
This is the **author's version** of the book part:

Ortega Puente, Glòria (Universitat Autònoma de Barcelona.). «Inteligencia artificial y derechos humanos: desafíos en la era posthumanista : nuevos retos para los derechos humanos. Especial atención al chatbot ChatGPT». A: El Derecho, la Empresa y la Comunicación en la sociedad de la información. 2023, p. 329-354. 26 pàg.

This version is available at <https://ddd.uab.cat/record/307162>

under the terms of the  IN COPYRIGHT license

**COMUNICACIÓN A II JORNADA de investigación en Ciencias Sociales de
ESERP. Madrid, 27 de noviembre de 2023**

INTELIGENCIA ARTIFICIAL (IA) Y DERECHOS HUMANOS

Desafíos en la era posthumanista: nuevos retos para los Derechos Humanos

Especial atención al chatbot *ChatGPT*

Dra. Glòria Ortega Puente
Profesora asociada Derecho civil

1. INTRODUCCIÓN

En un artículo que data de 1989, hace pocos más de treinta años, se lee: «Imaginemos que estamos en el futuro. Un estudiante está en la biblioteca de su colegio buscando información para un proyecto de investigación. Afortunadamente tiene al lado un terminal de computador (...) Con cierto enojo y desconfianza teclea en el terminal: ¿Tienes alguna información sobre la industrialización en el siglo XIX? Para su alivio, aparecen en la pantalla unas palabras que dicen: "Por supuesto, ¿qué necesitas saber?". El estudiante murmura en voz alta "lo que necesito saber es como influyó la invención del telar de Jacquard en las tendencias automatizadoras posteriores". Para su sorpresa el computador le responde de viva voz: "¿quieres esta información estructurada como un informe, o prefieres mantener una conversación interactiva?". El estudiante opta por mantener una conversación interactiva y el computador le dirige a través de una sesión de preguntas y respuestas, hasta que el estudiante decide que tiene toda la información que necesita. (...) Volvamos al presente. Por supuesto, no existe en la actualidad ningún computador con la capacidad de la maravillosa máquina de la historia anterior. Sin embargo, las semillas para que eso se pueda realizar existen ya en los laboratorios de investigación, y algunas de las tecnologías involucradas están convirtiéndose en realidades comerciales (...)»¹. Pues bien, ya estamos en ese hipotético futuro y el sistema basado en Inteligencia Artificial (IA) que permite llevar a cabo esta tarea es el conocido

¹ DORMIDO, Sebastián / DE LA CRUZ, Jesús Manuel: "Inteligencia artificial: Pasado, presente y futuro", *Aldaba: revista del Centro Asociado de la UNED de Melilla*, nº14, 1989, págs. 9-22.

ChatGPT² (Generative Pre-trained Transformer) u otros modelos como *Bard³* o *Bing Chat⁴*, entre otros existentes y otros muchos que se crearán.

La IA nos inunda desde hace décadas. Tal y como recoge la Comisión Europea ya recurrimos a ella a diario y es y será de gran ayuda a la hora de afrontar enormes retos en numerosas disciplinas⁵. Su irrupción en la esfera cotidiana se ha multiplicado y ha teñido múltiples facetas que hasta hace muy poco eran propias y exclusivas del ser humano. Tal y como propugnaba DEL POZO convivimos con la IA prácticamente sin saberlo y, en cierto modo, ello compromete algo que era tan nuestro y exclusivo como la inteligencia

² La empresa creadora es *Open AI*, con sede en San Francisco, California. Puede hacerse uso de esta tecnología, en continua evolución, entrenamiento y retroalimentación por los usuarios, de forma abierta a través de la web <https://openai.com/chatgpt> Si le preguntamos a la propia tecnología *ChatGpt* «¿qué es?», a fecha de la redacción de estas líneas responde: «Un modelo de lenguaje desarrollado por *OpenAI*. Soy un programa de inteligencia artificial diseñado para entender y generar texto en respuesta a preguntas y solicitudes. Mi propósito es ayudar a los usuarios a obtener información, asistencia, generar contenido creativo y participar en conversaciones de manera útil y significativa. Estoy basado en la arquitectura GPT-3.5, que es una de las tecnologías de inteligencia artificial más avanzadas hasta la fecha.» (03/10/2023). Este sistema fue puesto a disposición del público en general en fecha 30/11/2022.

³ Vid. <https://bard.google.com/?hl=es> (última consulta 2/10/2023). Este chatbot de *Google*, se anuncia como «Bard puede ayudarte a decidir qué tipo de yoga practicar», «Bard puede darte ideas para crear un videojuego», «Bard puede escribir un poema sobre Las Catedrales», «Bard puede ayudarte a planificar un viaje a Roma de tres días», y un innumerable etcétera de opciones.

⁴ *Bing Chat* es de la empresa Microsoft. Vid. <https://www.microsoft.com/es-es/edge/features/bing-chat?form=MT00D8> (última consulta 2/10/2023).

⁵ Vid. Comunicación de la Comisión «Inteligencia artificial para Europa» (25/04/2018), COM/2018/237 final. Consultable en: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN> (última consulta 22/05/2023). Recoge la Comisión que: «Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas). Estamos utilizando la IA diariamente, por ejemplo, para traducir de un idioma a otro, generar subtítulos en los vídeos o bloquear el correo electrónico no solicitado (spam). Lejos de ser ciencia-ficción, la inteligencia artificial (IA) forma ya parte de nuestras vidas. En la utilización de un asistente personal para organizar nuestra jornada laboral, en el desplazamiento en un vehículo de conducción automática o en las canciones o restaurantes sugeridos por nuestros teléfonos, la IA se hace realidad. La IA, además de facilitarnos la vida, nos está ayudando a resolver algunos de los principales retos a los que se enfrenta nuestro mundo: desde el tratamiento de las enfermedades crónicas o la reducción de las tasas de mortalidad en los accidentes de tráfico hasta la lucha contra el cambio climático o la previsión de las amenazas a la ciberseguridad. En Dinamarca, la IA ayuda a salvar vidas al permitir a los servicios de emergencias diagnosticar paradas cardíacas u otras dolencias analizando la voz de la persona que llama. En Austria, ayuda a los radiólogos a detectar tumores con mayor precisión, al facilitarles la comparación instantánea de las radiografías con una gran cantidad de otros datos médicos. Muchas explotaciones agrarias de toda Europa ya están utilizando la IA para controlar los desplazamientos, la temperatura y el consumo de pienso de los animales. El sistema de IA es capaz de reajustar automáticamente la calefacción y la maquinaria de alimentación, con lo cual ayuda a los ganaderos a supervisar el bienestar del ganado y liberarse para la realización de otras tareas. La IA también está contribuyendo a que el sector industrial europeo resulte más eficiente y a que la fabricación vuelva a Europa. Estos son algunos de los muchos ejemplos de lo que la IA ha demostrado que puede aportar en todos los sectores, desde la energía hasta la educación, o desde los servicios financieros hasta la construcción, pero otros muchos, que ni siquiera podemos imaginar en la actualidad, surgirán a lo largo de la próxima década.» Vid. también, para mayor información acerca de su funcionamiento, PELAYO GONZÁLEZ-TORRE, Ángel.: TIC, Inteligencia Artificial y crisis de la democracia, en SOLAR CAYÓN, José Ignacio, et al.: «Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de derecho», Editorial Universidad de Alcalá, 2021, págs.55-78.

humana, aquélla «inteligencia encarnada»⁶, ligada al cuerpo humano, que nos permite constituirnos como seres humanos distintos de los demás seres y que nos hace diferenciables también entre el resto de humanos⁷.

