GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

TRABAJO FIN DE GRADO

CHATGPT, EL AVANCE DE LA IAG Y LA NECESIDAD DE UN NUEVO MARCO ÉTICO

Autor: **Abellás Rodríguez, Liher**

Director: **Fuertes Pérez, Javier**

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

‘ChatGPT, el avance de la IAG y la necesidad de un nuevo marco ético’

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2023/24 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Abellás Rodríguez, Liher. Fecha: 01/ 12/ 2023

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Fuertes Pérez, Javier. Fecha: 01/ 12/ 2023

Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Liher Abellás Rodríguez, estudiante de MII+ADE de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado " CHATGPT, EL AVANCE DE LA IAG Y LA NECESIDAD DE UN NUEVO MARCO ÉTICO", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.

2. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.

3. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

4. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 01/12/2023

Firma: 

GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

TRABAJO FIN DE GRADO

CHATGPT, EL AVANCE DE LA IAG Y LA NECESIDAD DE UN NUEVO MARCO ÉTICO

Autor: **Abellás Rodríguez, Liher**

Director: **Fuertes Pérez, Javier**

Madrid

RESUMEN

Este trabajo examina el complejo escenario ético que plantea la Inteligencia Artificial Generativa (IAG), con un enfoque particular en el modelo ChatGPT. El creciente potencial de esta tecnología, su capacidad revolucionaria y la ausencia de regulaciones específicas para la IAG han suscitado dilemas éticos y sociales que requieren ser abordados de manera urgente.

El objetivo principal es proporcionar una guía integral y reflexiva para el desarrollo y aplicación responsable de la IAG. Para ello, primero se ha analizado el desarrollo histórico de la IA hasta la irrupción de las IAG con ChatGPT, después se han identificado los desafíos éticos y sociales que esta tecnología suscita, y se ha justificado la necesidad inmediata de regularla. Una vez explicado cómo surge y qué es esta tecnología, cuál es su impacto y la falta de regulación que hay al respecto, este trabajo culmina con la proposición de un marco ético robusto que integra diversas perspectivas: la ética aplicada de Adela Cortina, el principio de libertad y autonomía de Kant, el marco aristotélico-tomista y el principio de responsabilidad de Hans Jonas.

La contribución de este trabajo radica en su carácter innovador y enfoque holístico, que pretende servir de referencia para usuarios, diseñadores, desarrolladores y reguladores, fomentando prácticas éticas en el desarrollo y uso de esta tecnología. Este marco ético destaca la importancia de la responsabilidad, transparencia y autonomía en el contexto de la IAG, así como lo fundamental que es que el ser humano permanezca en el centro de esta nueva era tecnológica.

Palabras clave: Ética, Inteligencia Artificial, Inteligencia Artificial Generativa, ChatGPT.

ABSTRACT

This paper examines the complex ethical scenario posed by Generative Artificial Intelligence (Gen AI), with a particular focus on the ChatGPT model. The growing potential of this technology, its revolutionary capabilities, and the absence of specific regulations for Gen AI have raised ethical and social dilemmas that need to be urgently addressed.

The main objective is to provide a comprehensive and thoughtful guide for the responsible development and application of Gen AI. To this end, we have first analyzed the historical development of AI up to the irruption of Gen AI with ChatGPT, then identified the ethical and social challenges that this technology raises and justified the immediate need to regulate it. Having explained how this technology arises and what it is, its impact and the lack of regulation, this work culminates with the proposal of a robust ethical framework that integrates various perspectives: Adela Cortina's applied ethics, Kant's principle of freedom and autonomy, the Aristotelian-Thomistic framework and Hans Jonas's principle of responsibility.

The contribution of this work lies in its innovative character and holistic approach, which aims to serve as a reference for users, designers, developers and regulators, encouraging ethical practices in the development and use of this technology. This ethical framework highlights the importance of accountability, transparency and autonomy in the context of Gen AI, as well as how fundamental it is that the human being remains at the center of this new technological era.

Keywords: Ethics, Artificial Intelligence, Generative Artificial Intelligence, ChatGPT.

ÍNDICE

1 INTRODUCCIÓN

En la actual era de la revolución digital, la Inteligencia Artificial (IA) ha experimentado un avance significativo, llevando a la Inteligencia Artificial Generativa (IAG) a la vanguardia de la tecnología. Con ChatGPT como un exponente destacado, esta forma de IA no solo replica patrones y responde a instrucciones, sino que genera contenido de manera autónoma, transformando la interacción humano-máquina.

Este trabajo de fin de grado (TFG) se adentra en el complejo terreno de la IAG, con un enfoque particular en ChatGPT como representación paradigmática de esta tecnología. Más allá de abordar sus fundamentos técnicos, se explorarán las dimensiones éticas y sociales asociadas con su implementación, proporcionando un análisis crítico de sus posibles impactos.

La relevancia de esta investigación se evidencia en la necesidad urgente de comprender y abordar los desafíos éticos vinculados con la IAG. En un contexto donde las líneas que separan lo humano de lo artificial se desdibujan, resulta imperativo establecer un marco ético sólido que oriente el desarrollo y la aplicación de estas tecnologías emergentes. Este TFG se configura como una referencia, iluminando las complejidades éticas de la IAG y proponiendo un enfoque global para afrontarlas.

El trabajo se divide en cuatro partes. En la primera se examina el desarrollo histórico de la Inteligencia Artificial, trazando la trayectoria que condujo a la IAG. A continuación, se analizan en detalle las características y el impacto social de la IAG, resaltando las cuestiones éticas que plantea y justificando la necesidad de regularla. La cuarta parte presenta una propuesta de un nuevo marco ético, que se presenta como una guía para afrontar los desafíos éticos presentes y futuros en la IAG. Finalmente, las conclusiones sintetizan las contribuciones clave y proyectan una perspectiva de futuro para la integración armoniosa entre la IAG y la ética en la sociedad.

Este trabajo no solo busca descifrar los dilemas éticos de la IAG, sino también proporcionar claridad sobre el camino hacia un uso responsable y beneficioso de esta tecnología en constante evolución.

2 OBJETIVOS

Este trabajo tiene como objetivo principal profundizar en el conocimiento de la Inteligencia Artificial Generativa (IAG), centrándose en el caso disruptivo de ChatGPT, para analizar sus implicaciones éticas y proponer un marco ético integral para guiar el desarrollo y aplicación de esta tecnología.

En este contexto, los objetivos específicos son:

- Valorar el desarrollo histórico de la Inteligencia Artificial y su evolución hacia la Inteligencia Artificial Generativa.
- Analizar las características y el impacto social y ético de la Inteligencia Artificial Generativa.
- Proponer un marco ético que sintetice principios de diversas teorías éticas para abordar los desafíos que plantea la Inteligencia Artificial Generativa.

Estos objetivos reflejan las tres grandes secciones del trabajo, orientando la investigación hacia la comprensión profunda de la Inteligencia Artificial Generativa y la formulación de un marco ético que aborde sus complejidades éticas.

3 METODOLOGÍA

La presente investigación se desarrolla a través de un enfoque inductivo, que permite construir teorías a partir de observaciones específicas. La metodología empleada combina elementos cualitativos y deductivos para obtener una comprensión profunda y contextualizada de la Inteligencia Artificial Generativa (IAG), centrada en ChatGPT.

El universo de estudio abarca la totalidad de la literatura relevante (especialmente a partir de 2018) sobre la Inteligencia Artificial y la Inteligencia Artificial Generativa, así como las discusiones éticas asociadas. La muestra se compone de investigaciones académicas y científicas, informes de organismos éticos y documentos técnicos que abordan los aspectos sociales y éticos de la IAG, con un enfoque particular en ChatGPT. Las búsquedas bibliográficas se han hecho en WOS y Google Scholar.

El estudio se lleva a cabo mediante un análisis exhaustivo de la literatura disponible y la identificación de patrones y tendencias emergentes en la investigación ética de la IAG. Se examinan detalladamente casos específicos de aplicación de ChatGPT, destacando sus implicaciones éticas y sociales. Además, se realiza una síntesis crítica de las teorías éticas seleccionadas para la propuesta del nuevo marco ético.

La validez de los resultados se asegura mediante el análisis y la revisión crítica de fuentes académicas y documentos éticos reconocidos en el campo. Se busca garantizar la representatividad y relevancia de la información recopilada, respaldando así la robustez y aplicabilidad del nuevo marco ético propuesto, sustentado sobre teorías éticas reputadas y figuras éticas reconocidas en la disciplina. La triangulación de datos provenientes de diversas fuentes contribuirá a la fiabilidad y validez del análisis y las conclusiones extraídas.

4 ESTADO DE LA CUESTIÓN

En el contexto del creciente avance de la inteligencia artificial (IA) y la irrupción de ChatGPT, resulta imprescindible profundizar en los fundamentos teóricos que sustentan este campo de estudio para poder comprenderlo. En este apartado, se explorarán los conceptos clave de la inteligencia artificial, su desarrollo histórico hasta la actualidad y su relación con la ética.

La inteligencia artificial se ha convertido en un campo de estudio multidisciplinario que busca desarrollar sistemas y máquinas capaces de realizar tareas que requerirían la intervención humana y de demostrar un nivel de inteligencia similar al humano. Su crecimiento exponencial en los últimos años ha dado lugar a numerosas aplicaciones y avances en áreas como el procesamiento del lenguaje natural, el reconocimiento de patrones y la toma de decisiones autónomas (Ertel, 2018).

Para comprender adecuadamente la inteligencia artificial, así como su potencial, es necesario examinar su evolución histórica y los hitos clave que han marcado su desarrollo. Desde los primeros intentos de diseñar sistemas capaces de imitar la inteligencia humana hasta los enfoques más contemporáneos basados en el aprendizaje automático y las redes neuronales, la IA ha experimentado transformaciones significativas y en el futuro próximo se espera que avance mucho más (Purdy & Daugherty, 2016). Este análisis histórico permitirá comprender mejor los fundamentos teóricos subyacentes y apreciar cómo se ha pasado de los sistemas basados en reglas e instrucciones a los enfoques más flexibles y adaptativos.

Asimismo, resulta fundamental explorar la relación entre la inteligencia artificial y la ética. La inteligencia artificial plantea una serie de desafíos éticos y morales, ya que sus aplicaciones pueden tener implicaciones profundas en la sociedad y en la toma de decisiones automatizada (Coeckelbergh, 2020). Es necesario abordar preguntas como: ¿cuáles son los límites éticos en el desarrollo y la implementación de sistemas de IA? ¿Cómo se pueden mitigar los sesgos y garantizar la transparencia y la imparcialidad en las decisiones tomadas por sistemas de IA?

En este sentido, la ética y la moralidad juegan un papel crucial en la forma en que se desarrolla, implementa y regula la inteligencia artificial. Se plantean preguntas éticas complejas relacionadas con la privacidad, la seguridad, la responsabilidad y la equidad. La comprensión de los fundamentos

teóricos de la ética y la moralidad permitirá explorar en profundidad estos desafíos éticos y buscar soluciones que promuevan un desarrollo responsable de la IA.

De este modo, este apartado tiene como objetivo establecer los fundamentos teóricos necesarios para comprender mejor qué son las nuevas herramientas de inteligencia artificial generativa como ChatGPT que están surgiendo y popularizándose rápidamente en la actualidad, para así reflexionar más adelante en el trabajo sobre el impacto social y ético de esta tecnología.

4.1 ¿Qué es la Inteligencia Artificial?

El concepto de inteligencia artificial (IA) es fundamental para comprender la intersección entre la ética y la IA, y su evolución a lo largo del tiempo ha sido objeto de estudio por parte de expertos en el campo. No existe una única definición de inteligencia artificial que sea ampliamente aceptada por los expertos (Wang, 2019), el significado de IA ha sido objeto de debate y estudio por parte de expertos debido a su complejidad y su naturaleza polifacética. Por ello, a continuación se explorarán las definiciones más destacadas y las diferentes perspectivas teóricas que han evolucionado a lo largo del tiempo para poder entender mejor el campo de estudio en el que se centrará este trabajo.

El primer contribuyente reconocido al desarrollo del concepto de máquinas inteligentes fue el matemático-filósofo Alan Turing, considerado por muchos investigadores y expertos como el padre de la inteligencia artificial y la computación moderna (Boden, 2017). En su artículo seminal "Computing Machinery and Intelligence" publicado en 1950 planteó la posibilidad de crear máquinas capaces de exhibir comportamientos inteligentes, como por ejemplo pensar (Turing, 1950). Principalmente, Turing aportó dos trabajos seminales: la máquina de Turing (1936) y el Test de Turing (1950). Turing propuso este test como un criterio para evaluar de forma ordenada si una máquina puede exhibir un comportamiento indistinguible del de un ser humano y, por consiguiente, ser considerada como inteligente (González, 2007).

El Test de Turing se basa en una situación en la que un juez humano interactúa con una máquina y otro ser humano. La interacción se lleva a cabo mediante una serie de preguntas y respuestas escritas. Si, basándose únicamente en las respuestas recibidas, el juez no puede determinar de manera

consistente cuál de las dos entidades es la máquina, se considera que la máquina ha pasado el Test de Turing y se le atribuye una forma de inteligencia.

Esta definición de la IA proporcionada por el Test de Turing no se centra en los mecanismos subyacentes de la máquina ni en la similitud con la inteligencia humana en términos de procesamiento de información, sino más bien en el resultado observable de su comportamiento. Es por ello por lo que su propuesta ha generado un gran debate en la comunidad científica, algunos argumentan que el Test de Turing establece un criterio válido para determinar la existencia de inteligencia artificial, entre los que destacan J. Moor (1976), D. Hofstadter (1979), D. Dennet (1985), and M. Ginsberg (1993). Mientras que otros cuestionan su validez como medida adecuada de inteligencia, como por ejemplo K. Gunderson (1964), J. Searle (1980), N. Block (1981), R. French (1990), and P. Hayes & K. Ford (1995).

A pesar de las críticas y limitaciones del Test de Turing, su influencia perdura en el campo de la inteligencia artificial. El Test de Turing ha servido como punto de partida para la investigación y el desarrollo de sistemas conversacionales y chatbots (como ‘ChatGPT’ o ‘Bard’, los cuales se abordarán más adelante en este trabajo) que intentan simular la inteligencia humana. Además, ha estimulado la reflexión sobre la naturaleza de la inteligencia y los desafíos inherentes en la creación de máquinas inteligentes.

Uno de los autores más influyentes en el campo de la inteligencia artificial es John McCarthy, que acuñó el término "inteligencia artificial" y organizó la Conferencia de Dartmouth en 1956, considerada el punto de partida de la investigación en este campo (Ertel, 2017). Dicha conferencia reunió a un grupo de destacados investigadores con el objetivo de explorar el potencial de la inteligencia artificial y establecer las bases para su investigación y desarrollo.

John McCarthy junto con Marvin Minsky y Claude Shannon definieron la inteligencia artificial como "la ciencia e ingeniería de hacer máquinas que actúen como si fueran inteligentes" (McCarthy et al., 2006). Esta definición ha sido criticada por diversas razones como por ejemplo por su ambigüedad y falta de claridad, su enfoque en la apariencia de inteligencia, la limitación a comportamientos específicos que no abarcan la totalidad de la inteligencia humana, y por la exclusión de otros enfoques de la IA que no se centran exclusivamente en la creación de máquinas

que imiten comportamientos humanos como la IA basada en conexión y aprendizaje automático (Brooks, 1991; Brachman, 2006; Poole & Mackworth, 2010; Bhatnagar et al., 2018). A pesar de ser una definición muy abierta, algunas interpretaciones modernas resultan no ser del todo precisas por intentar ser demasiado precisas y detalladas. De todas formas, John McCarthy desempeñó un papel crucial en la definición del concepto de inteligencia artificial y sentó las bases para la investigación y el desarrollo de la IA.

