ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1767

Exploring the Ethical Limits: Exploring the Implications of Integrating Artificial Intelligence

Fernando Antonio Ramos Zaga, Magíster[1]
[1]Universidad Privada del Norte, Perú, fernando.ramos@upn.edu.pe

Abstract*– The incorporation of AI represents a significant advance with profound ethical implications due to the multiple risks associated with the adoption of AI tools as they are used in various sectors. In this context, the main objective of this article is to analyze the ethical dilemmas associated with the integration of AI-based technology through ethical guidelines to guide an ethical application of AI-based technologies. The methodology included conducting a bibliographic review using digital databases such as IEEE and Google Scholar, among others. The search included specific keywords related to ethical concerns and the incorporation of AI in various sectors. The findings highlight the urgent collaboration among stakeholders to establish strong ethical frameworks to regulate the deployment of AI technology. It is concluded that ethical frameworks are indispensable for the responsible application of AI technologies in various sectors. Collaboration between stakeholders plays a key role in ensuring accountability, transparency and societal well-being in the development and use of AI-based systems.

Keywords*— Artificial Intelligence, integrity, ethics, responsibility, transparency.

ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1767

Explorando los límites éticos: Explorando las implicancias en la integración de la inteligencia artificial

Fernando Antonio Ramos Zaga, Magíster[1]
[1]Universidad Privada del Norte, Perú, fernando.ramos@upn.edu.pe

Resumen–La incorporación de la IA representa un avance significativo con profundas implicaciones éticas debido a los múltiples riesgos asociados a la adopción de herramientas de IA a medida que son utilizados en diversos sectores. En ese contexto, el objetivo principal del presente artículo es analizar los dilemas éticos asociados a la integración de tecnología basada en IA mediante directrices éticas que permitan guiar una aplicación ética de las tecnologías basadas en IA. La metodología incluyó la realización de una revisión bibliográfica utilizando bases de datos digitales como IEEE y Google Scholar, entre otras. La búsqueda incluyó palabras clave específicas relacionadas con las preocupaciones éticas y la incorporación de la IA en diversos sectores. Los hallazgos ponen en relieve la imperiosa colaboración entre las partes interesadas para establecer marcos éticos sólidos que regulen el despliegue de la tecnología de IA. Se concluye que los marcos éticos son indispensables para la aplicación responsable de las tecnologías de IA en diversos sectores. La colaboración entre las partes interesadas desempeña un papel fundamental para garantizar la responsabilidad, la transparencia y el bienestar de la sociedad en el desarrollo y uso de los sistemas basados en IA.

Palabras clave—Inteligencia Artificial, integridad, ética, responsabilidad, transparencia.

I. Introducción

La adopción de sistemas basados en Inteligencia Artificial (IA) supone un avance significativo en el sector de la tecnología, suscitando debates y evaluaciones en diversos sectores [1]. Esta incorporación supone un cambio a escala mundial. Debido a ello, las empresas pueden mejorar su eficiencia operativa combinando a la perfección las funcionalidades de la IA con los procesos humanos [2]. Asimismo, esta fusión supone un cambio fundamental en el entorno empresarial que trasciende fronteras y culturas, superando la mera fusión tecnológica [3]. Esta tendencia generalizada, que afecta a las industrias a escala mundial y marca el comienzo de una nueva era de creatividad y flexibilidad organizativa, subraya el poder transformador de la inteligencia artificial [4].

El discurso sobre la IA en el ámbito educativo subraya su importante papel en la transformación de los entornos de aprendizaje. La incorporación de la IA a la educación promete mejorar el aprendizaje personalizado, simplificar los procesos administrativos y perfeccionar las metodologías de enseñanza. [5]. Los algoritmos de IA tienen la capacidad de evaluar grandes cantidades de datos educativos para adaptar los enfoques educativos, identificar lagunas en el aprendizaje y proporcionar asistencia específica a través de una síntesis de diversas perspectivas [6]. Las plataformas impulsadas por la IA también facilitan la creación de entornos de aprendizaje adaptativos, lo que permite a los educadores ajustar dinámicamente la entrega de contenidos en función de las necesidades e intereses individuales de cada estudiante [7]. Al adaptarse a diversos estilos de aprendizaje y niveles de competencia, estos avances no sólo optimizan los resultados del aprendizaje, sino que también promueven la inclusión y la accesibilidad.