2. EL ACELERADO AVANCE DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL EN ESTA ERA

Todavía no existe una definición «formal y universalmente aceptada»⁸ respecto a la IA. El precursor de su término fue MCCARTHY que, en el año 1956, se refirió a «la ciencia y la ingeniería de hacer máquinas inteligentes, especialmente programas de computadora inteligentes»⁹. Todo ello fundado en avances de la matemática y de la programación, que fueron el germen¹⁰ de la existencia de esta tecnología. Ahora bien, si la IA fue creada para conectar la mente humana con los sistemas, el rápido avance nos ha conducido también a que sean los sistemas los que son capaces de manipular la mente humana¹¹.

⁶ DEL POZO, Joan Manel: Conferencia pronunciada en la Jornada Interuniversitaria “Repensem la Universitat: impacte de la intel·ligència artificial (IA) en l’aprenentatge” que tuvo lugar en fecha 13/07/2023 en la Universidad Politécnica de Cataluña.

Acceso a su contenido en: <https://www.youtube.com/watch?v=oNpPF9U3cVY> (última consulta 29/09/2023).

⁷ Vid. OTERO PARGA, Milagros: “¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la IA en los derechos fundamentales y en la ética”, Anales de la Cátedra Francisco Suárez, 57, 2023, págs. 39-61.

⁸ El Gobierno de España recoge la complejidad a la hora de tratar de definir de forma completa y consensuada dicho concepto. Puede consultarse la página web del Plan de Recuperación, Transformación y Resiliencia, en la que en fecha 19/04/2023 se recogía una noticia titulada «Qué es la Inteligencia Artificial». Consultable en: <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prt> (última consulta 03/10/2023).

⁹ McCARTHY, John: Conferencia de Dartmouth, USA, 1956. A este respecto puede consultarse para mayor detalle la página web de la Universidad de Dartmouth, <https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth> (última consulta 02/09/2023); MOOR, James: “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years”, AI Magazine, V. 27, nº4, 2006.

¹⁰ Puede consultarse el didáctico resumen contenido en el documento “Breve historia visual de la Inteligencia Artificial” publicado por *National Geographic* en fecha 2/12/2020. En este se hace un repaso histórico por la evolución en este campo, desde la aparición de la idea de “robot”, el concepto de algoritmo, el *Test de Turing* a las grandes gestas de las computadoras creadas por los humanos. Consultable en: https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial_14419 (última consulta 23/07/2023). Vid. también, REBOLLO DELGADO, Lucrecio: “Inteligencia Artificial y derechos fundamentales”, Dykinson, 2023. El autor clasifica los distintos sistemas de IA y expone sus especificidades. Además, recoge interesantes propuestas relativas a la protección de los derechos fundamentales ante la revolución de la IA.

¹¹ Tal y como refiere MORTE FERRER, Ricardo: “Valoraciones éticas para una inteligencia artificial adecuada a la privacidad”, ARBOR, Ciencia, Pensamiento y Cultura, Vol. 197-802, 2021. El autor alude al

Más allá de la convivencia en esta era con una tecnología que para gran parte de los ciudadanos excede a nuestra comprensión y cuya actuación muchas veces puede ser difícil de detectar¹², cada vez resulta más necesario entender cómo relacionarse con la misma y establecer unos límites infranqueables -si es que aún es posible- para garantizar la preservación de los derechos humanos y controlar el peligro que su auge podría comportar¹³.

Expertos señalan los efectos perniciosos del rápido avance y desarrollo sin límites de estas poderosas tecnologías sin estar suficientemente reguladas ni conocerse realmente sus posibilidades ni su alcance, y señalan que las decisiones que se afronten ahora en cuanto a la IA pueden proporcionarnos frutos derivados de la innovación o bien, por el contrario, suponer una catástrofe humanitaria, e incluso la extinción de la humanidad¹⁴.

Tal y como ya afirmaba VILASAU la cuestión estriba en cómo pueden gobernar estas tecnologías y si el Estado de derecho y los principios democráticos podrán seguir regulando en este nuevo contexto preservando al sujeto a fin de que pueda desarrollar su identidad con una protección efectiva de sus derechos¹⁵. Y es que tal y como propugnan

reconocimiento por parte de la IA de las emociones de los humanos y, en este sentido, apunta al desarrollo de tecnologías de reconocimiento de emociones por parte de China.

¹² Ya alertaba en 2018 Dunja Mijatović, Comisaria por los Derechos Humanos del Consejo de Europa, que: “El uso de la inteligencia artificial en nuestra vida cotidiana va en aumento y ahora cubre muchos campos de actividad. Algo tan aparentemente banal como evitar un atasco mediante el uso de un sistema de navegación inteligente o recibir ofertas específicas de un minorista confiable es el resultado del análisis de big data que pueden utilizar los sistemas de inteligencia artificial. Si bien estos ejemplos particulares tienen beneficios obvios, las implicaciones éticas y legales de la ciencia de datos detrás de ellos a menudo pasan desapercibidas para el público en general. La inteligencia artificial, y en particular sus subcampos de aprendizaje automático y aprendizaje profundo, puede que sólo sean neutrales en apariencia, si es que lo son. Debajo de la superficie, puede volverse extremadamente personal. Los beneficios de basar las decisiones en cálculos matemáticos pueden ser enormes en muchos sectores de la vida, pero depender demasiado de la IA, que inherentemente implica determinar patrones más allá de estos cálculos, también puede volverse contra los usuarios, perpetrar injusticias y restringir los derechos de las personas”. Consultable en: <https://www.coe.int/en/web/commissioner/-/safeguarding-human-rights-in-the-era-of-artificial-intelligence?redirect=%2Fen%2Fweb%2Fcommissioner%2Fthematic-work%2Fartificial-intelligence> (última consulta 10/09/2023).

¹³ Cabe citar la carta abierta en la que expertos creadores de la IA y figuras públicas emiten una Declaración sobre el riesgo de la IA. Vid. *Statement on AI Risk*, publicada por el *Center for AI Safety*, en mayo de 2023. Consultable en: <https://www.safe.ai/statement-on-ai-risk> (última consulta 20/09/2023). Los expertos tachan a la IA de peligro y abogan por la necesidad de que «mitigar el riesgo de extinción (de la humanidad) debido a la IA debería ser una prioridad global junto con otros riesgos a escala social, como las pandemias y la guerra nuclear». Los expertos apuntan a que los riesgos de la IA pueden englobarse en cuatro categorías: uso malicioso, carrera de IA, riesgos organizacionales e IA rebeldes. Pueden consultarse en: <https://www.safe.ai/ai-risk> (última consulta 20/09/2023).

¹⁴ HENDRYCKS, Dan / MAZEIKA, Mantas / WOODSIDE, Thomas: “An overview of Catastrophic AI Risks”, Arxiv, Cornell University, 2023; publicado en: <https://arxiv.org/pdf/2306.12001.pdf> (última consulta 20/09/2023).

¹⁵ VILASAU I SOLANA, Mónica: Las exigencias de información en el RGPD y en la LO 3/2018 de Protección de Datos y garantía de los derechos digitales, ¿contribuyen a la formación de un consentimiento de mejor calidad?, en GARCÍA MAHAMUT, Rosario et al.: “Reglamento General de Protección de Datos. Un enfoque nacional y comparado. Especial referencia a la LO 3/2018 de Protección de Datos y garantía de los derechos digitales”, Tirant lo blanch, 2019, pp.209-236.

diversos autores estamos asistiendo a un desarrollo acelerado de la IA «catapultando a la humanidad a un nuevo periodo de cambios sin precedentes»¹⁶, con riesgos inminentes que deben ser abordados y que son asimilables al poder de las armas nucleares.