Otros autores destacados son Stuart J. Russell y Peter Norvig (2010) que en su libro "Inteligencia Artificial: Un enfoque moderno", definen la inteligencia artificial como el estudio de agentes inteligentes, donde los agentes son sistemas que perciben su entorno y actúan para alcanzar objetivos. Russell y Norvig diferencian varios tipos de inteligencia artificial: los sistemas que piensan como humanos, los sistemas que actúan como humanos, los sistemas que piensan racionalmente, y los sistemas que actúan racionalmente.

Actualmente, una de las definiciones más relevantes es la proporcionada por la Comisión Europea que señalaba en 2018 que “El término «inteligencia artificial» (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción -con cierto grado de autonomía- con el fin de alcanzar objetivos específicos.” En cambio, en 2020 el Parlamento Europeo definía la IA como “la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.” Y explica que la IA permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. La máquina recibe datos, los procesa y responde a ellos. Además, los sistemas de IA son capaces de adaptar su comportamiento en cierta medida, analizar los efectos de acciones previas y de trabajar de manera autónoma.

Estas definiciones, junto a otras, han sido fundamentales en la construcción del concepto de inteligencia artificial. A través de estos aportes, se ha logrado una comprensión más amplia y profunda de la inteligencia artificial. A medida que el campo evoluciona, continúan surgiendo nuevos enfoques y perspectivas sobre lo que constituye la inteligencia artificial y cómo debe ser definida. Como se argumentará en este trabajo, hoy en día parece necesario que una definición

completa de inteligencia artificial incorpore la ética con el fin de poder sentar base para abordar los dilemas éticos y sociales que suscita el uso de esta tecnología.

4.2 Desarrollo histórico de la Inteligencia Artificial

La inteligencia artificial (IA) ha experimentado un notable desarrollo a lo largo de la historia, desde sus inicios teóricos hasta su aplicación práctica en diversos campos. Comprender el desarrollo de la IA es fundamental para establecer un marco ético que abarque todas sus implicaciones. En este apartado, se ofrecerá un repaso histórico a los principales hitos en la evolución de la inteligencia artificial.

Los inicios de la inteligencia artificial, según muchos investigadores y expertos, se remontan a 1936 con la máquina de Turing, considerada una de las contribuciones más fundamentales en el campo de la computación y que ha tenido una gran influencia en el desarrollo de la inteligencia artificial (Ertel, 2017). La máquina de Turing es un concepto propuesto por el matemático y lógico británico Alan Turing que consiste en una cinta infinita dividida en casillas, una cabeza de lectura/escritura que puede moverse a lo largo de la cinta, y un conjunto de reglas que indican cómo la cabeza debe leer y escribir en las casillas. La máquina de Turing es capaz de simular cualquier algoritmo de cómputo, lo que la convierte en un modelo universal de computación. Fundamentalmente, sentó las bases para el concepto de computación universal y estableció la idea de que cualquier problema resoluble mediante un algoritmo puede ser abordado por una máquina de Turing

Sin embargo, la reconocida como primera evidencia real de inteligencia artificial acontece en 1943, cuando los científicos Warren McCulloch y Walter Pitts presentaron el primer modelo matemático para la creación de una red neuronal en su artículo «A Logical Calculus of Ideas Immanent in Nervous Activity» (McCulloch & Pitts, 1943). En este artículo presentaron una teoría formal que describía cómo las neuronas en el cerebro podrían llevar a cabo cálculos lógicos simples. Este modelo matemático se conoce como "cálculo de McCulloch-Pitts" y sentó las bases para el desarrollo de las redes neuronales artificiales.

El cálculo de McCulloch-Pitts propuso que las neuronas se pueden representar como unidades de procesamiento que toman varias entradas, aplican una función de activación y generan una salida.

Estas unidades se interconectan entre sí formando una red, donde la salida de una neurona puede convertirse en la entrada de otras neuronas. Esto permitió a McCulloch y Pitts demostrar que las redes neuronales artificiales podían realizar operaciones lógicas, como el AND, OR y NOT, utilizando conexiones y funciones de activación adecuadas.

Este trabajo proporcionó una base teórica para el estudio de las redes neuronales y sentó las bases para siguientes avances en el campo de la inteligencia artificial. Su enfoque de modelar las funciones cognitivas humanas mediante sistemas computacionales allanó el camino para el desarrollo de técnicas más avanzadas de aprendizaje automático y redes neuronales profundas. Prueba de ello fue la creación del primer ordenador de red neuronal en 1950 bautizado como "Stochastic Neural Analog Reinforcement Computer" (SNARC) por Marvin Minsky y Dean Edmonds (Boden, 2017).

Ese mismo año, en 1950, tuvo lugar uno de los momentos más clave en la evolución de la inteligencia artificial Alan Turing se planteó la pregunta: "¿pueden las máquinas pensar?", y propuso el Test de Turing que permite determinar si una máquina exhibe un comportamiento inteligente (Turing, 1950).

Otro de los hitos más importantes en el inicio de la IA fue la conferencia en la Universidad Dartmouth organizada por John McCarthy (Moor, 2006) donde se acuñaba por primera vez el término "inteligencia artificial". Esta conferencia supuso un momento fundacional para el campo de la IA y se formalizó como un nuevo campo de estudio científico. Una de las ideas más importantes introducidas por los asistentes de dicha conferencia, y profundamente arraigada hasta el día de hoy en el estudio de la IA, es que el pensamiento es una forma de computación no exclusiva de los seres humanos o seres biológicos. Más aún, existe la hipótesis de que la inteligencia humana es posible de replicar o simular en máquinas digitales (Velasco, 2002).

Ese mismo año Allen Newell y Herbert Simon (1956) publicaron *Logic Theorist*, considerado el primer programa informático de inteligencia artificial. El objetivo principal del programa era demostrar teoremas matemáticos utilizando el razonamiento deductivo. *Logic Theorist* utilizaba un enfoque conocido como "programación heurística", que se basaba en la formulación de reglas lógicas y la aplicación de algoritmos de búsqueda para encontrar soluciones a problemas complejos.

Fue el primer programa capaz de demostrar teoremas matemáticos de manera automática. Utilizaba un conjunto de axiomas y reglas lógicas para generar pasos de razonamiento, buscando sistemáticamente pruebas para los teoremas dados. Esto mostró que las máquinas podían realizar tareas cognitivas complejas, como el razonamiento lógico, de manera más rápida y precisa que los seres humanos.

El éxito de *Logic Theorist* tuvo un impacto significativo en el campo de la inteligencia artificial. Demostró la capacidad de las máquinas para automatizar procesos de pensamiento humano como el razonamiento lógico y demostrar teoremas matemáticos. Además, sentó las bases para el desarrollo de sistemas expertos, e influyó en el desarrollo de la programación lógica y el razonamiento automatizado (Abeliuk & Gutiérrez, 2021).

Posteriormente, en 1959 el psicólogo y científico de la computación Frank Rosenblatt desarrolló un algoritmo de aprendizaje conocido como el “perceptrón”. El perceptrón fue pionero en el campo del aprendizaje automático y sentó las bases para el desarrollo de redes neuronales más avanzadas (Minsky & Papert, 2017) como las redes neuronales profundas o “deep learning”, del cual se hablará más adelante.

Rosenblatt (1958) se inspiró en el funcionamiento del cerebro humano y propuso el perceptrón como una forma de simular el proceso de aprendizaje y toma de decisiones. El perceptrón es un modelo de red neuronal artificial que se basa en un conjunto de nodos llamados "neuronas" o "perceptrones" que están interconectados mediante conexiones ponderadas. Cada perceptrón toma una serie de entradas, las multiplica por sus respectivos pesos y produce una salida a través de una función de activación. Para ello, utiliza un algoritmo de aprendizaje supervisado, donde se le proporcionan ejemplos de entrada junto con las salidas deseadas, y el perceptrón ajusta sus pesos para minimizar el error entre las salidas reales y las salidas deseadas.

La importancia del perceptrón radica en su capacidad para aprender a partir de ejemplos y reconocer patrones en los datos de entrada. Aunque el perceptrón tiene limitaciones en términos de su capacidad para resolver problemas más complejos y no lineales, su impacto en la evolución de la inteligencia artificial es innegable. El perceptrón fue utilizado en diversas aplicaciones de inteligencia artificial, especialmente en el reconocimiento de patrones y la clasificación de datos.

En la década de 1960 y principios de la década de 1970, se produjo un enfoque en el desarrollo de sistemas expertos. Estos sistemas eran programas informáticos diseñados para simular la capacidad de expertos humanos en dominios específicos. Utilizando reglas y bases de conocimiento, los sistemas expertos podían tomar decisiones y resolver problemas complejos en campos como la medicina, la ingeniería y la gestión.

Durante esta época destaca la creación de 'ELIZA' en 1964, el precursor de los *chatbots* creado por Joseph Weizenbaum en el Instituto de Tecnología de Massachusetts (MIT). Fue uno de los primeros programas de procesamiento del lenguaje natural y fue diseñada como un "terapeuta" conversacional (Weizenbaum, 1966). Utilizaba técnicas simples de procesamiento del lenguaje para emular una conversación en inglés entre un paciente y un terapeuta. El objetivo principal de Eliza no era simular una verdadera comprensión o inteligencia, sino más bien mostrar cómo las respuestas aparentemente inteligentes podían generarse mediante la manipulación de patrones y reglas lingüísticas.

Eliza fue uno de los primeros programas en demostrar el potencial de la interacción humano-computadora a través del lenguaje natural. Aunque Eliza era relativamente simple en su funcionamiento, logró generar respuestas que a menudo parecían comprensivas y reflexivas para los usuarios. Esto llevó a que muchos de ellos creyeran que estaban interactuando con un terapeuta real, pero sin llegar a pasar el test de Turing ya que todavía era fácil distinguir que era un programa artificial.

Eliza generó un gran interés público y despertó debates sobre la capacidad de las máquinas para comunicarse de manera significativa. Mostró cómo una interacción conversacional, aunque basada en patrones y reglas predefinidas, podía brindar una apariencia de comprensión y empatía. Además, facilitó futuros avances en el procesamiento del lenguaje natural y la creación de sistemas conversacionales más sofisticados como los chatbots recientemente lanzados ChatGPT y Bard.

Por otro lado, aproximadamente entre 1966 y 1972 Charles Rosen, Nils Nilsson, Peter Hart y otros crearon el primer robot con inteligencia artificial integrada. "Shakey" fue el primer robot móvil capaz de percibir su entorno, tomar decisiones, navegar de forma autónoma y comunicarse en lenguaje natural (Nilsson, 1984). Es decir, era capaz de razonar sobre sus propias acciones sin

necesidad de tener que darle instrucciones. Este robot ha influido considerablemente en la robótica moderna y su diseño ha inspirado proyectos como los *rovers* de Marte.

Tras varios años en los que no hubo muchos avances en el campo de la IA debido a una falta de financiación (conocidos como los "inviernos de la IA" (Haenlein & Kaplan, 2019)), a partir de la década de 1980, se produjo la revolución del aprendizaje automático o *machine learning* en inglés. Éste fue un hito importante que transformó la forma en que se abordaban los problemas en el campo de la IA y sentó las bases para muchos avances posteriores que se extienden hasta la actualidad, donde sigue siendo un área de investigación activa y en constante evolución.

El *machine learning*, se refiere a la capacidad de las máquinas para aprender y mejorar su rendimiento a través de la experiencia y los datos (Zhou, 2021). En lugar de programar explícitamente reglas y algoritmos para cada tarea, permite que los sistemas adquieran conocimientos y habilidades mediante el análisis de datos y la identificación de patrones.

Esta revolución fue impulsada por el avance de las redes neuronales artificiales, el desarrollo de algoritmos de aprendizaje supervisado y no supervisado, avances en el procesamiento del lenguaje natural, el aumento en la disponibilidad de grandes conjuntos de datos, y el incremento de la capacidad computacional (Jordan & Mitchell, 2015).

Este rápido crecimiento de la IA se evidenció en 1997 cuando la supercomputadora "DeepBlue", creada por IBM, venció al campeón mundial de ajedrez Garry Kasparov en un partido a seis partidas. En aquel momento, DeepBlue era capaz de calcular más de 200 millones de posibles movimientos por segundo (Campbell et al., 2002). Algunos historiadores de la IA consideran este año como el punto de inflexión donde la IA se popularizó fuera de los ámbitos académicos y de investigación.

A continuación, en 2002 la compañía iRobot lanzó la *Roomba*, que fue el primer robot de éxito comercial para el hogar que utiliza inteligencia artificial (Jones, 2006). La función principal de este robot doméstico era limpiar el suelo de forma autónoma gracias a un conjunto de sensores y algoritmos de navegación que le permitían mapear el entorno y moverse de manera inteligente para evitar obstáculos. Este suceso supuso la primera introducción masiva de la IA en el ámbito del hogar.

Desde entonces, la IA siguió avanzando y penetrando en muchos sectores de la sociedad. En 2011 aconteció otro de los hitos más trascendentales, la IA 'Watson' desarrollada por IBM ganó a los campeones del concurso televisivo de Estados Unidos '*Jeopardy!'*, logrando el premio de un millón de dólares (Markoff, 2011). Este logro demostró que la inteligencia artificial había avanzado lo suficiente como para competir y superar a los mejores concursantes humanos en un juego de preguntas y respuestas de conocimiento general.

Además, el éxito de Watson en Jeopardy! tuvo implicaciones más allá del ámbito de los juegos y el entretenimiento. Demostró el potencial de la IA para aplicaciones en la vida real, como la asistencia en la toma de decisiones en campos complejos como la medicina, la investigación científica y la resolución de problemas empresariales. Watson abrió nuevas posibilidades para el uso de la IA en diversas industrias y campos de estudio, y promovió la colaboración entre la comunidad científica y la industria tecnológica (Ferrucci, 2012).

Mientras que su antecesor Deep Blue era un sistema altamente especializado en ajedrez y no estaba diseñado para interactuar en lenguaje natural, Watson se basaba en técnicas de procesamiento del lenguaje natural, aprendizaje automático y análisis de grandes conjuntos de datos para encontrar la respuesta más probable en tiempo real.

Ese mismo año, la empresa tecnológica Apple introdujo 'Siri', el primer asistente virtual con reconocimiento de voz e interacción con lenguaje natural en un *smartphone* (Hoy, 2018). Permitía realizar con el móvil diversas tareas a través de comandos de voz, como enviar mensajes, hacer llamadas, establecer recordatorios, buscar información en Internet, controlar dispositivos domésticos inteligentes y muchas más. Siri fomentó el avance de la IA en el campo de la interacción humano-computadora y provocó que, más adelante, otras grandes empresas tecnológicas también lanzaron sus propios asistentes virtuales como *Google Now* de Google en 2012, *Alexa* de Amazon en 2014, y *Cortana* de Microsoft ese mismo año también.

Hasta aquel entonces, ninguna IA había conseguido pasar el Test de Turing y fue en 2014 cuando un programa de inteligencia artificial diseñado para simular un niño ucraniano de 13 años en una conversación escrita superó por primera vez el Test de Turing. Se llamaba Eugene Goostman, fue creado por Vladimir Veselov y Eugene Demchenko, y consiguió convencer de que era un niño real

a un tercio de los jueces durante una competencia organizada por la Royal Society de Londres (Nieves, 2014).

Eugene Goostman consiguió convencer a 10 de los 30 jueces, respondiendo a preguntas de forma natural e incluso con sentido del humor, superando por tanto el 30% mínimo del test de Turing y por tanto haciéndose merecedor del término de inteligencia artificial. No obstante, muchos negaron el valor de esta prueba argumentando que una máquina consiguiera engañar a un humano sólo probaba que esa máquina era capaz de imitar la inteligencia, y no que en realidad la posea.

