Praxis Filosófica
ISSN (I): 0120-4688 / ISSN (D): 2389-9387

DOI: 10.25100/pfilosofica.v0i62S.15442

LA INTELIGENCIA ARTIFICIAL Y SUS IMPLICACIONES ÉTICAS EN LA ERA DE LOS GRANDES MODELOS DE LENGUAJE (LLM)

Artificial Intelligence and Its Ethical Implications in the Era of Large Language Models (LLM)

Julián Henao Henao
Luis Humberto Hernández Mora

Universidad del Valle, Cali, Colombia.

Universidad del Valle, Cali, Colombia.

Resumen

En este artículo abordaremos, en primera instancia, algunos aspectos básicos del desarrollo de la IA, tomando como punto de partida el modelo de lenguaje GPT-2, lanzado por la empresa OpenAI y considerado un hito crucial en la historia de los modelos de lenguaje. Su llegada demostró que los modelos de lenguaje a gran escala (LLM), podían generar textos largos, coherentes y de alta calidad sin un entrenamiento específico para cada tarea, lo que abrió la puerta al desarrollo de modelos posteriores como GPT-3, GPT-4 y el actual GPT-5. En segunda instancia, nos ocuparemos de algunas de las consecuencias éticas que surgen en el proceso de implementación de estas tecnologías y de la forma como se originó esta clase de reflexiones.

Palabras clave: *modelo de lenguaje a gran escala; inteligencia artificial; singularidad tecnológica; sesgo algorítmico; ética de datos.*

¿Cómo citar?: Henao Henao, J. y Hernández Mora, L. H. (2025). La inteligencia artificial y sus implicaciones éticas en la era de los grandes modelos de lenguaje (LLM). *Praxis Filosófica,* (62S), e20715442. https://doi.org/10.25100/pfilosofica.v0i62S.15442

Recibido: 15 de septiembre de 2025. Aprobado: 30 de octubre de 2025.

Artificial Intelligence and Its Ethical Implications in the Era of Large Language Models (LLM)

Julián Henao Henao[1]

Universidad del Valle, Cali, Colombia.

Luis Humberto Hernández Mora[2]

Universidad del Valle, Cali, Colombia.

Abstract

In this article, we will first address some basic aspects of AI development, taking as a starting point the GPT-2 language model, launched by OpenAI and considered a crucial milestone in the history of language models. Its arrival demonstrated that large-scale language models (LLMs) could generate long, coherent, and high-quality texts without specific training for each task, which opened the door to the development of subsequent models such as GPT-3, GPT-4, and the current GPT-5. Secondly, we will address some of the ethical consequences that arise in the implementation process of these technologies, and how this kind of reflections originated.

Keywords: *Large-scale language model; Artificial intelligence; Technological singularity; Algorithmic bias; Data ethics.*

---

[1] Estudiante de la Maestría en Filosofía de la Universidad del Valle.

[2] Profesor del Departamento de Filosofía de la Universidad del Valle.

LA INTELIGENCIA ARTIFICIAL Y SUS IMPLICACIONES ÉTICAS EN LA ERA DE LOS GRANDES MODELOS DE LENGUAJE (LLM)

Julián Henao Henao

Universidad del Valle, Cali, Colombia.

Luis Humberto Hernández Mora

Universidad del Valle, Cali, Colombia.

I. Introducción

La revolución de la inteligencia artificial (IA) ha tomado un nuevo impulso en los últimos años que se viene expresando en un variado número de nuevas aplicaciones, herramientas y proyectos de desarrollo que tendrán un impacto directo y disruptivo sobre muchos ámbitos de la sociedad. La Comisión Mundial de Ética del Conocimiento Científico y la Tecnología (COMEST) de la UNESCO caracteriza la inteligencia artificial de la siguiente manera:

> Los sistemas de IA son tecnologías de procesamiento de la información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones en entornos materiales y virtuales. Los sistemas de IA están diseñados para funcionar con diferentes grados de autonomía, mediante la modelización y representación del conocimiento y la explotación de datos y el cálculo de correlaciones. Pueden incluir varios métodos, como, por ejemplo, aunque no exclusivamente: i. El aprendizaje automático, incluido el aprendizaje profundo y el aprendizaje por[3] refuerzo; ii.

[3] En la cita hemos reemplazado la expresión "de" por la expresión "por", ya que "de" no es de uso habitual en el ámbito tecnológico. En el ámbito académico, la forma habitual es "aprendizaje por refuerzo". La expresión proviene del inglés *reinforcement learning* y en

> El razonamiento automático, incluidas la planificación, la programación, la representación del conocimiento y el razonamiento, la búsqueda y la optimización.
>
> Los sistemas de IA pueden utilizarse en los sistemas ciberfísicos, incluidos la Internet de las cosas, los sistemas robóticos, la robótica social y las interfaces entre seres humanos y ordenadores, que comportan el control, la percepción, el procesamiento de los datos recogidos por sensores y el funcionamiento de los actuadores en el entorno en que operan los sistemas de IA. (UNESCO, 2022, p. 10)

Desde diversos ámbitos se han expresado múltiples preocupaciones por la incertidumbre generada por las probables consecuencias negativas que puede provocar esta revolución, como la pérdida del empleo a causa de la automatización de diversas tareas, cambios importantes en la economía global, las implicaciones a nivel ético en la toma de decisiones, la autoría de artículos y trabajos académicos, las formas de evaluación, la desigualdad en el acceso a estas nuevas tecnologías y la dependencia tecnológica. A todo esto, se suma la falta de regulación y la incapacidad de los legisladores para implementar leyes que delimiten el desarrollo y el uso de la IA que permitan minimizar sus efectos adversos y que nos preparen como sociedad para adoptar, de manera responsable y provechosa, el uso de estas tecnologías en nuestra cotidianidad.

La aparición de estos nuevos desarrollos tecnológicos nos remite a los debates planteados por las corrientes de pensamiento transhumanistas con sus ambiciosas y, a veces, apocalípticas promesas, que plantean preocupaciones e incertidumbres de lo que será el futuro de la humanidad, entre ellas la promesa de la singularidad, es decir, el momento en el que la mente humana se una a las máquinas, transformado la condición biológica humana al integrar nuestras consciencias a la de una superinteligencia que será capaz de mejorarse a sí misma y dar nacimiento a nuevos sistemas inteligentes, también capaces de auto perfeccionarse y que terminarán conformando una entidad inteligente única que se expanda fuera de nuestro mundo (Diéguez, 2017, p. 18).

De los problemas mencionados, en este artículo abordaremos, en primera instancia, algunos aspectos básicos del desarrollo de la IA, tomando como punto de partida el modelo de lenguaje GPT-2 (Radford *et al.*, 2019), lanzado por la empresa OpenAI y considerado un hito crucial en la historia de

---

español se tradujo oficialmente como *aprendizaje por refuerzo*, porque describe el método mediante el cual se aprende (es decir, el aprendizaje a través de refuerzos).

los modelos de lenguaje. Su llegada demostró que los modelos de lenguaje a gran escala (LLM) podían generar textos largos, coherentes y de alta calidad sin un entrenamiento específico para cada tarea, lo que abrió la puerta al desarrollo de modelos posteriores como GPT-3 (Brown *et al*., 2020), GPT-4 (OpenAI, 2023) y el actual GPT-5 (OpenAI, 2025). En segunda instancia, nos ocuparemos de algunas de las consecuencias éticas que surgen en el proceso de implementación de estas tecnologías y de la forma como se originó esta clase de reflexiones.