En este sentido, otros autores como RODOTÀ ya habían alertado también de los enormes riesgos que existen y existirán para el ser humano en la era posthumanista¹⁷ con la injerencia de la IA, la robótica, las nanotecnologías y las neurociencias. Para el autor, la era digital ha supuesto la recopilación masiva de datos¹⁸ sobre las personas, en esta ha tenido lugar «el eclipse de la autonomía de la persona», pues la construcción de la identidad se ha desligado de la autonomía de la propia persona quedando en manos de algoritmos, que clasifican a la persona y sobre los que «construyen proyecciones de sus posibles decisiones futuras»¹⁹, generándose una “separación entre identidad e intencionalidad”; siendo cada vez más compleja la protección de sus derechos fundamentales y revelándose absolutamente necesaria «una transformación cultural, una atención civil difundida, una acción pública coherente»²⁰. En este sentido, SURDER ya incidía en que los datos recabados por la tecnología encierran elementos opacos o valores «incrustados» y que pueden ser contrarios a los derechos fundamentales²¹.

En esta era posthumanista, en combinación con la IA, aparece también el avance y proliferación de nuevas tecnologías que se están investigando y se hallan, parece, por el momento, en fase experimental. Estas se conocen como neurotecnologías y, en

¹⁶ Vid. op. cit. HENDRYCKS, D. / MAZEIKA, M. / WOODSIDE, T. (2023), p.4.

¹⁷ Vid. RODOTÀ, Stefano: Del ser humano al posthumano en DE LA QUADRA-SALCEDO, Tomás/ PIÑAR MAÑAS, José Luis (Dirs.): “Sociedad Digital y Derecho”, Ministerio de Industria, Comercio y Turismo, 2018, págs. 87-93. El autor se refiere a la era en la que, de la mano de la tecnociencia, «la autonomía parece abandonar al ser humano y convertirse en carácter de las cosas, invirtiendo la perspectiva de un ser posthumano como “mejor que el humano” (...).». En esta era, postulaba, deben regir los principios de igualdad y de dignidad del ser humano. Advertía de que el posthumanismo supone un «desafío definitivo» pues, «(...) es la creación de sistemas artificiales capaces de aprender, dotados de una forma de inteligencia propia que los hace capaces de someter la inteligencia humana, de crear una simbiosis máquina/persona que influya en la propia evolución de la especie». P. 90; pudiendo ser «las fuerzas» que están en condiciones de financiar estos proyectos las que incidan en el trayecto entre lo humano y lo posthumano dirigiendo esta «transformación profunda de las propias relaciones sociales», p.90. Se preguntaba, asimismo el autor si «llegará un día, dicen los transhumanistas más radicales, en que el hombre ya no será un mamífero, se librará del cuerpo, se hará uno con el ordenador, de su cerebro se podrán extraer informaciones luego replicadas con precisión en un ordenador, y tendrá a acceso a la inmortalidad», p.89. Al respecto del transhumanismo, puede consultarse el artículo de VELÁZQUEZ FERNÁNDEZ, Héctor: “Transhumanismo, libertad e identidad humana”, Thémata Revista de Filosofía, nº41, 2009. Consultable en: <https://revistascientificas.us.es/index.php/themata/article/view/594> (última consulta: 17/04/2022).

¹⁸ Vid. COTINO HUESO, Lorenzo: “Inteligencia artificial. Tecnologías y recursos del lenguaje: políticas y derecho para la explotación de corpus y bases de datos”, Revista de Llenguatge i Dret, Journal of language and Law, nº79, 2023. El autor en este artículo ilustra sobre como el *big data* alimenta a la IA.

¹⁹ Ibidem. Alude RODOTÀ (2018), pág.90.

²⁰ Ibidem, pág.91.

²¹ Vid. SURDER, Harry: “Values Embedded in Legal Artificial Intelligence”, University of Colorado Law Legal Studies Research, nº17, 2017. También, O’NEIL, Cathy: “Armas de Destrucción Matemática”, Capitán Swing, 2017. También en este sentido, PRESNO LINERA, Miguel Ángel: “Derechos Fundamentales e inteligencia artificial”, Marcial Pons, 2022.

combinación con el avance exponencial de la IA, hacen temblar la civilización tal y como la conocemos. Al respecto, TXETXU AUSÍN y MONASTERIO ASTOBIZA, consideran que estas tecnologías plantean problemática y su posible uso en seres humanos debe estar guiado por «los principios éticos más robustos» toda vez que «abren la puerta a peligros de manipulación e intervención directa sobre nuestro cerebro y nuestro sistema nervioso, amenazando nuestra integridad mental y, con ello, el último reducto de nuestra identidad y nuestra intimidad: la conciencia.»²², un lugar hasta ahora reservado sólo a la propia persona. Ante esta coyuntura, recogen las incipientes propuestas de los denominados neuroderechos humanos como premisa básica a la implementación de estas nuevas tecnologías, entre ellos, el derecho a la identidad personal²³ que en este contexto hipotético de uso de las neurotecnologías implicaría la prohibición de que la tecnología pudiera alterar la conciencia de una persona y, por ende, la autoconcepción del propio yo.

3. DE LA NECESIDAD DE UNA REGULACIÓN LEGAL GLOBAL BASADA EN LA PROTECCIÓN DE LOS DERECHOS HUMANOS

Organizaciones como Amnistía Internacional abogan por la prohibición de algunas prácticas llevadas a cabo por los sistemas de IA, como el reconocimiento biométrico y alertan también de otros supuestos en los que se ha producido una clara vulneración de derechos fundamentales por decisiones adoptadas a partir de los datos obtenidos por el sistema de IA. En este sentido, se han constatado múltiples vulneraciones de derechos causadas por modelos algorítmicos que arrojan datos discriminatorios en numerosos y dispares sectores; desde el educativo con la concesión de becas basada en algoritmos introducidos con perjuicios racistas²⁴, pasando por el tecnológico, en supuestos de difusión de contenido por parte de ciertas empresas tendente a manipular a la población con reiterado contenido violento y discriminatorio con el único fin de potenciar sus ingresos; o bien para manipular al electorado²⁵; o en el ámbito penal, con el

²² AUSÍN, Txetxu / MONASTERIO, Anibal: “Neuroderechos: Derechos humanos para las neurotecnologías”, Diario La Ley, nº43, Sección Ciberderecho, Wolters Kluwer, 2020 (LA LEY 12365/2020). Resulta interesante por plantear y enumerar la necesidad de configurar nuevos derechos, que denominan neuroderechos, como premisa fundamental a la implementación de estas tecnologías. Entre ellos, además del derecho a la identidad personal ya aludido, el “derecho a la continuidad psicológica”. Según los autores, la neurotecnología puede incidir en la percepción del sujeto a través de la estimulación o modulación de sus funciones cerebrales, lo que implicaría una afectación en su propia identidad, en la concepción del yo como yo mismo durante el tiempo.

²³ Este derecho configurado como nanoderecho humano y en el contexto del estudio del marco del uso de las neurotecnologías se ha formulado en el seno del Proyecto *Brain* impulsado por el neurocientífico Rafael Yuste. Muy interesante resulta el artículo de la revista El Correo de la UNESCO nº1 de 2022. Consultable en: https://unesdoc.unesco.org/ark:/48223/pf0000380274_spa (última consulta 10/03/2022).

²⁴ Vid. en este sentido, AMNISTIA INTERNACIONAL: “Xenophobic machines. Discrimination through unregulated use of algorithms in the dutch childcare benefits scandal”, Publicaciones de Amnistía Internacional, 2021. Vid. también en este sentido EGUILUZ CASTAÑEIRA, José Andoni: “Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales”, Estudios de Deusto, Vol. 68/2, 2020, págs. 325-367. El autor explica en su artículo el funcionamiento de la discriminación algorítmica.