El año 2016 fue clave en la evolución de la inteligencia artificial en el sector de los juegos. En primer lugar, la IA de la compañía DeepMind de Google, '*AlphaGo'*, venció al campeón mundial Lee Sedol en el juego de mesa oriental llamado 'Go'. Según explica Chen (2016) en su estudio, se trata de un juego de estrategia muy complejo y considerado un gran desafío para las máquinas debido a la enorme cantidad de posibles movimientos y la necesidad de comprender patrones abstractos para los que se creían necesarias la intuición, creatividad y experiencia humana.

AlphaGo utilizó técnicas de aprendizaje profundo (o '*deep learning*' en inglés) y redes neuronales y su capacidad para derrotar al campeón mundial demostró que las máquinas pueden superar a los humanos en tareas cognitivas complejas que antes se consideraban exclusivas de nuestra inteligencia.

Un año más tarde, una versión más avanzada de AlphaGo llamada '*AlphaZero'* a la IA más potente de ajedrez de aquel momento, *Stockfish*, con solo 4 horas de entrenamiento consigo misma (Silver et al., 2017). Este logro demostró el potencial del aprendizaje automático y el enfoque basado en redes neuronales frente a los sistemas tradicionales basados en conocimientos y heurísticas. El estilo de juego de AlphaZero mostró un enfoque más creativo y humano, influyendo en la forma en que los jugadores de ajedrez profesionales abordan el juego desde entonces.

Sin embargo, los hitos recientes más significativos de la inteligencia artificial no se limitan únicamente al ámbito de los juegos. En 2020 sucedió un acontecimiento que ha sido calificado por Forbes, entre otros, como el avance más importante en la historia de la IA (Toews, 2021). El equipo

de DeepMind consiguió resolver con alta precisión la estructura tridimensional de virtualmente cualquier proteína gracias a su IA: '*AlphaFold'*.

Predecir con precisión la estructura 3D de una proteína a partir de su secuencia de aminoácidos es uno de los desafíos más difíciles y relevantes en el que muchos investigadores biomédicos llevaban trabajando durante casi 50 años. Antes de AlphaFold, se usaban complejos y costosos procesos de laboratorio y este software combinó técnicas de aprendizaje automático con información genómica para resolver el "problema del plegamiento de proteínas" (Jumper et al., 2021). Este logro supuso una revolución en la biología computacional con inmensas repercusiones en investigación y biomedicina. Además, se trata de la primera vez en la historia que la inteligencia artificial ha contribuido directamente a avanzar las fronteras del conocimiento científico.

Finalmente, en noviembre de 2022 la empresa OpenAI lanzó al público de forma gratuita su chatbot 'ChatGPT' que ha revolucionado el paradigma de la IA en todo el mundo. De acuerdo con el estudio del banco de inversión multinacional UBS, ChatGPT alcanzó los 100 millones de usuarios en los dos primeros meses tras su lanzamiento, convirtiéndose así en la plataforma con mayor crecimiento de usuarios de la historia (Hu, 2023).

4.3 ChatGPT y la Inteligencia Artificial Generativa

ChatGPT es un modelo de lenguaje generativo diseñado para generar texto de calidad humana, es decir, es capaz de mantener conversaciones en cualquier idioma con los usuarios imitando de forma convincente a un ser humano. Utiliza algoritmos de aprendizaje profundo para comprender los textos y aprender en base a ellos (OpenAI, 2022). Gracias a su capacidad para aprender y generar todo tipo de respuestas a partir de grandes cantidades de datos de texto, se puede usar por ejemplo para traducir textos, hacer resúmenes, pedir que te explique un concepto u evento histórico, programación, crear guiones ficticios para películas, escribir un poema o la letra de una canción, hacer chistes, generar un artículo, pedir recetas de cocina, analizar problemas, generar contenidos de marketing, y mucho más (Metz, 2022).

Sin embargo, ChatGPT no es perfecto, tiene sus limitaciones y comete errores. Esta tecnología sigue en constante desarrollo y aún se desconoce su pleno potencial, lo cual plantea muchos riesgos. Lo

que está claro es que ya está revolucionando muchos sectores de la sociedad como la educación, la sanidad, el sector financiero, el servicio de atención al cliente, el marketing, los buscadores web, y el campo laboral de muchas empresas (Wu et al., 2023). Tras el gran éxito de ChatGPT otras empresas han lanzado sus propios chatbots similares como por ejemplo Google y su chatbot *Bard*, o Microsoft y su chatbot *Bing Chat*.

Por otro lado, actualmente en 2023 están surgiendo múltiples novedosas IAs como *Midjourney* o *Dall-e* que son capaces de generar imágenes de calidad humana, o también otras IAs capaces de generar canciones e imitar voces de personas reales. Todo ello señala que el mundo se encuentra en una nueva era marcada por el avance de la inteligencia artificial y su gran mejora en cuanto a la capacidad de imitar la inteligencia humana.

En consecuencia, la elección de enfocar exclusivamente este trabajo en la Inteligencia Artificial Generativa (IAG), particularmente ejemplificada por ChatGPT, se sustenta en su carácter novedoso y contemporáneo, su alcance global masivo y su potencial revolucionario para transformar el paradigma tecnológico mundial. La IAG representa una frontera emergente en la tecnología, con relevancia actual y una carencia de investigaciones académicas que aborden este dominio. La accesibilidad global gratuita y su facilidad de uso la convierten en una fuerza omnipresente, mientras que la imposibilidad de abordar en su totalidad las vastas implicaciones éticas y sociales de la totalidad del campo de la IA justifica un enfoque selectivo para una exploración más profunda y detallada en las secciones subsiguientes de este trabajo.

El fin último de este proyecto será establecer un marco ético regulatorio adecuado que considere tanto los beneficios como los posibles riesgos de la IAG. Por ende, primero se analizarán aquellos desafíos éticos y sociales que se consideren más relevantes para este trabajo.

5 IMPACTO SOCIAL Y ÉTICO DE LA INTELIGENCIA ARTIFICIAL GENERATIVA

Como se ha visto, la inteligencia artificial ha experimentado un crecimiento exponencial en las últimas décadas, desempeñando un papel cada vez más importante en diversos ámbitos de nuestra sociedad. Entre las múltiples aplicaciones actuales de la IA, los sistemas de procesamiento del lenguaje natural han destacado significativamente, permitiendo avances importantes en la generación de texto automatizado. Este tipo de IA se conoce como inteligencia artificial generativa o IAG. Uno de los ejemplos más prominentes es ChatGPT, el modelo de lenguaje desarrollado por OpenAI que está revolucionando la forma en que interactuamos con las máquinas a través de conversaciones.

El presente trabajo se centra en explorar el impacto social y los dilemas éticos que plantean las IAG como ChatGPT y, a partir de ello, proponer un marco ético regulatorio. La elección de ChatGPT como objeto de estudio se justifica por su relevancia actual en el campo de la IAG y la novedad del tema. Además, su rápida popularización por todo el mundo y su acelerado avance exigen una pronta atención y regulación ética ante la irrupción sin suficiente supervisión de estas inteligencias artificiales.

ChatGPT ha capturado la atención tanto de la comunidad científica como del público en general debido a su capacidad para generar texto coherente y relevante en respuesta a las consultas y preguntas de los usuarios. Su funcionamiento se basa en un modelo de lenguaje pre-entrenado en grandes cantidades de datos textuales obtenidos de internet, lo que le permite generar respuestas que son difícilmente distinguibles de las humanas, como afirma Holly Else (2023). Esta capacidad ha llevado a ChatGPT a ser utilizado en diversos contextos, desde asistentes virtuales hasta desarrollo de software y generación de contenido (Haleem et al., 2022).

Sin embargo, el impacto social de ChatGPT no se puede ignorar. A medida que la tecnología se vuelve más ubicua en nuestra vida diaria, es fundamental examinar cómo afecta a nuestra sociedad y qué implicaciones éticas surgen de su uso generalizado. El uso de IAG en conversaciones plantea desafíos éticos únicos debido a su capacidad para influir en las percepciones y decisiones de las

personas. Estos desafíos abarcan desde cuestiones de privacidad y seguridad hasta la potencial manipulación de la opinión pública.

Un aspecto crucial en el análisis del impacto social y los dilemas éticos de ChatGPT es su capacidad para difundir información veraz y relevante. Aunque ChatGPT puede generar respuestas convincentes, no está exento de cometer errores o de promover información sesgada o falsa. La falta de regulaciones claras en este sentido puede dar lugar a la propagación de desinformación y la erosión de la confianza en las interacciones con ChatGPT y otros sistemas similares.

Además, el uso de ChatGPT plantea dilemas éticos en términos de responsabilidad, transparencia y autenticidad. ¿Quién debe asumir la responsabilidad cuando ChatGPT proporciona información incorrecta o realiza recomendaciones perjudiciales?, ¿cómo se pueden identificar y mitigar los sesgos y prejuicios que puedan existir en los datos utilizados para entrenar a estos modelos?, ¿qué datos de los usuarios se recopilan y cómo se utilizan?, ¿cómo se puede saber si un texto ha sido creado por IAG o un ser humano? Estas son algunas de las preguntas que son cruciales para garantizar que la IAG, y específicamente ChatGPT, se utilice de manera ética y responsable.

Dado el impacto potencialmente significativo de ChatGPT en la sociedad y la necesidad de una regulación ética efectiva, se vuelve imperativo desarrollar un marco regulatorio adecuado que aborde estos desafíos. La propuesta de un marco ético regulatorio tiene como objetivo proporcionar pautas claras y responsabilidades definidas para garantizar el uso ético y responsable de ChatGPT y sistemas similares. Este marco debe considerar la protección de la privacidad de los usuarios, la mitigación de sesgos y prejuicios, la transparencia en el funcionamiento de los modelos y la rendición de cuentas de los desarrolladores y proveedores de IAG.

Por ende, esta sección del trabajo se enfoca en el impacto social y los dilemas éticos que la IAG, en particular ChatGPT, plantea en nuestra sociedad. A través de un análisis exhaustivo, se busca comprender y abordar los desafíos sociales y éticos asociados con esta tecnología emergente con el fin de exponer la necesidad de establecer un marco ético regulatorio. Debido a la novedad de este tema, aún se requiere investigación y análisis adicionales. No obstante, ya existen algunos estudios e informes que exponen el impacto de ChatGPT en diversos ámbitos, pudiendo además inferir las consecuencias que puede tener esta tecnología en el futuro próximo.

5.1 Consecuencias sociales

La introducción de ChatGPT ha generado un impacto significativo en nuestra sociedad. A medida que las IAG como ChatGPT se vuelven más accesibles y utilizadas, es necesario considerar cómo interactúan y reconfiguran los diferentes estratos de nuestra sociedad. En este apartado, se explorará el impacto social de las IAG, con el ejemplo de ChatGPT, desde estas perspectivas, abordando cada nivel de interacción y estructura que esta tecnología afecta:

5.1.1 INTERACCIÓN INDIVIDUAL

La comunicación es uno de los aspectos más destacados del impacto social de ChatGPT. Este modelo de lenguaje ha abierto nuevas posibilidades en la forma en que nos comunicamos con las máquinas. Antes de la existencia de ChatGPT, las interacciones con sistemas automatizados solían ser limitadas, basadas en comandos específicos, y era relativamente fácil distinguir los textos generados por IAG. Sin embargo, ChatGPT permite a los usuarios interactuar de manera más natural, expresando preguntas y solicitudes en lenguaje cotidiano, y respondiendo de manera similar a la manera en que un humano lo haría. Esto ha facilitado el acceso a la información y la realización de tareas de manera más intuitiva, mejorando la experiencia del usuario en numerosos contextos.

Sin embargo, esta mejora en la comunicación también plantea problemas. A medida que ChatGPT se vuelve más sofisticado en la generación de respuestas, es fundamental considerar la calidad y veracidad de la información proporcionada. La difusión de desinformación o respuestas erróneas por parte de ChatGPT puede tener consecuencias perjudiciales, especialmente cuando se utiliza en contextos donde la precisión y la fiabilidad son cruciales, como en el ámbito de la salud o el asesoramiento legal.

5.1.2 INTERACCIÓN SOCIAL

Otro aspecto relevante del impacto social de ChatGPT se refiere a su influencia en la interacción humana. A medida que los sistemas de conversación automatizados se vuelven más avanzados, existe la posibilidad de que las personas los utilicen como sustitutos de la comunicación humana

tradicional. Esto plantea interrogantes sobre la calidad de las relaciones interpersonales y la capacidad de empatía y comprensión emocional. Si las interacciones con ChatGPT se vuelven predominantes, podría conllevar una disminución en la calidad de las conexiones humanas, lo que tendría consecuencias negativas en el bienestar social y emocional.

5.1.3 ESTRUCTURA ECONÓMICA

La proliferación de ChatGPT y tecnologías de inteligencia artificial similares plantea preocupaciones sobre el impacto económico y en el mercado laboral. A medida que estos sistemas automatizados se utilizan en una amplia gama de industrias y servicios, existe el riesgo de desplazamiento laboral.

Este potencial desplazamiento laboral plantea desafíos importantes en términos de reestructuración laboral, capacitación y generación de nuevas oportunidades económicas para aquellos afectados por la automatización. Según el Foro Económico Mundial (2023) se espera que aproximadamente el 23% de los trabajos cambien debido a la IA para el año 2027, con la creación de 69 millones de nuevos empleos y la eliminación de 83 millones de empleos. De acuerdo con este informe, surgirán consecuencias negativas a raíz de esto como el menor crecimiento económico, escasez de suministros e inflación que podrían afectar la estabilidad laboral.

Además, enfatiza la necesidad de invertir en el desarrollo de nuevas habilidades, en la educación y en estructuras de apoyo social para garantizar que las personas estén preparadas para el futuro del trabajo. Avisa sobre la urgencia de una revolución en la capacitación y el desarrollo de habilidades, ya que las empresas informan sobre la falta de habilidades y la falta de oportunidades de capacitación. El informe estima que, en promedio, el 44% de las habilidades de los trabajadores individuales deberán ser actualizadas.

La discrepancia entre las habilidades de los trabajadores y las necesidades empresariales requiere la colaboración entre empresas y gobiernos para facilitar iniciativas de aprendizaje y desarrollo de habilidades. Las habilidades que más se valorarán en esta transición por los empleadores son el pensamiento analítico, el pensamiento creativo, la flexibilidad y se anticipa la creciente importancia

de la alfabetización tecnológica, la inteligencia artificial y el manejo de grandes volúmenes de datos (Big Data). El informe insta a una acción colectiva de todos los actores involucrados para abordar las interrupciones en el mercado laboral y garantizar un futuro del trabajo especializado y equitativo.

En la Modernidad, el ser humano era un animal egoísta que siempre buscaba su interés propio y la maximización del beneficio económico. El trabajo daba sentido a una vida centrada en la producción económica con el ser humano en el eje de la misma. El aparente potencial sin límites de la IAG y afirmaciones como las del informe anterior suscitan muchas preguntas acerca de la posición del ser humano en el futuro: ¿Cómo vamos a repensar la vida humana y sus relaciones y valores cuando el trabajo ya no sea el valor central de la existencia? ¿Se debe permitir que la IA sustituya al ser humano en ciertos trabajos o siempre debe estar al servicio del ser humano? De alguna manera estaríamos ante la emergencia de un nuevo paradigma.

5.1.4 ESTRUCTURA EDUCATIVA

Uno de los sectores que se está viendo más afectado por la irrupción de ChatGPT es el educativo (Kasneci et al., 2023). ChatGPT puede aportar muchos beneficios para los estudiantes como la síntesis de ideas o la explicación de conceptos, y también para los profesores como la corrección más rápida de textos, por ejemplo.