II. La inteligencia artificial

Desde la conferencia de Dartmouth en 1956, en la que John McCarthy introdujo el término inteligencia artificial para nombrar el ambicioso proyecto de crear máquinas o sistemas inteligentes, surgieron promesas y visiones rimbombantes de un futuro donde las máquinas llegarían a imitar la mente humana. En las actas de dicha conferencia se evidenciaba lo ambicioso del proyecto que apenas nacía:

> Se intentará averiguar cómo lograr que las máquinas utilicen el lenguaje, establezcan abstracciones y conceptos, resuelvan tipos de problemas que ahora están reservados para los seres humanos y se perfeccionen a sí mismas. Creemos que se puede realizar un avance notable en uno o más de estos problemas si un grupo de científicos cuidadosamente escogido trabaja de manera conjunta en ellos durante un verano. (Larson, 2022, p. 68)

La meta a la que se refiere la cita anterior había sido establecida diez años antes por Alan Turing, a partir de lo que se conoce hoy como el "test de Turing" que pretende evaluar, de manera concluyente, la habilidad de una máquina para imitar el pensamiento humano. La superación de esta prueba nos permitiría afirmar de forma contundente que la inteligencia artificial puede imitar la inteligencia humana. El test de Turing consiste en someter a una computadora y a un ser humano a una serie de preguntas escritas por un juez humano. A partir de las respuestas, el juez debe tratar de identificar adecuadamente cuales fueron creadas por la computadora, si el juez no logra hacerlo, entonces tendríamos el primer sistema artificial capaz de imitar la inteligencia humana. A partir de este momento para muchos de los pioneros de la IA, el objetivo fundamental consistía en lograr que las máquinas comprendieran nuestro lenguaje: "Todo ordenador que pudiera mantener una conversación sostenida y convincente con una persona estaría […] haciendo algo para lo que es necesario el pensamiento." (Larson, 2022, p. 15)

Sin embargo, para los científicos de Dartmouth esta no sería la única manera de alcanzar el objetivo. Pensaron en un camino alternativo que consistía en lograr la programación de una máquina que tuviera la capacidad de entender el lenguaje natural. Si esto se consigue, asistiríamos al nacimiento de una "IA fuerte", una forma de inteligencia artificial similar a la humana.[4] Entonces, dado que esto llegara a ocurrir, habitaríamos en un mundo donde las computadoras dominarían el lenguaje natural y lograrían una inteligencia general como la nuestra.

Para los años 70, la IA se había propuesto lograr la traducción automática, lo que permitiría procesar textos en un idioma y transcribirlos de manera automatizada a otro idioma (Larson, 2022, p. 69). Pero a pesar de que las computadoras hicieron posible este objetivo, la calidad de los primeros resultados fue muy baja, presentando fallas aun en sistemas que tenían enfoques específicos en ciertas áreas del conocimiento, como la literatura biomédica.

Entonces surgieron estrategias que buscaron hallar la estructura sintáctica de las frases con la ayuda de las gramáticas transformacionales desarrolladas por Noam Chomsky. Pero esta tarea se enfrentó a problemas no previstos, como el doble sentido de ciertas palabras y su dependencia contextual en una frase determinada donde las otras palabras, no necesariamente cercanas, podían cambiar su significado; o problemas causados por la semántica y figuras literarias como la anáfora y la metáfora.

Los informes de Yehoshua Bar-Hillel, investigador del MIT, señalaron las grandes dificultades de la traducción automatizada, lo que tuvo un gran impacto en la comunidad de investigadores, debido a que evidenciaban los obstáculos a los que se enfrentaba la traducción automática, la falta de sentido común y de entendimiento del mundo real. Para esos tiempos, no era posible ofrecer a las máquinas un conocimiento sobre el mundo y su cotidianidad y

[4] En este sentido resulta pertinente distinguir tres tipos de inteligencia artificial: *La inteligencia artificial débil*, diseñada para llevar a cabo tareas específicas y limitadas. Por ejemplo, un algoritmo de entrenamiento de imágenes desarrollo con el propósito identificar gatos en fotografías, que solo está pensado para realizar esta tarea específica. *La inteligencia artificial fuerte*, que no se limita a realizar tareas específicas, sino que adquiere una versatilidad que sería parecida a la que tienen los seres humanos. Así estaría en condiciones, como todo ser humano, de aprender cualquier tema y actuar con gran autonomía. *La inteligencia artificial superinteligente*, que podría superar a la inteligencia humana en ámbitos diversos como la creatividad, resolución de problemas, la toma de decisiones y la adaptabilidad emocional. Hay que tener en cuenta, sin embargo, que esta distinción resulta poco precisa para dar cuenta de las diferencias intermedias que se pueden establecer entre distintas clases de inteligencia artificial que, aunque no se pueden considerar como inteligencia artificial fuerte, son más versátiles que la inteligencia artificial débil. (Morillas *et. al*., 2024, p. 33)

estos conocimientos, que para los seres humanos eran un asunto cotidiano, se tornaban confusos para las máquinas (Larson, 2022, p. 71).

Otra consecuencia de los hallazgos de Bar-Hillel, fue la suspensión de la financiación que la National Research Council proporcionaba al proyecto de IA, que ya había sobrepasado los 20 millones de dólares, una decisión justificada por los obstáculos que enfrentaba este proyecto. La consecuencia fue el estancamiento de la investigación en IA, que solo acabaría con la llegada de la internet, lo que significó un aumento exponencial de los datos, que sirvió de aliciente para que el proyecto de IA resurgiera y que las estrategias adoptadas en el pasado empezaran a mostrar resultados relevantes (Larson, 2022, pp. 71-79). Empezó a ser notable su utilidad y posibles aplicaciones en la detección de fraudes, correo spam, la clasificación de textos, el reconocimiento facial y las traducciones, lo que permitió que se reanudara la financiación.

Este nuevo impulso que tomó la investigación en IA contribuyó al avance de varias de sus ramas, entre la más relevantes: el aprendizaje profundo[5], el aprendizaje automático[6], la visión por computadora[7], la robótica[8] y el procesamiento de lenguaje natural o NLP (Natural Language Processing)[9] enfocado en la comunicación humano-máquina a través del lenguaje natural (Jurafsky y Martin, 2025). Con diversas aplicaciones como el reconocimiento y el entendimiento del lenguaje humano y la traducción automática (Khurana *et al.*, 2023) que habían constituido problemas para los proyectos de IA anteriores (Hutchins, 2014; Parra Escartín, 2012).

El entendimiento del lenguaje natural desafió a los investigadores, debido a que se encontraron frente a "un sistema complejo e intricado de expresiones humanas, regido por reglas gramaticales" (Zhao *et al*., 2023, p. 1). Para lograr el entendimiento y la generación de lenguaje por parte de las máquinas los estudios de las últimas dos décadas se centraron en el modelado del lenguaje. Pasando de modelos de lenguaje estadísticos hasta modelos de lenguaje neuro-

[5] "Subset del machine learning basado en redes neuronales artificiales con múltiples capas."

