Inteligencia artificial: promesas, riesgos y regulación. ¿Algo nuevo bajo el sol?[1]

Luís Roberto Barroso

Presidente del Supremo Tribunal Federal. Profesor titular de la Universidadd del Estado do Río de Janeiro (Uerj) y del Centro Universitário de Brasília (Uniceub). Doctor y profesor asociado de la Uerj. Máster de la Yale Law School, EE. UU. *Visiting scholar* en la Facultad de Derecho de Harvard. *Senior fellow* en la Harvard Kennedy School, EE. UU.

Patricia Perrone Campos Mello

Secretaria de Estudios Avanzados del Supremo Tribunal Federal de Brasil. Profesora de Derecho Constitucional en la Uerj y el Uniceub. Doctora y máster en Derecho por la Uerj. Realizó estudios posdoctorales como *visiting scholar* en el Instituto Max Planck de Derecho Público Comparado e Internacional (Alemania) y en la Harvard Kennedy School, EE. UU. Procuradora del Estado de Río de Janeiro.

[1] Artículo publicado en la Revista Direito e Práxis, v. 15, n. 4, 2024.

Nota inicial[2]

Hay algo nuevo bajo el sol [3]. Muchas de nuestras creencias y certezas podrían tener los días contados. Así como los antiguos navegantes contemplaban la inmensidad de los océanos, llenos de promesas, misterios y peligros, nos enfrentamos de nuevo a un mundo desconocido. Se siente en el aire que se avecina una profunda transformación. Una revolución, quizás. Algo grandioso como la invención de la imprenta con tipos móviles, que difundió exponencialmente el conocimiento humano, o la Ilustración, que transformó la vida social, la cultura y la política (Kissinger; Schmidt; Huttenlocher, 2023). El futuro nunca ha parecido tan cercano e impredecible [4].

Ante las aparentemente infinitas posibilidades de la tecnología, solo existe una carta de navegación segura: los valores que desde hace mucho tiempo han guiado el avance de la civilización y la evolución de la condición humana en la Tierra. Ya sean seculares o míticos, provienen de Grecia e incluyen la Torá, los Evangelios, Buda, Tomás de Aquino, Kant y muchos otros que han forjado el patrimonio ético de la humanidad. Pero hay un punto dramático aquí: el vertiginoso progreso científico que hemos presenciado, acumulativamente, a lo largo de los siglos, no ha ido acompañado de una correspondiente evolución ética, e incluso espiritual, de la condición humana. La bondad, la justicia real y la solidaridad a menudo se descuidan en un mundo de extrema pobreza en muchos lugares, desigualdades injustas, guerras y un orden nacional e internacional en el que algunos ganan todo y otros siempre pierden. Es en este escenario que surge el tema de la inteligencia artificial (en adelante, IA) y su potencial para mejorar el mundo. O empeorarlo. O incluso aniquilarlo [5].

Quizás ningún tema en la historia de la civilización haya suscitado tanta reflexión simultánea. En los medios de comunicación, en los bares, en las universidades, en los grandes eventos internacionales y en las reuniones de expertos, un tema se ha vuelto omnipresente: la inteligencia artificial. No hay aspecto de sus implicaciones que no haya sido explorado por las mentes más brillantes y los ciudadanos más comunes. El siguiente texto forma parte de esta profusión de escritos que buscan captar el espíritu de la época, trazar rutas e impulsar la historia en la dirección correcta, evitando los abismos que pondrían en riesgo, si no nuestras vidas, al menos nuestra humanidad tal como la conocemos. La fe en la ciencia, como toda fe en este mundo, no puede conducir al fanatismo. Necesitamos definir rumbos y límites. Este es solo un intento más de hacerlo.

---

[2] Los autores agradecen a Pedro Henrique Ribeiro Morais e Silva por su ayuda con la investigación y sugerencias en la preparación del texto, y también a Frederico Alvim, por los importantes comentarios y recomendaciones bibliográficas.

[3] Hay un pasaje muy conocido en la Biblia, en Eclesiastés 1:9, que dice: «Lo que fue, eso será; lo que se hizo, se volverá a hacer. No hay nada nuevo bajo el sol». El significado de esta frase es que, con el paso de los siglos, los seres humanos lidian con las mismas preguntas existenciales. Sin embargo, quizás estén surgiendo algunas nuevas preguntas.

[4] Esta imprevisibilidad se explica, en parte, por la progresiva autonomía adquirida por las soluciones basadas en el aprendizaje automático y, en parte, por la naturaleza acelerada de las innovaciones en este campo, que se suceden (o se renuevan) a un ritmo que dificulta su comprensión completa. Nina Schick (2020, p. 11) señala que transcurrieron cuatro siglos entre la invención de la imprenta y el desarrollo de la fotografía, por ejemplo, pero en tan solo tres décadas, pasamos de la aparición de internet a los smartphones, y de ahí a la plataformización de las vidas en las redes sociales, con graves implicaciones para el régimen de la información. Según la autora, estos cambios tan rápidos en segmentos vitales conllevan un alto componente de incertidumbre, que debe ser considerado por la sociedad en su conjunto.

[5] La investigación en IA evita el alarmismo, lo que no significa que los desarrollos potencialmente catastróficos no se consideren hipótesis serias. Véase el trabajo de Stuart Russell sobre el tema, «Inteligencia artificial en nuestro nombre. Cómo mantener el control sobre la tecnología». Las preocupaciones más relevantes giran en torno a la inteligencia artificial general, también conocida como superinteligencia artificial, que define un estado en el que las computadoras superarían considerablemente las capacidades humanas, lo que daría lugar, según Kai-Fu Lee (2019, pp. 159-173), a «problemas de control» y «problemas de alineación».

Este artículo se estructura de la siguiente manera. La introducción presenta algunas nociones básicas sobre el tema. La Parte I explora el potencial positivo de la IA. La Parte II busca catalogar los principales riesgos que conlleva. La Parte III identifica algunos principios que deberían regir la regulación del tema. Y, finalmente, una conclusión busca disipar nuestras preocupaciones sobre el futuro.

CONTENIDO: Nota inicial; Introducción: El amanecer de la cuarta revolución industrial; 1. Un mundo feliz y nuevo; 2. Qué es la inteligencia artificial; 3. Aprendizaje automático, modelos fundacionales y otros conceptos relevantes; Parte I: La inteligencia artificial y sus beneficios; 1. Mejor capacidad de decisión en muchas áreas; 2. Automatización; 3. Lenguaje; 4. Investigación e innovación; 5. Aplicaciones en la medicina; 6. Aplicaciones en el sistema de justicia; 7. Educación y cultura; 8. Otras aplicaciones útiles de la IA; 8.1 Utilidades prácticas del día a día; 8.2 Protección del medio ambiente; 8.3 Personalización de las relaciones comerciales y otras; Parte II: La inteligencia artificial y sus riesgos; 1. Impacto sobre el mercado laboral; 2. Uso con fines bélicos; 3. Masificación de la desinformación; 4. Violación de la privacidad; 5. Discriminación algorítmica; 6. Cuestiones sobre propiedad intelectual y derechos de autor; Parte III: Algunos principios para la regulación de la inteligencia artificial; 1. Complejidades de la regulación; 2. Algunos esfuerzos de regulación; 3. Algunas directrices; 3.1 Defensa de los derechos fundamentales; 3.2 Protección de la democracia; 3.3 Promoción de la buena gobernanza; Conclusión; Referencias.

Resumen: El presente artículo trata sobre las potencialidades y riesgos de la inteligencia artificial (IA). Con ese objetivo, sitúa a la IA en el amanecer de la IV Revolución Industrial, explicando sus categorías esenciales y su modo de funcionamiento. Aborda los beneficios que aporta esta nueva tecnología: ampliación de la capacidad de decisión humana, automatización, avances en investigación e innovación, medicina y educación, entre otros. Examina los riesgos que genera, entre ellos: impactos sobre el mercado laboral, uso con fines bélicos, masificación de la desinformación y violación de derechos fundamentales. Propone principios para la regulación de la IA. Demuestra que se trata de una tecnología con gran potencial, cuyos efectos reales dependerán, sobre todo, del uso que hagamos de ella. En tales condiciones, el derecho tiene el importante desafío de crear un diseño institucional que fomente su buen uso y contenga su desviación.

Palabras clave: inteligencia artificial; derechos fundamentales; democracia; trabajo; riesgos; regulación.

Introducción
El amanecer de la cuarta revolución industrial

1. Un mundo feliz

Una nueva revolución industrial se vislumbra. La primera ocurrió a mediados del siglo XVIII y está representada por el uso del vapor como fuente de energía. La segunda revolución industrial, a finales del siglo XIX y principios del XX, está simbolizada por la electricidad y el motor de combustión interna. La tercera tuvo lugar en las últimas décadas del siglo XX y culminó con la sustitución de la tecnología analógica por la digital. Conocida como la Revolución Tecnológica o Revolución Digital, permitió la universalización de las computadoras personales y los teléfonos inteligentes, y está simbolizada por internet, que conecta a miles de millones de personas en todo el mundo (Barroso, 2019, p. 1262). La cuarta revolución industrial, que comienza a invadir nuestras vidas, llega con la combinación de la inteligencia artificial, la biotecnología y la expansión del uso de internet, creando un ecosistema interconectado que abarca personas, objetos e incluso mascotas, en una internet de las cosas y los sentidos.

En este desafiante nuevo mundo que se está desarrollando, las nuevas tecnologías pueden liberarnos de las actividades cotidianas más sencillas, así como de realizar tareas altamente complejas. Pueden limpiar entornos, regular la temperatura y, pronto, conducir vehículos autónomos (Manyika, 2022, p. 12). Prometen restaurar los movimientos corporales perdidos (Caczan, 2023), proporcionar diagnósticos médicos más precisos (Dilsizian; Siegel, 2014, p. 441), remediar deficiencias neurológicas, mejorar las capacidades cognitivas (Schmidt, 2017, p. 6-10), crear el «gemelo virtual» de alguien[6], reproducir a una persona que ya ha fallecido[7], permitir reencuentros con seres queridos

[6] El sitio invita a los usuarios a «duplicarse virtualmente» para «mejorar su productividad, salud mental y longevidad» (Mindbank, 2024).