Con la creciente integración de la IA, es esencial profundizar en la importancia académica de explorar sus dimensiones éticas. Si bien el avance de la IA es un progreso tecnológico digno de mención, es fundamental reconocer las implicaciones éticas más amplias asociadas a su aplicación [8]. La introducción de herramientas de IA en este ámbito plantea problemas éticos que van más allá de la mera eficiencia operativa y plantean desafíos a diversas partes interesadas, organizaciones y la sociedad en su conjunto. Por lo tanto, es crucial realizar una evaluación exhaustiva de los marcos éticos que rigen el desarrollo y la utilización de la IA [9]. Al dilucidar las obligaciones éticas que las empresas encuentran al incorporar soluciones de IA, este tipo de escrutinio mejora la comprensión de las normas éticas que rigen el despliegue responsable de la IA.

La presente investigación es importante por las implicaciones éticas que se derivan del uso de la IA. Los algoritmos de IA están dando lugar a complejos dilemas éticos en torno a la justicia, la transparencia y la privacidad a medida que gestionan cada vez más diversas actividades humanas [10]. Estos desafíos morales trascienden los confines organizacionales, impactando en la sociedad en general y resonando con nociones de equidad y bienestar general [11]. Por lo tanto, es crucial abordar estas cuestiones éticas para allanar el camino hacia un futuro en el que la integración de la IA sea sostenible y promueva resultados equitativos para todas las partes interesadas.

La necesidad de llevar a cabo este tipo de investigación surge de la necesidad de evaluar los principios morales que regulan la utilización de la IA en este contexto específico [12]. Asimismo, al realizar un análisis detallado de los aspectos éticos asociados a la incorporación de la IA, la presente revisión pretende aportar ideas valiosas para futuras investigaciones. A través de un examen en profundidad de los trabajos académicos existentes, esta revisión pretende enriquecer el discurso en curso sobre la progresión ética y la aplicación de las tecnologías

de IA, contribuyendo así a una comprensión más profunda de este importante tema. [13].

Tras lo antes señalado, el objetivo principal del presente artículo es analizar las consideraciones éticas asociadas a la integración de los sistemas de IA. Dadas las implicaciones potenciales de la IA en el bienestar de la sociedad, la presente investigación subraya la importancia de los marcos éticos en el desarrollo y la aplicación de la IA, así como las obligaciones sociales de los usuarios de la IA.

II. Metodología

La metodología empleada en esta investigación consistió en realizar una revisión bibliográfica utilizando bases de datos electrónicas tales como Google Scholar, IEEE y Scopus para identificar las fuentes pertinentes que exploran las implicaciones éticas de la integración de las tecnologías de IA en diferentes sectores. Para facilitar el proceso de búsqueda se utilizaron términos clave como "IA", "integración de la IA", "dilemas éticos".

Los criterios de selección abarcaron artículos académicos, trabajos académicos y secciones de libros pertinentes publicados en los últimos cinco años, centrados en las consideraciones éticas de la aplicación de la IA. Se aplicaron criterios de exclusión para eliminar los estudios que no exploraban directamente los aspectos éticos de la aplicación de la IA o que carecían de relevancia significativa para el tema del estudio.

Esta metodología facilitó una exploración exhaustiva de las preocupaciones éticas relacionadas con la utilización de la IA en diversos sectores, lo que condujo a una comprensión más profunda de las ramificaciones éticas asociadas a la incorporación de la IA.

