los que adopten decisiones que establezcan prioridades de servicios de intervención (policía, médicos, bomberos, etc.) en situaciones críticas de emergencia⁹⁰. Por último, contempla también los sistemas empleados por las autoridades públicas para la aplicación de la ley que impliquen grave desequilibrio de poder y riesgo alto para los derechos del ciudadano, por ejemplo, en el campo de la administración de justicia, en la gestión de la migración, el asilo y control fronterizo, en los procesos democráticos, etc.⁹¹

Para el resto de sistemas, de bajo o mínimo riesgo⁹², se establece una serie de obligaciones de transparencia de los propios sistemas (por ejemplo, advertencia de que se interactúa con un sistema de IA, que el sistema reconoce emociones o datos biométricos, etc.) y para los operadores de los mismos⁹³. En estos supuestos se insta

tificación o enjuiciamiento de los autores o sospechosos de delitos susceptibles de penas superiores a tres años, debiendo obtenerse en estos casos la autorización judicial previa si es posible.

⁸⁶ Arts. 1.c) y 6 a 15 de la Propuesta de Ley de Inteligencia Artificial UE.

⁸⁷ Arts. 16 a 29 y 52 de la Propuesta de Ley de Inteligencia Artificial UE. Se establecen obligaciones para proveedores, fabricantes, importadores, distribuidores y usuarios.

⁸⁸ Propuesta de Ley de Inteligencia Artificial UE, Considerando 34.

⁸⁹ Propuesta de Ley de Inteligencia Artificial UE, Considerandos 35 y 36.

⁹⁰ Propuesta de Ley de Inteligencia Artificial UE, Considerando 37.

⁹¹ Propuesta de Ley de Inteligencia Artificial UE, Considerandos 38-40.

⁹² Art. 1.d) de la Propuesta de Ley de Inteligencia Artificial UE.

⁹³ Propuesta de Ley de Inteligencia Artificial UE, Considerando 70.

a los proveedores de los sistemas a aprobar códigos de conducta con los que se comprometan voluntariamente a cumplir requisitos, bien los ya establecidos para los sistemas de alto riesgo u otros relativos “por ejemplo, a la sostenibilidad medioambiental, la accesibilidad para las personas con discapacidad, la participación de las partes interesadas en el diseño y el desarrollo de sistemas de IA, y la diversidad de los equipos de desarrollo”⁹⁴.

4. Una propuesta: una necesidad

Los riesgos para la dignidad humana derivados del uso de la IA no sólo resultan evidentes, sino que crecen cada día que pasa con las innovaciones tecnológicas en un clima de opacidad, de ahí que numerosos Estados hayan comenzado ya a modificar su legislación interna para introducir limitaciones con carácter vinculante. Pero no basta con unas legislaciones internas que pretendan únicamente tranquilizar a los ciudadanos.

Si todos los seres humanos estuviéramos comprometidos con el respeto de las exigencias éticas universales, los riesgos de la IA no deberían preocuparnos, pero desgraciadamente no es así en una sociedad en la que imperan criterios utilitaristas y económicos. La Recomendación de la UNESCO supone un paso significativo, pero insuficiente. Creemos que en la actualidad son los Instrumentos internacionales vinculantes los que constituyen el único bastión de garantía de los derechos humanos, por lo que habría que acometer de forma inmediata una Convención internacional sobre diseño, creación y aplicación de la IA que sirva de freno a las actuaciones de particulares, empresas y, en especial, de los poderes públicos en clara vulneración de los derechos humanos y que debilitan el Estado de Derecho y la democracia. Para su elaboración podría servir de modelo la Recomendación, pero también la línea seguida por la UE con la prohibición de sistemas de IA que impliquen un riesgo inaceptable y el establecimiento de requisitos exigentes para los sistemas de alto riesgo.