[7] El objetivo del sistema es replicar la personalidad de los usuarios, su forma de pensar, hablar y otras características, para poder incluso interactuar con sus seres queridos tras la muerte de la persona duplicada (Ramírez, 2023).

que han fallecido (Here.After [20--]), cuidar a los ancianos (Horowitz, 2023), encontrar al amigo o pareja romántica ideal (Inner Circle, [20--]; Tinder, [20--]), escribir textos en los idiomas más diversos (ChatGPT, [20--]), distribuir ayuda social a los más vulnerables, dirigir los servicios públicos esenciales a los lugares más necesitados (Katyal, 2022, p. 327; Urueña, 2023). También pretenden predecir la práctica o reincidencia de delitos (Eubanks apud Eubanks, 2015), mejorar la vigilancia ambiental, promover la planificación de ciudades inteligentes (Galaz et al, 2021, p. 2), estimar el desempeño de los candidatos a puestos de trabajo, la probabilidad de pago de financiación, así como el desarrollo de enfermedades graves (Silberg; Manyika, 2019), entre otras cuestiones[8].

Hay más. Se estima que las mismas tecnologías pueden revelar la orientación sexual de una persona (Morrison, 2021), prever y reportar la intención de abortar (Cox, 2022), reemplazar a cientos de extras y actores en Hollywood (Beckett, 2023), crear o eliminar miles de trabajos mecánicos o creativos (Manyika, 2022, p. 20), manipular o falsificar información, sonidos, imágenes, creencias y deseos (Hacker; Engel; Mauer, 2023, p. 1 y 2), generar adicciones (Mohammad; Jan; Alsaedi, 2023; Becket; Paul, 2024), interferir en los comportamientos de los consumidores (Makhnoumi, 2024), influir en el resultado de los procesos electorales (Heawood, 2018, p. 429-434; Berghel, 2018, p. 84-89), provocar comportamientos violentos (Pauwels, 2020), fortalecer agendas extremistas (Vlachos, 2022), agravar la desigualdad y la discriminación contra grupos minoritarios (Angwin et al., 2016; Eubanks apud Eubanks, 2015), alterar y adquirir libre albedrío (Hutson, 2023)[9], activar armas de destrucción masiva, poner en riesgo la vida, la salud y la seguridad de las personas (Manyika, 2022, p. 21 y 27).

La lista es interminable y puede llevarnos a lo sublime o al horror, a la libertad o a la esclavitud. A la afirmación generalizada de los derechos humanos o a su supresión. Como se intuye, el problema no reside en la tecnología en sí, sino en el uso que haremos de ella y, sobre todo, en cómo pretendemos distribuir los beneficios que generará. El reto, por tanto, reside en crear un diseño institucional que fomente el buen uso de la Inteligencia Artificial y contenga su distorsión, impidiendo la automatización de la producción de injusticias (Degli-Esposti, 2023, p. 10) y la multiplicación de los riesgos existentes (Coeckelbergh, 2023, p. 167).

2. Qué es la inteligencia artificial

En una definición simple, se puede afirmar que la inteligencia artificial consiste en programas (software) que transfieren capacidades humanas a las computadoras. Estas capacidades incluyen tareas cognitivas y la toma de decisiones, generalmente basadas en los datos, instrucciones y objetivos con los que se alimentan[10]. Sin embargo, no existe una convergencia total en el concepto

---

[8] Los resultados positivos son realmente impresionantes, lo que lleva a algunas corrientes a considerar el uso de las nuevas tecnologías para transformar los mecanismos de gobernanza, en favor del establecimiento de una «democracia algorítmica», supuestamente neutral y eficaz. Sin embargo, la neutralidad algorítmica no existe, y la legitimidad democrática está necesariamente relacionada con la representación basada en la voluntad popular. En este sentido, la Asamblea Parlamentaria del Consejo de Europa entiende que la definición de objetivos políticos y sociales no puede dejarse en manos de algoritmos. Por el contrario, debe permanecer en manos de seres humanos que se someten a un sistema de rendición de cuentas política y legal (Unión Europea, 2022).

[9] Se teme que la capacidad de la IA para aprender de forma autónoma la lleve a adquirir una inteligencia sobrehumana, volviéndola incontrolable. Este fenómeno se denomina «singularidad».

[10] La expresión «inteligencia artificial» se atribuye a un taller celebrado en 1956, en Dartmouth, con el objetivo de buscar desarrollar máquinas capaces de resolver problemas resueltos por humanos y mejorarse a sí mismas (MacCarthy et al., 1955; Manyika, 2022, p. 15).

técnico de IA y su alcance[11]. Numerosas entidades e instituciones, como la OCDE[12] y la UNESCO[13], buscan establecer sus límites. Es posible señalar algunos rasgos comunes en estos intentos de definición: son sistemas con la capacidad de procesar datos e información de forma similar a la inteligencia humana, lo que incluye el aprendizaje, el razonamiento, la percepción y la comunicación a través del lenguaje. Consultado, ChatGPT4 proporcionó la siguiente definición:

> La inteligencia artificial (IA) es una rama de la informática dedicada a crear sistemas capaces de realizar tareas que tradicionalmente requieren inteligencia humana. Estas tareas incluyen el aprendizaje (la capacidad de mejorar el rendimiento con la experiencia), el razonamiento (la capacidad de resolver problemas mediante métodos lógicos), la percepción (la capacidad de interpretar datos sensoriales para comprender aspectos del mundo) y la interacción lingüística (la capacidad de comprender y producir lenguaje natural).

En su etapa actual (Degli-Esposti, 2023, p. 10; Rebollo Delgado, 2023, p. 24)[14], la inteligencia artificial no es consciente de sí misma, no tiene discernimiento del bien o del mal, ni tiene emociones, sentimientos, moralidad o incluso sentido común. En otras palabras, depende completamente de la inteligencia humana para alimentarse, incluidos los valores éticos. Las computadoras no tienen voluntad propia (Winston, 2018; Lenharo, 2023). Aunque esta es la sabiduría convencional sobre el tema, algunos experimentos revelan una sorprendente capacidad de aprendizaje, lo que plantea nuevas preocupaciones. Uno de ellos fue Alpha Zero, un programa de IA desarrollado por Google que derrotó a Stockfish, hasta entonces el programa de ajedrez más poderoso del mundo. A diferencia de los programas anteriores, Alpha Zero no se alimentó con movimientos previamente diseñados por el hombre. En otras palabras, no se basó en el conocimiento, la experiencia o las estrategias humanas. Solo se le dieron las reglas del juego. Alpha Zero se entrenó jugando consigo mismo, desarrolló sus propios movimientos y estrategias, originales y poco ortodoxos, con su propia lógica (Kissinger; Schmidt; Huttenlocher, 2021, p. 7 y ss. y 26).

Dos perspectivas han competido por la primacía en la investigación de la inteligencia artificial a lo largo de los años. La primera se inspiró en el funcionamiento de la mente humana, buscando imitar la forma en que formulamos preguntas y desarrollamos razonamiento lógico. Esta primera perspectiva dominó los experimentos de IA hasta la década de 1980. La segunda perspectiva se inspiró en el funcionamiento de las estructuras del cerebro humano. Proponía conectar unidades de procesamiento de información, equivalentes a neuronas, para simular su funcionamiento (Dreyfus; Dreyfus, 1988, p. 15-44). Esta es la perspectiva que se ha vuelto dominante en el ámbito de la IA, denominada «enfoque conexionista» (*connectionist approach*). No busca reproducir la forma en que

---

11 Las organizaciones que representan a empresas de IA abogan por la formulación de un concepto más restrictivo de inteligencia artificial, mientras que las organizaciones de derechos humanos abogan por la expansión del concepto a otras tecnologías, que también pueden tener efectos adversos sobre los derechos humanos. En este contexto, el alcance del propio contexto de la IA depende, en parte, de cuánto se pretenda regularlo (Madiega, 2023, págs. 6-8).

12 Un sistema de IA es un sistema basado en máquinas que, con fines explícitos o implícitos, infiere a partir de la información que recibe cómo generar resultados como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad tras su implementación (OCDE, 2019; Russel; Perset, Marko, 2023).

13 «Por lo tanto, esta Recomendación aborda los sistemas de IA como sistemas que tienen la capacidad de procesar datos e información de una manera que se asemeja al comportamiento inteligente, y típicamente incluye aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control» (Unesco 2021).

14 Esta salvedad es necesaria dado que no se puede descartar que la IA del futuro otorgue a las máquinas intensas dosis de autonomía y conciencia, en un escenario en el que las aplicaciones inteligentes adquieran racionalidad propia, persiguiendo objetivos imprevistos.

la mente humana racionaliza. Al contrario, busca establecer correlaciones y patrones entre miles de datos y ciertos resultados. Sus principales pilares son la estadística y la neurociencia.

Los sistemas de inteligencia artificial se basan en datos y algoritmos. Cuanto mayor sea el conjunto de datos al que tienen acceso, mayor será el número de correlaciones confirmadas y descartadas y, naturalmente, más precisos tienden a ser los resultados (Dreyfus; Dreyfus, 1988, p. 15-44). Un universo dado de datos o características correlacionadas lleva a la IA a identificar un perro o un gato, un deudor bueno o malo, una persona con tendencias depresivas o un niño en riesgo. Establecer correlaciones entre estos elementos puede parecer aleatorio o irracional para la forma de conocimiento de la mente humana. Pero recuerde, el modelo se basa en la estadística, no en la lógica.