[6] "Subcampo de la IA que permite a las máquinas aprender de los datos sin ser programadas explícitamente."

[7] "Campo de la IA que entrena a las computadoras para interpretar y entender el mundo visual."

[8] "Campo que combina ingeniería y computación para diseñar y construir máquinas automatizadas."

[9] "Capacidad de un sistema de IA para comprender, interpretar y generar lenguaje humano." (Aguilar *et al*., 2024, p. 463).

nal y más recientemente a modelos de lenguaje preentrenado (PLM). Los PLMs son modelos que han sido entrenados con grandes cantidades de datos antes de otorgarles una aplicación específica. Su entrenamiento se logra a través de algoritmos denominados Transformers. Los datos masivos son pasados a los Transformers como entrenamiento para lograr que aprendan patrones complejos del lenguaje humano (Zhao *et al*., 2023, p. 1).

III. ¿Qué es el modelo de lenguaje GPT-2 y cómo funciona?

Un modelo de lenguaje en IA es un algoritmo empleado para procesar y comprender el lenguaje humano, que puede ser utilizado en gran cantidad de aplicaciones que involucran el lenguaje (traducción automática, respuesta a preguntas, generación de texto, etc.). Estos modelos se entrenan con grandes cantidades de texto, obtenido de la internet, que utilizan para aprender a reconocer patrones y entender el significado de las palabras y frases. El 14 de febrero de 2019, la empresa OpenAI anunciaba el lanzamiento de Generative Pre-trained Transformer 2 (GPT-2), un modelo de lenguaje a gran escala que superaba en gran medida a los modelos que existían hasta el momento (Radford *et al.,* 2019). En la página web de OpenAI se publicó un artículo que contenía los detalles sobre el modelo de lenguaje GPT-2:

> Hemos entrenado un modelo de lenguaje no supervisado de gran escala que genera párrafos coherentes de texto, alcanza un rendimiento de vanguardia en muchos benchmarks de modelado de lenguaje y realiza comprensión de lectura rudimentaria, traducción automática, respuesta a preguntas y resumen, todo esto sin necesidad de entrenamiento específico de tareas. (Open AI, 2023, traducción propia)

GPT-2 constituye uno de los modelos de lenguaje más avanzados desarrollados hasta la fecha que, además, muestra versatilidad en múltiples tareas de procesamiento de lenguaje natural sin un entrenamiento específico.

GPT-2 no es una máquina física, razón por la que no puede participar en el test de Turing como lo haría una computadora o un robot. Sin embargo, al ser un modelo de procesamiento del lenguaje natural avanzado, puede establecer conversaciones escritas que pueden parecer convincentemente humanas a algunas personas. La cuestión de si esto es suficiente para aprobar el test de Turing, depende de cómo se lo interprete y aplique.

Como ocurre con muchas tecnologías nuevas los riegos comienzan a evidenciarse. La IA puede generar textos extensos a partir de una oración o

un párrafo, con una coherencia y similitud que se puede equiparar a los que puede lograr un ser humano. Y uno de los primeros peligros se relaciona con la creación y difusión de noticias falsas. Como estos modelos de lenguaje se entrenan utilizando enormes cantidades de datos de texto provenientes de la Internet, asumen, en las respuestas que ofrecen, los errores o la información mentirosa e imprecisa que se encuentra allí. No hay que olvidar que estos modelos de lenguaje no tienen la capacidad de verificar la veracidad o precisión de la información que proporcionan. Una investigación realizada por la Universidad de Cornell (EE. UU.), determinó que los lectores consideraban convincentes y confiables las noticias falsas creadas a través de GPT-2 (Kreps *et al*., 2020).

Un año después, en junio de 2020, OpenAI lanzó GPT-3, que superaba de manera exponencial a su antecesor y que nos colocaba en la antesala de la era de los Large Language Models (LLM), en una de sus posibles traducciones Grandes Modelos de Lenguaje. GPT-3 había sido entrenado con millones de textos y un gran porcentaje de todo el contenido de la internet. Una vez fue liberado, los usuarios de internet plagaron las redes con muchos ejemplos que daban evidencia de la adaptabilidad de esta IA y se abrió el debate sobre si estábamos finalmente frente a una inteligencia artificial general (IAG). Pero realmente la gran novedad de GPT-3 frente a GPT-2 era la diferencia exponencial en la cantidad de datos que habían usado para entrenarla. Mientras que GPT-2 contaba con 1.500 millones de parámetros, la cantidad de parámetros de GPT-3 era de 175.000 millones (Heaven, 2020).

A pesar de que GPT-3 no era, como algunos pensaban, una inteligencia artificial general y, además, en ocasiones, cometía errores evidentemente ridículos, los temores de la comunidad de internet se hacían cada vez más evidentes. Algunas profesiones, como la de los programadores, se enfrentaron a una tecnología que, aun con limitaciones, podía generar texto y código en distintos lenguajes de programación.

El 30 de noviembre de 2022, la empresa OpenAI presentó al chat Generative Pre-trained Transformer (ChatGPT), el chatbot creado a partir de Large Language Model (LLM), un asistente virtual basado en IA que puede interactuar a través del lenguaje natural con los seres humanos con una elocuencia superior a cualquier modelo previamente creado. Se preguntó al ChatGPT lo siguiente: ¿Qué eres? ¿Quién te creo? ¿Hasta dónde llega tu conocimiento? La respuesta fue la siguiente: "Soy ChatGPT, un modelo de lenguaje de inteligencia artificial desarrollado por OpenAI, basado en la arquitectura GPT-4. Fui creado para entender y generar texto de manera coherente y útil en una amplia gama de temas y contextos. Fui creado

por OpenAI, una organización de investigación en inteligencia artificial fundada por Elon Musk, Sam Altman y otros en diciembre de 2015. Mi conocimiento se extiende hasta septiembre de 2021, lo que significa que estoy familiarizado con eventos, desarrollos y descubrimientos hasta esa fecha. No tengo información sobre eventos o avances que hayan ocurrido después de ese punto en el tiempo" (Generado por ChatGPT, 2023).

Según Eloundou (Eloundou *et al.,* 2023), GPT es un modelo de lenguaje de deep learning de tipo LLM. Los LLM como GPT están en la capacidad de analizar y generar datos secuenciales diversos, no se limitan a las aplicaciones de lenguaje natural, sino que pueden generar código en distintos lenguajes de programación computacional, como ensamblador, Python, JavaScript, etc.; e incluso generar secuencias de proteínas y partidas de ajedrez. Cabe aclarar que ChatGPT es solo una de las muchas aplicaciones que tienen los GPT.