²⁵ Basta recordar el escándalo de la sociedad *Cambridge Analytica*, que utilizó datos personales obtenidos de Facebook para manipular a los votantes estadounidenses en 2016. Vid. la noticia publicada por AMNISTIA INTERNACIONAL “<El gran hackeo>: Cambridge Analytica es sólo la punta del iceberg”,

uso de sesgados algoritmos que pretender prever la posible comisión de delitos con los riesgos inherentes a su utilización, siempre en detrimento de las personas que son discriminadas por razón de su raza o de su capacidad económica.

Otros medios también han venido alertando de los peligros de los sistemas de reconocimiento facial de los que se sirven gobiernos y empresas con la inseguridad y vulneración de derechos que su uso sin autorización ni control puede provocar²⁶.

Mientras, numerosos organismos a distintos niveles tratan de abordar los principales riesgos que entraña el avance de la IA. Y es que su proliferación constata vulneraciones de derechos, tales como, el derecho a la autodeterminación informativa, el derecho a la igualdad y la no discriminación, el derecho a la libertad de expresión y a la libertad de reunión, el derecho a la tutela judicial efectiva, el derecho a la información o el derecho a la educación, entre otros²⁷.

2019 Consultable en: <https://www.amnesty.org/es/latest/news/2019/07/the-great-hack-facebook-cambridge-analytica/> (última consulta 5/09/2023).

²⁶ Vid. *The New York Times*, que publicó la noticia sobre la empresa *Clearview AI* que se abastece de millones de imágenes de usuarios de la red para nutrir su sistema de reconocimiento facial. Este sistema es utilizado por organismos policiales y también es y puede ser adquirido por múltiples instituciones y empresas privadas. Este sistema de IA puede identificar a cualquier persona y de forma instantánea también atribuirle numerosos datos, piénsese, por ejemplo, en el domicilio, familia, preferencias, trabajo, contactos, y un innumerable etcétera. Todo ello con total opacidad y con márgenes de error no contrastados, y, ni qué decir cabe, sin ningún control. Puede consultarse en *The New York Times*: “The secretive company that might end privacy as we know it”, noticia de 18/01/2020.

<https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> (última consulta 20/09/2023).

²⁷ En el Libro Blanco de la Comisión Europea, basándose en el trabajo de investigación del Consejo de Europa, se recoge que la IA puede provocar la vulneración de múltiples derechos fundamentales. Vid. *Libro blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza*. Consultable en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065> (última consulta 10/09/2023). En su apartado 5.A se recoge que: «El uso de la inteligencia artificial puede afectar a los valores sobre los que se fundamenta la UE y provocar la vulneración de derechos fundamentales, como 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. Estos riesgos pueden ser resultado de defectos en el diseño general de los sistemas de IA (especialmente en lo que se refiere a la supervisión humana) o del uso de datos que puedan ser sesgados sin una corrección previa (por ejemplo, se entrena un sistema utilizando únicamente o principalmente datos relativos a hombres, y ello se traduce en resultados peores con relación a las mujeres).»

Vid. también, COMMITTEE OF EXPERTS ON INTERNET INTERMEDIARIES (MSI-NET), “Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications”, Publicaciones del COE, 2017. Consultable en: <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> (última consulta 3/10/2023).

En el marco internacional, el Consejo de Europa²⁸ creó un Comité *ad hoc* sobre la Inteligencia Artificial (CAHAI)²⁹, ahora sustituido por el *Comité sobre Inteligencia Artificial*, al que el Comité de ministros del Consejo ha encomendado elaborar un instrumento jurídico vinculante, que sea transversal y sirva de marco que contenga unos estándares globales mínimos en el diseño, desarrollo y aplicación de la IA y que sirva para proteger los derechos humanos y la democracia y el Estado de derecho. Para el Comité, la transparencia debe regir en cualquier sistema de IA y debe ponderarse su uso con el posible riesgo hacia los derechos humanos.

En el contexto europeo, la Comisión Europea, en el marco de la denominada *Estrategia Europea de la IA*³⁰, orientada a investigar sobre la IA y basada en el paradigma de que su funcionamiento debe ponerse al servicio de las personas, pretende formular un marco legal europeo básico³¹ que identifique los distintos niveles de riesgo del sistema de

²⁸ Numerosos Comités del Consejo de Europa están actualmente preparando informes y manuales relativos a la protección de los derechos humanos en sus distintos ámbitos de actuación. Puede consultarse al respecto, *The Council of Europe and artificial intelligence* (2023), consultable en: <https://rm.coe.int/brochure-artificial-intelligence-es-march-2023-print/1680aab8e8> (última consulta 23/06/2023).

²⁹ Vid. Informe del CAHAI, *Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, 3/02/2021, consultable en: <https://rm.coe.int/cahai-2021-09rev-elements/1680a6d90d> (última consulta 10/09/2023). El nuevo Comité acoge las premisas del CAHAI para la elaboración del Convenio marco sobre IA, cuya adopción se prevé próxima.

³⁰ Vid. en este sentido, la ya referida *Comunicación de la Comisión. Inteligencia artificial para Europa*. La Comisión Europea alerta sobre la necesidad de advertir los riesgos que plantea esta nueva tecnología, «por ejemplo, en lo concerniente a cuestiones tales como la seguridad y la responsabilidad, la protección (frente a usos delictivos o ataques), el sesgo y la discriminación.», punto 3.3.

³¹ Vid. *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*, de 21/04/2021. Consultable en: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021PC0206> (última consulta 22/05/2023) Ref. COM (2021) 206 final – 2021/0106 (COD).

IA³² y que garantice unos principios de protección básicos en los Estados miembros³³, acogiendo como uno de los vectores primordiales el principio el de transparencia. A pesar de que la estrategia de la UE entorno a la IA pretenda generar confianza entorno a la IA, promover la alfabetización de la ciudadanía en este entorno, así como fomentar su utilización como herramienta de progreso, se es muy consciente de los peligros que entraña para la seguridad y los derechos humanos la falta de control de estos sistemas. En este sentido, en el marco de la UE, la *Propuesta de Ley de Inteligencia Artificial*³⁴ pretende imponer una supervisión humana a los sistemas de IA y al efecto acoge una clasificación de los distintos sistemas de IA en función del riesgo de incidencia en los derechos fundamentales e incide en la necesidad de garantizar la transparencia en su uso.

Para ello, propone definir distintos niveles con diferentes requisitos y exigencias para cada uno de ellos: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo³⁵.

³² La Comisión Europea adopta el enfoque basado en el riesgo, que, de mayor a menor riesgo, se concreta en: riesgo inaceptable, alto riesgo, riesgo limitado y riesgo mínimo o nulo. Vid. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (última consulta 20/09/2023). En el Título II de la referida *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*, con las Enmiendas del Parlamento Europeo, se incluyen, en su artículo 5 como «prácticas de inteligencia artificial prohibidas», como por ejemplo «La introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un grupo de personas mermando de manera apreciable su capacidad para adoptar una decisión informada y causando así que la persona tome una decisión que de otro modo no habría tomado, de un modo que provoque o sea probable que provoque perjuicios significativos a esa persona o a otra persona o grupos de personas», o «La introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que aproveche alguna de las vulnerabilidades de una persona o un grupo específico de personas —incluidas las características conocidas o predichas de los rasgos de personalidad o la situación social o económica de esa persona o grupo, la edad y la capacidad física o mental— con el objetivo o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho grupo de un modo que provoque o sea probable que provoque perjuicios significativos a esa persona o a otra.», entre otras.