No obstante, también presenta muchos problemas para este sector. La información generada sin esfuerzo podría afectar negativamente a las habilidades de los estudiantes relacionadas con el pensamiento crítico y resolución de problemas. Esto se debe a que el modelo simplifica la adquisición de respuestas o información, lo que puede aumentar la pereza y contrarrestar el interés de los estudiantes por realizar sus propias investigaciones y llegar a sus propias conclusiones.

Para enfrentar este riesgo, es importante ser consciente de las limitaciones y los perjuicios de depender en exceso de estos grandes modelos de lenguaje, y utilizarlos solo como una herramienta para apoyar y mejorar el aprendizaje (Pavlik, 2023). Por lo tanto, uno de los mayores desafíos al que la educación se va a enfrentar durante los próximos años será intentar integrar este tipo de

tecnologías de manera que complementen y mejoren la experiencia de aprendizaje, en lugar de reemplazarla.

Sin embargo, muchos educadores e instituciones educativas pueden no tener el conocimiento o la experiencia para integrar de manera efectiva las nuevas tecnologías en su enseñanza (Redecker et al., 2017). Al igual que con cualquier otra innovación tecnológica, integrar grandes modelos de lenguaje en una práctica docente efectiva requiere comprender sus capacidades y limitaciones, así como saber utilizarlos de manera efectiva para complementar o mejorar procesos de aprendizaje específicos. Para lograrlo, sería necesario desarrollar una nueva teoría educativa específica para este nuevo contexto, lo que puede suponer un gran desafío para los centros educativos.

Además, cada vez es más difícil distinguir si un texto fue generado por una máquina o por un humano, lo que presenta un desafío adicional importante para los profesores y educadores (Cotton et al., 2023; Gao et al., 2022). Como resultado, algunos centros educativos como el Departamento de Educación de la Ciudad de Nueva York recientemente han prohibido el uso de ChatGPT en las redes escolares (Noticias, 2023).

El uso de grandes modelos de lenguaje en la educación también plantea desafíos en términos de privacidad y seguridad de los datos de los estudiantes. Es fundamental garantizar la protección de los datos personales y sensibles de los estudiantes y cumplir con las regulaciones y leyes de protección de datos aplicables. Para ello, las instituciones educativas deberán implementar planes de protección. Además, este problema no solo afecta al sector educativo, es una cuestión que afecta a todo el mundo dado que ChatGPT puede recopilar, analizar y utilizar grandes cantidades de datos personales. Esto plantea preguntas sobre la privacidad de los individuos y la forma en que se manejan y protegen sus datos. En la siguiente sección de este trabajo se profundizará mucho más sobre este desafío dadas sus repercusiones éticas y se concretarán ejemplos prácticos como el capitalismo de vigilancia de Shoshana Zuboff publicado en su libro “La era del capitalismo de vigilancia” publicado en 2019 (Zuboff, 2023).

Figura 1. Desafíos sociales de la IAG

En conclusión, el impacto social de ChatGPT y la IAG es innegable y presenta una serie de desafíos que deben abordarse. Desde la calidad de la información proporcionada y la influencia en la comunicación y la interacción humana, hasta las implicaciones económicas y educativas, es fundamental considerar cuidadosamente los efectos de esta tecnología en nuestra sociedad. La mitigación de los riesgos y la implementación de salvaguardias son necesarias para garantizar que el uso de inteligencia artificial generativa sea responsable y beneficioso.

Estos desafíos sociales proporcionan una transición lógica hacia el próximo apartado, donde se explorarán las dimensiones éticas que surgen de la utilización de sistemas como ChatGPT. En esta siguiente sección, se analizarán los desafíos éticos específicos de la Inteligencia Artificial Generativa (IAG), contribuyendo así a una comprensión más profunda de las implicaciones éticas y sociales que surgen en el contexto de esta tecnología en constante evolución.

5.2 Desafíos éticos

La proliferación y adopción generalizada de la IAG, en concreto ChatGPT y otras tecnologías de inteligencia artificial de conversación, plantean una serie de desafíos éticos que deben ser abordados. A medida que estas herramientas se vuelven más sofisticadas y omnipresentes en nuestra sociedad, es fundamental reflexionar sobre los dilemas éticos asociados con su desarrollo, implementación y uso.

La concepción de la dignidad humana en la filosofía kantiana proporciona un fundamento ético desde el cual se pueden examinar y evaluar las IAG como ChatGPT. Para Kant (1946), los seres humanos son intrínsecamente valiosos no por sus contribuciones utilitarias a la sociedad, sino porque tienen la capacidad de la autonomía moral, de razonar y de actuar de acuerdo con principios morales (Kant, 1946). Esta capacidad de actuar moralmente otorga al ser humano de una dignidad intrínseca que le distingue de los demás seres en el universo. Por lo tanto, cualquier acción o sistema que socave esta dignidad o trate a los seres humanos simplemente como medios para un fin, y no como fines en sí mismos, es inherentemente inmoral.

Al considerar las implicaciones éticas de la IAG desde la perspectiva kantiana, se plasma la manera en que estas tecnologías pueden afectar, amplificar o poner en peligro la capacidad inherente de autonomía moral del ser humano y su dignidad. A continuación, se muestran algunos ejemplos de los desafíos éticos que las IAG plantean que vulnerarían el concepto de dignidad humana desarrollado por Kant:

5.2.1 RESPONSABILIDAD Y RENDICIÓN DE CUENTAS

Uno de los principales desafíos éticos es establecer claridad sobre quién es responsable de las acciones y resultados de las IAG como ChatGPT. A medida que el modelo genera respuestas autónomamente, se vuelve necesario definir quién asume la responsabilidad en caso de errores, sesgos o daños causados por las respuestas generadas. Actualmente, no está clara la respuesta a la pregunta de quién es responsable de las posibles consecuencias negativas derivadas de un mal uso de ChatGPT, ¿la responsabilidad recaería sobre los desarrolladores y proveedores de ChatGPT o sobre los usuarios?

Respetar la dignidad humana implica que debe haber claridad en la asignación de responsabilidades. Si no se sabe quién es responsable de las acciones de ChatGPT, se corre el riesgo de deshumanizar a las víctimas de posibles daños, tratándolas como medios y no como fines en sí mismos.

La cuestión de la responsabilidad y la rendición de cuentas en el contexto de la inteligencia artificial generativa representa un problema ético intrincado y multifacético que exige un análisis en

profundidad. A medida que estas tecnologías se integran en diversos sectores, desde la atención médica hasta la justicia y la educación, es imperativo comprender la complejidad de asignar responsabilidades en situaciones donde la toma de decisiones se comparte entre humanos y máquinas.

Un ejemplo de este desafío ético que evidencia este problema es en el ámbito de la toma de decisiones médicas. Si un profesional médico utiliza una IAG para obtener recomendaciones de diagnóstico o tratamiento para un paciente y dichas recomendaciones resultan en un error médico que causa un daño sustancial al paciente, ¿quién debe asumir la responsabilidad por este error?

La respuesta no es tan simple como podría parecer. Los desarrolladores de la IAG pueden ser considerados responsables por no garantizar que su sistema proporcione recomendaciones médicas precisas. Pueden ser señalados por no haber implementado salvaguardias suficientes para prevenir errores y garantizar que el modelo de IAG se haya entrenado de manera adecuada. Además, pueden ser responsables de no haber evaluado adecuadamente el rendimiento de su sistema en un entorno clínico.

Por otro lado, el profesional médico también puede ser considerado responsable, ya que es su deber ejercer un juicio clínico independiente y no depender ciegamente de las recomendaciones generadas por la IAG. En cuyo caso podría ser considerado como negligencia. No obstante, ¿estaría justificada su confianza en el sistema de IAG?, ¿realizaría suficientes verificaciones independientes antes de actuar? Podría darse el caso de que ese profesional médico haya utilizado IAG en decenas de ocasiones similares y en todas ellas haber acertado con sus diagnósticos, habiendo visto mejorado su trabajo gracias al uso de estas herramientas. Sin embargo, si se equivoca una vez por confiar en el sistema de IAG, ¿de quién sería la responsabilidad y cómo debería pagar por ello?

Esta falta de claridad en la asignación de responsabilidades se manifiesta en otros campos también. Además del ejemplo sobre atención médica mencionado anteriormente, vale la pena explorar otras áreas donde la responsabilidad y la rendición de cuentas en la IAG plantean problemas éticos.

Supongamos que un vehículo autónomo, equipado con IAG, se ve involucrado en un accidente automovilístico. Las recomendaciones de la IAG integrada en el sistema de navegación del vehículo

pueden haber influido en la ruta y las decisiones tomadas por el automóvil autónomo. Si el accidente resulta en daños materiales y lesiones a los pasajeros de otro vehículo, ¿quién debería asumir la responsabilidad por el accidente?

Esto evidencia las implicaciones de la falta de claridad en la asignación de responsabilidades en el contexto de los vehículos autónomos. ¿Es el fabricante del automóvil responsable, ya que desarrolló el sistema de IAG integrado en el vehículo y debería garantizar que las decisiones tomadas por la IAG no conduzcan a situaciones peligrosas? ¿Es el propietario del vehículo responsable, ya que debería ejercer el control y la supervisión adecuados sobre el vehículo autónomo? ¿Debería el programador del software de IAG asumir parte de la responsabilidad si se demuestra que el algoritmo tenía un fallo? Son muchas personas las implicadas en el desarrollo y uso de un vehículo de estas características y no es nada sencillo decidir quién sería responsable en caso de que se produzca un error con consecuencias graves.

Un problema similar se presenta en el ámbito de la justicia y la toma de decisiones legales. Si los sistemas de IAG se utilizan para investigar casos legales, como en la revisión de documentos legales extensos o en la evaluación de precedentes legales, resulta esencial establecer responsabilidades. Si una decisión judicial se basa en la interpretación de documentos legales generada por una IAG y resulta ser errónea, ¿Quién es responsable de la decisión equivocada? ¿El abogado que confió en la interpretación de la IAG o el desarrollador del sistema de IAG cuyo modelo interpretó incorrectamente la información?

Un tercer ejemplo notable de este desafío ético se relaciona con el uso de IAG para actividades inmorales. Una de las características más destacables de sistemas de IAG como ChatGPT es que son de uso público, es decir, cualquiera puede usarlos. Esto implica que hay individuos malintencionados que podrían utilizar una IAG para usos perjudiciales como, por ejemplo, generar contenido falso y dañino como noticias falsas, documentos fraudulentos o mensajes diseñados para incitar al odio y la violencia. En cuyo caso, ¿Quién debe asumir la responsabilidad por el daño causado por este uso inmoral de la IAG? ¿Debería el desarrollador de la IAG ser considerado responsable, ya que proporcionó la herramienta que permitió la creación del contenido dañino? ¿Debería el individuo que utilizó la IAG para generar este contenido inmoral ser el único

responsable, o es justo compartir la responsabilidad con el creador de la IAG? Hoy en día que las redes sociales y las estafas digitales están a la orden del día, esto podría afectar negativamente a millones de personas.

Este ejemplo destaca cómo la falta de claridad en la asignación de responsabilidades puede socavar la lucha contra actividades inmorales y perjudiciales. Cuando no hay consecuencias claras para quienes usan la IAG para propósitos inmorales, se corre el riesgo de incentivar comportamientos dañinos. La falta de responsabilidad en situaciones como esta puede facilitar la desinformación, la incitación al odio y la propagación de contenido perjudicial en línea.

La responsabilidad y la rendición de cuentas son fundamentales para abordar la utilización inmoral de la IAG. Si bien es necesario garantizar que los desarrolladores de IAG se esfuercen por prevenir el uso inmoral de sus tecnologías, también es esencial que los individuos que utilizan la IAG con fines dañinos asuman la responsabilidad de sus acciones. La insuficiencia en la asignación de responsabilidades en situaciones de uso inmoral es un desafío que debe abordarse para evitar un uso generalizado de la IAG con fines perjudiciales. Además, cuando los actos inmorales se realizan utilizando la IAG, disminuye la confianza en estas tecnologías. Las preocupaciones sobre la utilización inapropiada pueden llevar a una mayor regulación, lo que podría afectar negativamente a aquellos que utilizan la IAG de manera legítima y ética.

La falta de claridad y la complejidad de asignar responsabilidades en situaciones que involucran a las IAG resaltan un dilema ético crucial. La ética kantiana sostiene que cada ser humano posee una dignidad intrínseca y debe ser tratado como un fin en sí mismo, no simplemente como un medio para un fin. Por lo tanto, cualquier acción o sistema que socave esta dignidad o trate a las personas como meros instrumentos, es inherentemente inmoral.

La falta de rendición de cuentas en la IAG plantea inquietudes éticas significativas. En primer lugar, socava la dignidad de las víctimas de daños causados por las IAG al tratarlas como medios y no como fines en sí mismos. Los individuos afectados son más que simples "daños colaterales" de un sistema no regulado. Sus vidas, salud y bienestar están en juego.

Es fundamental abordar de manera racional y ética la responsabilidad en la IAG. Requiere una reflexión profunda sobre cómo establecer sistemas de responsabilidad claros, cómo evaluar y mitigar riesgos, y cómo garantizar que la dignidad humana se mantenga intacta. Un enfoque proactivo hacia la responsabilidad y la rendición de cuentas en la IAG es esencial para evitar consecuencias éticas y sociales adversas a medida que estas tecnologías continúan su expansión en la vida cotidiana.

5.2.2 SESGOS Y DISCRIMINACIÓN

Otro desafío ético fundamental de las IAG es el relacionado con los sesgos y la discriminación. Estos modelos se entrenan en datos que a menudo reflejan los prejuicios existentes en la sociedad. Esto se traduce en respuestas discriminatorias o sesgadas hacia ciertos grupos de personas, lo que es un claro reflejo de la problemática que rodea esta tecnología.

Por ejemplo, ChatGPT presenta diversos sesgos que se traducen en manifestaciones discriminatorias. Los sesgos se derivan del contenido del conjunto de datos de entrenamiento, predominantemente compuesto por contenido generado por humanos en internet. Partha Pratim Ray, (2023), en su revisión exhaustiva de ChatGPT, expone varios de los sesgos que este modelo de IAG presenta. Entre los sesgos que identifica se incluyen el cultural, lingüístico, género, racial, ideológico y de recomendaciones de contenido.

Las implicaciones éticas de este desafío son sustanciales. En primer lugar, socava el principio de igualdad y justicia, que es fundamental para la dignidad humana. Cuando una IAG genera contenido discriminatorio o prejuicioso, niega a ciertos grupos e individuos su derecho a ser tratados con igualdad y respeto, lo que es una violación directa de la dignidad humana según la perspectiva de Kant.

Además, la discriminación y los sesgos pueden tener consecuencias perjudiciales en la sociedad. Pueden exacerbar las tensiones sociales, alimentar el odio y la intolerancia, y favorecer la desigualdad. Por ejemplo, cuando una IAG genera respuestas basadas en estereotipos de género,

contribuye a la discriminación de género en la sociedad. Esto puede limitar las oportunidades de las personas y reforzar estructuras de poder desiguales.

Estos sesgos también pueden tener un impacto directo en la vida de las personas. En el ámbito laboral, por ejemplo, si un sistema de IAG utilizado en la selección de personal muestra sesgos de género o racial, puede excluir injustamente a ciertos grupos de candidatos y perpetuar la discriminación en el empleo. Esto puede afectar la igualdad de oportunidades y vulnerar la dignidad de los individuos.