Apenas unos meses después de la publicación de ChatGPT, el 14 de marzo de 2023, OpenAI hacía un nuevo lanzamiento, GPT-4. Un modelo notablemente superior a los ya existentes capaz de procesar no solo texto sino también imágenes. Produciendo salidas de texto. Su potencial abarca la elaboración de resúmenes de texto, sistemas de diálogo y traducción automática. Las capacidades de GPT-4 fueron puestas a prueba a través de la realización de múltiples exámenes aplicados normalmente a humanos, superando, por ejemplo, en un 10 % los resultados obtenidos por humanos en un examen especializado para abogados, sumado a su gran desempeño en 24 idiomas (OpenAI, 2023). Más recientemente, el 7 de agosto de 2025, OpenAI lanzó su nuevo modelo GPT-5, dejando claro que la investigación en el campo de la IA se enfocará hacia modelos multimodales aún más generalistas, con la capacidad de integrar texto, imágenes, audio y video además de incorporar mejoras en el razonamiento a largo plazo y en la planificación de tareas complejas (OpenAI, 2025).

IV. Las consecuencias

La calidad de las respuestas generadas en particular por el ChatGPT potenció el debate sobre sus alcances y el papel de los seres humanos frente a esta nueva tecnología, en una sociedad donde los sistemas de inteligencia artificial, a diferencia de las décadas pasadas, evolucionan de manera exponencial en solo cuestión de días o semanas. ChatGPT es solo la punta del iceberg, ya que la denominada inteligencia artificial generativa incluye también sistemas de procesamiento y generación de imágenes y audio (Eloundou *et al.*, 2023).

El 27 de marzo de 2023 OpenAI publicó un estudio sobre el impacto de los LLM como los GPTs en el mercado laboral estadounidense, centrado en la evaluación de las ocupaciones que se verían más impactadas por estos sistemas, dando como resultado que el 80% de la fuerza laboral vería afectadas sus labores en al menos un 10% por la introducción de los LLMs y el 19 % de los empleados al menos un 50% de impacto en sus tareas. El estudio también indica que la disposición de un LLM podría acelerar al menos un 15% de las tareas de los trabajadores estadounidenses. Pero con la inclusión de software y herramientas basadas en LLM, el porcentaje podría llegar hasta un 56% del total de las tareas. Según el estudio, los GPTs, y en general los LLMs, poseen características de sistemas de propósito general que pueden afectar la economía, la sociedad y las cuestiones políticas de manera directa (Eloundou *et al.*, 2023).

Frente a las preocupaciones que generó el surgimiento de los LLM en cabeza de GPT-4, acompañado de la avalancha de aplicaciones y herramientas de IA que surgieron en los meses posteriores, el 22 de marzo de 2023 se publicó una carta abierta a través de la organización Future of Life Institute, titulada "Pause on Giant AI Experiments". Estaba firmada, entre otras personalidades, por Elon Musk, CEO de SpaceX, Tesla y Twitter; Steve Wozniak, cofundador de Apple; y Yuval Noah Harari, autor, historiador y profesor de la Universidad Hebrea de Jerusalén. Como su título lo indica, la carta pide una pausa de al menos seis meses en el entrenamiento de los LLMs superiores a GPT-4, pero sin afectar las investigaciones sobre la IA. Se pide prudencia debido al profundo impacto que causarán en la sociedad y, en general, en los sistemas de IA con inteligencia competitiva y al hecho de que no se están llevando a cabo la gestión y planeación debidas. La carta pide que la pausa se haga pública y pueda verificarse y que, en caso tal de que no se realice, los gobiernos intervengan con legislaciones que exijan la moratoria. Sin embargo, los laboratorios de IA se han lanzado a una carrera para desarrollar y desplegar sistemas inteligentes cada vez más potentes (Future of Life Institute, 2023).

La carta también plantea interrogantes frente a los sistemas inteligentes que ya compiten con los humanos en la realización de tareas generales y cuya respuesta no debe quedar en manos de líderes tecnológicos no electos: "¿Deberíamos permitir que las máquinas inunden nuestros canales de información con propaganda y mentiras? ¿Deberíamos automatizar todos los trabajos, incluso los gratificantes? […] ¿Deberíamos arriesgarnos a perder el control de nuestra civilización?" (Future of Life Institute, 2023, párr. 2).

Una de las preguntas, no incluida en la cita anterior, encierra una relación directa con las proyecciones esperadas de algunos transhumanistas como

Raymond Kurzweil: "¿Deberíamos desarrollar mentes no humanas que eventualmente podrían superarnos en número, inteligencia, obsolescencia y reemplazarnos?" (Future of Life Institute, 2023, párr. 2).

El término singularidad, que más que plantearnos un futuro parece un apocalipsis tecnológico, reaparece con más fuerza en los debates y en los temores de la humanidad frente a los sistemas inteligentes. Pero el origen de este término en el contexto de la evolución tecnológica "imparable" surgió en los años cincuenta en boca de John Von Neumann, en una conversación que sostuvo con el también matemático Stanislaw Ulam. Neumann mencionó un posible punto de inflexión tecnológica que cambiaría por completo la vida humana como la conocemos. Con la aparición de la bomba atómica, proyecto en el que Neumann participaba, las posibilidades de un futuro distópico provocado por la tecnología y su evolución aparentemente imparable tomaron mayor fuerza. La singularidad, que en matemáticas hace referencia a un punto que se vuelve indefinido, un valor que se hace infinito fue la metáfora usada por Neumann para preguntar a Ulam si el avance de la tecnología llegaría a ese punto infinito donde las ideas, los planes y cualquier tipo de acción no sería frutífera frente a su imparable progreso. Pero quien introdujo el término en el ámbito informático y específicamente en el campo de la inteligencia artificial fue el científico Vernor Vinge a través de su obra de ciencia ficción *Marooned in Realtime* y, posteriormente, a través de un artículo técnico para la Nasa en el que según Vinge, treinta años más tarde, los medios tecnológicos permitirían la creación de una inteligencia superhumana que una vez apareciera daría fin a la humanidad como la que conocemos. Este acontecimiento sería una singularidad. Raymond Kurzweil profundizará en la idea de la singularidad a través de algunas de sus publicaciones, en los años noventa con *La era de las máquinas inteligentes* y *La era de las máquinas espirituales* y, en 2005, con su libro *La singularidad está cerca: cuando los humanos trascendamos la biología* (Larson, 2022, pp. 60-66).

En este último libro Kurzweil predice que en el año 2029 el test de Turing será superado por la máquina indicando que, finalmente, la IA se habrá equiparado a la humana, la máquina entrará en una especie de estado de preparación, evolución y perfeccionamiento, que acabará en el "advenimiento de la singularidad" en el año 2045. Cuando esto ocurra, nuestro apreciado "dominio" sobre el mundo habrá terminado, una superinteligencia tomará control sobre todo lo existente, incluidos los materiales y recursos energéticos que usará para garantizar su continua evolución hasta lograr su expansión por el cosmos. Entonces, será el fin de la era humana y el surgimiento de una era denominada por Diéguez "postbiológica" (Diéguez, 2017, p. 48).

Pero es necesario aterrizar un poco estas ideas. Tales predicciones quizás no lleguen a ocurrir en lo absoluto y, en cambio, temores latentes y reales empiezan a volverse una realidad tal vez aun evitable. En un reciente artículo, Slavoj Žižek expone varias cuestiones, a pesar de la carta de moratoria mencionada con anterioridad. Žižek se plantea los siguientes interrogantes: ¿Cuál es la razón del pánico de las élites tecnológicas? Durante la pausa propuesta para evaluar los riesgos de las IA ¿quién defenderá a la humanidad? o ¿quién va a representarla? ¿cómo lograr un gran debate público global, que Žižek considera difícil de alcanzar, que garantice que los laboratorios de India, Rusia y China también hagan la aclamada pausa?