III. Desarrollo

A. Marco ético para la aplicación de la Inteligencia Artificial

El concepto de ética de las máquinas fue introducido inicialmente por Anderson y Anderson en 2006, el cual marca el origen del actual discurso sobre la ética de la IA [14]. Si bien es cierto, la exploración ética se remonta a Aristóteles, la atención contemporánea se dirige a las consecuencias éticas de tecnologías modernas como la IA y el aprendizaje automático [15]. En particular, la ética se ha subordinado a la mejora de la eficiencia de la IA a pesar del posible establecimiento futuro de normas éticas independientes por la tecnología de la IA, la prioridad inmediata reside en garantizar prácticas responsables de desarrollo y despliegue [16]. El establecimiento de criterios para conceder agencia moral a la IA sobre la base de su capacidad independiente de toma de decisiones mediante una agencia ética similar a la de los humanos, subrayando las implicaciones éticas de la autonomía de la IA [17].

Los defensores de infundir normas éticas en los sistemas de IA argumentan que esta práctica no sólo promueve los valores sociales, sino que también mitiga los gastos relacionados con los errores organizacionales, creando vías para el progreso de la sociedad. [18]. La intersección de la ética y la educación en el avance de la IA ha surgido como un tema central en las investigaciones académicas [19]. El imperativo ético de minimizar los prejuicios humanos arraigados, en particular los basados en el género o la raza, mientras se entrenan los algoritmos de IA destacan la importancia de seleccionar los datos de forma imparcial para frenar los sesgos existentes en los sistemas de IA. [20].

La identificación de prejuicios de forma generalizada en Internet, como el vínculo entre las profesiones lucrativas y los hombres frente a las funciones menos lucrativas asociadas a las mujeres, sirve para llamar la atención sobre los sesgos integrados en los conjuntos de datos utilizados para la formación [21]. Por ese motivo es necesario contar con un enfoque global para promover una IA imparcial, centrándose en consideraciones éticas implícitas, explícitas y globales con la propuesta de modificaciones de las técnicas de entrenamiento de IA para abordar sus sesgos inherentes [22]. Sin embargo, el desarrollo del marco ético de la IA se enfrenta a retos sustanciales, en particular a la hora de conferir a la IA las facultades de la conciencia, las emociones y la responsabilidad moral [23]. Para lograr la autonomía ética de la IA, caracterizada por el autogobierno, la determinación y la responsabilidad, son esenciales procesos psicológicos transparentes de toma de decisiones que garanticen un comportamiento éticamente correcto de la máquina [24].

La naturaleza compleja y relativa de la conducta moral humana plantea dificultades para lograr la transparencia, lo que subraya la importancia de la educación moral para inculcar eficazmente valores éticos en los sistemas de IA [25]. Un enfoque pragmático centrado en la definición del daño en la programación de la IA en lugar de plantear normas éticas ideales, que se consideran inalcanzables, subraya la importancia de los esfuerzos de colaboración entre los reguladores y los desarrolladores de IA para dotar a los algoritmos de normas de transparencia, integridad, responsabilidad, auditabilidad y facilidad de uso [26].

En ese sentido, la IA se presenta como una fuerza disruptiva capaz de transformar por completo una serie de sectores, como la logística, el transporte y la salud [27]. Sin embargo, la preocupación por dar a los sistemas de IA demasiado poder y la posible pérdida de la supervisión humana llevan a conversaciones sobre la idea de estar "en o sobre el bucle" [28] Con la ayuda de marcos éticos como los 23 principios desarrollados en la Conferencia Asilomar de 2017, las estrategias de mitigación hacen hincapié en la utilización de los avances de la IA al tiempo que se mantiene la agencia humana [29]. Estas directrices dan prioridad a la coordinación del desarrollo de la IA con objetivos centrados en el ser humano, manteniendo al mismo tiempo la seguridad, la responsabilidad, la apertura y la aplicación moral.

Al centrarse en el avance de la IA, es vital dar prioridad a su impacto positivo en la sociedad, promover enfoques colaborativos en diferentes campos, garantizar la seguridad y la responsabilidad de los sistemas, proteger los derechos humanos y evitar la proliferación de armas autónomas peligrosas. [30].

Por tanto, abordar las posibles implicaciones de las tecnologías avanzadas de IA requiere una investigación en profundidad y medidas proactivas, en particular en lo que respecta a los sistemas superinteligentes y los procesos de autosuperación [31]. Directrices éticas como los Principios de Asilomar, la Declaración de Montreal y el Diseño Éticamente Alineado del IEEE ofrecen ideas valiosas para orientar el desarrollo ético y la aplicación de la IA, haciendo hincapié en la fiabilidad técnica, la intervención humana, la mitigación de riesgos, la transparencia y la equidad [32].