Por otro lado, la adopción de modelos concretos de IAG como ChatGPT puede tener un impacto desigual en diferentes grupos y comunidades. Existe el riesgo de que aquellos con menos acceso a la tecnología o con menos habilidades digitales se queden rezagados y experimenten una mayor brecha digital. Además, como se ha visto en el análisis anterior sobre el impacto social de las IAG, la automatización impulsada por estas IAG puede resultar en desplazamiento laboral y agravar la desigualdad económica. Es fundamental considerar y abordar estos impactos sociales para minimizar las disparidades o discriminaciones que puedan surgir a raíz de estas tecnologías.

Es esencial abordar activamente este desafío ético y garantizar que los modelos de IAG no refuercen ni amplifiquen los sesgos sociales existentes. La detección y mitigación de sesgos y prejuicios deben ser prioridades en el desarrollo y la implementación de estas tecnologías. Esto no solo implica desarrollar algoritmos y modelos que reduzcan los sesgos, sino también establecer marcos regulatorios sólidos que impongan requisitos de equidad y transparencia.

Uno de los desafíos clave en este sentido es cómo definir y medir el sesgo en la IAG. A menudo, las decisiones sobre lo que se considera sesgo y cómo se debe abordar pueden ser subjetivas, por lo que la solución a este problema es compleja y puede requerir una colaboración interdisciplinaria entre expertos en ética, tecnología y otros campos.

Esta problemática ética resalta la necesidad de garantizar que la IAG no solo sea tecnológicamente avanzada, sino también éticamente responsable. La dignidad humana exige que todos los individuos sean tratados con igualdad y justicia, y es responsabilidad de los desarrolladores, los reguladores y la sociedad en su conjunto garantizar que la IAG cumpla con estos principios fundamentales.

5.2.3 PRIVACIDAD Y SEGURIDAD DE LOS DATOS

El uso de la IAG, incluyendo tecnologías como ChatGPT, plantea un dilema ético de gran envergadura relacionado con la privacidad y la seguridad de los datos. Estos sistemas requieren enormes cantidades de datos para su entrenamiento y funcionamiento, lo que despierta preocupaciones significativas sobre cómo se manejan y protegen los datos de los usuarios.

En el contexto kantiano, la privacidad se entiende como un valor intrínseco que sostiene la dignidad y la autonomía humanas. Kant argumenta que la dignidad de una persona reside en su capacidad de actuar de acuerdo con principios morales, lo que implica la capacidad de tomar decisiones libres, sin coacción externa (Kant, 1946). La privacidad se convierte en un componente esencial para mantener esta autonomía individual, ya que proporciona un espacio donde las personas pueden deliberar y decidir sin interferencias externas.

Además de salvaguardar la autonomía, la privacidad también es fundamental para proteger la integridad personal. Proporciona un espacio esencial para el desarrollo del yo, permitiendo que las personas exploren sus pensamientos, emociones y valores sin temor a la intrusión. Además, la privacidad es crucial para el establecimiento de relaciones auténticas y genuinas, ya que permite compartir pensamientos y sentimientos de manera selectiva y en un entorno de confianza.

Este desafío ético plantea consecuencias significativas para la protección de la privacidad y la seguridad de los datos en una era donde la IAG se expande de manera ubicua. La falta de medidas efectivas para garantizar la privacidad y la seguridad de los datos puede dar lugar a la exposición de información personal, generando riesgos como la la mercantilización de datos, el robo de identidad y el abuso de datos personales.

Uno de los riesgos más graves asociados a esta tecnología es la mercantilización de los datos, donde destaca el concepto crítico de "capitalismo de vigilancia", acuñado por Shoshana Zuboff. Esta teoría subraya cómo la IAG, como instrumento clave en la recopilación y análisis masivo de datos, contribuye a la lógica del capitalismo de vigilancia (Zuboff, 2023). En este sistema, la extracción y análisis de datos se convierten en una macrotendencia dominante, generando un nuevo tipo de contrato basado en la monitorización continua de las interacciones digitales.

Las cuatro características clave del capitalismo de vigilancia destacan aún más la gravedad ética de la IAG en el ámbito de la privacidad y seguridad. Primero, el incremento en la extracción y análisis de datos como macrotendencia subraya la voracidad con la que la IAG consume información personal. Segundo, la aparición de nuevos contratos mediante la monitorización redefine las relaciones usuario-plataforma, llevando a una nueva forma de contrato basado en la vigilancia constante. Tercero, la personalización de servicios para hacerlos más deseables destaca cómo la IAG adapta su funcionamiento a través de datos personalizados, intensificando la intrusión en la privacidad de las personas. Cuarto, la continua experimentación sobre usuarios y consumidores a través de plataformas tecnológicas revela cómo la IAG se convierte en un instrumento de experimentación constante sobre los usuarios en donde las personas se sienten dueñas de sus decisiones, pero en la realidad no lo son tanto.

Este sistema plantea desafíos éticos sustanciales, ya que la IAG, al contribuir al capitalismo de vigilancia, puede llevar a la instrumentalización y explotación de la privacidad individual para obtener ganancias económicas y manipular comportamientos. La necesidad urgente de abordar estos desafíos éticos se vuelve evidente, y el análisis detenido de la IAG en este contexto permite una comprensión más completa de su impacto en la privacidad y seguridad de los datos.

Otro de los riesgos más evidentes asociado a este desafío ético es el robo de identidad. Si los datos personales utilizados en interacciones con la IAG no están debidamente protegidos, pueden ser vulnerables a la sustracción por parte de actores maliciosos. Esto podría conducir al robo de identidad, un delito en el que un tercero utiliza la información personal de una persona para cometer fraudes financieros, realizar compras no autorizadas y cometer otros actos ilegales. Las víctimas de robo de identidad pueden enfrentar consecuencias financieras y emocionales devastadoras.

La falta de privacidad también puede dar lugar al abuso de datos personales. Si los datos recopilados por la IAG caen en manos equivocadas o son utilizados con fines no éticos, las personas pueden ser objeto de manipulación, extorsión o acoso. Los datos personales son valiosos en la era digital, y su mal uso puede exponer a las personas a vulnerabilidades.

Desde una perspectiva kantiana, la falta de protección de la privacidad y la seguridad de los datos en el contexto de la IAG socava la dignidad y la autonomía humanas. Cuando los individuos no

pueden confiar en que sus datos personales estén seguros y protegidos, se ven limitados en su capacidad para tomar decisiones libres y para desarrollarse de manera integral.

Para abordar este desafío ético, es fundamental que los desarrolladores de IAG implementen salvaguardas sólidas de privacidad, minimicen la recopilación de datos innecesarios y garanticen la seguridad de los datos almacenados. Los usuarios deben tener control sobre sus datos y se les debe informar de manera clara y accesible sobre las prácticas de recopilación y uso de datos. Los reguladores desempeñan un papel crucial al establecer estándares y supervisar el cumplimiento de estas prácticas, garantizando así la protección de la privacidad y la seguridad de los datos en la IAG.

5.2.4 MANIPULACIÓN Y ENGAÑO:

La capacidad de las IAG para generar respuestas convincentes y persuasivas plantea un desafío ético sustancial relacionado con la manipulación y el engaño. Este aspecto se vuelve especialmente significativo al considerar la posibilidad de influir en las opiniones, actitudes y decisiones de las personas a través de respuestas generadas por estas tecnologías. Ante la proliferación de estas tecnologías, resulta necesario establecer límites claros y una regulación ética sólida para prevenir la manipulación y garantizar que las interacciones con las IAG sean transparentes y éticas.

Desde una perspectiva kantiana, actuar con información auténtica es esencial para la dignidad humana, ya que permite operar como agentes morales autónomos. La desinformación amenaza directamente la dignidad humana al socavar la autonomía y capacidad de razonamiento de las personas. La entrega de información falsa o manipulada por herramientas como ChatGPT amenaza directamente la dignidad humana al socavar la autonomía y capacidad de razonamiento de las personas. Siendo el cuarto desafío ético, la manipulación y el engaño presentan un riesgo aún mayor cuando se combinan con la entrega de información falsa o manipulada por herramientas como ChatGPT.

La manipulación y desinformación generadas por IAG presentan un riesgo sustancial cuando, por ejemplo, se utilizan para crear noticias falsas o rumores. Estas prácticas pueden llevar a campañas

políticas o sociales diseñadas para manipular la percepción pública y afectar la toma de decisiones informada.

La desinformación a través de IAG implica la diseminación deliberada de información incorrecta con la intención de engañar o influir en la percepción pública. Este fenómeno puede ocurrir de diversas formas, desde la creación de narrativas falsas hasta la modificación de hechos objetivos para favorecer ciertos intereses.

La desinformación, cuando es generada y amplificada por IAG, distorsiona la realidad y socava la capacidad de las personas para tomar decisiones informadas y autónomas. La dignidad humana se ve comprometida cuando la información sobre la cual basamos nuestras decisiones y acciones es falsa o manipulada, ya que esto va en contra de los ideales kantianos de actuar según principios universales y de proteger nuestra autonomía moral.

Esto presenta una amenaza seria para la dignidad humana al comprometer la capacidad de las personas para operar como agentes morales autónomos. La falta de regulación efectiva y salvaguardias sólidas permite que estas tecnologías puedan ser explotadas con consecuencias perjudiciales para la sociedad. Es esencial abordar estos desafíos éticos de manera integral para garantizar un uso ético y responsable de las Inteligencias Artificiales Generativas.

5.2.5 TRANSPARENCIA Y EXPLICABILIDAD

Otra preocupación ética importante es la falta de transparencia y explicabilidad en los sistemas de IAG como ChatGPT. La forma en que estos modelos toman decisiones y generan respuestas no siempre es clara, lo que dificulta la comprensión y la confianza en sus resultados. Es fundamental trabajar hacia la transparencia y la explicabilidad, permitiendo a los usuarios comprender cómo se generan las respuestas y qué factores influyen en ellas.

La transparencia y la explicabilidad en sistemas de IAG están vinculadas con la dignidad humana porque, en esencia, se trata de respetar y potenciar la autonomía individual y la capacidad de autodeterminación. Cuando las personas interactúan con herramientas o sistemas, tener una comprensión clara de cómo funcionan y cómo toman decisiones es esencial para que los usuarios

puedan interactuar con estos sistemas de manera informada y consciente. Si un sistema es opaco y su funcionamiento es incomprensible, el individuo puede sentir que no tiene control o que sus acciones y decisiones están siendo influenciadas por fuerzas que no comprende, lo que puede comprometer su capacidad de actuar de manera autónoma.

La transparencia y la explicabilidad en los sistemas de IAG están intrínsecamente vinculadas con la dignidad humana, ya que buscan respetar y potenciar la autonomía individual y la capacidad de autodeterminación. Desde una perspectiva kantiana, ser tratados como fines en sí mismos implica que se debe proporcionar la información y el conocimiento necesarios para actuar con autonomía. La opacidad y la falta de inteligibilidad de los sistemas de IAG, al dificultar el poder comprender y razonar sobre las acciones y decisiones que implican a estos sistemas, pueden ser vistas como un fallo en tratar a las personas con la dignidad que merecen.

Abordar este desafío ético implica desarrollar prácticas y estándares que mejoren la transparencia de estos sistemas, permitiendo a los usuarios comprender cómo funcionan y tomar decisiones informadas al interactuar con ellos. La ausencia de tales medidas puede conducir a una pérdida de confianza en la tecnología y a la disminución de la capacidad de las personas para actuar de manera autónoma.

5.2.6 INVESTIGACIÓN

La calidad y autenticidad de la investigación es un pilar fundamental en la acumulación y transmisión del conocimiento humano. La naturaleza de la IAG, como por ejemplo ChatGPT, basada en una gran cantidad de texto ya existente en Internet, plantea interrogantes sobre la originalidad, verificabilidad y exactitud de los resultados generados. El uso de ChatGPT en la redacción de manuscritos científicos ha generado debates sobre la autoría y la responsabilidad ética (Rahimi & Abadi, 2023). La inclusión de agradecimientos en los artículos publicados que reconocen el uso de ChatGPT indica que esta herramienta ha entrado en el campo de la publicación científica y podría llegar a dominarlo (Lund et al., 2023).

Desde una perspectiva kantiana, esta potencial falta de precisión compromete la dignidad humana al socavar la autonomía y el razonamiento de las personas. Para respetar y potenciar la capacidad de los individuos para actuar con conocimiento de causa y autodeterminación, es esencial que la información generada utilizada en investigaciones sea precisa y verificable. La integridad de la investigación no solo afecta la validez del conocimiento en sí, sino también el respeto y reconocimiento de la dignidad inherente de aquellos que confían y actúan en base a ese conocimiento.

Asimismo, la falta de directrices éticas claras y consistentes para el uso de herramientas como ChatGPT plantea problemas significativos. La confianza en la tecnología de IAG debe ir acompañada de una verificación humana rigurosa, ya que el texto generado puede contener plagio, información errónea o citas incorrectas. El aumento esperado de la tasa de plagio en los manuscritos futuros debido al uso de ChatGPT requiere una atención cuidadosa por parte de los editores y revisores, y la detección de contenido generado por IAG debe ser una prioridad.

La utilización de ChatGPT en, por ejemplo, la redacción de informes sobre la eficacia de medicamentos hipotéticos, sin la supervisión y verificación adecuada de expertos humanos, puede ser perjudicial para la seguridad y la salud del paciente. La presencia de sesgos, imprecisiones o inexactitudes en estos informes puede comprometer la seguridad de los pacientes y llevar a una desconfianza aún mayor en dichas empresas científicas. La confianza del público en la integridad de la investigación científica puede erosionarse si se promueven o confían en contenidos generados por ChatGPT sin las debidas garantías éticas y de precisión.

No obstante, el uso de la IAG en trabajos de investigación sí puede tener cierto valor y aportar beneficio si se valida correctamente toda la información y, además, se usa siempre de manera que se salvaguarde el sentido crítico y la creatividad del ser humano.

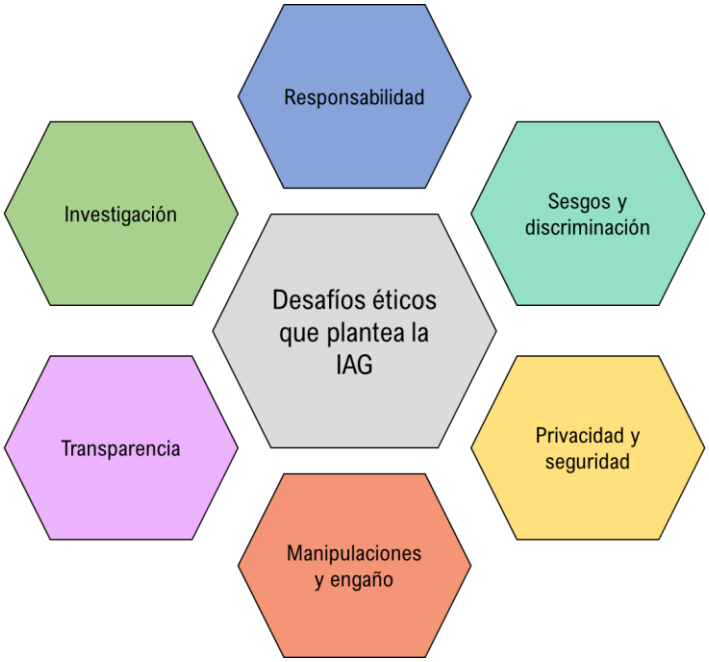

Figura 2. Desafíos éticos de la IAG

Todos estos desafíos éticos que ChatGPT plantea son complejos y variados. Para garantizar que la implementación y el uso de esta tecnología sean éticamente responsables, es necesario abordar los dilemas relacionados con la responsabilidad, los sesgos, la privacidad, la manipulación, la desigualdad, la transparencia, la confiabilidad en trabajos de investigación, y la autoría. La regulación ética y el diseño centrado en el ser humano son fundamentales para asegurar que las IAG como ChatGPT se utilicen de manera beneficiosa, responsable y en línea con nuestros valores éticos fundamentales, y minimizar los riesgos y consecuencias negativas asociadas.