Nos enfrentamos, según Žižek, a ciencias post-humanas cuyo fin no es la dominación. La sorpresa es el nuevo credo, pues no se sabe a ciencia cierta qué propiedades no deseadas e imprevistas se gestarán y emergerán de los sistemas de IA y que, dadas las características de caja de negra de estas tecnologías, son invisibles para nosotros (Žižek, 2023).

El Instituto Internacional para la Educación Superior de la UNESCO (2023), realizó una investigación sobre el impacto de las inteligencias artificiales en la educación, particularmente el ChatGPT.

Esta investigación plantea que debemos afrontar la llegada de esta tecnología disruptiva y preguntarnos si el camino indicado es la prohibición del uso de tecnologías como los GPT o, por el contrario, si es necesario promocionar su uso, pero acompañado de un marco normativo que regule y establezca límites adecuados para que el control siga estando del lado humano y que la calidad educativa no se vea comprometida de manera negativa. Según el IESALC debemos cuestionarnos y tomar acciones frente a los siguientes aspectos que surgen del uso de una tecnología como esta:

—Falta de regulación
—Protección de los datos
—Sesgo cognitivo
—Género y diversidad
—Accesibilidad
—Comercialización (UNESCO, 2023).

V. Consideraciones éticas sobre la inteligencia artificial

En términos generales, la ética de la IA se ocupa del comportamiento moral de los seres humanos cuando diseñan, fabrican y usan sistemas artificialmente inteligentes, y también de las consecuencias que surgen de su implementación. Además, esta disciplina contribuye a la creación de principios éticos y líneas de investigación que tratan sobre los riesgos y el impacto sociotecnológico de la inteligencia artificial (Arenas *et al*., 2020).

Los problemas planteados por la ética de la IA pueden ser abordados desde la perspectiva tanto de los usuarios, como de los desarrolladores; en el caso de estos últimos se abordan cuestiones relacionadas con los prejuicios, sesgos y creencias que se reflejan en los algoritmos que diseñan y que, por lo tanto, no son neutrales. Esto se ve ejemplificado en los programas de identificación de rostros en los que se han detectado sesgos de sexismo, machismo, xenofobia y otros prejuicios peligrosos.

Los esfuerzos para desarrollar una ética de la IA que esté apoyada en principios clásicos como la prudencia, la autonomía restringida y la responsabilidad, deben ser complementados por mecanismos jurídicos, de control político y sistemas educativos apropiados. De no ser así, se reduce la ética a un conjunto de gestos simbólicos incapaces de enfrentar, adecuadamente, los desafíos que plantea el surgimiento y desarrollo de esta nueva tecnología. Esto significa que estamos frente a problemas que trascienden el ámbito de lo ético y que incorporan componentes de orden social, educativo, cultural e histórico (Maldonado, 2024).

La ética de la IA adquiere relevancia académica tan solo a partir del año 2016, lo que se explica a partir de varios eventos. Desde el punto de vista académico es importante señalar la publicación del artículo de Luciano Floridi y Mariarosaria Taddeo (2016), titulado: "What is data ethics?", en el que se expresa la necesidad de sentar las bases para una reflexión sobre la ética de datos, con el propósito de evaluar problemas morales relacionados con la implementación, la generación, el registro, la conservación, el procesamiento, la difusión, el intercambio y el uso de algoritmos. Estas dificultades incluyen la inteligencia artificial, los agentes artificiales, el aprendizaje automático y los robots, además de las prácticas correspondientes como la innovación responsable, la programación, la piratería informática y los códigos profesionales, que tienen como objetivo formular y apoyar soluciones moralmente buenas como, por ejemplo, la integración de conductas o valores correctos en el desarrollo de estas tecnologías.

> La ciencia de datos ofrece enormes oportunidades para mejorar la vida privada y pública, así como nuestro medio ambiente (pensemos en el desarrollo de ciudades inteligentes o los problemas causados por las emisiones de carbono). Desafortunadamente, estas oportunidades también conllevan importantes desafíos éticos. El uso extensivo de cada vez más datos —a menudo personales, si no sensibles (big data)— y la creciente dependencia de algoritmos para analizarlos con el fin de definir opciones y tomar decisiones (incluido el aprendizaje automático, la inteligencia artificial y la robótica), así como la reducción gradual de la intervención humana o incluso de la supervisión de muchos procesos automáticos, plantean cuestiones urgentes de equidad, responsabilidad y respeto de los derechos humanos, entre otras. (Floridi y Taddeo, 2016, p. 2)

También se destaca el papel desempeñado por publicaciones de divulgación científica como "Weapons of Math Destruction" (Armas de destrucción matemática) de Cathy O'Neil (2017), que contribuyeron a crear conciencia sobre los efectos sociales de la inteligencia artificial. Otro evento importante fue el cuestionamiento del uso de ciertos algoritmos como COMPAS (Perfiles de Gestión de Delincuentes Correccionales para Sanciones Alternativas), herramienta de aprendizaje automático empleada en el sistema de justicia penal de los Estados Unidos para evaluar el riesgo de reincidencia de una persona. Se trata de un caso que se ha vuelto paradigmático en la discusión de problemas éticos que surgen cuando se aplican algoritmos en ámbitos de gran incidencia pública y que tienen profundos impactos sociales, que se concretan en asuntos tan sensibles como la disparidad de raza y género, la discriminación sistemática, las definiciones de justicia y la influencia de los sesgos.

También se destaca el reporte, publicado en el 2018, del Instituto AI Now (Whittaker *et al.*, 2018), orientado por la Universidad de Nueva York, en el que se plantean problemas relacionados con la falta de privacidad, por ejemplo, brechas de seguridad y exposición de datos de Facebook; sesgos y discriminación, un algoritmo que sugiere masivas deportaciones en Reino Unido; daño físico como sucedió con los accidentes de Tesla y Uber; y daño moral como en los casos de Cambridge Analytica y su influencia en las elecciones del Brexit en el Reino Unido y presidenciales en los estados Unidos. Simultáneamente a estos debates surgieron las primeras iniciativas legales y normativas que proponían directrices para guiar estas nuevas tecnologías, como el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, la ley de privacidad del consumidor en California, en los Estados Unidos, y las peticiones de Microsoft para regular el reconocimiento

facial, expresadas, en julio de 2018, al gobierno de los Estados Unidos (Russell y Norvig, 2004).