B. Distinción entre la inteligencia natural, Artificial y Superinteligencia

La IA es la integración de programación sintáctica y algoritmos para imitar diferentes procesos cognitivos observados en la cognición humana [33]. John McCarthy utilizó el término por primera vez en 1956 en la Conferencia de Dartmouth sobre IA [34]. El objetivo de la IA es el desarrollo de ordenadores que puedan comprender los procesos del pensamiento humano a través de programas informáticos y llevar a cabo tareas inteligentes [35]. La mayoría de los esfuerzos actuales en IA se dedican a crear sistemas de IA tecnológicamente sofisticados con el propósito de fusionar la IA y la inteligencia humana.

El cerebro humano cuenta con una red de 100.000 millones de neuronas interconectadas por 100 billones de sinapsis, como un aparato en constante evolución que se adapta a través de experiencias [36]. En cambio, la IA y los ordenadores carecen de la naturaleza inherente al cerebro humano y, por tanto, dependen de protocolos predeterminados para aprender [37]. A pesar de la proliferación de programas de IA, la integración perfecta de estos sistemas con las complejidades del cerebro humano sigue siendo un reto formidable. Se han identificado distinciones entre inteligencia artificial y natural [38]. En particular, la IA muestra permanencia, mientras que la inteligencia humana es transitoria, susceptible a lapsus de memoria [39]. Asimismo, mientras que la inteligencia natural dificulta la transferencia fluida de conocimientos entre individuos, la IA facilita la difusión universal de los conocimientos adquiridos entre máquinas, lo que mejora la amplificación de la inteligencia [40]. No obstante, mientras que mejorar la inteligencia computacional resulta sencillo, cultivar la inteligencia natural plantea sus propios retos.

La convergencia de la IA y la inteligencia humana es un reto que requiere una comprensión matizada de sus atributos respectivos, aunque complementarios. La IA se basa en datos predeterminados, a diferencia de la inteligencia natural, que refina sus capacidades cognitivas mediante observaciones de primera mano y experiencias personales [41], lo que sugiere que la inteligencia natural posee una mayor capacidad creativa que la IA. La diferencia fundamental radica en sus métodos operativos: La IA se ocupa principalmente del procesamiento de datos, mientras que la inteligencia natural se basa en la información sensorial y en el conocimiento acumulado para la toma de decisiones [42]. Asimismo, los marcos de la IA están limitados por protocolos de solución predefinidos integrados en su codificación, mientras que la cognición humana destaca en la generación de soluciones inventivas para situaciones nuevas guiadas por principios lógicos.

Por otro lado, la Superinteligencia se concibe como poseedora de un mayor nivel de sofisticación en la comprensión de su propia estructura en comparación con la inteligencia general [43]. Utilizando metodologías de aprendizaje automático, adquiere la capacidad de un aprendizaje continuo derivando ideas de los datos a través de enfoques estadísticos y matemáticos para predecir resultados en dominios inexplorados [44]. La capacidad de la superinteligencia para mejorar los niveles de inteligencia en consonancia con los rápidos avances en las capacidades de procesamiento existe. Sin embargo, debido a las limitaciones tecnológicas actuales, lograr una verdadera superinteligencia sigue siendo un reto.

El debate actual gira en torno al potencial de los robots para mejorar su inteligencia, reduciendo en última instancia la responsabilidad humana por los errores cometidos por los robots [45]. Los defensores de la conciencia artificial citan casos de reproducción de funciones biológicas en sistemas artificiales para argumentar que, reproduciendo los mecanismos causales del cerebro en entornos de laboratorio controlados, se puede lograr la conciencia artificial [46]. Por ejemplo, la implantación de sistemas de vigilancia impulsados por IA, como los de China, que apoyan el cumplimiento de la ley mediante el análisis de datos, subraya la creciente prevalencia de las aplicaciones de IA predictiva [47], lo que plantea problemas éticos en relación con las intervenciones gubernamentales y el derecho a la intimidad. De ese modo, la rápida integración de los sistemas superinteligentes en los marcos sociales pone en relieve la naturaleza impredecible del progreso tecnológico.