Un algoritmo, a su vez, es un concepto fundamental en informática. El término identifica el conjunto de instrucciones, reglas y parámetros que guían a las computadoras para realizar las tareas que se les asignan. Son fórmulas, códigos y scripts que seleccionan, procesan y almacenan datos con el fin de obtener un resultado determinado. Los datos seleccionados (input) y sus correlaciones permiten obtener los resultados que busca el programa (output), que pueden ser muy variados. Por ejemplo: si el resultado da lugar a la diferenciación entre objetos y seres vivos, hablamos de IA discriminativa; si el resultado es la predicción de comportamientos —de consumo, financieros o políticos—, tenemos IA predictiva; si se trata de la generación de contenido —textos, imágenes o sonidos—, decimos que es IA generativa (Hacker; Engel; Mauer, 2023, págs. 1-3 y 13)[15].

3. Aprendizaje automático, modelos fundacionales y otros conceptos relevantes

En términos de modo operativo, los sistemas de inteligencia artificial más avanzados actualmente son aquellos capaces de desarrollar aprendizaje automático. El aprendizaje automático se refiere a la capacidad de un modelo para adquirir conocimiento de forma autónoma, sin programación explícita previa, basándose en la identificación de correlaciones entre grandes volúmenes de datos, como se describió anteriormente. Cabe destacar también que, para conceptos más restringidos de IA, la capacidad de aprendizaje automático es lo que diferencia la inteligencia artificial de la mera automatización, que sería un fenómeno más amplio (Nunes; Andrade, 2023, p. 4; Brown, 2021). El aprendizaje automático es el proceso que sustenta la mayoría de los servicios de IA que utilizamos hoy en día, como los sistemas de recomendación de contenido en plataformas como Netflix, YouTube y Spotify, los modelos de selección y clasificación de resultados en buscadores como Google, Bing y Baidu, así como los feeds y sistemas de recomendación de contactos en redes sociales como Facebook y X (anteriormente Twitter) (Hao, 2018; Nunes; Andrade, 2023, p. 5).

El aprendizaje automático se basa en algoritmos y redes neuronales artificiales. Las «redes neuronales» artificiales (*neural networks*) se inspiran en las redes neuronales humanas. Son modelos matemáticos que imitan nuestro sistema nervioso (Porto; Araújo; Gabriel, 2024, p. 37). A través de ellas, diferentes procesadores de datos trabajan juntos para procesar dichos datos. El aprendizaje profundo de máquinas (*deep learning*) es la técnica que amplía la capacidad del aprendizaje automático mediante el establecimiento de varias capas de redes neuronales artificiales, simulando el complejo funcionamiento del cerebro humano, de modo que estas múltiples redes y procesadores gestionen la información y establezcan correlaciones simultáneamente (Hao, 2018; Re; Solow-Niederman, 2019, págs. 244-246). Tiene aplicaciones en el reconocimiento de imágenes y voz, la

[15] Las calificaciones anteriores no son exhaustivas. Se hace referencia, por ejemplo, a la IA adversaria, cuyo objetivo es impedir el funcionamiento de otra IA con fines de protección de datos, entre otros (Hacker; Engel; Mauer, 2023, p. 13).

traducción automática y el procesamiento de textos. Existen tres tipos de aprendizaje automático: supervisado (*supervised*), no supervisado (*unsupervised*) y por refuerzo (*reinforcement*) (Brown, 2021)[16].

En términos de uso o propósito, los sistemas de IA se clasifican en numerosas categorías, que a veces se superponen. Modelos fundacionales (*foundational models*) se entrenan con grandes cantidades de datos, preparados para adaptarse a múltiples tareas. ChatGPT (Chat Generative Pre-trained *Transformer*), una IA generativa, es un ejemplo de un modelo fundacional (también llamado de «propósito general») y un «modelo de lenguaje grande» (*large language model*), con la capacidad de generar arte, imágenes, textos y sonidos (Jones, 2023)[17]. Cabe destacar que los programas de IA generativa pueden crear contenido nuevo, no solo analizar o clasificar el existente. Si alguien busca en Google cómo funciona un coche eléctrico, accederá a un enlace que lleva a un sitio web de terceros. Sin embargo, ChatGPT explica con sus propias palabras cómo funciona un coche eléctrico (Ariyaratne, 2023, págs. 4 y 5).

Los modelos fundacionales se consideran un paso hacia la denominada IA de propósito general (*general purpose AI*), que aún no se ha logrado plenamente, pero que ya es capaz de realizar un gran número de tareas, sin limitarse a objetivos específicos (Bommasani et al., 2022, págs. 4-12; Hacker; Engel; Mauer, 2023, págs. 2-4; Uuk; Gutierrez; Tamkin, 2023). A su vez, los modelos de propósito fijo o propósito estricto (*fixed-purpose AI systems* o *narrow AI*), como su nombre indica, están diseñadas para un propósito específico y, por lo tanto, se definen de forma más restrictiva. Se entrenan con una base de datos más específica. Esta categoría incluye la mayoría de los sistemas de IA actualmente en uso, como asistentes de voz como Siri, Alexa y el Asistente de Google, diseñados para comprender y ejecutar comandos de voz; sistemas de recomendación de contenido en plataformas de streaming, filtros de spam, sistemas de predicción meteorológica y software de reconocimiento facial, entre otros. Todos tienen un propósito muy específico.

Finalmente, la expresión IA fuerte (*strong AI*), también conocida como inteligencia artificial general (*AGI – artificial general intelligence*), designa sistemas con capacidad de comprensión, aprendizaje y aplicación concreta equivalente a la de los seres humanos. Por lo tanto, sería capaz de razonar, resolver problemas y tomar decisiones por sí misma. Este tipo de IA constituye un concepto teórico, aún no alcanzado, pero posible de alcanzar en pocos años, según algunos investigadores (Taulli, 2020, p. 218).

Sería posible seguir explorando los múltiples tecnicismos del tema. Por ejemplo, la cadena de valor de la IA, con sus diferentes fases: diseño, desarrollo, implementación y mantenimiento del sistema (Hacker; Engel; Mauer, 2023, p. 8-11). Cada una de estas fases presenta sus propios desafíos y riesgos, involucrando a diferentes actores, entre ellos el desarrollador (*developer*),[18] el implementador

---

16 El aprendizaje supervisado, el principal tipo de aprendizaje utilizado actualmente, permite definir el resultado (*output*) adecuado. Este es el caso de los sistemas de IA centrados en la categorización (en gatos, perros, medios de transporte). El aprendizaje no supervisado, menos común, busca identificar patrones, correlaciones y agrupaciones sin definir previamente el resultado adecuado. Finalmente, el aprendizaje por refuerzo busca entrenar y mejorar las máquinas mediante un sistema de prueba y error, para enseñarles a tomar las mejores decisiones (Brown, 2021).

17 No toda la IA generativa constituye un modelo fundacional. Puede modelarse para fines muy específicos. Las capacidades generativas pueden incluir la manipulación y el análisis de texto, imágenes y vídeos, así como la producción de voz. Las aplicaciones generativas incluyen chatbots, filtros de fotos y vídeos, y asistentes virtuales.

18 Este sería el caso de OpenAI con respecto a GPT-4 (modelo fundacional).

(*deployer*)[19], el usuario[20] y el destinatario final (*recipient*)[21]. Estos agentes tienen diferentes conocimientos y habilidades, y pueden generar distintas contribuciones, daños y responsabilidades (Hacker; Engel; Mauer, 2023, p. 8-11). Esta variedad de roles, como es lógico, añade cierta dificultad a la regulación del asunto.

Es hora, sin embargo, de avanzar y explorar las implicaciones de la inteligencia artificial que trascienden las cuestiones estrictamente técnicas.

PARTE I
LA INTELIGENCIA ARTIFICIAL Y SUS BENEFICIOS

La inteligencia artificial se ha incorporado cada vez más a nuestra vida cotidiana, a veces con tanta naturalidad que ni siquiera asociamos ciertas utilidades con ella. Lo cierto es que la vida analógica está quedando atrás. Es cierto que siempre habrá quienes prefieran objetos fabricados a la antigua usanza, como algunos relojes de marcas muy caras, que celebran un pasado artesanal, aunque están atrasados, adelantados y funcionan mucho peor que sus equivalentes digitales. Pero siguen atrayendo compradores, lo que demuestra que la especie humana no se rige únicamente por la razón y el pragmatismo. Sin embargo, salvo extravagancias e idiosincrasias, lo cierto es que hoy investigamos cualquier tema mediante algoritmos de búsqueda. Elegimos productos, lecturas, viajes y alojamientos mediante algoritmos de recomendación. Optamos por desplazamientos más rápidos o decidimos qué ponernos basándonos en sistemas inteligentes de medición del tráfico y la temperatura. En resumen, la IA aporta muchas cosas positivas que nos hacen la vida mejor y más fácil.

Considerando el gran impacto de la inteligencia artificial, sus usos y potencialidades son tan numerosos que ni siquiera resulta fácil seleccionarlos y sistematizarlos. A continuación, se presentan algunos ejemplos significativos.

1. Mejor capacidad de decisión en muchas áreas

En muchos ámbitos, la IA tendrá una mayor capacidad de toma de decisiones que los humanos, por diversas razones. Primero, porque puede almacenar una cantidad mucho mayor de información que el cerebro humano. Segundo, porque puede procesarla mucho más rápido. Tercero, porque puede establecer correlaciones dentro de un volumen masivo de datos, más allá de las capacidades de una sola persona o incluso de un equipo. Dichas correlaciones pueden revelar asociaciones entre factores que desconocemos, debido a su complejidad o sutileza. Sin embargo, como ya se mencionó, la eficiencia de la IA dependerá de la cantidad y calidad de los datos que se le proporcionen. Además, en el estado actual de la técnica, las herramientas de IA generativa pueden producir información inventada o absurda, en una desviación conocida como «alucinación» (*hallucination*)[22]. Cabe señalar

[19] En el caso de los productos a que se refiere la nota anterior, el usuario es quien genera el texto con ChatGPT o consulta la aplicación de voz, vía Virtual Volunteer.