El desarrollo de la IA plantea importantes dilemas éticos y riesgos que deben ser abordados por los investigadores y la sociedad. A continuación, se señalan algunos de estos problemas éticos:

- **La pérdida de empleos provocada por la automatización:** La IA está cambiando aceleradamente el mercado laboral, lo que hace necesario entender los riesgos y oportunidades que genera. Algunos estudios indican que podría desplazar una gran cantidad de puestos de trabajo, entre los que se cuentan: programadores, desarrolladores de software, contables, abogados, analistas de datos, médicos, científicos de datos, ingenieros, radiólogos, personas dedicadas a la ciberseguridad.
- **El efecto de la presión laboral sobre el tiempo de ocio:** no ha sido posible, como se esperaba, que la tecnología disminuya la jornada laboral; lo que ha ocurrido, por el contrario, es que en sectores relacionados con el conocimiento la exigencia por trabajar más ha aumentado a causa de la competencia y la aceleración de la innovación.
- **La pérdida del sentido del carácter único de los seres humanos:** la IA cuestiona la idea de que los humanos son únicos, al indicar que la inteligencia no es un atributo exclusivo de las personas. Algo parecido había ocurrido con Copérnico cuando sostuvo que la tierra no estaba en el centro del universo, o con Darwin quien colocó al hombre al mismo nivel que las otras especies del planeta.
- **La disminución de la privacidad:** a medida que la IA sigue desarrollándose, surgen riesgos de privacidad que se vuelven cada vez más complejos y que son ocasionados, principalmente, por la falta de control sobre el uso, sin aprobación, de la información personal. Muchas tecnologías como, por ejemplo, el reconocimiento de voz y los sistemas de vigilancia masiva amenazan seriamente las libertades civiles y la privacidad individual.
- **El surgimiento de problemas relacionados con la pérdida de responsabilidad legal:** a medida que la IA se vuelva más autónoma, uno de los problemas más importantes será si se podrá considerar una entidad legal independiente. Se trata de una pregunta que ya se ha formulado con respecto a las Organizaciones Autónomas Descentralizadas, que en esencia se construyen y operan sobre reglas automatizadas codificadas en contratos inteligentes, almacenadas y ejecutadas en blockchains, en las que disminuye la intervención

humana. El nivel de desarrollo alcanzado por la IA, que le permite imitar los procesos de pensamiento y las tareas humanas, plantea la pregunta de si tales tecnologías deben estar sujetas a protección legal. Aunque se trata de un asunto que no ha sido abordado por los tribunales, los gobiernos y los reguladores parecen no aceptar tal posibilidad. Por ejemplo, el Parlamento Europeo rechazó la propuesta de otorgar personalidad jurídica a la IA, argumentando que cualquier cambio legal debería *comenzar con la aclaración de que los sistemas de IA no tienen personalidad jurídica ni conciencia humana* (Parlamento Europeo, 2023). Por ejemplo, el uso de sistemas expertos en medicina y agentes inteligentes pone en cuestión la atribución de la responsabilidad en caso de errores o daños, lo que tiene implicaciones legales, muchas de las cuales están aún por precisar.

- **La amenaza existencial proveniente del desarrollo de una IA avanzada:** la posibilidad de que se produzca una singularidad tecnológica, a la que se hizo referencia en párrafos anteriores, consistente en la aparición una superinteligencia que supere a la inteligencia humana plantea serias dificultades sobre el futuro de la humanidad y la relación con las máquinas conscientes.
- **Los derechos y la ética para robots conscientes:** en el caso de que los robots pudieran adquirir conciencia, entonces sería necesario considerar sus derechos y responsabilidades morales, un asunto que hasta el momento se ha mantenido en los límites de la ciencia ficción y que, sin embargo, podría volverse una realidad.

En el ámbito de la ética de la IA, han surgido un conjunto de subdisciplinas que abordan diversos aspectos que han resultado del desarrollo de esta tecnología. A continuación, señalamos algunas de ellas:

- *Ética de los datos*: trata de la recopilación, uso, almacenamiento, manipulación y distribución de los datos que se emplean en los sistemas de IA, y que dan lugar a problemas como: la protección de la privacidad, el consentimiento, la transparencia, la equidad, la seguridad, la gobernanza, los sesgos, y la responsabilidad.
- *Ética de las máquinas (ética computacional o moralidad artificial)*: busca implementar principios morales y preferencias en los procesos de toma de decisiones de sistemas artificiales y máquinas inteligentes. Aborda problemas como: el estatus moral de máquinas y la IA, la creación de agentes morales artificiales, como sistemas de armas, predicciones, y decisiones automatizadas; justificación ética de las

acciones, adaptación a diferentes contextos éticos, evitar daños y promover beneficios.

- *Ética y riesgo de singularidad*: como ya se señaló la singularidad es un concepto propuesto por John Von Neumann y desarrollado por pensadores como Ray Kurzweil, que hace referencia a un futuro en el que la evolución tecnológica produce una aceleración exponencial del progreso, generando transformaciones impredecibles y profundas en la sociedad. Este tipo de ética reflexiona sobre tales transformaciones y los riesgos existenciales que provocaría para la humanidad el desarrollo de una superinteligencia.
- *Interacción humano-robot*: es un campo interdisciplinario que reflexiona sobre la manera cómo los seres humanos y los robots se comunican, colaboran y coexisten. Trata de entender el contexto y las capacidades de toma de decisiones de los robots para mejorar la comunicación y la colaboración con los seres humanos. Plantea el problema de la necesidad y posibilidad de derechos para los robots.
- *Ética de algoritmos*: se ocupa del estudio y la aplicación de principios morales a los algoritmos y sistemas de IA, con el propósito de minimizar los daños y maximizar los beneficios para la sociedad. Aborda problemas como: equidad y no discriminación, transparencia y explicabilidad, responsabilidad y rendición de cuentas, impacto social, privacidad y protección de datos, supervisión humana, sostenibilidad, los sesgos, la replicabilidad, la interpretabilidad (Arenas *et al*, 2020).

VI. El problema del sesgo algorítmico

El "sesgo algorítmico" se produce cuando los algoritmos de aprendizaje automático generan resultados injustos o discriminatorios provocados por errores sistemáticos que, por lo regular, benefician a un grupo de individuos frente a otro y reflejan o refuerzan prejuicios que se encuentran en los datos de entrenamiento o en el proceso de diseño del algoritmo, lo que puede conducir a resultados distorsionados y potencialmente dañinos (Ferrante, 2021).

La internet se ha convertido en una máquina de recopilación masiva de información personal de sus usuarios, que se emplea para personalizar contenidos y publicidad, lo que tiene un efecto sobre la privacidad, la libertad y el pensamiento crítico de quienes la utilizan. También aumenta la vulnerabilidad ante los ciberataques y la desconfianza y el riesgo que suponen el comercio de datos que realizan diversas empresas.

Acceder al enorme cúmulo de información que nos ofrece la internet, implica encontrarnos, en cada sitio que visitamos, una red de seguimiento constante, que se operacionaliza a través de una docena de los denominados cookies de rastreo, es decir, pequeños archivos de texto que los sitios web guardan en el navegador del usuario para almacenar información sobre sus visitas y sus diversas preferencias, posibilitando de esta manera la personalización de sus contenidos, a través de la recopilación de información sobre las necesidades y las intenciones de la gente, a partir de los hábitos de navegación, los patrones de movilidad, las búsquedas en internet, etc. Esta tarea se lleva a cabo en los grandes centros de almacenamiento, donde los datos obtenidos son analizados por la inteligencia artificial y el aprendizaje automático. Este proceso de aprendizaje posibilita localizar patrones en el conjunto de datos, lo que permite emplearlos para la realización de predicciones. Estos datos pueden estar constituidos por imágenes, sonidos, texto escrito, redes, posiciones de un GPS, tablas o cualquier otra representación. La información así procesada permite potenciar el comercio electrónico, a través de la publicidad y el marketing, a través de recomendaciones dirigidas a los usuarios sobre diversos productos de electrónica, películas, ropa, etc. (Storm *et al.*, 2016). Este manejo de los datos crea un problema de seguridad, ya que los expone a situaciones de espionaje, chantajes o abusos. Además, agrupar datos en centros y conectar infraestructuras críticas a internet aumentan la amenaza de ciberataques que pueden tener impactos sociales graves (Ferrante, 2021).