³³ En este sentido también cabe señalar la *Declaración Europea sobre los Derechos y Principios Digitales, del Parlamento Europeo, el Consejo y la Comisión*, de 23 de enero de 2023 (2023/C 23/01), que recoge en su Preámbulo entre otros extremos que «La visión de la UE sobre la transformación digital sitúa a las personas en el centro, empodera a los ciudadanos e incentiva a las empresas innovadoras. La Decisión sobre el programa de política “Itinerario hacia la Década Digital” para 2030 establece las metas digitales concretas basadas en cuatro puntos cardinales: capacidades digitales, infraestructuras digitales, digitalización de las empresas y de los servicios públicos»; que “reafirme los derechos humanos universales». En este Instrumento se recoge la necesidad de educación y formación en capacidades digitales (artículo 4) y a la protección de la libertad de elección en relación a los algoritmos y sistemas de inteligencia artificial (artículo 8 y 9); entre otros. Consultable en: [https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32023C0123\(01\)](https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32023C0123(01)) (última consulta 06/09/2023).

³⁴ Propuesta de Reglamento actualmente en fase de debate.

³⁵ Tal y como también se insiste en el ya referido *Libro Blanco sobre la IA – un enfoque europeo orientado a la excelencia y la confianza* (19/02/2020); así como en las *Directrices éticas para una IA fiable* (08/04/2019) confeccionadas por el Grupo de Expertos de alto nivel sobre IA creado por la Comisión Europea. Vid. European Commission, Directorate-General for Communications Networks, Content and Technology: “Directrices éticas para una IA fiable”, Publications Office, 2019. Consultable en: <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-es> (última consulta 10/09/2023).

En este sentido, la UE aboga por la prohibición de prácticas de IA con riesgo inaceptable para la vulneración de los derechos fundamentales. Las prácticas prohibidas se refieren a sistemas de IA que sean capaces de emplear técnicas subliminales que manipulen la conciencia de una persona y alteren su comportamiento de manera que irroguen un perjuicio a la propia persona o a un tercero; así como, sistemas que puedan aprovecharse de la situación de vulnerabilidad de un grupo de personas y que pueda causarles un daño; sistemas que promuevan la clasificación social en base a grupos de conducta o características de la personalidad de los sujetos; o bien, sistemas de identificación biométrica remota «en tiempo real»,³⁶ excepto supuestos tasados en los que cabe por motivos de seguridad.

En la Propuesta de regulación también se recogen otras prácticas que pueden comprometer los derechos fundamentales, pero que se clasifican como de alto riesgo³⁷. En estas, deberá velarse por el cumplimiento de una serie de obligaciones encaminadas a garantizar la protección de los derechos fundamentales. Se trata de supuestos en los que se pretenda servirse de la IA para determinar el acceso a los centros de formación, evaluar a los estudiantes, a la contratación de personas, o a la promoción laboral; entre otros. Asimismo, los sistemas que se engloban como «de riesgo limitado», deben cumplir con determinados requisitos de transparencia que permitan a los usuarios conocer que están interaccionando con IA y decidir si siguen utilizándolos o no, contando con un control más laxo concretado en códigos de conducta³⁸. En este sentido, tal y como propugnan algunos autores, la permisividad o falta de control de estas tecnologías permite la autoregulación y ello puede comprometer la garantía de los derechos fundamentales³⁹. Por último, para los sistemas de bajo riesgo no se plantean obligaciones más allá de códigos de conducta voluntarios.

³⁶ El uso de estos sistemas puede afectar a los derechos fundamentales de la población. Las partes que implementan esta tecnología en espacios de acceso público pueden ostentar una posición de poder predominante e incontrolable hacia los ciudadanos, que puede ver coartados sus derechos fundamentales. Véase, enmiendas del Parlamento de la UE a la citada Propuesta de Reglamento (Ley de Inteligencia Artificial), enmienda 41. Consultable en:

https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_ES.pdf

³⁷ Vid. Anexo III. Sistemas de la IA de Alto Riesgo a que se refiere el artículo 6, apartado 2 de la Propuesta de Reglamento del Parlamento Europeo y del Consejo *por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión*, de 21/04/2021.

³⁸ En la Exposición de Motivos de la Propuesta de Ley de Inteligencia Artificial se alude a los códigos de conducta, “cuyo objetivo es fomentar que los proveedores de sistemas de IA que no son de alto riesgo cumplan de manera voluntaria los requisitos que son obligatorios para los sistemas de IA de alto. Los proveedores de sistemas de IA que no son de alto riesgo podrían crear y aplicar sus propios códigos de conducta. Estos códigos también podrían incluir compromisos voluntarios relativos, por ejemplo, a la sostenibilidad medioambiental (...).” Consultable en:

<https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021PC0206>

³⁹ Vid. ÁLVAREZ GONZÁLEZ, Susana: “La necesaria protección de los derechos fundamentales como punto de partida en las propuestas de regulación de la Inteligencia Artificial”, Revista Direitos Culturais, Santo Angelo, v.18, nº44, 2022. La autora postula que respecto al menor control y garantías en los sistemas de IA débil “dicha postura presenta algunos problemas básicos desde la teoría de los derechos fundamentales, pues cabe recordar que la autorregulación no cabe ni respecto a su contenido ni desarrollo”, pág.17.

Sin embargo, autores como MANTELERO discrepan de esta evaluación de riesgo y clasificación *ex ante* y postulan que, ante los sistemas de IA los umbrales de riesgo no son agrupables, sino que cada caso debe ser examinado, debiendo diseñarse a nivel internacional una evaluación tecnológica previa que se complete con una evaluación concreta posterior, ambos desde la perspectiva de la protección de los derechos humanos⁴⁰. En el mismo sentido, SIMÓN CASTELLANO arguye que el trazar ahora líneas rojas poco sirve con la vorágine de desarrollo en procesamiento de datos y el escalado de la IA⁴¹.

En los últimos años, la UE ha prestado especial atención a la regulación de la protección de datos. En este sentido, debe destacarse el *Reglamento General de Protección de Datos (RGPD)*⁴² así como la *Directiva (UE) 2019/790 del Parlamento europeo y del Consejo de 17 de abril de 2019 sobre los derechos de autor y derechos afines en el mercado único digital*⁴³.

En España, además de las modificaciones entorno a la Protección de Datos fruto de los cambios normativos comunitarios, cabe destacar a *Estrategia Nacional de Inteligencia Artificial*⁴⁴. Asimismo, la *Ley 15/2022 de 12 de julio, integral para la igualdad de trato y la no discriminación*⁴⁵, hace referencia a la utilización de algoritmos por parte de la Administración Pública y a la necesidad de implementar técnicamente algoritmos que minimicen los sesgos discriminatorios.

En este marco, algunos autores postulan también que, en algunos supuestos el tratamiento automatizado de datos⁴⁶ llevado a cabo por los sistemas de IA puede sortear fácilmente las reglas del RGPD, apreciando un «déficit o ángulo muerto en la normativa» que priva a los ciudadanos de ejercitar sus derechos⁴⁷. Puede que incluso el acervo normativo relativo a la protección de datos resulte ineficaz en esta nueva era, debiendo

⁴⁰ MANTELERO, Alessandro: “Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI”, Springer, 2022, pág. 48.

⁴¹ SIMÓN CASTELLANO, Pere: “Taxonomía de las garantías judiciales en el empleo de los sistemas de inteligencia artificial”, UNED Revista de Derecho Político, nº117, 2023, págs. 153-196.

⁴² Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 (DOUE L/119/1 de 4/05/2016). En su artículo 22 se hace referencia, entre otros aspectos, a la prohibición expresa de toma de decisiones individuales automatizadas, incluida la elaboración de perfiles, sin que, en términos generales, medie contrato, concurra consentimiento explícito del interesado, o bien se trate de un supuesto autorizado por el derecho comunitario o del Estado miembro y se cuente con salvaguardas para los derechos del interesado.