[20] En el caso de ChatGPT (un modelo de propósito fijo), OpenIA es tanto desarrollador como implementador. Be My Eyes, a su vez, es únicamente implementador de Virtual Volunteer (también un modelo de propósito fijo), adaptado de GPT-4.

[21] Quién usa el texto y la guía de voz. Tenga en cuenta que las noticias que se emiten pueden ser ciertas o no. La voz puede ser una buena guía o un *deep fake*, creado para engañar al receptor.

[22] Cabe destacar que las alucinaciones de IA pueden tener un impacto drástico en procesos sociales relevantes. Por ejemplo, según un informe del laboratorio de IA Forense, el chatbot Bing de Microsoft proporcionó información

que en áreas que dependen de la inteligencia emocional, los valores éticos o la comprensión de los matices del comportamiento humano, la intervención humana será indispensable y su capacidad de toma de decisiones, superior.

2. Automatización

La IA permite la automatización de numerosas tareas, tanto rutinarias como complejas, aumentando la productividad y la eficiencia en diversas áreas de actividad. Las tareas repetitivas, agotadoras o extenuantes para los humanos pueden ser realizadas por máquinas, como en las líneas de producción industrial. Además, se reduce el margen de error y es posible eliminar riesgos en trabajos como la minería, el desarme de bombas, la reparación de cables en el fondo del océano o los viajes espaciales. Además, la IA puede trabajar de forma continua las 24 horas, todos los días de la semana, produciendo a mayor escala, con mayor precisión y a menor coste. No se cansa, no se enferma, no cambia de humor y no hay riesgo de presentar una demanda laboral. El impacto negativo que todo esto puede tener en el mercado laboral se analizará más adelante.

3. Lenguaje

El impacto de la IA en el campo del lenguaje ha sido profundo y multifacético, especialmente a través del uso del Procesamiento del Lenguaje Natural (*Natural Language Processing*). La calidad de las traducciones realizadas por Google Translate, ChatGPT y DeepL, por nombrar algunos ejemplos, ha mejorado significativamente, volviéndolas bastante precisas y fluidas. Esto ha derribado muchas de las barreras lingüísticas en la comunicación humana. Herramientas como Siri, Alexa y el Asistente de Google responden a comandos de voz. Otras herramientas transforman texto en voz. Los chatbots ayudan a resolver preguntas y problemas de consumidores y clientes. Y la IA generativa, que ha estado arrasando en todo el mundo, se comunica con los usuarios mediante texto, sonidos e imágenes. Los avances en este campo son extraordinarios.

4. Investigación e innovación

La IA ha expandido las fronteras de la investigación y la innovación en casi todas las áreas de la actividad humana, desde la física y la química hasta las industrias automotriz y espacial. El volumen de la ciencia generada mediante IA ha crecido exponencialmente. La IA puede simplificar y acortar la investigación clínica y las pruebas de nuevos fármacos, materiales y productos. El análisis de grandes cantidades de datos acelera el proceso de descubrimiento científico. Cabe destacar la reducción de costes y plazos en el desarrollo de nuevos fármacos, así como de vehículos autónomos, con la promesa de reducir el número de accidentes. Se espera que la IA, en su relación con la investigación y la innovación, pueda ayudar a afrontar numerosos desafíos de la humanidad, como el cambio climático, la lucha contra el hambre, el control de pandemias, la sostenibilidad de las ciudades y enfermedades como el cáncer y el Alzheimer (Unión Europea, 2023a). El movimiento que promueve la exploración beneficiosa del potencial de la IA se conoce como Data for good (Muñoz Vela, 2022, p. 65).

incorrecta en el 30% de las consultas básicas sobre temas electorales en Alemania y Suiza. También se descubrió que el problema se agravaba cuando las preguntas se formulaban en idiomas distintos del inglés (Guadián, 2024).

5. Aplicaciones en la medicina

La medicina es una de las áreas donde la inteligencia artificial tendrá el mayor impacto en la vida y la salud de las personas. Tecnologías como el aprendizaje automático y el procesamiento del lenguaje natural mejorarán la calidad de la atención al paciente y reducirán los costos. Diagnósticos mejorados, análisis de imágenes, cirugías robóticas, planificación y personalización del tratamiento, telemedicina, predicción de enfermedades futuras y gestión de datos de pacientes son algunos de los muchos beneficios que pueden surgir. La IA no hará prescindible el trabajo de los médicos, pero puede cambiar algunas de las funciones que desempeñan, transfiriendo responsabilidades del plano técnico al plano humano de empatía y motivación. También habrá implicaciones éticas y legales, como errores cometidos por los equipos de IA (Davenport; Kalacota, 2019).

6. Aplicaciones en el sistema de justicia

La IA ofrece la perspectiva de profundas transformaciones en el ejercicio de la abogacía y la administración de justicia. En un entorno donde los precedentes cobran mayor importancia, su valor para la investigación jurisprudencial eficiente es enorme. La posibilidad de que los abogados redacten escritos, el Ministerio Público emita dictámenes y los jueces tomen decisiones con base en borradores investigados y preparados por IA simplificará la vida y acortará los plazos de tramitación. Por supuesto, todo estará bajo estricta supervisión humana, ya que cada uno de estos profesionales sigue siendo responsable. En los tribunales, los programas de IA que agrupan los casos por materia, así como aquellos que pueden resumir casos extensos, optimizan el tiempo y la energía de los jueces. Asimismo, la digitalización de los casos —en Brasil, hoy en día, casi todos los casos y su tramitación son electrónicos—, la automatización de ciertos procedimientos y la resolución de litigios en línea tienen el potencial de agilizar y hacer más eficiente la justicia. En Brasil, en el ámbito de los diversos tribunales del país, existen más de cien proyectos para el uso de la IA en la administración de justicia.

Existe un punto controvertido y particularmente interesante: el uso de la IA para respaldar la redacción de decisiones judiciales. Muchos temen, con razón, los riesgos de sesgo, discriminación, falta de transparencia y explicabilidad. Por no mencionar la falta de sensibilidad social, empatía y compasión. Pero es importante no olvidar que los jueces humanos también están sujetos a estos mismos riesgos. Por ello, existe otra cara de la moneda: la posibilidad de que la IA esté más preparada, sea más imparcial y esté menos sujeta a intereses personales, influencias políticas o intimidación. Esto puede ocurrir en cualquier lugar, pero especialmente en países menos desarrollados, con menor grado de independencia judicial o mayor grado de corrupción (Ariel Gustavo, 2021; Sustein, 2022). En cualquier caso, en el estado actual de la civilización y la tecnología, la supervisión de un juez humano es esencial, aunque puede que se le imponga una mayor carga de argumentación en los casos en que pretenda producir un resultado diferente al propuesto por la IA.

7. Educación y cultura

La inteligencia artificial transformará el panorama educativo mundial, tanto en métodos de enseñanza como en posibilidades de aprendizaje. Inicialmente, la internet, impulsada por la IA, ha expandido exponencialmente el acceso al conocimiento y la información, ampliando los horizontes de todas las personas con acceso a la red informática mundial. Además, la educación a distancia ha derribado las barreras del tiempo y el espacio, permitiendo aprender en cualquier momento y desde cualquier lugar. Las bibliotecas digitales eliminan la necesidad de desplazamientos físicos y permiten la consulta de repertorios ubicados en cualquier lugar del mundo. Desde la perspectiva del

profesorado, puede ayudar en la preparación de clases, la redacción de preguntas e incluso la corrección de trabajos, además de realizar tareas administrativas que liberan al profesorado para dedicar más tiempo a las actividades académicas. Todo, siempre, cabe reiterar, bajo supervisión humana.

Desde la perspectiva de los estudiantes, la IA, especialmente la generativa, facilita la investigación, puede resumir textos largos, corregir errores gramaticales y sugerir mejoras en la escritura, además de ayudar a superar las barreras lingüísticas, como se vio anteriormente. También permite una enseñanza personalizada, adaptada a las necesidades de los estudiantes, incluidas las personas con discapacidad, ya que puede, por ejemplo, transformar texto en voz o viceversa [23]. Naturalmente, para aportar beneficios que se distribuyan equitativamente entre la población, el uso de la IA presupone una conectividad de calidad para todos (inclusión digital). Aquí, no debemos descartar algunas disfunciones que pueden surgir del uso de la IA en la educación, que van desde el plagio hasta la creatividad y el pensamiento crítico limitados. Por esta misma razón, organizaciones internacionales como la OCDE (2023) y la UNESCO (2021) han elaborado documentos relevantes, con principios y directrices para el uso de la IA en la educación.

El impacto de la IA en la cultura también será inmenso. Como aspecto positivo, abrirá caminos a la creatividad, en sinergia con músicos, pintores, escritores, arquitectos, diseñadores gráficos e innumerables otros agentes creativos. La IA generativa puede asistir en la composición de sinfonías, obras literarias, poemas, narraciones, etc., incrementando la creatividad y el universo estético, pero también planteando innumerables cuestiones éticas sobre la propiedad intelectual y los derechos de autor (Beiguelman, 2021, p. 59). Merece la pena reflexionar sobre la afirmación de Yuval Noah Harari de que la IA ya ha pirateado el sistema operativo de la cultura humana, es decir, el lenguaje. Y se pregunta: ¿qué significará para los seres humanos vivir en un mundo donde un porcentaje de novelas, música, imágenes y leyes, entre muchas otras creaciones, sean generadas por inteligencia no humana? (Harari, 2023)

8. Otras aplicaciones útiles de la IA

8.1 Utilidades prácticas del día a día

La tecnología de IA está presente en ordenadores personales y smartphones en múltiples aplicaciones, como Google Maps, Waze, Uber, Spotify, Zoom, Facebook e Instagram. También en asistentes personales como Siri y Alexa. La IA también desempeña un papel importante en la industria del entretenimiento a través del streaming (Netflix, Amazon Prime, HBO Max) y los videojuegos. Sin olvidar las aplicaciones que permiten realizar transacciones bancarias y pagos con tarjeta de crédito, entre otras innumerables utilidades.