Esta manera de comerciar ha generado un nuevo tipo de economía de datos, en el que han aparecido nuevos actores como los "revendedores de información", también conocidos como data brokers, que se dedican a recopilar información minuciosa acerca de lo que compran las personas, su etnia, sus preocupaciones económicas y de salud, sus actividades en la web y en las redes sociales. Otras empresas se especializan en elaborar rankings de consumidores empleando algoritmos de IA para hacer valoraciones de los usuarios de internet, identificando a los que se podrían denominar consumidores "valiosos", porque cumplen ciertas condiciones que permiten ofrecerles beneficios, como tarjetas de crédito, por ejemplo. Al utilizar diversos medios electrónicos, como el internet, los consumidores van dejando una huella digital que se traduce en datos con los que se comercia para obtener beneficio económico, condicionándolos y creando potenciales sesgos de información (Storm *et al*., 2016).

Consideremos algunas de las razones por las que estos sistemas pueden generar predicciones sesgadas.

Los modelos desarrollados a través de la IA pueden adquirir un sesgo que les permite desempeñarse de forma discriminatoria con respecto a grupos caracterizados por distintos atributos demográficos, lo que se origina en el tratamiento de los datos que se utilizan para entrenarlos. Por ejemplo, para el desarrollo de los modelos más extendidos de clasificación de imágenes se emplea ImageNet, una de las bases de datos de imágenes etiquetadas más grandes del planeta. La cuestión radica en que más del 45% de estas imágenes provienen de Estados Unidos, lo que significa que hacen referencia a una realidad cultural cuyas representaciones corresponden al hemisferio norte. Esto permite comprender la compleja relación que existe entre los datos, los modelos y las personas que los diseñan. Las decisiones relacionadas con la población utilizada para construir las muestras y las variables que se miden desempeñan un papel fundamental en el tratamiento que se hace de los datos. Por lo tanto, es muy difícil que se puede considerar que esta forma de organizar los datos se pueda considerar neutral, ya que las personas encargadas de diseñar estos sistemas poseen sus propias visiones del mundo, prejuicios, valoraciones de los hechos, intereses y sesgos adquiridos a lo largo de su experiencia de vida, que se refleja en el diseño y la definición de criterios empleados para evaluar estos modelos.

El mundo que habitamos nos sumerge, cada vez más, en un entramado de sistemas informacionales o computacionales que permean casi todas las actividades que llevamos a cabo diariamente a través de correos electrónicos, llamadas, navegación en internet, chats, operaciones bancarias, viajes en avión, compras con tarjetas de débito o de crédito. Para dar cuenta de esta situación, hablamos de la *huella digital* de una persona, que consiste en el registro de datos que va dejando en la web. Esto ha hecho posible el acceso a los contenidos de cada dispositivo tecnológico y al conocimiento de las creencias, los gustos, los comportamientos, los hábitos, las relaciones, los temores de cualquier persona. Este proceso de digitalización al que nos vemos abocados a través del manejo y procesamiento de datos (analítica de datos), ha transformado profundamente la vida de los seres humanos en la actualidad, convirtiendo los datos y la información en mecanismos de control y manipulación.

Esta invasión digital de nuestras vidas ha provocado que las personas estén siendo clasificadas y manipuladas de diversas maneras, a nivel comercial, político, económico, financiero, militar y diplomático. En realidad, estos sistemas de clasificación constituyen sistemas de exclusión. La información que se obtiene de las personas a través de los diversos dispositivos tecnológicos puede guardarse indefinidamente y cruzarse con otros datos, para ser explotada en el mercadeo de productos, sin el

consentimiento de quienes son los verdaderos dueños de tales datos. En términos técnicos, el fundamento teórico de esta tecnología es la estadística bayesiana, que clasifica variables en función del resultado deseado con mayor impacto; es decir, el análisis de la probabilidad de que un evento suceda por la ocurrencia de otro evento o suceso. Por ejemplo, si alguien busca en internet la palabra "casa", hay una gran probabilidad de que se interese por la palabra "edificio". Si alguien ha tenido determinado comportamiento X, es muy probable que más adelante tenga un comportamiento sucesivo de tipo Y (Maldonado, 2024).

En cierto sentido, los datos, la IA e internet están diseñando la vida de las personas. La internet nos sumerge en un mundo de publicidad, comercio y consumo, que está potenciando más nuestra atención que nuestras intenciones. Quienes diseñan estas tecnologías saben que el tiempo de las personas es limitado, por lo que ofrecen información comprimida y de corta duración: vídeos que se reproducen solos o una avalancha de notificaciones. Por esta razón también se maximizan objetivos como "tiempo pasado en el sitio", "número de visualizaciones del vídeo", "número de visitas a la página", etc. Se trata de un entorno que potencia cierto tipo de eficiencia con respecto a la información que se ofrece y se recibe, pero que al mismo tiempo crea adicciones, dependencias y acondicionamientos que impiden realizar razonamientos a largo plazo y pensar libremente. Lo que pasa por ser mera diversión, como videoclips, mensajes, servicios de datos, etc., puede alejar a las personas de sus propias necesidades, sus propios objetivos, la capacidad de reflexionar y liberarse de lo menos importante, para concentrarse en asuntos más valiosos (Rodríguez, 2018).

Algunas investigaciones han señalado los efectos de la internet sobre la memoria,[10] detectando la aparición de una tendencia que los investigadores denominan "descarga cognitiva" o externalización de la memoria. Se ha descubierto que cada vez que utilizamos los motores de búsqueda y otras herramientas para acceder a la información, aumenta nuestra confianza en esta tecnología como una extensión de nuestra memoria. Como consecuencia,

[10] Estudio realizado por los investigadores Benjamin Storm, Sean Stone y Aaron Benjamin de las Universidades de California e Illinois. Estos investigadores diseñaron experimentos para establecer nuestra probabilidad de utilizar dispositivos como una computadora o un teléfono inteligente para contestar preguntas. Los resultados establecieron que la mayor parte de los participantes utilizó Google para responder a las preguntas que se les plantearon, y que muy pocos confiaron en su memoria. El 30% de los participantes que utilizó la internet ni siquiera intentó responder a una sola pregunta simple apelando a su memoria. En un entorno donde obtener información depende de un simple clic, se reduce la necesidad de recordar hechos triviales, como figuras y números, para desempeñarnos en la vida cotidiana (Storm *et al*., 2016).

el cerebro podría estar adaptándose para almacenar menos información en la memoria a largo plazo (Rodríguez, 2018).