De igual modo, sería de vital importancia, en un campo tan dinámico por las continuas innovaciones, la creación de una Relatoría Especial y de un Comité de expertos en el marco de Naciones Unidas que velaran por el cumplimiento de las exigencias contenidas en la Convención, pudieran hacer un seguimiento del cumplimiento de la misma y atender las reclamaciones por vulneración de los derechos

⁹⁴ Propuesta de Ley de Inteligencia Artificial UE, Considerando 81. El control de cumplimiento de la normativa estaría encomendado conjuntamente al futuro Comité Europeo de Inteligencia Artificial y a las autoridades independientes que deben ser creadas en cada Estado. Arts. 56 a 59 de la Propuesta de Ley de Inteligencia Artificial UE.

humanos mediante sistemas de IA. Sin embargo, no está de más que sigan desarrollándose marcos éticos que complementen el legal y que podrían servir de orientación en aspectos de riesgo más difíciles de concretar.

Por último, creemos acertado mantener algo importante en lo que ambas propuestas analizadas coinciden. En primer lugar, que se trabaje con una concepción de la IA siempre al servicio de la humanidad y no al revés, de modo que se evite cualquier trato indigno hacia la persona. Y, en segundo lugar, que nunca se reconozca personalidad jurídica a los sistemas de IA, de modo que sea siempre posible hacer responsable de la vulneración de derechos humanos a personas (físicas o jurídicas) y que respondan por ello⁹⁵.

Bibliografía

Alonso, A., Carbó, J.M., *Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation*. Banco de España. Documento de Trabajo N° 2105, Madrid, 2021

Álvarez Rubio, J., “Inteligencia Artificial y protección jurídica de los consumidores”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Univ. de Alcalá, Alcalá de Henares, 2020, pp. 275-333

Andrés Segovia, B., “El reinicio tecnológico de la Inteligencia Artificial en el servicio público de salud”, *Ius et Scientia* vol. 7, n° 1, 2021, pp. 327-356

Anzalone, A., *Lo humano de los derechos humanos*. Dykinson, Madrid, 2021

Cárcar Benito, J.E., “La Inteligencia Artificial (IA): aplicación jurídica y regulación en los servicios de salud”, *Derecho y Salud (DS)* vol. 29, n° extra 1, 2019, pp. 265-277.

Comisión Iberoamericana de Ética Judicial (CIEJ), Noveno Dictamen, *Sobre el uso por el juez de las nuevas tecnologías: ventajas y desafíos éticos*, 12 de marzo de 2020

Corral Hernández, D., “Inteligencia Artificial en la batalla. Un guerrero preciso y letal sin alma ni corazón”, *Documento de Opinión IEEE* (M° de Defensa) 79, 2021, pp. 1-16

Cotino Hueso, L., “Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales”, *Itemata* 24, 2017, pp. 131-150

⁹⁵ Algunos órganos de la UE propusieron en un principio reconocer personalidad a los sistemas de IA. Los últimos documentos abogan por lo contrario. Vid., por ejemplo, Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial (2020/2014(INL)), n. 7. En el mismo sentido se pronuncia la UNESCO en el *Anteproyecto de Recomendación sobre la Ética de la Inteligencia Artificial*, cit., n. 68, al afirmar que La responsabilidad debe “recaer siempre en última instancia en personas físicas o jurídicas y no se debe otorgar personalidad jurídica a los sistemas de IA”. Vid. sobre esta cuestión Lacruz Mantecón, L., *Robots y personas: una aproximación jurídica a la subjetividad cibernética*. Reus, Madrid, 2020.