8.2 Protección del medio ambiente

La IA desempeñará un papel cada vez más crucial en la protección del medio ambiente, el análisis de datos, la predicción de fenómenos y la monitorización de situaciones. Entre los ejemplos se incluyen: el análisis de datos sobre el cambio climático, el uso de imágenes satelitales y de drones, la

[23] En el ámbito educativo, la IA también puede beneficiar a los estudiantes superdotados, dado que las habilidades de percepción, reconocimiento y recomendación permiten supervisar, comprender y adaptar el proceso de aprendizaje de cada estudiante, además de liberar a los docentes para dedicar más tiempo a la instrucción individual (Lee, 2019, p. 149).

monitorización de los niveles de contaminación del aire, el agua y el suelo, la racionalización de la distribución y el consumo de energía y agua, la predicción de desastres naturales (como huracanes, terremotos e inundaciones), la asistencia a la agricultura sostenible mediante sensores de suelo y otros instrumentos, la reducción del uso de pesticidas, la orientación del riego y la planificación de la reforestación[24].

8.3 Personalización de las relaciones comerciales y otras

La IA permite a la industria, el comercio, los servicios, los medios de comunicación y las plataformas digitales dirigir información, noticias y anuncios a sus consumidores que coinciden con sus intereses. Esto optimiza naturalmente el tiempo de las personas y facilita la compra de productos, libros, la planificación de viajes y un sinfín de otras opciones y decisiones que deben tomarse. Las recomendaciones de películas, música u otras formas de entretenimiento provienen de este uso de la inteligencia artificial. Sin embargo, no se deben ignorar aquí los aspectos negativos asociados a una cierta tribalización de la vida, debido al sesgo de confirmación resultante del envío de materiales que, en general, reiteran preferencias y convicciones. Este fenómeno reduce la pluralidad de opiniones, genera nuevas formas de control social[25] y puede conducir a la polarización y al radicalismo (Barroso; Barroso, 2023). En términos de relaciones personales, la investigación muestra que los matrimonios resultantes de relaciones iniciadas en línea, con la ayuda de algoritmos, han demostrado ser ligeramente más satisfactorios que los matrimonios en los que las parejas se conocieron a través de métodos convencionales, fuera de línea (Tropiano, 2023; Harms, 2013).

No podemos enumerar indefinidamente todos los usos y beneficios de la inteligencia artificial, que, además, se expanden cada día. Estos incluyen el desarrollo de vehículos autónomos, equipos de monitorización para detectar posibles fallos de infraestructura, la detección de fraudes, especialmente financieros, la mejora de la ciberseguridad y los controles de aviación. Es hora de centrar nuestra atención en los problemas, riesgos y amenazas que pueden surgir del uso a gran escala de la inteligencia artificial.

PARTE II
LA INTELIGENCIA ARTIFICIAL Y SUS RIESGOS

Toda nueva tecnología tiene un efecto disruptivo en las relaciones de producción y consumo, así como en el mercado laboral, impactando la vida social. Además, como muchos aspectos de la vida, las innovaciones pueden tener un lado negativo o ser apropiadas por actores sociales deshonestos. El telar dejó sin trabajo a costureras y artesanas; la impresión offset eliminó los trabajos de linotipia. La informatización redujo la necesidad de empleados bancarios en el sistema financiero. Las

[24] Sin embargo, tampoco debe ignorarse el gasto energético resultante de la alimentación, operación y mantenimiento de las IA, así como sus impactos sistémicos sobre diferentes ecosistemas, como ya han observado algunos investigadores (García-Martín et al., 2019, p. 75-88; Centeno et al., 2021, p. 1-10).

[25] Sobre el uso de la IA como forma de control social: «Los motores de búsqueda presentan otro desafío: hace diez años, cuando se basaban en la minería de datos (en lugar del aprendizaje automático), si una persona buscaba «restaurante gourmet» y luego «ropa», su última búsqueda sería independiente de la primera. En ambos casos, un motor de búsqueda agregaría la mayor cantidad de información posible y les ofrecería opciones [...]. Las herramientas contemporáneas, en cambio, se guían por el comportamiento humano observado. [...] Una persona puede estar buscando ropa de diseñador. Sin embargo, existe una diferencia entre elegir entre diversas opciones y realizar una acción —en este caso, realizar una compra; en otros casos, adoptar una postura o ideología política [...]— sin haber visto nunca el abanico inicial de posibilidades o implicaciones, simplemente confiando en que una máquina configurara las opciones de antemano (traducción propia)» (Kissinger; Schmidt; Huttenlocher, 2023, p. 20).

plataformas digitales allanaron el camino para la polarización extremista (Fisher, 2023, p. 20), la desinformación (Kakutani, 2018, p. 17) y el discurso de odio (Campos Mello, 2020; Williams, 2021, p. 207). Aún más grave: la invención de la carabela permitió el comercio transoceánico, pero también la trata de esclavos (Acemoglu; Simon; 2023, p. 4-5)[26].

Por estas razones, es necesario prestar atención a los efectos adversos del uso de la inteligencia artificial y buscar neutralizarlos o mitigarlos. Estos impactos negativos de la IA pueden tener implicaciones sociales, económicas y políticas, o incluso socavar la paz mundial. A continuación, enumeramos algunas consecuencias, riesgos y amenazas que conlleva la inteligencia artificial.

1. Impacto sobre el mercado laboral

Este es el efecto más obvio y predecible, resultante de lo que normalmente ocurre cuando una nueva tecnología altera el modo de producción anterior. Con el avance de la automatización, el panorama del mercado laboral cambiará profundamente, lo que requerirá que los trabajadores de diferentes áreas de la economía se adapten a nuevos empleos. Esta transición no siempre es fácil. Cabe señalar que, en el caso de la IA, el impacto no solo se dará en trabajos más mecánicos, sino que también afectará a roles más calificados y creativos [27]- [28]. Es cierto que las nuevas tecnologías también tienden a generar nuevos mercados y, en consecuencia, nuevos empleos. Sin embargo, existe un problema de tiempo y escala en esta consideración. Es poco probable que se generen nuevos empleos espontáneamente al mismo ritmo y volumen (Keynes, 1930)[29]. Este es un desafío importante, que requerirá que los gobiernos inviertan en protección social y capacitación de los trabajadores. Vale la pena recordar que la expansión de la vulnerabilidad económica tiende a impactar la esfera de protección democrática, dado que históricamente ha surgido como un factor potencial de desestabilización.

2. Uso con fines bélicos

Existe relativamente poca literatura sobre el uso de la IA con fines militares, en parte debido al secretismo que normalmente se impone al tema por razones de seguridad. Sin embargo, a lo largo de la historia, las nuevas tecnologías se originan en la investigación con fines militares o se dirigen rápidamente a ese fin. No es difícil imaginar a países como Estados Unidos y China compitiendo por emplear la IA con fines militares, utilizando nuevas tecnologías y robots. De hecho, los drones automatizados (*automated drones*) operados a distancia se han utilizado desde hace tiempo con este fin, en misiones de reconocimiento, vigilancia, entrega de equipos o incluso ataques aéreos. Un tema que ha suscitado gran preocupación es el de las armas letales autónomas (*autonomous lethal weapons*), que pueden entrar en combate y atacar objetivos por sí solas, sin control humano. Existen debates en curso sobre el control estricto de su uso mediante leyes internacionales (Klare, 2023). Las

---

[26] Los autores señalan algunos inventos que, en los últimos mil años, no necesariamente han traído prosperidad a todos.

[27] Se estima que los bancos y algunas empresas tecnológicas destinan entre el 60% y el 80% de sus nóminas, o incluso más, a trabajadores con alta probabilidad de verse afectados por las nuevas tecnologías (Lohr, 2024). En la misma línea, véase Maheshwari (2024) sobre el mercado de la publicidad y el marketing.

[28] Otros estudios indican que los roles que requieren inteligencia social (relaciones públicas), creatividad (biólogos y diseñadores), percepción y manipulación fina (cirujanos) tienden a verse más afectados (Frey; Ousborne, 2017). También evalúan que las tareas de análisis, previsión y estrategia serán las más afectadas (Webb, 2019).

[29] Como observó Keynes (1930) hace casi un siglo, los avances tecnológicos tienden a generar al menos un desequilibrio temporal en términos de trabajo, hasta que se identifican nuevas oportunidades laborales.

implicaciones éticas de este tipo de armamento son drásticas, por lo que es imperativo regular estrictamente su uso o, quizás mejor, prohibirlo.

Además, las tecnologías de la comunicación y la información han movilizado esfuerzos militares desde hace algún tiempo, presentando tácticas recurrentes en el contexto de las «guerras híbridas» (*hybrid warfares*). Se trata de nuevas formas de agresión que implican, además de la destrucción por medios físicos, campañas de influencia y desinformación (*cognitive warfares*), además de ciberataques con el propósito de comprometer sistemas informáticos vitales, como las estructuras de suministro de energía (Alvim; Zilio; Carvalho, 2023, p. 69).

3. Masificación de la desinformación

Desde al menos 2016, la difusión de información a través de plataformas digitales y aplicaciones de mensajería ha representado un grave problema para el proceso democrático y electoral. Estudios documentan que la circulación de falsedades y el radicalismo en línea ocurre a un ritmo más rápido y con mayor participación que la difusión de discurso veraz y moderado. Lo emotivo, improbable y alarmante genera mayor participación y movilización. El *deep fake* empeora las cosas, ya que simula que las personas dicen cosas que nunca dijeron, adulterando el contenido y la realidad de una manera imperceptible para los ciudadanos (Barroso; Barroso, 2023; Campos Mello; Rudolf, 2023, p. 53-78). Este escenario no es hipotético y los precedentes son preocupantes. La influencia que la difusión de desinformación tuvo en eventos históricos como la salida del Reino Unido de la Unión Europea (Brexit), las elecciones en Estados Unidos, ambas en 2016, y las elecciones brasileñas de 2018 se ha vuelto notoria. La democracia presupone la participación informada de los ciudadanos y, naturalmente, se ve seriamente comprometida por la circulación generalizada de mentiras deliberadas, la destrucción de reputaciones y las teorías conspirativas.