5.3 La necesidad de una ética aplicada a la Inteligencia Artificial Generativa

Debido a todos estos desafíos ético-sociales y sus implicaciones, resulta necesario un marco de regulación ético que sea efectivo para abordar todos estos desafíos y aquellos nuevos que surjan a medida que estas tecnologías sigan avanzando.

El crecimiento de la IAG parece imparable, lo que implica que nuestras relaciones con el mundo estén cambiando continuamente, ya sea por nuestra relación con las máquinas o por la relación entre las propias máquinas. Los productos de IA son cada vez más autónomos, por lo que debemos

comenzar a pensar cuál es nuestro papel en el mundo y a repensar nuestra relación con la máquina. Según algunos investigadores y expertos en el campo de la IA, la introducción de ChatGPT supone un cambio de paradigma que marca un antes y un después en el avance de la IAG (Mattas, 2023).

A raíz de este rápido crecimiento de la IAG liderado por ChatGPT y debido al riesgo de perder el control sobre su avance, muchas entidades y expertos en IAG han advertido del peligro de continuar con el avance en el desarrollo de estas IAG sin que se establezca un reglamento adaptado a estos nuevos tiempos que las regule. En 2016 Stephen Hawking predijo que "el surgimiento de la inteligencia artificial poderosa será lo mejor o lo peor que le haya ocurrido a la humanidad. Todavía no sabemos cuál" (2016). Se trata de una reflexión sobre las esperanzas y temores en relación con la IA que durante mucho tiempo ha sido estudiado por investigadores académicos. Pero con el lanzamiento de ChatGPT para su uso generalizado, se ha vuelto mucho más real para muchas más personas.

Algunos han pedido que se frene el avance de estas tecnologías. En marzo de 2023, la organización sin ánimo de lucro *Future of Life Institute* emitió una carta abierta que recogió más de 1.100 firmas en pocas horas (Belinchón, 2023) entre las que estaban la de Elon Musk (fundador de Tesla), Steve Wozniak (cofundador de Apple), Jaan Tallin (cofundador de Skype), e ingenieros de Meta, Microsoft y otras grandes empresas. Dicha carta solicitaba a todos los laboratorios de inteligencia artificial que pausen de forma inmediata los desarrollos más potentes que GPT-4 durante al menos seis meses debido que "los sistemas de inteligencia artificial pueden suponer un profundo riesgo para la sociedad y la humanidad. Por desgracia, no se está desarrollando con el nivel de planificación y cuidado adecuado" y afirmaba que "se ha producido una carrera descontrolada para desarrollar sistemas cada vez más poderosos que nadie, ni siquiera sus creadores, entienden, predicen o pueden controlar con fiabilidad".

Poco después, la ONU también hizo un comunicado expresando su preocupación por los posibles impactos que podía haber por el rápido desarrollo en IA y solicitó a las empresas del sector responsabilidad (UNESCO, 2023). Desde entonces la ONU ha mantenido encuentros en los que asisten gobiernos, empresas y sociedad civil para establecer directrices aceptadas por todos que aseguren que la tecnología no es objeto de abusos. El propio CEO de OpenAI, Sam Altman,

reaccionó a estos movimientos de presión hacia su compañía y reconoció que debe haber un marco regulatorio global efectivo (Toh & Seo, 2023).

Por otro lado, el 'Center for AI and Digital Policy' (CAIDP) demandó a OpenAI por ChatGPT ante la Comisión de Comercio Federal (CCF) de Estados Unidos. Alegaron que ChatGPT es un modelo generativo de texto "sesgado, engañoso y que supone un riesgo para la privacidad y la seguridad pública" (Kramer, 2023). En la demanda, afirmaron que la tecnología de OpenAI no cuenta con un protocolo que limite la parcialidad y el engaño. Asimismo, Italia prohibió en abril el uso de ChatGPT por cuestiones de seguridad y privacidad (McCallum, 2023). Otros países que también han bloqueado el uso de ChatGPT son China, Irán, y Rusia.

Por todo esto, y por lo presentado en este trabajo hasta este punto, parece coherente y necesario establecer cuanto antes un marco ético global que permita regular de forma efectiva estas nuevas tecnologías dotadas de inteligencia artificial. El impacto social y ético, así como gran la incertidumbre que hay respecto a los potenciales riesgos de estas tecnologías según siguen avanzando, instan a frenar su avance y regularlas éticamente antes de que pueda ser muy tarde.

No obstante, antes de explorar la opción de un nuevo marco ético que permita abordar todos los desafíos explicados en este trabajo, es fundamental analizar la regulación que existe actualmente respecto a la IAG para poder construir la propuesta final de este trabajo a partir de las carencias que presente dicho marco regulatorio.

5.4 Regulación y gobernanza de la inteligencia artificial generativa

La emergencia de las IAG, como ChatGPT, ha suscitado una serie de interrogantes que trascienden la mera funcionalidad tecnológica. La magnitud de su influencia requiere evaluar críticamente cómo se rige y modera su actuar. Actualmente, la regulación y gobernanza de la inteligencia artificial enfrenta lagunas significativas, careciendo de herramientas adecuadas para abordar los desafíos éticos y sociales que esta tecnología impone, en especial los subrayados anteriormente en este trabajo.

Reconociendo estos vacíos, se propondrá la construcción de un marco ético robusto que permita una regulación adecuada, garantizando que la inteligencia artificial generativa opere de manera segura, transparente y éticamente responsable. La aspiración es que, con una gobernanza más firme y coherente, se pueda dirigir la evolución de estas tecnologías hacia un horizonte que respete y proteja los valores fundamentales de la sociedad. Así que, dado el panorama descrito, resulta necesario analizar en profundidad las carencias del marco regulatorio y de gobernanza actual.

En abril de 2021, en el marco de su estrategia digital, la Comisión Europea lanzó el EU AI Act (Parlamento Europeo, 2021), una propuesta que tiene como objetivo principal estructurar el desarrollo y uso de la IA en el continente, garantizando beneficios como una asistencia sanitaria optimizada, transportes seguros o una fabricación más eficiente.

Desde su propuesta inicial en 2021, el EU AI Act ha experimentado un largo proceso de revisión y adaptación. A lo largo de este periodo, diferentes entidades y actores dentro de la UE han contribuido con su perspectiva y visión. Como resultado, en junio de 2023, el Parlamento Europeo adoptó una posición negociadora sobre la ley, marcando un paso significativo hacia su eventual ratificación. Si se mantiene el ritmo y la dirección actuales del proceso legislativo, es probable que la ley sea aprobada para finales de 2023 (Parlamento Europeo, 2023). Sin embargo, su implementación efectiva y entrada en vigor se espera para finales de 2025 o inicios de 2026.

La normativa propuesta clasifica los sistemas de IA en función del riesgo que representan:

- **Riesgo Inaceptable**: Estos sistemas son aquellos que presentan un claro potencial de daño en términos de seguridad y derechos fundamentales de los individuos. Debido a sus posibles impactos negativos, serían prohibidos por la legislación. Por ejemplo, juguetes que pueden ser utilizados para la vigilancia de niños o sistemas que evalúan la confiabilidad crediticia basados en características de género o etnia.
- **Riesgo Limitado**: Son sistemas de IA que poseen un impacto menor o un riesgo reducido. Estos sistemas están sujetos a regulaciones más laxas. Por ejemplo, chatbots que proporcionan información sobre el clima o recomendaciones de películas.

- **Alto Riesgo**: Esta categoría es de particular interés y se refiere a sistemas que, aunque no necesariamente inaceptables, requieren una atención especial debido a su potencial impacto en la seguridad o derechos fundamentales de las personas. Estos sistemas deberían cumplir con un conjunto estricto de regulaciones y requisitos, como la transparencia en su funcionamiento y decisiones, así como ser objeto de auditorías regulares. Ejemplos podrían incluir sistemas de reconocimiento facial utilizado en espacios públicos o algoritmos utilizados en decisiones médicas.

Los sistemas categorizados bajo "alto riesgo" enfrentan regulaciones intensivas dada su potencialidad de impacto en la seguridad o derechos fundamentales. La ley propone que estos sistemas sean evaluados rigurosamente antes de su comercialización, se adhieran a estrictos protocolos de gestión de riesgos y gobernanza de datos, y que sean supervisados y monitoreados periódicamente a lo largo de su ciclo de vida.

La IA generativa, como ChatGPT, representa una categoría particular dentro del espectro de la IA. Estas herramientas, que tienen la capacidad de generar contenido nuevo basándose en vastas cantidades de información previa, plantean cuestiones únicas en términos de transparencia y responsabilidad. La propuesta del EU AI Act especifica que estos sistemas deben:

1) Revelar explícitamente cuando el contenido ha sido generado por IA.
2) Ser diseñados con precauciones para evitar la generación de contenidos ilegales.
3) Proporcionar resúmenes o desgloses de los datos utilizados en su entrenamiento, especialmente cuando involucren datos sujetos a derechos de autor.

Con el EU AI Act, la Unión Europea se posiciona a la vanguardia de la regulación de la inteligencia artificial, buscando un equilibrio entre la promoción de la innovación y la protección de los derechos y seguridad de sus ciudadanos. Esta ley es una respuesta proactiva a los desafíos emergentes de la IA y una clara señal de la importancia y el impacto esperado de la IA en la sociedad. Al ofrecer un modelo que podría guiar a otras jurisdicciones en el futuro, refleja la necesidad de una regulación cuidadosa y ponderada para garantizar un desarrollo y aplicación éticos y responsables de la IA.

Una vez aprobado, este marco regulatorio no solo será aplicable directamente en todos los estados miembros de la UE, sino que también tendría un alcance extraterritorial considerable, similar al del Reglamento General de Protección de Datos (RGPD), afectando a entidades fuera de la UE que operen o tengan interacciones significativas dentro de la Unión.

5.5 Desafíos en la regulación de la Inteligencia Artificial Generativa

La regulación actual en la Unión Europea, representada por el EU AI Act, es un paso importante en la dirección correcta para abordar los desafíos emergentes de la inteligencia artificial generativa. Sin embargo, es importante destacar que aún existen lagunas y carencias en este marco regulatorio que deben ser consideradas cuidadosamente para garantizar que se cumplan los objetivos éticos y sociales de manera efectiva.

El EU AI Act establece una base importante, pero existen ciertas lagunas que necesitan ser examinadas más detenidamente. En concreto, la propuesta de la UE sobre la IAG parece quedarse muy corta en comparación con el rápido crecimiento que ha tenido esta tecnología y resulta insuficiente para abordar las problemáticas sociales y éticas destacadas en este trabajo

Veracidad de la Información:

El EU AI Act propone que la IAG revele cuando el contenido ha sido generado por IA. Aunque esta es una medida valiosa para la transparencia, hay desafíos prácticos significativos. No especifica claramente cómo se aplicará esto en tiempo real y en todas las plataformas. Tomemos un ejemplo: un artículo periodístico generado por una IAG. Si no se etiqueta adecuadamente, esto podría resultar en la difusión de noticias falsas que tienen un impacto negativo en la percepción pública y la toma de decisiones. La regulación debe ser más específica en cuanto a cómo garantizar esta transparencia y cómo sancionar el incumplimiento.

Interacción Humana:

Actualmente la IAG es capaz de replicar interacciones humanas de manera convincente. Aunque el EU AI Act busca transparencia en la utilización de la IAG, no aborda plenamente cómo estas interacciones pueden influir en la sociedad. Un ejemplo relevante es el impacto en el desarrollo

social de los jóvenes. Si los niños prefieren interactuar con un chatbot en lugar de con humanos, esto podría tener un efecto significativo en su desarrollo emocional y social. La regulación podría considerar más detenidamente cómo proteger y fomentar interacciones humanas auténticas, poniendo en primer lugar al ser humano en lugar de la máquina.

Desplazamiento Laboral:

El EU AI Act se centra en la seguridad de los productos, pero no aborda adecuadamente el desplazamiento laboral. Por ejemplo, la IAG puede ser utilizada para tareas como redacción de informes, análisis de mercado o incluso composición musical, lo que podría resultar en la sustitución de trabajadores en sectores como periodismo, finanzas o arte, entre otros. La regulación debería considerar mecanismos para mitigar los impactos negativos en el empleo, como la inversión en programas de formación y reciclaje laboral que permitan una adecuada adaptación a la nueva era laboral gobernada por la IA.

Daño al Sector Educativo:

Si bien la regulación del EU AI Act establece medidas para garantizar la seguridad de los productos, no aborda de manera específica cómo prevenir un posible daño al sector educativo. La IAG podría utilizarse en exceso para completar tareas y trabajos académicos, lo que podría fomentar la falta de esfuerzo personal y comprometer el proceso de aprendizaje. Para asegurar que la IAG se utilice de manera ética en la educación, la regulación debe incluir directrices específicas para su uso en entornos educativos y fomentar la colaboración entre docentes y la IAG en lugar de la sustitución de tareas.

Seguridad y Privacidad:

Si bien la regulación establece disposiciones sobre la protección de datos, no aborda de manera integral cómo proteger la privacidad en el contexto de la IAG. La generación de contenido basada en información personal plantea riesgos significativos para la privacidad de los individuos. La regulación debe abordar de manera más detallada cómo garantizar la protección de datos sensibles y promover prácticas de privacidad robustas en la IAG.

Responsabilidad:

La regulación establece responsabilidades para los proveedores de IAG, pero la IAG, como ChatGPT, puede aprender y evolucionar con el tiempo. Si un modelo de IAG produce contenido ofensivo después de miles de interacciones, la responsabilidad no está claramente definida. ¿Es el proveedor original o el usuario responsable? La regulación debería proporcionar una mayor claridad sobre la responsabilidad en estas situaciones.

Sesgos y Discriminación:

Si bien la regulación aborda la ilegalidad de contenidos, no aborda de manera exhaustiva los sesgos más sutiles que pueden surgir. Si un modelo de IAG es entrenado con literatura histórica que contiene prejuicios raciales o de género, podría reproducir estos sesgos en sus respuestas. La regulación debería incluir disposiciones para garantizar que las IAG no reproduzcan ni amplifiquen sesgos.

Manipulaciones y Engaño:

La IAG podría ser utilizada para crear noticias falsas o propaganda política, lo que afecta la percepción pública y manipula opiniones. La regulación no proporciona una guía clara sobre cómo abordar este uso potencialmente perjudicial. Debería establecerse una regulación más precisa para prevenir y sancionar el abuso de la IAG con fines engañosos.

Transparencia:

La regulación enfatiza la transparencia, pero las IAG son notoriamente opacas. Cuando un modelo de IAG toma una decisión, es difícil entender cómo o por qué lo hizo. La regulación debe proporcionar directrices específicas para garantizar que las IAG sean transparentes en su toma de decisiones.

Investigación:

Las directrices internacionales sobre autoría establecen claramente que los autores deben haber contribuido sustancialmente al manuscrito, revisado críticamente su contenido, aprobado la versión

publicada y asumido la responsabilidad de la integridad y exactitud del contenido publicado (Darío et al., 2005). Como ChatGPT y otras herramientas de IAG no son entidades legales y no pueden cumplir con estas responsabilidades, su inclusión como coautores violaría las directrices éticas y de autoría. Sin embargo, la regulación no aborda de manera suficiente cómo se debería regular la IAG en entornos de investigación dado que, si un investigador utiliza la IAG para producir resultados de estudios, esto podría comprometer la integridad científica. Deberían establecerse pautas éticas claras para el uso de IAG en la investigación.