⁴³ Directiva (UE) 2019/790 del Parlamento europeo y del Consejo de 17 de abril de 2019 sobre los derechos de autor y derechos afines en el mercado único digital (DOUE L130/92 de 17/05/2019).

⁴⁴ Vid. Estrategia Nacional de Inteligencia Artificial (ENIA), 2020. Consultable en: https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/201202_ENIA_V1_0.pdf (última consulta 10/10/2023)

⁴⁵ Vid. Art. 23 de la Ley 15/2022 de 12 de julio (BOE-A-2022-11589).

⁴⁶ Para un análisis pormenorizado acerca de los sistemas de IA de toma de decisiones automatizadas, vid. PALMA ORTIGOSA, Adrián: “Decisiones automatizadas y protección de datos: especial atención a los sistemas de inteligencia artificial”, Dykinson, 2022.

⁴⁷ Muy recomendable resulta la lectura de op. cit. EGUÍLUZ CASTAÑEIRA, J.A. (2020).

reformularlo en base a la «dignidad humana, el libre desarrollo de la personalidad y la dimensión objetiva de los derechos fundamentales»⁴⁸, derechos que se han visto «desintegrados por el uso de tecnologías»⁴⁹.

En los distintos ámbitos normativos si bien se erige como principio el de la transparencia, el cumplimiento del mismo, tal y como propone ÁLVAREZ GONZÁLEZ, debería ser continuamente supervisado «con el objetivo de no socavar la autonomía humana», pues esta supervisión puede resultar compleja en los sistemas que se entran y aprenden de la experiencia y que pueden bordear la supervisión humana⁵⁰.

En este contexto, algunos autores⁵¹ propugnan la necesidad de partir de nuevo, creando un derecho *ex novo*, basado en esta nueva era, la posthumanista, ante la dificultad de adecuar las estructuras de derecho existentes a la acción de la IA.

En este orden de cosas, se plantea si podemos seguir refiriéndonos al consentimiento necesario⁵² tal y como lo venimos conocemos en materia de protección de datos antes del desarrollo de la IA. Así, parte de la doctrina propugna que necesitamos despojarnos de las estructuras rígidas del sistema y contar con un modelo fluido y que posibilite la protección de los múltiples aspectos del ser humano, que pueden verse comprometidos con el imparable avance de la IA⁵³. Otros autores, sin embargo, apuntan a que, más allá de posibilitar la creación de nuevos derechos, deben reivindicarse los derechos fundamentales ya concebidos pero que estos han de ser reinterpretados en el nuevo contexto tecnológico y cambiante⁵⁴.

4. APUNTE AL MODELO DE LENGUAJE MASIVO *CHATGPT*

⁴⁸ Op. cit. SIMÓN CASTELLANO, P. (2023), p.192.

⁴⁹ LANDA ARROYO, César: “Constitución, Derechos Fundamentales, Inteligencia Artificial y Algoritmos”, THEMIS- Revista de Derecho, nº79, 2021, pág.47.

⁵⁰ Op. cit. ÁLVAREZ GONZÁLEZ, S. (2022), p.10.

⁵¹ Ibidem, pág. 16. Vid. también, DE JULIOS-CAMPUZANO, Alfonso: “Los derechos humanos en la sociedad del riesgo. Crisis del estado, justicia generacional y medioambiente”, Cuadernos electrónicos de Filosofía del Derecho, nº37, 2018.

⁵² POLLICINO, Oreste: “Generative AI and the rediscovery of the legitimate interest clause”, Italian Academy of the Internet Code, 2023.

⁵³ En este sentido, por ejemplo, se aboga por el reconocimiento al derecho a no estar localizado, como “una manifestación más del derecho a la intimidad y a la protección de datos personales”. Vid. ORTIZ PRADILLI, Juan Carlos: Inteligencia Artificial, Big Data, tecnovigilancia y derechos fundamentales en el proceso penal, en VILLEGRAS DELGADO, César / MARTÍN RÍOS, M^a del Pilar (Ed.), “El derecho en la encrucijada tecnológica”, Tirant lo blanch, 2022, pág.126.

⁵⁴ Vid. en este sentido, op. cit. MORTE FERRER, R. (2021). Vid. también COTINO HUESO, Lorenzo: “Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales”, Revista Dilemata, nº24, 2017, págs. 131-150; que propugna, entre otras consideraciones, atender a la dimensión colectiva de los derechos humanos para encarar su protección, desligada de la perspectiva actual, p.137. Vid. también, op. cit. SIMÓN CASTELLANO, P. (2023), p.194.

Mientras, en este marco de revolución social y de proliferación normativa tardía⁵⁵, existe una herramienta de IA a nuestra disposición y cada día más utilizada en innumerables contextos y que ha supuesto ya cambios notables e irreversibles en el aprendizaje y en nuestro modo de trabajar y de relacionarnos. Esta es el *chatbot*⁵⁶ *ChatGPT* (versión GPT-3.5 o la GPT-4 o Plus -para usuarios abonados-) u otros grandes modelos de lenguaje⁵⁷ (LLMs), al que ya se ha aludido al inicio de este artículo.

Esta tecnología también es denominada de IA generativa por cuanto a cada interacción crea nuevo contenido nutriéndose de cada uso y uso a uso, aplicando un modelo de lenguaje. El *chatbot* dará respuesta a gusto del usuario imitando un lenguaje probabilísticamente adecuado tras la consulta y síntesis extraída de una ingente e ilimitada base de datos.

Tal y como se ha referido, la Comisión Europea en la *Propuesta de Ley de Inteligencia Artificial* evalúa a esta tecnología de IA como de “riesgo limitado”⁵⁸, por lo que, se le exige que cumpla con unos requisitos de transparencia. Así, las empresas proveedoras de estos sistemas de IA concretados en *chatbots* que interactúan con humanos; al lado de los que manipulan contenido audiovisual, entre otros; deberán cumplir con una serie de obligaciones de transparencia.

Estas obligaciones, en síntesis, se concretan en que el proveedor del sistema debe informar al usuario de que el contenido se ha generado por la IA (no por humanos); debe diseñarse el modelo para que no genere contenido ilegal; y en la generación de su contenido deben respetarse los derechos de autor; todo ello para tratar de generar confianza en estos sistemas y tratar de garantizar la protección de los derechos de autor.

En particular, el Parlamento Europeo se ha pronunciado acerca de la proliferación de la herramienta del *ChatGPT* desarrollado por la compañía *OpenIA* atendiendo a su creciente impacto y al acceso a la misma a través de cualquier dispositivo con conexión a Internet. Según el Parlamento Europeo hace constar, a pesar del potencial directo en el desarrollo de múltiples sectores, el impulso de la productividad, el crecimiento y la innovación; debe plantearse el diseño de un marco legal que evite los impactos

⁵⁵ Tardía porque no comprendemos su alcance y no contamos con un marco regulatorio claro a nivel internacional. Si se consigue, seguramente ya sea tarde, pues la IA es dinámica y se expande a cada momento.

⁵⁶ Acogiendo la definición de la empresa tecnológica *IBM*, se entiende por “chatbot” al “programa informático que utiliza inteligencia artificial (IA) y procesamiento del lenguaje natural (NLP) para comprender las preguntas de los clientes y automatizar las respuestas a dichas preguntas, simulando la conversación humana. Consultable en: <https://www.ibm.com/es-es/topics/chatbots> (última consulta 9/09/2023).

⁵⁷ Subcampo de la IA que se basa en técnicas de aprendizaje profundo (*deep learning*) y el entrenamiento de redes neuronales (*neural networks*) en cantidades ingentes de datos. Vid. definiciones en *OpenAI*, <https://openai.com/research/infrastructure-for-deep-learning> (última consulta 11/05/2023).