4. Violación de la privacidad

El modelo de negocio de las plataformas que utilizan IA se basa en la recopilación de la mayor cantidad posible de datos personales de los individuos, lo que convierte la privacidad en una mercancía (Morozov, 2018, p. 36). A partir de estos datos, algoritmos complejos y múltiples capas neuronales establecen correlaciones profundas que permiten obtener sus datos genéticos, sus sistemas psíquicos, vulnerabilidades y comportamientos de consumo, políticos, financieros, sexuales y religiosos (Huq, 2020, p. 37). Con estos datos y correlaciones, la IA puede realizar predicciones, recomendaciones, manipular intereses y producir los resultados deseados por el algoritmo. Por lo tanto, el acceso a datos privados, tanto de personas como de empresas, es fundamental para el modelo de negocio de la IA tal como está establecido (Zuboff, 2022, p. 1-79)[30]. No es casualidad que, en el ámbito académico, los datos se hayan considerado el petróleo del siglo XXI (Rebollo Delgado, 2023, p. 17).

Hay al menos tres aspectos que requieren atención con respecto al tema de la privacidad. El primero es la recopilación de datos de los usuarios de Internet sin su consentimiento por parte de plataformas digitales y sitios web. Dicha información se utiliza para la venta comercial, para la información y publicidad dirigidas, o incluso para manipular la voluntad de los usuarios, como lo demuestra la

[30] Muchas de las otras restricciones a los derechos se derivan de restricciones a la privacidad, como las relacionadas con: daños físicos, reputacionales, relacionales, psicológicos (emocionales), económicos, discriminatorios y relacionados con la autonomía humana (coerción, manipulación, desinformación, distorsión de expectativas, pérdida de control, entre otros) (Citron; Solove, 2022, p. 793-863; Huq, 2020).

investigación en neurociencia. Un segundo aspecto se refiere a la vigilancia y el seguimiento por parte del gobierno y las autoridades policiales, utilizando tecnologías de reconocimiento facial y herramientas de ubicación. Aunque el propósito legítimo es combatir el crimen, los riesgos de abuso son muy altos. Estos riesgos, como intuitivamente se sugiere, se agravan en el caso de gobiernos autoritarios. Finalmente, un tercer punto es que los sistemas de IA requieren la recopilación de grandes cantidades de datos para entrenar sus modelos, con el riesgo de filtraciones y ciberataques por parte de actores maliciosos, por ejemplo, en actividades de *spear phishing* (Muñoz Vela, 2022, p. 64)[31] y el *doxxing* (Prado, 2023, p. 162)[32], que a menudo alimentan prácticas de acoso, violencia política, *malinformation* y desinformación.

5. Discriminación algorítmica

Los algoritmos se entrenan con datos existentes, que a su vez expresan comportamientos humanos pasados y presentes, llenos de sesgos y prejuicios, profundamente determinados por circunstancias históricas, culturales y sociales (Prado, 2023, p. 162; Horta, 2019, p. 85-122). Por esta razón, tienden a reproducir estructuras sociales actuales y pasadas de inclusión y exclusión. En esta medida, los datos sobre empleabilidad muestran una menor contratación de mujeres, negros e indígenas, una tendencia que no está relacionada con su capacidad y productividad, pero que puede llevar a la reproducción de comportamientos futuros (Dastin, 2018)[33]; los datos sobre seguridad pública registran una mayor propensión a la reincidencia y la violencia que involucra a personas negras, no necesariamente porque sean más violentas, sino posiblemente porque viven en contextos sociales más adversos (Larson, 2016); los datos sobre costos de salud tienden a sobreestimar el gasto de algunos grupos y minimizar el gasto de otros, por razones no necesariamente relacionadas con sus condiciones físicas[34]; los datos sobre el riesgo crediticio aumentarán los riesgos y, en consecuencia, los costos de financiamiento de las personas con un estatus económico y social más bajo, incluso cuando hayan logrado mejorar sus condiciones, dependiendo de las circunstancias de la recopilación de datos (Pasquale, 2016). En esta medida, se encuentra que algunos algoritmos de contratación pueden tender a descartar a las mujeres, criminalizar a los hombres negros y dificultar el acceso al crédito para los más pobres. En tales condiciones, la forma en que funciona la IA puede reforzar profundamente las desigualdades existentes, en detrimento de los grupos más vulnerables de la sociedad (Huq, 2020, p. 29-34; Silberg; Manyika, 2019, p. 3).

---

31 Se trata de correos electrónicos o mensajes maliciosos, personalizados para un destinatario específico, con apariencia de credibilidad, orientados a obtener información sensible (contraseñas, por ejemplo) o instalar *malware*, que son programas maliciosos con graves efectos en los sistemas afectados.

32 *Doxxing* significa la eliminación maliciosa de información sobre alguien, ya sea de archivos públicos o mediante *hack* de computadoras, con fines de hostigar, intimidar o extorsionar, entre otros.

33 De hecho, Amazon dejó de usar un sistema de selección de personal tras descubrir que discriminaba a las mujeres. La empresa descubrió que la discriminación se debía a que el sistema de IA se había entrenado con datos de contratación recopilados durante los últimos 10 años, cuando las mujeres tenían una menor presencia en el mercado laboral. El sistema interpretó la menor presencia de mujeres como una preferencia por los hombres y descartó a las candidatas.

34 En este caso, los datos utilizados para entrenar el algoritmo estaban incompletos. Se utilizaron datos sobre los costos de los pacientes blancos y negros para estimar el alcance de sus necesidades de salud. Los recursos utilizados para los pacientes negros fueron menores que para los blancos, no porque sus necesidades fueran menores, sino porque tenían mayor dificultad para acceder al servicio. Como resultado, la IA subestimó erróneamente las necesidades de los pacientes negros (Obermeyer, 2019).

6. Cuestiones sobre propiedad intelectual y derechos de autor

El modelo de negocio de la IA plantea importantes preguntas sobre los derechos de autor y la propiedad intelectual. ¿Quién posee los derechos de autor del vasto universo de canciones, películas, noticias y contenido recopilado por los grandes medios de comunicación? ¿Empresas tecnológicas con el propósito de alimentar a sus IA? ¿A sus autores y creadores o a quienes comenzaron a emplearlos y explotarlos mediante algoritmos? La IA generativa se alimenta de una increíble cantidad de datos. Sin embargo, las respuestas a las preguntas que se le formulan no identifican la fuente ni al autor. Los debates sobre este tema se han intensificado y han llegado a los tribunales. Tomemos el ejemplo de la prensa. El contenido producido por las empresas de noticias es recopilado por empresas de IA, que lo utilizan para entrenar aplicaciones que compiten con la prensa en la producción de información[35]. La cuestión es objeto de una acción judicial presentada por el periódico The New York Time contra OpenAI y Microsoft (Grynbaum; Mac, 2023). Una demanda similar involucra a Getty Images, empresa de medios visuales y proveedor de imágenes, y a Stability AI, empresa de inteligencia artificial (Vincent, 2023)[36].

No es posible explorar exhaustivamente los riesgos que conlleva el desarrollo de la IA, ya que existen innumerables posibilidades a considerar, sin mencionar aquellas que ni siquiera somos capaces de imaginar o anticipar. Pero hay una última preocupación que merece especial consideración. Se trata de la llamada «singularidad», un término utilizado para identificar el riesgo de que las computadoras adquieran consciencia, voluntad propia y dominen la condición humana. Esto se debe a que, al ser capaces de procesar un volumen de datos mucho mayor a una velocidad igualmente mucho mayor, si tienen consciencia y voluntad, se volverán superiores a todos nosotros. El temor surge del hecho de que los sistemas de IA pueden automejorarse, alcanzando la «superinteligencia», dominando el conocimiento científico, la cultura general y las habilidades sociales que los situarían por encima de los mejores cerebros humanos.

Alguien escéptico del potencial humano podría incluso asumir que una superinteligencia extrahumana tendría mayor capacidad para abordar algunos de los grandes problemas no resueltos de la humanidad, como la pobreza, la desigualdad o la degradación ambiental. Pero nunca se sabe si esta inteligencia descontrolada serviría a la causa y los valores de la humanidad. Por esta misma razón, la gobernanza de la IA, tanto nacional como internacional, necesita establecer protocolos de seguridad y parámetros éticos diseñados para gestionar y mitigar este riesgo. Si la tecnología logra alcanzar este punto —algo que muchos científicos dudan—, el futuro mismo de la civilización y la humanidad estará en juego.

Yuval Noah Harari (2023) hace un comentario interesante sobre el tema. Según él, en 2022, se preguntó a unos 700 de los científicos e investigadores de IA más importantes sobre los peligros de que esta tecnología afecte a la existencia humana o cause una pérdida significativa de poder. La mitad respondió que el riesgo sería del 10% o superior. Ante esto, plantea la crucial pregunta: ¿te subirías a un avión si los ingenieros que lo construyeron te dijeran que hay un 10 % de riesgo de estrellarse? Si eso es cierto, no podrás dormir tranquilo.

---

35 Sobre la crisis del modelo de negocio de la prensa y su impacto en la democracia, véase: Minow, 2021, p. 35; Jackson, 2022, p. 280 y siguientes; Barroso; Barroso, 2023; Campos Mello; Rudolf apud Cunha França; Casimiro, 2023.

36 Getty Images afirma que Stability utilizó las imágenes que produjo para entrenar un sistema de IA generador de imágenes llamado Stable Diffusion, sin autorización, violando sus derechos de propiedad intelectual y los derechos de autor de sus colaboradores, con el propósito de ofrecer servicios similares a los suyos.