Cotino Hueso, L., “Ética en el diseño para el desarrollo de una Inteligencia Artificial, robótica y big data confiables y su utilidad desde el Derecho”, *Revista Catalana de Dret Públic* 58, 2019, pp. 29-48

De Asís Pulido, M., “La incidencia de las nuevas tecnologías en el derecho al debido proceso”, *Ius et Scientia* vol. 6, nº 2, 2020, pp. 186-199

De Asís, R., *Una mirada a la robótica desde los derechos humanos*. Marcial Pons, Madrid, 2015

García San José, D., “Implicaciones jurídicas y bioéticas de la Inteligencia Artificial (IA). Especial consideración al marco normativo internacional”, *Cuadernos de Derecho Transnacional* vol. 13, nº 1, 2021, pp. 255-276

González Arencibia, M., Martínez Cardero, D., “Dilemas éticos en el escenario de la Inteligencia Artificial”, *Economía y Sociedad* vol. 25, nº 57, 2020, pp. 1-17

Juan, G.R., “Inteligencia Artificial y Filosofía del Bioderecho: una tesis crítica y una propuesta ética”, *Ius et Scientia* vol. 6, nº 2, 2020, pp. 96-110

Lacruz Mantecón, L., *Robots y personas: una aproximación jurídica a la subjetividad cibernética*. Reus, Madrid, 2020

Llano Alonso, F.H. (ed.), *Inteligencia Artificial y Derecho. El jurista ante los retos de la era digital*. Cizur Menor, Aranzadi, 2021

Pelayo González-Torres, A., “TIC, Inteligencia Artificial y crisis de la democracia”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Univ. de Alcalá, Alcalá de Henares, 2020, pp. 55-78

Petit, M., “Por una crítica de la razón algorítmica. Estado de la cuestión sobre la inteligencia artificial, su influencia en la política y su regulación”, *Quaderns del CAC* 44, vol. XXI, 2018, pp. 5-15

Piedra Alegría, J., “El despertar de los robots (sociales)”, en Bermúdez Vázquez, M., (coord.), *Luces en el camino: filosofía y ciencias sociales en tiempos de desconcierto*. S.I., Dykinson, 2021, pp. 29-51

Ramón Fernández, F., “Robótica, inteligencia artificial y seguridad: ¿cómo encajar la responsabilidad civil?”, *Diario La Ley* nº 9365, Sección Doctrina, 2019, pp. 1-13

Robles Carillo, M., “La gobernanza de la Inteligencia Artificial: contexto y parámetros generales”, *REEI (Revista Electrónica de Estudios Internacionales)* 39, 2020, pp. 1-27

Rodríguez Puerto, M.J., “¿Puede la Inteligencia Artificial interpretar normas jurídicas? Un problema de razón práctica”, *Cuadernos Electrónicos de Filosofía del Derecho (CEFD)* 44, 2021, pp. 74-96

Sánchez Martínez, M.O., “Desafíos democráticos en el ecosistema digital”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Universidad de Alcalá, Alcalá de Henares, 2020, pp. 79-119

San Miguel Caso, C., “La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales?”, *Ius et Scientia* vol. 7, nº 1, 2021, pp. 286-303

Santos González, M.J., “Regulación legal de la robótica y la Inteligencia Artificial: retos de futuro”, *Revista Jurídica de la Universidad de León* 4, 2017, pp. 25-50

Solar Cayón, J.I., “Inteligencia Artificial en la justicia penal: los sistemas algorítmicos de evaluación de riesgos”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Universidad de Alcalá, Alcalá de Henares, 2020, pp. 125-172

Téllez Carbajal, E., “Derechos humanos, ética y transparencia algorítmica”, *Ius et Scientia* vol. 7, nº 1, 2021, pp. 370-386

Terrones Rodríguez, A.L., “Inteligencia Artificial y ética de la responsabilidad”, *Cuestiones de Filosofía* vol. 4, nº 22, 2018, pp. 141-170

Tomillo Urbina, J., “La responsabilidad por la prestación de servicios a través de plataformas de intermediación *on line*”, en Solar Cayón, J.I. (ed.), *Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de Derecho*. Universidad de Alcalá, Alcalá de Henares, 2020, pp. 333-364

Jerez de la Fra., octubre de 2021