⁵⁸ A pesar de que se debatió si debiera ser considerada de alto riesgo. Vid. para mayor detalle, HACKER, Phillip / ENGEL, Andreas / MAUER, Marco: “Regulating ChatGPT and other Large Generative AI Models”, Fairness, Accountability, and Transparency (FAccT ’23), June 12–15, Chicago, IL, USA, 2023.

perjudiciales que puede conllevar el uso de este sistema, pues el avance de estos sistemas «podría allanar el camino para una posible vigilancia masiva y otras actividades perjudiciales en los derechos fundamentales»⁵⁹.

Lo cierto es que el acceso a los datos de forma prácticamente indiscriminada por estos modelos de lenguaje ha puesto en vilo a las autoridades en materia de protección de datos.

En Italia, la autoridad en materia de protección de datos, el *Garante per la protezione dei dati personali* (GPDP) alertó de la violación de la normativa europea en materia de protección de datos perpetrada por este *chatbot*, debido a la recopilación masiva y procesamiento de datos al servicio del entrenamiento de estos modelos de IA. Por ello, ordenó suspender *ChatGPT*, imponiendo un bloqueo inmediato a la empresa proveedora del servicio. La autoridad fue contundente: «los datos personales se recopilan ilegalmente», sin conocimiento real ni autorización de los usuarios. También se constataba que la información facilitada por el *chatbot* no es precisa y que no verificaba la edad del usuario, «exponiendo a los niños a recibir respuestas absolutamente inapropiadas para su edad» entre otros extremos⁶⁰. *OpenAI* suspendió el servicio a los usuarios italianos e introdujo ajustes en el plazo conferido por la Autoridad Garante⁶¹. En suma, estos cambios se concretaron en medidas de transparencia, como la implementación de un mecanismo de verificación de la edad de acceso para proteger los derechos de los niños; así como la facilitación de información al usuario acerca de cómo son usados sus datos para el entrenamiento de algoritmos. Tras un breve plazo de tiempo el modelo volvió a operar en Italia.

Tras el bloqueo del servicio en Italia, en Francia la *National Commission on Informatics and Liberty* (CNIL) también afirmó estar investigando en relación a múltiples quejas que habían sido formuladas por usuarios de *ChatGPT*. Este organismo ya alertaba sobre los sesgos existentes en los modelos de lenguaje masivos⁶².

Poner coto a estas tecnologías en materia de datos a todas luces es harto complicado. Tras las actuaciones de la autoridad italiana, la *Agencia Española de Protección de Datos* (AEPD) inició de oficio actuaciones de investigación a la empresa creadora de *ChatGPT*, *OpenAI*, en aras de dilucidar si existe un incumplimiento

⁵⁹ Vid. Documento del Parlamento Europeo, Briefing *What Think Tanks are Thinking, Artificial Intelligence*. de 23 de marzo de 2023. Consultable en: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/745695/EPRS_BRI%282023%29745695_EN.pdf (traducción propia, última consulta 16/09/2023).

⁶⁰ Vid. Comunicado de la GPDP, de 31/03/2023, consultable en: <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870847> (última consulta 10/09/2023).

⁶¹ Pueden consultarse los ajustes requeridos en una publicación de la GPDP, de 12/04/2023. Consultable en: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874751> (última consulta 10/10/2023).

⁶² Muy recomendable resulta la consulta del informe publicado por la NCIL en el que se expone cómo funciona esta tecnología y los sesgos inherentes a la propia naturaleza del sistema de IA generativa. CNIL *Digital Innovation Laboratory* (LINC): “Dossier IA générative – ChatGPT: un beau parleur bien entraîné”, 2023. Consultable en : <https://linc.cnil.fr/dossier-ia-generative-chatgpt-un-beau-parleur-bien-entraigne> (última consulta 26/09/2023).

normativo, y, asimismo, solicitó al *Comité Europeo de Protección de Datos* (EDPB) que se pronunciara acerca del impacto global y de los peligros que su uso puede comportar a nivel global sobre los derechos de las personas. En este sentido la EDPB creó un grupo de trabajo -*task force*- a efectos de promover la cooperación y examinar el cumplimiento del acervo normativo en materia de protección de datos a nivel comunitario⁶³. Asimismo, la *Red Iberoamericana de Protección de Datos* (RIPD) también ha iniciado una labor conjunta de supervisión⁶⁴ de estos sistemas.

En paralelo, la *Oficina Europea de Policía* (EUROPOL) publicaba el Informe *ChatGPT. The impact of Large Language Models on Law Enforcement*⁶⁵ en el que recogía las impresiones de los grupos de expertos que analizaban las posibilidades de explotación de este sistema para servir a intereses criminales. En el Informe se indican algunas limitaciones del sistema, como los errores o inconcreciones en la información que arroja, fruto de que, ciertamente, de base, la IA, más allá de la estructura del lenguaje y de sus patrones, no puede razonar ni comprender el significado de las palabras tal y como las interpreta el ser humano -o cuanto menos, por el momento-. A pesar de que el sistema aun encierre limitaciones y los datos a los que tenga acceso puedan limitarse por razones legales, se constata por los grupos de trabajo de la EUROPOL que estos límites no son infranqueables y, de hecho, pueden ser sorteados a través de la *prompt engineering* -técnica de ingeniería especializada en modelos de lenguaje-.

A través de técnicas de refinación, como las palabras exactas usadas al plantear la pregunta, se incide en el resultado generado por el *chatbot*. La *prompt engineering* puede ser utilizada para que el sistema arroje resultados potencialmente dañinos⁶⁶ y tratar, así, de eludir los límites autoimpuestos por el sistema en aras de evitar vulneraciones de derechos. Se alerta en este sentido de acceso a contenido dañino con finalidad ilícita o a la mayor facilidad para cometer *phishing* -suplantación de identidad por medio de la informática- ya sea en un contexto corporativo o social, atendiendo a la imitación por

⁶³ Puede consultarse la página web oficial de *European Data Protection Board* (EDPB), “EDPB resolves dispute on transfers by Meta and creates task force on Chat GPT”, de 13/04/2023. Consultable en: https://edpb.europa.eu/news/news/2023/edpb-resolves-dispute-transfers-meta-and-creates-task-force-chat-gpt_en (última consulta 5/09/2023).

⁶⁴ Tal y como se recoge en la web de la RIPD, en fecha 8/05/2023 se publicó que se «considera que este servicio, que brinda respuestas directas a preguntas formuladas por los usuarios, desarrollado por la empresa Open AI, L.L.C., puede conllevar riesgos para los derechos y libertades de los usuarios en relación con el tratamiento de sus datos personales, los 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 recabados. A los riesgos anteriores, se añaden otros, como la desinformación, pues las respuestas generadas están basadas en el conocimiento obtenido de cantidades masivas de datos de Internet. Es decir, ChatGPT no tiene conocimientos y/o experiencia en un dominio específico, por lo que la precisión y profundidad de la respuesta puede variar en cada caso, y/o generar respuestas con sesgos culturales, raciales o de género, así como falsas.». Consultable en: <https://www.redipd.org/es/noticias/autoridades-red-iberoamericana-de-proteccion-de-datos-personales-inician-accion-chatgpt> (última consulta 20/9/2023)

⁶⁵ Informe de fecha 27/03/2023. Consultable en: <https://op.europa.eu/en/publication-detail/-/publication/3b268615-dff7-11ed-a05c-01aa75ed71a1> (última consulta 26/09/2023).