A todo lo anterior se suma el hecho de que ese entramado de sistemas informacionales o computacionales, en el que estamos sumergidos, genera una ruptura profunda entre el saber técnico que sustenta este sistema y las condiciones de utilización de los objetos técnicos que caracterizan nuestro papel como usuarios de la tecnología. Los dispositivos con los que tratamos cotidianamente están basados en un saber técnico cada vez más complejo, al que, como usuarios, tenemos menos acceso y comprensión. Esto tiene como consecuencia una devaluación de la autonomía que deberíamos tener, como usuarios, con respecto a dichas tecnologías.

La apropiación de los datos personales de los usuarios de la internet, por parte de las grandes compañías que desarrollan estas tecnologías, está creando una sociedad en la que la gente alcanza cierto bienestar a través del consumo al costo de ceder la intimidad y el control de su vida, es decir, sacrificando su privacidad. Por esta razón, se hace necesario reivindicar el derecho a disponer de una vida privada y del momento y la razón para no hacerla pública, impulsando la comprensión sobre el precio y el valor de la privacidad (Rodríguez, 2018).

Referencias

Aguilar, I., Alepuz, V., Alfaro, J., Bañón, J. J., Botti, V., Despujol, I., Giménez, J., Linares, J., Linares, J. M., Majadas, V., Martínez, J., Monsoriu, M., Montesa, E., Morillas, C., Muñoz, J. M., Ortega, J., Ortuño, A., Peñarrubia, J. P., Plasencia, A., Rieta, J., Sales, M. y Segarra, R. (Eds.) (2024). *Guía básica de la IA*. Smart Digital

Arenas, M., Arriagada, G., Mendoza, M. y Prieto, C. (2020). *Una breve mirada al estado actual de la Inteligencia Artificial.* Centro de Desarrollo Docente, Pontificia Universidad Católica de Chile. Recuperado de: https://desarrollodocente.uc.cl/wp-content/uploads/2020/09/Una-breve-mirada-al-estado-actual-de-la-Inteligencia-Artificial.pdf

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners*. Arxiv. https://doi.org/10.48550/arXiv.2005.14165

Diéguez, A. (2017). *Transhumanismo. La búsqueda tecnológica del mejoramiento humano*. Herder.

Eloundou, T., Manning, S., Mishkin, P. y Rock, D. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*. Arxiv. https://doi.org/10.48550/arXiv.2303.10130

Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos. ¿Por qué deberían importarnos? *Nueva sociedad*, (294), 27-36.
Floridi, L. y Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A, 374*(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360
Future of Life Institute. (22 de marzo de 2023). *Pause Giant AI Experiments: An Open Letter*. Recuperado de: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
Heaven, W. D. (20 de julio de 2020). *OpenAI's new language generator GPT-3 is shockingly good — and completely mindless*. MIT Technology Review. Recuperado de: https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/
Hutchins, J. (2014). *The history of machine translation in a nutshell*. ACL Anthology. Recuperado de: https://aclanthology.org/www.mt-archive.info/10/Hutchins-2014.pdf
Jurafsky, D. y Martin, J. H. (2025). *Speech and Language Processing* (3rd ed. draft). Stanford University. Recuperado de: https://web.stanford.edu/~jurafsky/slp3/
Khurana, D., Koli, A., Khatter, K. y Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications, 82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4
Kreps, S., Miles McCain, R. y Brundage, M. (2020). All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science, 9*(1), 104-117. https://doi.org/10.1017/XPS.2020.37
Larson, E. J. (2022). *El mito de la inteligencia artificial*. Shackleton Books.
Maldonado, C. E. (2024). *Inteligencia artificial y ética*. Ediciones desde abajo.
Morillas, C., Despujol, I. y Montesa, E. (2024). Tipos de inteligencia artificial. En I. Aguilar, V. Alepuz, J. Alfaro, J. J. Bañón, V. Botti, I. Despujol, J. Giménez, J. Linares, J. M. Linares, V. Majadas, J. Martínez, M. Monsoriu, E. Montesa, C. Morillas, J. M. Muñoz, J. Ortega, A. Ortuño, J. P. Peñarrubia, A. Plasencia, J. Rieta, M. Sales, y R. Segarra (Eds.), *Guía básica de la IA* (pp.33-42). Smart Digital.
O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.
OpenAI (2023). *GPT-4 Technical Report*. Arxiv. https://doi.org/10.48550/arXiv.2303.08774
OpenAI (7 de agosto de 2025). *Introducing GPT-5*. Recuperado de: https://openai.com/es-419/index/introducing-gpt-5/
Parlamento Europeo. (12 de junio de 2023). *Ley de IA de la UE: primera normativa sobre inteligencia artificial*. Recuperado de: https://www.europarl.europa.eu/topics/es/article/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial
Parra Escartín, C. (2012). Historia de la traducción automática. *La Linterna del Traductor*, (6), 85-91.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. y Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. OpenAI. Recuperado de: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Rodríguez, P. (2018). *La inteligencia artificial. Cómo cambiará el mundo (y tu vida)*. Titivillus, epulibre.

Russell, S. J. y Norvig, P. (2004). *Inteligencia artificial. Un enfoque moderno* (2a. Ed.). Pearson.

Storm, B. C., Stone, S. M. y Benjamin, A. S. (2016). Using the Internet to access information inflates future use of the Internet to access other information. *Memory, 25*(6), 717-723. https://doi.org/10.1080/09658211.2016.1210171

UNESCO (2022). *Recomendaciones sobre la ética de la inteligencia artificial*. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNESCO (2023). *ChatGPT e inteligencia artificial en la educación superior: guía de inicio rápido*. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000385146_spa

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Mysers West, S., Richardson, R., Schultz, J. y Schwartz, O. (2018). *AI Now Report 2018*. AI Now Institute at New York University.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Jiang, J., Chen, Z., Zhang, J., Chen, Y., Han, X., Song, R., Tang, Y., Li, X., Wang, J., Xie, X., Zhang, M., Sun, M., … Wen, J.-R. (2023). *A Survey of Large Language Models*. Arxiv. https://doi.org/10.48550/arXiv.2303.18223

Žižek, S. (9 de abril de 2023). *The Post-Human Desert*. The Daily Star. Recuperado de: https://www.thedailystar.net/opinion/views/project-syndicate/news/the-post-human-desert-3292651

Datos de financiación del artículo

Los autores declaran que no recibieron financiación para este artículo.

Implicaciones éticas

Los autores no tienen ningún tipo de implicación ética que se deba declarar en la escritura y publicación de este artículo.

Declaración de conflicto de interés

Los autores declaran que no tienen ningún conflicto de interés en la escritura o publicación de este artículo.

Contribuciones de los autores

Julián Henao Henao: escritura (borrador original), escritura (revisión del borrador y revisión/corrección).
Luis Humberto Hernández Mora: escritura (borrador original), escritura (revisión del borrador y revisión/corrección).

Autor de correspondencia

Luis Humberto Hernández Mora. luis.h.hernandez@correounivalle.edu.co. Ciudad Universitaria Meléndez, Calle 13 # 100-00 Santiago de Cali, Valle del Cauca, Colombia.

Declaración de uso de inteligencia artificial

Los autores declaran que no utilizaron ningún programa o aplicación de inteligencia artificial.