PARTE III
ALGUNOS PRINCIPIOS PARA LA REGULACIÓN DE LA INTELIGENCIA ARTIFICIAL

1. Complejidades de la regulación

De todo lo explicado hasta ahora, es evidente que regular la inteligencia artificial se ha vuelto esencial. Sin embargo, la tarea no es sencilla y enfrenta desafíos y complejidades. A continuación, buscamos identificar algunos de ellos.

La regulación debe hacerse mientras el tren está en movimiento. En marzo de 2023, más de mil científicos, investigadores y emprendedores firmaron una carta abierta pidiendo una pausa en el desarrollo de los sistemas de IA más avanzados, dados los «profundos riesgos para la sociedad y la humanidad» que planteaban. La pausa, de al menos seis meses, tendría por objeto introducir «un conjunto de protocolos de seguridad compartidos» (Future for Life Institute, 2023). Las preocupaciones estaban plenamente justificadas, pero la suspensión de la investigación no se produjo. El tren siguió avanzando a gran velocidad. Sobre todo porque los avances en esta área se han convertido en objeto de disputa entre naciones, investigadores y emprendedores. Sin embargo, la carta reforzó las demandas de gobernanza, regulación, seguimiento y atención a los impactos sociales, económicos y políticos de las nuevas tecnologías.

La velocidad del cambio es asombrosa. Esto dificulta enormemente predecir el futuro y comprender las nuevas realidades de las normas jurídicas, que corren el riesgo de quedar obsoletas en poco tiempo. No es difícil ilustrar este punto. El teléfono fijo tradicional tardó 75 años en alcanzar los 100 millones de usuarios. El teléfono móvil, 16 años. Internet, 7 años. Pues bien, ChatGPT alcanzó los 100 millones de usuarios en dos meses (The Feed, 2023). No es fácil que la legislación y la regulación sigan el ritmo de la innovación.

Riesgos de la sobrerregulación. La regulación se ha vuelto esencial, como se mencionó anteriormente, pero también conlleva riesgos. Cabe destacar dos de ellos. El primero es que las restricciones y la responsabilidad civil no pueden ser tan onerosas como para inhibir el impulso a la innovación. El segundo es que una regulación desproporcionada puede crear una reserva de mercado para las empresas consolidadas, creando una brecha entre ellas y la competencia, agravando la concentración económica en manos de las grandes empresas. La opinión generalizada es que la regulación debe centrarse en los resultados, no en la investigación en sí.

Comentado [PS1]: Este período está irregular: tem forma de titulo mas está dentro do paragrafo.

Asimetría de información y poder entre empresas y reguladores. La tecnología de IA está controlada, sobre todo, por las empresas involucradas en su desarrollo, que poseen mayor conocimiento que los posibles reguladores. A esto se suma el hecho de que las empresas tecnológicas conocidas como *big techs* se encuentran entre las más valiosas del mundo y gozan de un poder económico que puede transformarse fácilmente en poder político. Este poder quedó en evidencia cuando el Congreso Nacional de Brasil votó un proyecto de ley que regula la desinformación en redes sociales. Algunas empresas tecnológicas lanzaron una intensa campaña contra la medida, tanto en sus propias

Comentado [PS2]: Este período está irregular: tem forma de titulo mas está dentro do paragrafo.

plataformas como ejerciendo presión en el Congreso Nacional, logrando que el proyecto de ley se eliminara de la agenda (Rezende, 2024; Poder 360, 2023a; Poder 360, 2023b)[37].

Necesidad de armonización global de la regulación. La IA es una tecnología predominantemente privada que no respeta las fronteras nacionales. Las empresas operan globalmente y, por lo general, ni siquiera tienen sus sedes en los principales centros de su negocio. Los datos pueden recopilarse e incorporarse al entrenamiento de sistemas en diferentes partes del mundo. En tales condiciones, la forma en que opera la IA pone en tela de juicio algunos elementos esenciales del derecho, tal como lo practicamos. Estos elementos son: la exigibilidad de los derechos fundamentales y humanos frente a los Estados (y no específicamente frente a agentes privados) y el alcance de las jurisdicciones nacionales, que están limitadas por la soberanía de otros países. Además, el tratamiento regulatorio heterogéneo del tema en diferentes países puede provocar la fuga de inversiones y obstáculos al desarrollo tecnológico en Estados restrictivos y representar una invitación a la violación generalizada de derechos en lugares más permisivos.

Comentado [PS3]: Este período está irregular: tem forma de titulo mas está dentro do paragrafo.

2. Algunos esfuerzos de regulación

A nivel internacional, algunas iniciativas implican propuestas no vinculantes (*soft law*) fueron notables. Entre ellas, destacan: a) la Recomendación del Consejo sobre Inteligencia Artificial de la OCDE (Organización para la Cooperación y el Desarrollo Económicos), de 2019 (OCDE, 2019)[38]; y b) la Recomendación sobre Ética en Inteligencia Artificial de la UNESCO, de 2021 (UNESCO, 2023)[39], [40]. Ambos documentos buscan responder a los riesgos ya indicados, son convergentes y complementarios, y reúnen principios muy generales sobre IA, que se detallarán en la normativa nacional de los respectivos países.

A nivel nacional, Estados Unidos de América publicó, a finales de 2023, una larga *Executive Order* (EO) sobre IA[41]. Se trata de una amplia regulación que abarca múltiples áreas de riesgo tecnológico, mediante la cual el presidente estadounidense se dirigió a las agencias federales, según su experiencia, para ordenarles establecer estándares y medidas para probar, garantizar la seguridad y fiabilidad de la tecnología, prevenir el fraude, la discriminación algorítmica y la vulneración de los derechos fundamentales de los ciudadanos, consumidores, competidores y estudiantes. La OE también dispuso la identificación del contenido producido por IA con marcas de agua. Estableció la definición de buenas prácticas y la realización de estudios sobre los impactos de la IA en las relaciones laborales, con medidas para mitigarlos. Incluyó financiación para investigación y apoyo a pequeñas empresas para acceder a asistencia técnica, recursos y al mercado de la IA, así como la

---

37 Se estima que no hay gran interés en contener las noticias falsas ni en moderar el contenido. Cuantas más noticias falsas, mayor es la participación del usuario y mayor la interacción en redes; por lo tanto, mayor es la producción de datos, materia prima para las *big techs*.

38 Esta recomendación también fue adoptada por el G-20.

39 El documento enumera los siguientes 10 principios: 1 – Proporcionalidad y no daño; 2 – Seguridad; 3 – Justicia y no discriminación; 4 – Sostenibilidad; 5 – Privacidad y protección de datos; 6 – Supervisión y determinación humana; 7 – Transparencia y explicabilidad; 8 – Comprensión y educación; 9 – Rendición de cuentas y control; 10 – Pluralidad de participantes, gobernanza adaptativa y colaboración.

40 Las Naciones Unidas también adoptaron Principios para el Uso Ético de la IA en el Sistema de las Naciones Unidas, que son bastante similares a los incluidos en la recomendación de la UNESCO (Ceb, 2022).

41 Una *Executive Order* es un documento normativo, un tipo de directiva, emitido por el presidente de los Estados Unidos de América, dirigido a la gestión del gobierno federal. Presenta limitaciones en cuanto a su posibilidad de estandarización, ya que no es una ley emitida por la Legislatura y puede ser modificada por decisión del siguiente presidente.

atracción de nuevos talentos mediante medidas migratorias. Determinó que los desarrolladores de modelos fundacionales que puedan presentar riesgos para la seguridad nacional, la economía nacional y la salud pública deben notificar a las Autoridades Públicas al entrenar sus sistemas y compartir con ellas los resultados de sus pruebas de seguridad (red-team safety-tests). Y pidió al Congreso aprobar una ley que proteja el derecho a la privacidad y proteja los datos de los ciudadanos.

La Unión Europea (UE) aprobó la Ley de Inteligencia Artificial (EU AI Act) en marzo de 2024. La regulación propuesta en el seno de la UE, a diferencia de lo que ocurre con el Ejecutivo La Orden de EE. UU. se caracteriza por establecer directamente normas y sanciones relativas al desarrollo, la implementación y el funcionamiento de la IA. También prevé la actuación concentrada de ciertos organismos en su supervisión e implementación. Sin embargo, dichas normas son proporcionales al riesgo que plantea la tecnología (*risk-based approach*) para personas y bienes (Unión Europea, 2021, 2023). Los sistemas se clasifican en tres niveles: a) sistemas sujetos a riesgos inaceptables, cuya implementación está prohibida[42]; b) sistemas de alto riesgo, cuya implementación está permitida siempre que cumplan con los estándares obligatorios[43]; y c) sistemas de IA que no presentan alto riesgo, para los cuales se ofrecen incentivos para la adopción voluntaria de códigos de conducta, un tipo de autorregulación (Unión Europea, 2024).

En Brasil, los Proyectos de Ley n.º 21/2020 y n.º 2.338/2023 se encuentran actualmente en trámite en el Congreso Nacional, con una tendencia a acercarse a los estándares establecidos en las propuestas de regulación de la Unión Europea (Brasil, 2023). En términos generales, las propuestas buscan: a) garantizar los derechos de las personas directamente afectadas por los sistemas de IA; b) establecer responsabilidades según los niveles de riesgo que imponen los sistemas y algoritmos basados en este tipo de tecnología; y c) establecer medidas de gobernanza aplicables a las empresas y organizaciones que exploran este campo.

3. Algunas diretrices

A la luz de todo lo explicado hasta ahora, es posible extraer algunos valores, principios y objetivos que deberían guiar la regulación de la IA, para que estas tecnologías sirvan a la causa de la humanidad, potenciando sus beneficios y minimizando sus riesgos. Dicha regulación debe centrarse en la defensa de los derechos fundamentales, la protección de la democracia y la promoción de la buena gobernanza. A continuación, se presentan algunos elementos y aspectos relacionados con cada uno de estos propósitos.