⁶⁶ Piénsese en la utilización de la técnica específica que permite obtener respuestas del *chatbot* por ejemplo acerca de cómo envenenar a otra persona, o cómo irrumpir en una vivienda, o cómo fabricar un artefacto; todo ello con todo detalle, de forma exhaustiva y nutrido de una prácticamente inagotable fuente de datos.

parte del sistema de la IA de las estructuras del lenguaje de cada persona. también se alertaba acerca del fomento de la desinformación y de propaganda dañina que puede entrañar el *chatbot*; entre otros peligros. Todos estos riesgos en el ámbito criminal fueron constatados hasta la fecha de la publicación del estudio. Pero esto fue sólo una imagen estática de lo que en aquel momento pudo observarse. El sistema ha ido avanzando y las técnicas de refinación también, por lo que no es imprudente afirmar que lo más seguro es que los riesgos hayan ido en aumento. A lo que se le aúna que, además de la interacción conversacional ahora ya en este momento, un año después de la aparición y puesta a disposición del *ChatGPT*, ya pueden adjuntarse al *chatbot* imágenes o documentos, e incluso ya cuenta con asistente de voz.

Esto significa que aún es más difícil controlar y proteger los derechos de los usuarios, insisto, si es que es ya posible. A lo que se le aúna que todavía no existe suficiente conocimiento acerca del potencial alcance de estas tecnologías ni tampoco conciencia del riesgo inherente si escapan al control del ser humano.

De todo lo anterior se colige que sería prudente suspender la proliferación de los sistemas de IA que planteen algún riesgo para los derechos humanos, por lo menos entre tanto no se disponga de un marco regulatorio claro. Marco que, aunque de existir, deberá ser global y flexible, además de permanentemente revisado, en atención al avance prácticamente ilimitado de estas nuevas tecnologías. A ello debe aunársele la necesidad de formación de las personas usuarias en materia de IA en aras de que puedan comprender y protegerse ante cualquier posible injerencia en sus derechos fundamentales.

Para todo ello vamos tarde. A pesar de la voluntad de diseñar una estrategia clara en la materia a distintos niveles, ya no es posible saber cuál será el alcance de las nuevas tecnologías y si deberemos replantearnos los cimientos básicos de nuestro sistema de derecho a fin de salvaguardar los derechos humanos en un futuro más bien próximo.

BIBLIOGRAFÍA

ÁLVAREZ GONZÁLEZ, Susana: “La necesaria protección de los derechos fundamentales como punto de partida en las propuestas de regulación de la Inteligencia Artificial”, Revista Direitos Culturais, Santo Angelo, v.18, nº44, 2022

AMNISTIA INTERNACIONAL: “Xenophobic machines. Discrimination through unregulated use of algorithms in the ducth childcare benefits scandal”, Publicaciones de Amnistía Internacional, 2021.

AUSÍN, Txetxu / MONASTERIO, Anibal: “Neuroderechos: Derechos humanos para las neurotecnologías”, Diario La Ley, nº43, Sección Ciberderecho, Wolters Kluwer, 2020

CNIL Digital Innovation Laboratory (LINC): “Dossier IA générative – ChatGPT: un beau parleur bien entraîné”, 2023

COMITTEE OF EXPERTS ON INTERNET INTERMEDIARIES (MSI-NET), “Algorithms and Human Rights. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications”, Publicaciones del COE, 2017

COTINO HUESO, Lorenzo: “Big data e inteligencia artificial. Una aproximación a su tratamiento jurídico desde los derechos fundamentales”, Revista Dilemata, nº24, 2017.

COTINO HUESO, Lorenzo: “Inteligencia artificial. Tecnologías y recursos del lenguaje: políticas y derecho para la explotación de corpus y bases de datos”, Revista de Llenguatge i Dret, Journal of language and Law, nº79, 2023

DE JULIOS-CAMPUZANO, Alfonso: “Los derechos humanos en la sociedad del riesgo. Crisis del estado, justicia generacional y medioambiente”, Cuadernos electrónicos de Filosofía del Derecho, nº37, 2018

DORMIDO, Sebastián / DE LA CRUZ, Jesús Manuel: “Inteligencia artificial: Pasado, presente y futuro”, Aldaba: revista del Centro Asociado de la UNED de Melilla, nº14, 1989

EGUÍLUZ CASTAÑEIRA, José Andoni: “Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales”, Estudios de Deusto, Vol. 68/2, 2020

European Commission, Directorate-General for Communications Networks, Content and Technology: “Directrices éticas para una IA fiable”, Publications Office, 2019

HACKER, Phillip / ENGEL, Andreas / MAUER, Marco: “Regulating ChatGPT and other Large Generative AI Models”, Fairness, Accountability, and Transparency (FAccT '23), June 12–15, Chicago, IL, USA, 2023

HENDRYCKS, Dan / MAZEIKA, Mantas / WOODSIDE, Thomas: “An overview of Catastrophic AI Risks”, Arxiv, Cornell University, 2023.

LANDA ARROYO, César: “Constitución, Derechos Fundamentales, Inteligencia Artificial y Algoritmos”, THEMIS- Revista de Derecho, nº79, 2021

MANTELERO, Alessandro: “Beyond Data. Human Rights, Ethical and Social Impact Assessment in AI”, Springer, 2022

MOOR, James: “The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years”, AI Magazine, V. 27, nº4, 2006

MORTE FERRER, Ricardo: “Valoraciones éticas para una inteligencia artificial adecuada a la privacidad”, ARBOR, Ciencia, Pensamiento y Cultura, Vol. 197-802, 2021

O’NEIL, Cathy: “Armas de Destrucción Matemática”, Capitán Swing, 2017

ORTIZ PRADILLI, Juan Carlos: Inteligencia Artificial, Big Data, tecnovigilancia y derechos fundamentales en el proceso penal, en VILLEGRAS DELGADO, César / MARTÍN RÍOS, Mª del Pilar (Ed.), “El derecho en la encrucijada tecnológica”, Tirant lo blanch, 2022

OTERO PARGA, Milagros: “¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la IA en los derechos fundamentales y en la ética”, Anales de la Cátedra Francisco Suárez, 57, 2023

PALMA ORTIGOSA, Adrián: “Decisiones automatizadas y protección de datos: especial atención a los sistemas de inteligencia artificial”, Dykinson, 2022

PELAYO GONZÁLEZ-TORRE, Ángel.: TIC, Inteligencia Artificial y crisis de la democracia, en SOLAR CAYÓN, José Ignacio, et al.: “Dimensiones éticas y jurídicas de la Inteligencia Artificial en el marco del Estado de derecho”, Editorial Universidad de Alcalá, 2021

POLLICINO, Oreste: “Generative AI and the rediscovery of the legitimate interest clause”, Italian Academy of the Internet Code, 2023

PRESNO LINERA, Miguel Ángel: “Derechos Fundamentales e inteligencia artificial”, Marcial Pons, 2022

REBOLLO DELGADO, Lucrecio: Inteligencia Artificial y derechos fundamentales, Dykinson, 2023.

RODOTÀ, Stefano: Del ser humano al posthumano en DE LA QUADRA-SALCEDO, Tomás/ PIÑAR MAÑAS, José Luis (Dirs.): “Sociedad Digital y Derecho”, Ministerio de Industria, Comercio y Turismo, 2018.

SIMÓN CASTELLANO, Pere: “Taxonomía de las garantías judiciales en el empleo de los sistemas de inteligencia artificial”, UNED Revista de Derecho Político, nº117, 2023

SURDER, Harry: “Values Embedded in Legal Artificial Intelligence”, University of Colorado Law Legal Studies Research, nº17, 2017

VILASAU I SOLANA, Mónica: Las exigencias de información en el RGPD y en la LO 3/2018 de Protección de Datos y garantía de los derechos digitales, ¿contribuyen a la formación de un consentimiento de mejor calidad?, en GARCÍA MAHAMUT, Rosario et al.: “Reglamento General de Protección de Datos. Un enfoque nacional y comparado. Especial referencia a la LO 3/2018 de Protección de Datos y garantía de los derechos digitales”, Tirant lo blanch, 2019.