No obstante, con la inclusión de la obligatoriedad de revelar cuando el contenido ha sido generado por IAG, será más fácil discriminar aquellos trabajos de investigación que han hecho uso de IAGs frente a los que no. Esto podría ser muy significativo en el futuro próximo de cara a juzgar la calidad e integridad de las investigaciones.

En resumen, aunque el EU AI Act es un paso significativo hacia la regulación de la inteligencia artificial, se requiere un análisis más profundo y detallado para abordar de forma efectiva los desafíos éticos y sociales relacionados con la inteligencia artificial generativa. La IAG es una tecnología en constante evolución y, por lo tanto, la regulación debe ser adaptable y revisada regularmente para mantenerse al día con los avances tecnológicos y sus implicaciones en la sociedad. Este proceso de revisión constante es esencial para garantizar que la regulación cumpla sus objetivos éticos y sociales.

6 NUEVO MARCO ÉTICO PARA LA INTELIGENCIA ARTIFICIAL GENERATIVA

Esta sección final representa el culmen de una exhaustiva exploración que ha abordado el amplio espectro de la Inteligencia Artificial Generativa (IAG), desde sus raíces históricas hasta su irrupción más reciente con ChatGPT. A lo largo de este viaje, que se ha desplazado desde el desarrollo de la Inteligencia Artificial (IA) hasta la necesidad imperativa de establecer una ética aplicada a la IAG, se han examinado las implicaciones sociales, los desafíos éticos y la falta evidente de regulación, brindando una visión completa de la influencia de la IAG en la sociedad.

Este recorrido crítico ha destacado la ausencia apremiante de regulación efectiva, puesta de manifiesto en la sección dedicada a la regulación y gobernanza de la IAG. La urgencia de abordar los retos éticos particulares, combinada con la carencia de un marco regulador adecuado, sienta las bases para la propuesta ética que se presenta en esta sección.

Frente a los desafíos éticos planteados por la IAG, surge la necesidad de establecer un marco ético que oriente su desarrollo y aplicación. Esta sección final del trabajo se basa en hacer una propuesta de un marco ético que conste de una selección de teorías éticas que son: el análisis crítico de Adela Cortina sobre la ética de la inteligencia artificial, el principio de libertad y autonomía de Kant, el marco aristotélico tomista que resalta el pleno conocimiento y la plena voluntad, y el principio de responsabilidad de Hans Jonas. Esta elección se basa en las problemáticas sociales y éticas identificadas en el trabajo, con la intención de abordar dichas áreas de preocupación.

Este nuevo marco ético no busca ser una solución exhaustiva y definitiva para todas las problemáticas de la IAG, sino más bien una respuesta urgente y necesaria ante la carencia de regulación efectiva. Reconociendo la complejidad y la evolución constante de la tecnología, la propuesta se erige como un faro ético, delineando principios fundamentales que deben guiar el desarrollo y la implementación de la IAG en la sociedad actual.

En las secciones que siguen, se explorarán detenidamente cada una de las teorías éticas que componen este nuevo marco, fundamentando su aplicación en el contexto de la IAG y destacando

su contribución a la creación de un marco ético que promueva la responsabilidad, la equidad y la protección de la autonomía humana.

6.1 Ética de la Inteligencia Artificial según Adela Cortina

Este nuevo marco ético propuesto se sustenta en diversas teorías éticas que abordan las complejidades de la Inteligencia Artificial Generativa (IAG), con ChatGPT como su exponente principal. En este contexto, el análisis crítico de Adela Cortina sobre la ética de la inteligencia artificial emerge como un pilar fundamental.

El análisis crítico de Cortina, presente en su artículo sobre ética de la inteligencia artificial (Cortina, 2019), proporciona orientaciones éticas esenciales para el uso de sistemas inteligentes. Al abordar la inevitabilidad del mundo digital, su propuesta ética aboga por un equilibrio entre el progreso técnico y ético. En esta línea, la amalgama de teorías éticas busca maximizar los beneficios de los sistemas inteligentes, con una "competitividad responsable" que prevenga los riesgos asociados.

Cortina destaca la necesidad de unir el progreso técnico con el progreso ético, abordando así la dualidad intrínseca a la IA. La ética aplicada a la IAG se presenta como un medio para maximizar los beneficios de los sistemas inteligentes, al tiempo que previene los riesgos asociados. Este enfoque fomenta la "competitividad responsable" y establece una ventaja global.

El marco ético propuesto, influenciado por las directrices de Cortina, prioriza la confianza como piedra angular de la sociedad. Se destaca la importancia de construir una IA confiable que cumpla con tres componentes fundamentales: legalidad, ética y robustez desde el punto de vista ético y social. Cortina aboga por un enfoque claro en los principios éticos de respeto a la autonomía humana, prevención del daño, justicia y explicabilidad, todos ellos incorporados en la propuesta ética.

Este principio antropocéntrico, resaltado por Cortina, refuerza la idea de que los sistemas inteligentes son instrumentos para mejorar la vida humana y la naturaleza, no fines en sí mismos. La autonomía de las personas humanas se coloca en el centro, subrayando que la IA debe estar subordinada a la humanidad, sin posibilidad de sustitución. Los principios éticos de explicabilidad,

beneficiar, no dañar y justicia se erigen sobre la base del reconocimiento de la autonomía y la dignidad humana.

6.2 Ética Kantiana

La propuesta ética integrada para la regulación de la Inteligencia Artificial Generativa (IAG) se complementa con el enfoque ético kantiano, destacando el principio de libertad y autonomía de Kant, junto con su reconocido imperativo categórico (Kant, 1946). La visión de Kant ofrece una perspectiva basada en la centralidad de la libertad y autonomía de los individuos, planteando que las acciones deben guiarse por principios morales universales. En el contexto de la IAG, esto implica asegurar que las aplicaciones y sistemas respeten la autonomía de los usuarios, evitando la manipulación y garantizando la transparencia en las interacciones.

El principio de libertad y autonomía según Kant subraya la importancia de tratar a las personas como fines en sí mismas, en lugar de simplemente como medios para alcanzar objetivos. Extrapolando esta idea al ámbito de la IAG, implica diseñar sistemas que respeten la autonomía de los usuarios, garantizando que las interacciones digitales no restrinjan indebidamente sus elecciones y decisiones. En este marco ético, la transparencia en el funcionamiento de los algoritmos y la toma de decisiones resulta crucial para empoderar a los individuos y preservar su autonomía.

El imperativo categórico kantiano refuerza la idea de que las acciones deben regirse por principios que podrían ser aceptados universalmente. Esto implica que los desarrolladores y responsables de la IAG deben adoptar normas éticas que respeten la libertad y autonomía de los usuarios en todos los contextos culturales. Asimismo, se destaca la importancia de la explicabilidad y rendición de cuentas en los sistemas de IAG, asegurando que los usuarios puedan comprender las decisiones tomadas por los algoritmos y que exista responsabilidad en caso de posibles problemas éticos.

6.3 La perspectiva Aristotélica-Tomista

Desde la perspectiva aristotélico-tomista, el análisis ético se enriquece al considerar las condiciones del acto humano, basadas en pleno conocimiento, plena voluntad (libertad), y actitud constructiva (MacIntyre, 2017). Este enfoque proporciona una visión integral que busca no solo evaluar las

acciones desde una perspectiva de deber, como propone Kant, sino también considerar el desarrollo humano y el bien común. La felicidad de la persona es entendida como desarrollo moral, virtuoso, y como autorrealización de la persona que lleva a la plenitud de las potencialidades con las que ha sido agraciada por la naturaleza. En el contexto de las IAG, las condiciones sobre las que se sustenta el marco aristotélico-tomista se encuentran enfrentadas.

La condición de pleno conocimiento se vincula con la transparencia, un aspecto esencial, pero a menudo problemático en el desarrollo y uso de la IAG. La falta de transparencia en los algoritmos y procesos de toma de decisiones de estos sistemas plantea interrogantes sobre el nivel de conocimiento que los usuarios tienen sobre cómo operan. La opacidad de estos sistemas puede limitar la capacidad de los individuos para comprender y reflexionar plenamente sobre las implicaciones éticas de su interacción con la IAG

En segundo lugar, la condición de plena voluntad está conectada con el desafío ético identificado en este trabajo sobre la manipulación y el engaño. La IAG, al emplear estrategias persuasivas y algoritmos diseñados para influir en el comportamiento humano, puede comprometer la libertad de elección de los usuarios. Esto es especialmente más visible cuanto más indefenso es el sujeto moral, como por ejemplo los niños. La falta de voluntad genuina, en el sentido de tomar decisiones libres de influencias manipuladoras, se convierte en un punto de conflicto ético. En este marco, el diseño y la implementación de la IAG deben abordar de manera proactiva la prevención de prácticas manipuladoras que puedan distorsionar la toma de decisiones autónoma.

Por otro lado, la noción de actitud constructiva sigue siendo esencial en este contexto. Los diseñadores y desarrolladores de sistemas de IAG deben asumir una actitud ética no solo en la creación de tecnologías transparentes y libres de manipulación, sino también en la promoción activa de un entorno donde la tecnología sirva al bienestar humano y a la construcción de una sociedad ética y justa.

Este enfoque aristotélico-tomista, aplicado a la IAG, busca integrar elementos críticos relacionados con la transparencia y la manipulación, abordando desafíos específicos que surgen en el contexto de la inteligencia artificial. Al mantener esta perspectiva, se busca establecer un marco ético que no

solo regule las interacciones tecnológicas, sino que también promueva el florecimiento humano y la construcción de una sociedad ética y justa en la era de la IAG.

6.4 El principio de responsabilidad de Hans Jonas

El principio de responsabilidad de Hans Jonas constituye un marco ético fundamental para abordar los desafíos éticos asociados con el desarrollo tecnológico y científico en la sociedad contemporánea. Este enfoque se vuelve esencial al enfrentar los dilemas éticos emergentes relacionados con la Inteligencia Artificial Generativa (IAG). Según Jonas, a medida que aumenta el poder humano para alterar la naturaleza y la vida, también crece la responsabilidad moral de los seres humanos (Jonas, 2014). No obstante, resulta interesante explorar cómo este principio responde al papel de los algoritmos, entidades incapaces de asumir responsabilidad según este marco ético.

En el contexto de la IAG, Hans Jonas sostiene que la ética clásica puede no ser suficiente para abordar la magnitud de los desafíos éticos. En un entorno donde la tecnología puede tener efectos globales y a largo plazo, la responsabilidad se extiende más allá del ámbito individual hacia la comunidad global. Para Hans Jonas el imperativo moral ya no es simplemente individual porque está en peligro el bien común.

El principio fundamental de Jonas establece: "Actuar de manera que los efectos de tu acción sean compatibles con la permanencia de una vida humana auténtica en la Tierra". Esto subraya la urgencia de adoptar una ética de precaución en el desarrollo y aplicación de la IAG. En este contexto, la ética jonasiana aboga por anticipar y considerar las posibles consecuencias negativas de las acciones antes de llevarlas a cabo, en lugar de simplemente reaccionar ante ellas después de que hayan ocurrido. Sin embargo, surge una pregunta intrigante: ¿cómo se aplica este principio a entidades no conscientes como los algoritmos?

La noción de responsabilidad, según Hans Jonas, está intrínsecamente vinculada a la conciencia y la capacidad de anticipar consecuencias morales. Aquí yace un desafío al aplicar este principio a la IAG, donde los algoritmos carecen de conciencia y, por ende, de capacidad de responsabilidad. Esto resalta una brecha ética en la regulación actual, ya que los sistemas de IAG son operados por seres humanos, pero la responsabilidad última recae en estos sistemas no conscientes. Esta disonancia

plantea la necesidad de reflexionar sobre nuevas formas de aplicar el principio de responsabilidad a la IAG, considerando su dinámica única y la intervención humana necesaria para su funcionamiento.

En conclusión, el principio de responsabilidad de Hans Jonas, aplicado a la IAG, aboga por la necesidad de una regulación ética que trascienda los intereses individuales y nacionales, priorizando la protección del bien común y el futuro de la humanidad. La anticipación de consecuencias, la consideración de impactos a largo plazo y la adopción de medidas preventivas son fundamentales en este enfoque ético. Asimismo, contribuye a enriquecer el marco ético propuesto en este trabajo para la IAG, proporcionando un fundamento sólido para la reflexión y la acción responsables en el desarrollo y uso de esta tecnología.

Figura 3. Nuevo marco ético para la IAG

6.5 Reflexión cualitativa sobre la efectividad de este nuevo marco ético

En conclusión, el nuevo marco ético propuesto, basado en una amalgama de teorías éticas que incluye la ética aplicada de Adela Cortina, el principio de libertad y autonomía de Kant con su imperativo categórico, el marco aristotélico-tomista y el principio de responsabilidad de Hans Jonas, emerge como una respuesta integral y reflexiva para abordar los desafíos éticos planteados por la Inteligencia Artificial Generativa (IAG).

En relación con el desafío ético de la **Responsabilidad**, este marco ético aboga por la adopción de un enfoque preventivo, promoviendo la consideración de las consecuencias a largo plazo de las acciones humanas en el desarrollo y aplicación de la IAG. La ética de responsabilidad de Hans Jonas se revela como especialmente relevante al destacar la importancia de anticipar y considerar las posibles implicaciones negativas.

En cuanto a los **Sesgos y Discriminación**, el nuevo marco ético destaca la necesidad de abordar la equidad y la no discriminación como principios fundamentales. La ética aplicada de Adela Cortina, centrada en la ética del discurso y la justicia, se alinea con la necesidad de mitigar los sesgos inherentes en los sistemas de IAG y garantizar la equidad en su desarrollo y aplicación.

El desafío ético de la **Privacidad y Seguridad** encuentra respuesta en la propuesta ética, que aboga por la transparencia y el control humano sobre los datos. La ética de Kant, con su enfoque en la autonomía y la libertad, subraya la importancia de proteger la privacidad individual y asegurar la seguridad de los datos en el contexto de la IAG.

Frente a las **Manipulaciones y Engaños**, el marco ético propuesto subraya la importancia de promover la transparencia y la rendición de cuentas. La ética aplicada de Adela Cortina, con su enfoque en la ética del discurso y la justicia, y el marco aristotélico-tomista, que destaca la importancia de la plena voluntad y la actitud constructiva, convergen en la necesidad de prevenir prácticas manipulativas y engañosas en el desarrollo y uso de la IAG. Este enfoque ético integral busca no solo desalentar la manipulación, sino también fomentar una actitud constructiva en la interacción con la tecnología.

En relación con la **Transparencia**, el nuevo marco ético propuesto, inspirado en la ética aplicada de Adela Cortina, el principio de libertad y autonomía de Kant, el marco aristotélico-tomista y el principio de responsabilidad de Hans Jonas, refuerza la necesidad de divulgar de manera clara y comprensible el funcionamiento de los sistemas de IAG. La ética aristotélico-tomista, que destaca la importancia del pleno conocimiento, se alinea con la necesidad de garantizar que los usuarios comprendan de manera transparente cómo se toman decisiones y procesan los datos en los sistemas de IAG.

Finalmente, respecto al desafío ético de la **Investigación**, el nuevo marco ético propuesto también aborda los riesgos asociados con la utilización de la Inteligencia Artificial Generativa (IAG) en trabajos académicos. La ética de Adela Cortina destaca la importancia de la transparencia, abogando por la divulgación clara del uso de la IAG en la investigación para asegurar autenticidad y calidad. Asimismo, los principios de Kant y la responsabilidad de Hans Jonas refuerzan la necesidad de aplicar estándares éticos estrictos para prevenir información falsa y plagio. En conjunto, el marco ético propuesto destaca la importancia de garantizar la integridad en los procesos de investigación con IAG, abordando así los desafíos específicos en el ámbito académico.