3.1 Defensa de los derechos fundamentales

el) Privacidad. El uso de IA debe respetar los datos personales de personas físicas y jurídicas, sin que puedan utilizarse sin consentimiento. La vigilancia invasiva (*invasive surveillance*), como el reconocimiento facial, la biometría y el monitoreo de ubicación, debe utilizarse de forma restringida y controlada. Además, dada la enorme cantidad de datos que alimentan la IA, deben existir mecanismos de seguridad adecuados contra filtraciones.

---

42 Esta categoría incluye tecnologías de puntuación social (*social scoring*), identificación biométrica en lugares públicos con fines de aplicación de la ley (con excepciones específicas), así como prácticas subliminales de manipulación de personas y/o explotación de las vulnerabilidades de grupos vulnerables.

43 Esta categoría incluye tecnologías de puntuación social, identificación biométrica en lugares públicos con fines de aplicación de la ley (con excepciones específicas), así como prácticas subliminales de manipulación de personas y/o explotación de las vulnerabilidades de grupos vulnerables.

b) Igualdad (no discriminación). La igualdad de todas las personas, en sus dimensiones formal, material y de reconocimiento, es uno de los pilares más valiosos de la civilización contemporánea. Los peligros de la discriminación algorítmica ya se han advertido aquí. La regulación de la IA debe evitar que las personas sean tratadas de forma desigual basándose en categorías sospechosas que exacerban vulnerabilidades, como el género, la raza, la orientación sexual, la religión, la edad y otras características. Existe un historial negativo en este sentido (Dastin, 2018; Larson et al., 2016; Obermeyer et al., 2019; Heaven, 2021).

c) Libertades. En cuanto a la autonomía individual, el uso de la neurociencia y la publicidad dirigida (microtargeting) tiene el poder de manipular el comportamiento y la voluntad de las personas, mediante sentimientos de miedo, prejuicio, euforia y otros sesgos cognitivos, induciéndolas a comprar bienes, contratar servicios o adoptar comportamientos contrarios a sus intereses, vulnerando su libertad cognitiva o autodeterminación mental. Además, el derecho a la información, el pluralismo de ideas y la libertad de expresión pueden verse comprometidos por algoritmos de recomendación o moderación, que filtran, dirigen y eliminan contenido, en una conducta equivalente a la censura privada.

3.2 Protección de la democracia

el) Combate a la desinformación. La democracia es un régimen de autogobierno colectivo, que presupone la participación informada y esclarecida de la ciudadanía. Por esta misma razón, la circulación de desinformación y teorías conspirativas engaña o genera temores infundados en las personas, comprometiendo su discernimiento y sus decisiones. Como ya se ha señalado, todo esto se ve agravado por las *deep fakes*, falsificaciones que simulan vídeos y discursos inexistentes con la apariencia de la realidad. Todos estamos educados para creer lo que vemos y oímos. Manipulaciones de esta naturaleza rompen los paradigmas de la experiencia (Filimowicz apud Filimowicz, 2022, p. x e xi)[44] y son destructivas para la democracia.

b) Combate a los discursos de odio. Desde la consagración histórica del sufragio universal, la democracia ha implicado la participación igualitaria de todas las personas. El discurso de odio consiste en ataques a grupos vulnerables y declaraciones racistas, discriminatorias o capacitistas, contra personas negras, homosexuales, personas con discapacidad, indígenas, entre otras. Al intentar descalificar, debilitar o silenciar a ciertos grupos sociales, el discurso de odio socava la protección de la dignidad humana y debilita la democracia.

c) Combate a los ataques a las instituciones democráticas. Las redes sociales, con la ayuda de la IA, han sido fundamentales para orquestar ataques a las instituciones democráticas con el objetivo de desestabilizarlas. Actos insurreccionales como los del 6 de enero de 2021 en Estados Unidos o el 8 de enero en Brasil, con intentos de golpe de Estado para desacatar los resultados electorales, ponen en riesgo la democracia y son intolerables (Ramonet, 2022).

3.3 Promoción de la buena gobernanza

A la luz de las recomendaciones y actos normativos internacionales, regionales y nacionales antes mencionados, y del amplio debate público actualmente en curso en el ámbito académico, la sociedad

[44] La experiencia acumulada nos lleva a entender los vídeos y las fotografías como elementos que atestiguan la veracidad de la información. Sin embargo, la IA basada en tecnologías de aprendizaje profundo permite la producción de contenido multimedia fraudulento pero ultrarrealista que confunde las mentes y permite la producción sintética de representaciones falsas de la realidad.

civil y la prensa, es posible extraer algunos consensos superpuestos con respecto a la gobernanza de la IA, que se describen en las cinco directrices que se establecen a continuación.

el) Centralidad del bien común. La IA debe desarrollarse y orientarse hacia el bienestar de las personas, los países y el planeta. Sus beneficios deben distribuirse equitativamente entre todos y sus impactos negativos deben mitigarse mediante legislación y regulación[45].

b) Gobernanza plural. La gobernanza de la IA debe considerar, en sus diferentes etapas y con la proporcionalidad adecuada, la participación de un conjunto diverso de actores, incluyendo el gobierno, científicos e investigadores, la sociedad civil, el mundo académico, las empresas y las organizaciones de derechos humanos. La diversidad de perspectivas y la ponderación de valores e intereses son fundamentales para la legitimidad de las decisiones y la aplicación de regulaciones adecuadas.

c) Transparencia y explicabilidad. La transparencia define el conocimiento mínimo del usuario sobre el funcionamiento del sistema y la información con la que interactúa en un sistema de IA. La explicabilidad implica hacer inteligibles las razones de las decisiones tomadas, incluyendo la posibilidad de cuestionar los resultados. Ambos requisitos se combinan para mitigar las preocupaciones sobre la precisión e imparcialidad de los algoritmos, así como para fomentar el uso responsable de las tecnologías de automatización (Bender, 2022, p. 12).

d) Seguridad. Los sistemas de IA deben ser seguros internamente para evitar errores que produzcan resultados indeseables, y también deben estar protegidos contra ataques externos. La seguridad en el uso de la IA incluye el análisis de impacto, la calidad de los datos y la ciberseguridad, así como el mapeo de los procesos y las decisiones que conforman el ciclo de vida de la IA (traceability).

e) Control y responsabilidad. La supervisión o el control humano son esenciales para que la IA opere dentro de los límites de la legalidad, la ética y la justicia. A pesar de la relativa autonomía en sus procesos de toma de decisiones, la responsabilidad siempre recaerá en una persona física o jurídica. En caso de uso indebido o malicioso, una o ambas estarán sujetas a responsabilidad civil, administrativa y penal.

Conclusión

El papel del conocimiento es consolar a los afligidos y afligir a los consolados (Shedden, 2014)[46]. Este artículo pretende cumplir esta función. La inteligencia artificial, como se demuestra aquí, presenta potencial y riesgos en casi todas las áreas en las que puede aplicarse. En el ámbito político, puede ayudar a mejorar el sistema representativo y a captar mejor los sentimientos y la voluntad de los ciudadanos. Pero también puede difundir desinformación, discursos de odio y teorías conspirativas, engañando a los votantes, debilitando a grupos vulnerables o difundiendo miedos infundados, sacando a relucir lo peor de las personas.

---

45 La regulación debe considerarse una condición necesaria, pero insuficiente. En este sentido, abordar los riesgos asociados a la inteligencia artificial trasciende la dimensión jurídica y alcanza también otros ámbitos, entre los que destaca la ética aplicada a la economía y la programación. Para Lucrecio Rebollo: «Concebir el derecho como la única vía para ordenar e igualar la sociedad digital es un grave error. El derecho debe ser, como siempre lo ha sido, una vía para resolver los conflictos sociales desde una perspectiva de bien común, pero en todos los casos requiere la colaboración de otras áreas del conocimiento, de todos los elementos que conforman la estructura social» (Rebollo Delgado, 2023, p. 52).

46 Esta frase es una paráfrasis de *The job of the newspaper is to comfort the afflicted and afflict the comfortable*, atribuida a un personaje ficticio –Mr. Dooley–, creado por el periodista Finley Peter Dunne, del Chicago Evening Post.

En el ámbito económico, la IA puede contribuir al aumento de la productividad en diversas áreas, desde la agroindustria hasta la industria, y a mejorar significativamente el sector servicios. Sin embargo, también puede concentrar la riqueza en los sectores más favorecidos y en las naciones más ricas, incrementando la desigualdad mundial. En el ámbito social, puede ser una herramienta importante para resolver problemas relacionados con la pobreza y las desigualdades injustas, pero también puede provocar desempleo entre una gran cantidad de trabajadores. Existen, además, dualidades éticas. Una mayor comprensión de la naturaleza humana puede elevar el nivel humanístico o espiritual del mundo, pero no se puede descartar la pérdida de la centralidad de la persona humana.

En resumen, vivimos en una era de ambigüedades y decisiones decisivas. En opinión de los autores, la historia del mundo ha sido un flujo constante, aunque no lineal, hacia la bondad, la justicia y el avance de la civilización. Hemos pasado de épocas de penurias, sacrificios humanos y despotismo, hasta llegar a la era de los derechos humanos. Por ello, es posible tener una visión y una actitud constructivas hacia la inteligencia artificial. Sin miedos paralizantes, pero también sin ingenuidad ni fantasías. Necesitaremos legislación, regulación y, sobre todo, educación y concienciación entre científicos, empresas y ciudadanos para no perdernos en el camino. Y, como ya se mencionó, la brújula, la dirección que indican las estrellas, son los valores que conducen a una buena vida: la virtud, la razón práctica y la valentía moral. Si perdemos las referencias de la bondad, la justicia y la dignidad humana, entonces sería el momento de dejar que las máquinas tomen el control y apostar por ellas.

Pero no tiene por qué ser así. Quizás, paradójicamente, la inteligencia artificial pueda ayudarnos a rescatar y profundizar nuestra propia humanidad, valorando la empatía, la fraternidad, la solidaridad, la alegría, la capacidad de amar y otros atributos que siempre nos diferenciarán de las máquinas.