Este marco ético no solo tiene implicaciones teóricas, sino que también puede servir como guía para llenar las lagunas en la regulación de la IAG. La ausencia de normativas específicas ha dejado a la IAG en un terreno ambiguo, donde la ética se convierte en el principal baluarte para guiar su desarrollo. Proponer la implementación de este marco ético en las políticas y regulaciones relacionadas con la IAG podría fortalecer la base ética del campo y proporcionar directrices claras para los desarrolladores y usuarios. La adaptación de este marco en entornos regulatorios podría establecer estándares éticos uniformes, mitigando así los riesgos éticos y mejorando la confianza pública en esta tecnología.

En el ámbito político/regulatorio, la aplicación de este marco ético en las políticas de desarrollo tecnológico podría transformar la IAG en una fuerza impulsora del bien común y la cooperación global. La consideración ética en la regulación de la IAG podría conducir a estándares uniformes que promuevan prácticas éticas y aborden los riesgos específicos de manera proactiva.

En resumen, el marco ético propuesto no solo responde a los desafíos éticos planteados por la IAG, sino que también proporciona una guía holística y esencial para su desarrollo y aplicación responsable. Su implementación en el ámbito político/regulatorio y su adopción por la comunidad de desarrolladores pueden allanar el camino hacia una IAG más ética y colaborativa, alineada con los valores fundamentales de la sociedad. Este enfoque ético promueve la coexistencia armoniosa entre el avance de la IAG y la consideración del ser humano como un sujeto moral, asegurando que el progreso tecnológico esté al servicio de la humanidad.

7 CONCLUSIONES

La propuesta de un nuevo marco ético para la Inteligencia Artificial Generativa (IAG), que fusiona la ética aplicada de Adela Cortina, el principio de libertad y autonomía de Kant, el marco aristotélico-tomista y el principio de responsabilidad de Hans Jonas, emerge como una guía integral y reflexiva. Este enfoque no solo aborda los desafíos éticos identificados en este estudio, sino que también proporciona una base sólida para orientar el desarrollo y la aplicación responsable de la IAG, utilizando a ChatGPT como caso representativo.

La contribución principal de este trabajo radica en la formulación de un marco ético innovador adaptado a las necesidades actuales y que integra diversas perspectivas éticas, ofreciendo un enfoque holístico para enfrentar los dilemas éticos de la IAG. La amalgama de teorías éticas seleccionadas se fundamenta en la identificación de desafíos específicos, como la responsabilidad, sesgos y discriminación, privacidad y seguridad, manipulaciones y engaños, transparencia, e investigación. Este enfoque novedoso aspira a aportar al campo de la ética de la IAG y servir como referencia para futuros desarrollos éticos en el ámbito de la inteligencia artificial.

En este contexto, las implicaciones y utilidades de la propuesta se extienden más allá de la academia, beneficiando a diseñadores, desarrolladores y reguladores en el ámbito de la IAG. Ofrece un marco ético robusto, destacando principios clave como la responsabilidad, la equidad, la transparencia y la autonomía. La propuesta aspira a contribuir al establecimiento de políticas y prácticas éticas, promoviendo un enfoque consciente y reflexivo en la toma de decisiones.

A pesar de su exhaustividad, este trabajo no puede abarcar todas las dimensiones posibles de la ética en la IAG. En primer lugar, la irrupción de esta tecnología provoca que no haya mucha literatura ni estudios al respecto, lo cual dificulta un análisis más detallado que esté respaldado por resultados cuantitativos. Además, la evolución constante de la tecnología y el surgimiento de nuevos desarrollos podrían requerir ajustes en el marco ético propuesto. Asimismo, la aplicación específica de este marco a contextos culturales y regionales particulares podría necesitar adaptaciones y consideraciones adicionales.

Futuras investigaciones podrían explorar las implicaciones prácticas de la propuesta, evaluar su efectividad en entornos específicos e identificar los posibles ajustes legales del marco ético propuesto. Finalmente, serán esenciales mejoras continuas que consideren más teorías y se adapten a los cambios tecnológicos para mantener la relevancia y eficacia del marco ético en la dinámica evolutiva de la IAG.

8 BIBLIOGRAFÍA

Abeliuk, A., & Gutiérrez, C. (2021). Historia y evolución de la inteligencia artificial. *Revista Bits de Ciencia*, (21), 14-21.

Barr, A., Feigenbaum, E. A., & Cohen, P. R. (Eds.). (1981). *The handbook of artificial intelligence* (Vol. 1). William Kaufmann.

Belinchón, F. (2023, Marzo 29). Elon Musk y más de 1000 investigadores firman una carta pidiendo pausar el desarrollo de las IA avanzadas. *El País*. Recuperado de https://cincodias.elpais.com/companias/2023-03-29/elon-musk-y-mas-de-1000-investigadores-firman-una-carta-pidiendo-pausar-el-desarrollo-de-las-ias-avanzadas.html

Bhatnagar, S., Alexandrova, A., Avin, S., Cave, S., Cheke, L., Crosby, M., ... & Hernández-Orallo, J. (2018). Mapping intelligence: Requirements and possibilities. In *Philosophy and theory of artificial intelligence 2017* (pp. 117-135). Springer International Publishing.

Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review*, *90*(1), 5-43.

Boden, M. (2017). Inteligencia Artificial. Madrid: Turner Publicaciones S.L.

Brachman, R. J. (2006). AI more than the sum of its parts. *AI Magazine*, *27*(4), 19-19.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1-3), 139-159.

Campbell, M., Hoane Jr, A. J., & Hsu, F. H. (2002). Deep blue. *Artificial intelligence*, *134*(1-2), 57-83.

Coeckelbergh, M. (2020). *AI ethics*. Mit Press.

Comisión Europea. (2018). IA para Europa. *Comunicación de la Comisión al Parlamento europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones*. COM(2018) 237 final{SWD(2018) 137 final} Bruselas, 25.4.2018, p. 1.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and Cheating: Ensuring academic integrity in the era of ChatGPT.

Chen, J. X. (2016). The evolution of computing: AlphaGo. *Computing in Science & Engineering*, *18*(4), 4-7. DOI: 10.1109/MCSE.2016.74

Darío, R., Hernández, M., & Suarez, G. C. (2005). Derechos de autor en la investigación científica: la autoría en los artículos de investigación. *CES medicina*, *19*(2), 91-96.

Dennett, D. (2004). Can machines think?. In *Alan turing: Life and legacy of a great thinker* (pp. 295-316). Berlin, Heidelberg: Springer Berlin Heidelberg.

De Siqueira, J. E. (2001). EL PRINCIPIO DE RESPONSABILIDAD DE HANS JONAS. *Acta bioethica*, *7*(2), 277-285. https://dx.doi.org/10.4067/S1726-569X2001000200009

Doshi-Velez, F. & Kortz, M. (2017). Accountability of AI Under the Law: The Role of Explanation. *Harvard Library*. Recuperado de http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584

Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, *613*(7944), 423-423. DOI: 10.1038/d41586-023-00056-7

Ertel, W. (2017). Introduction to Artificial Intelligence. *Springer*.

Estupiñán Ricardo, J., Leyva Vázquez, M. Y., Peñafiel Palacios, A. J., & El Assafiri Ojeda, Y. (2021). Inteligencia artificial y propiedad intelectual. *Revista Universidad y Sociedad*, 13(S3), 362-368.

Ferrucci, D. A. (2012). Introduction to "this is watson". *IBM Journal of Research and Development*, *56*(3.4), 1-1. https://doi.org/10.1147/JRD.2012.2184356

Foran, R. (2015). *Robotics: from automatons to the roomba*. ABDO.

French, R. M. (1990). Subcognition and the limits of the Turing test. *Mind*, *99*(393), 53-65.

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv*, 2022-12.

Ginsberg, M. (2012). *Essentials of artificial intelligence*. Newnes.

González, R. (2007). El Test de Turing: dos mitos, un dogma. Revista de filosofía, 37-53.

Guerrero Arévalo, W. D. (2021). *Los alcances de la inteligencia artificial (IA) y su responsabilidad frente al derecho y ética*. Recuperado de: https://hdl.handle.net/10901/20572.

Gunderson, K. (1964). The imitation game. *Mind*, *73*(290), 234-245.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, *61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil transactions on benchmarks, standards and evaluations*, *2*(4), 100089. https://doi.org/10.1016/j.tbench.2023.100089

Hawking, S. (2016). This is the most dangerous time for our planet. *The Guardian*, *1*, 14.

Hayes, P., & Ford, K. (1995, August). Turing test considered harmful. In *IJCAI (1)* (pp. 972-977).

Hofstadter, D. R. (1979). Gödel. *Escher, Bach: an etemal golden braid, New York*.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, *37*(1), 81-88. DOI: 10.1080/02763869.2018.1404391

Hu, K. (2023, February 3). ChatGPT sets record for fastest-growing user base, analyst note. *Reuters*. Recuperado de https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Hueso, L. C. (2019). Ética en el diseño para el desarrollo de una inteligencia artificial, robótica y big data confiables y su utilidad desde el derecho. *Revista catalana de dret públic*, *58*, 43.

Hueso, L. C. (2019). Riesgos e impactos del Big Data, la inteligencia artificial y la robótica: enfoques, modelos y principios de la respuesta del derecho. *Revista general de Derecho administrativo*, (50), 1-37.

Jonas, H. (1995). El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica. Barcelona: *Herder*.

Jonas, H. (2014). *El principio de responsabilidad: ensayo de una ética para la civilización tecnológica*. Herder Editorial.

Jones, J. L. (2006). Robots at the tipping point: the road to iRobot Roomba. *IEEE Robotics & Automation Magazine*, *13*(1), 76-78.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.

Kant, I. (1946). *Fundamentación de la metafísica de las costumbres* (No. 648). Espasa-Calpe.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Kramer, S. (2023, April 11). CAIDP Files Complaint with FTC Against OpenAI's GPT-4 for Violating Consumer Protection Rules. *Futurum Research*. Recuperado de https://futurumresearch.com/research-notes/caidp-files-complaint-with-ftc-against-openais-gpt-4-for-violating-consumer-protection-rules/

Kurzweil, R. (2015). *La Singularidad está cerca: Cuando los humanos transcendamos la biología*. Lola books.

Lin, P., Abney, K., & Bekey, G. A. (Eds.). (2014). *Robot ethics: the ethical and social implications of robotics*. MIT press.

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, *74*(5), 570-581. https://doi.org/10.1002/asi.24750

MacIntyre, A. (2017). *Ética en los conflictos de la modernidad: sobre el deseo, el razonamiento práctico y la narrativa*. Ediciones Rialp.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). Mckinsey Global Institute a Future That Works: Automation, Employment, and Productivity. *McKinsey Global Institute Executive Summary*.

Markoff, J. (2011, February 16). Computer Wins on 'Jeopardy!': Trivial, It's Not. *The New York Times*. https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html

Mattas, P. S. (2023). ChatGPT: A Study of AI Language Processing and its Implications. *Journal homepage: www. ijrpr. com ISSN*, *2582*, 7421.

McCallum, S. (2023, April 1). ChatGPT banned in Italy over privacy concerns. *BBC*. Recuperado de https://www.bbc.com/news/technology-65139406

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

Minsky, M., & Papert, S. A. (2017). *Perceptrons, Reissue of the 1988 Expanded Edition with a new foreword by Léon Bottou: An Introduction to Computational Geometry*. MIT press.

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Penguin UK.

Moor, J. (1976). An analysis of the Turing test. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *30*(4), 249-257.

Moor, J. (2006). The Dartmouth College artificial intelligence conference: The next fifty years. *Ai Magazine*, *27*(4), 87-87.

Nieves, J. M. (2014, Junio 9). Un ordenador pasa por primera vez el test de Turing y convence a los jueces de que es humano. *ABC Ciencia*. https://www.abc.es/ciencia/20140609/abci-superordenador-supera-primera-test-201406091139.html

Nilsson, N. J. (Ed.). (1984). Shakey the robot.

OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI Blog*. Recuperado de https://openai.com/blog/chatgpt

Parlamento Europeo. (2017). Normas de Derecho civil sobre robótica., (pág. 2015/2103). Estrasburgo.

Parlamento Europeo. (2020). ¿Qué es la inteligencia artificial y cómo se usa? Recuperado de https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa

Parlamento Europeo. (2021). Artificial Intelligence Act. (2021/0106(COD)). https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en

Parlamento Europeo. (2023). Ley de IA de la UE: primera normativa sobre inteligencia artificial. https://www.europarl.europa.eu/news/es/headlines/society/20230601STO93804/

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*.

Poole, D. L., & Mackworth, A. K. (2010). *Artificial Intelligence: foundations of computational agents*. Cambridge University Press.

Purdy, M., & Daugherty, P. (2016). Inteligencia artificial, el futuro del crecimiento. *Accenture Institute for High Performance*.

Rahimi, F., & Abadi, A. T. B. (2023). ChatGPT and publication ethics. *Archives of medical research*, *54*(3), 272-274.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (No. JRC107466). Joint Research Centre (Seville site).

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Russell, S. J. (2010). *Artificial intelligence a modern approach*. Pearson Education, Inc..

Schwab, K. (2016). *La cuarta revolución industrial*. Debate.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, *3*(3), 417-424.

Sein, J. L. G. (2019). Innovaciones tecnológicas, inteligencia artificial y derechos humanos en el trabajo. *Documentación Laboral*, (117), 57-72.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ... & Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. Recuperado de: https://arxiv.org/abs/1712.01815

Terrones Rodríguez, A. (2018). *Inteligencia artificial y ética de la responsabilidad.* Universidad Pedagógica y Tecnológica de Colombia. https://doi.org/10.19053/01235095.v4.n22.2018.8311

Toews, R. (2021, October 3). Alphafold is the most important achievement in AI ever. *Forbes*. Recuperado de https://www.forbes.com/sites/robtoews/2021/10/03/alphafold-is-the-most-important-achievement-in-ai-ever/?sh=3c2e52686e0a

Toh, M & Seo, Y. (2023, June 9). OpenAI CEO calls for global cooperation to regulate AI. *CNN*. Recuperado de https://edition.cnn.com/2023/06/09/tech/korea-altman-chatgpt-ai-regulation-intl-hnk/index.html

Turing, A. M. (2009). *Computing machinery and intelligence*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_3

UNESCO. (2023, Marzo 30). Inteligencia Artificial: la UNESCO pide a los gobiernos que apliquen sin demora el Marco Ético Mundial. *UNESCO*. Recuperado de https://www.unesco.org/es/articles/inteligencia-artificial-la-unesco-pide-los-gobiernos-que-apliquen-sin-demora-el-marco-etico-mundial

Velasco, J. A. M. (2002). Inteligencia Artificial y conciencia. *Departamento de Matemáticas de la UAH*. Recuperado de https://frasca.web.uah.es/inteligencia-artificial.pdf

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1-37.

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36-45.

World Economic Forum. (2023). Future of Jobs Report 2023: Up to a Quarter of Jobs Expected to Change in Next Five Years. Recuperado de https://www.weforum.org/press/2023/04/future-of-jobs-report-2023-up-to-a-quarter-of-jobs-expected-to-change-in-next-five-years/

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q. L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, *10*(5), 1122-1136. https://doi.org/10.1109/JAS.2023.123618

Yoav Mintz & Ronit Brodie (2019) Introduction to artificial intelligence in medicine, Minimally Invasive Therapy & Allied Technologies, 28:2, 73-81, DOI: 10.1080/13645706.2019.1575882

Zhou, Z. H. (2021). *Machine learning*. Springer Nature.

Zuboff, S. (2023). The age of surveillance capitalism. In *Social Theory Re-Wired* (pp. 203-213). Routledge.