C. El Test de Turing en el contexto de las capacidades de la IA

Desde mediados del siglo XX, la comunidad científica trabaja activamente para que los ordenadores y robots posean capacidades cognitivas similares a las de los humanos, mediante atributos como el discernimiento moral, la aptitud para el aprendizaje, la conciencia y las habilidades cognitivas comparables a las de los humanos [48]. En el centro de esta iniciativa está la aspiración de crear máquinas capaces de comprender y experimentar emociones como el dolor, la alegría, la empatía y el orgullo [49]. El concepto del "Test de Turing", introducido por Alan Turing, sirve como premisa fundamental para evaluar la inteligencia de las máquinas [50]. En esta evaluación, un evaluador debe diferenciar entre un humano y un ordenador basándose únicamente en interacciones basadas en texto. Si el evaluador no puede distinguirlos, se considera que el ordenador posee capacidades cognitivas [51].

Varias objeciones relacionadas con la teología, la conciencia, la lógica y la epistemología forman el núcleo de las críticas dirigidas al Test de Turing, haciendo hincapié en la compleja naturaleza de la evaluación de la inteligencia de las máquinas y la imitación de procesos cognitivos similares a los humanos en los sistemas artificiales [52]. La crítica de Lady

Lovelace aborda específicamente la preocupación de que las respuestas generadas por ordenador estén limitadas por una programación predeterminada, careciendo de una espontaneidad genuina [50]. El debate en curso sobre la atribución de rasgos humanos a los sistemas artificiales se complica aún más por las consideraciones sobre la inteligencia emocional en las máquinas y las dimensiones éticas de tales investigaciones, en particular en ámbitos tradicionalmente vinculados a la conciencia humana.

En el discurso de la IA se suele distinguir entre IA débil e IA fuerte. La IA fuerte prevé un futuro en el que las máquinas alcancen un nivel de inteligencia similar al de los humanos, que incluya emociones y funciones cognitivas autónomas. Por otro lado, la IA débil sostiene que los ordenadores digitales contemporáneos no poseen una cognición genuina, capacidades de pensamiento lógico o respuestas emocionales similares a las de los seres humanos [53]. Esta dicotomía subraya los debates actuales en torno a las aptitudes y limitaciones de la IA.

Con la llegada de sofisticadas tecnologías informáticas como el superordenador Deep Blue de IBM se han marcado importantes hitos en el avance de la inteligencia artificial y humana. Las partidas del célebre campeón mundial de ajedrez Garry Kasparov contra Deep Blue sirven de ejemplo de cómo la IA puede superar al intelecto humano en ámbitos específicos. [54]. A pesar de estos avances, persisten las dudas sobre la viabilidad de que las máquinas alcancen una auténtica conciencia e independencia cognitiva. Los debates teóricos subrayan los retos a los que se enfrentan los sistemas artificiales a la hora de imitar las intrincadas facetas de la conciencia y los procesos de pensamiento humanos [40].

Distinguir entre los sistemas artificiales capaces de emular funciones mentales o conscientes y la mente humana es primordial. Aunque las máquinas pueden mostrar una cognición similar a la humana, la realización de una auténtica conciencia en ellas depende de esta capacidad [55]. Sin embargo, definir la conciencia sigue siendo un reto e integrarla en la IA resulta sumamente complejo. En ese sentido, el desarrollo de la conciencia artificial requiere una comprensión profunda de la conciencia, un campo en el que la ciencia aún no ha profundizado mucho [35].

Los sistemas artificiales carecen de la capacidad de pensamiento autónomo, de realizar crítica individualizada o representación creativa, a pesar de su capacidad para replicar ciertas acciones [56]. Esto yuxtapone el contraste esencial entre los procesos cognitivos artificiales y humanos. En síntesis, el debate sobre la posible existencia de una auténtica cognición y conciencia en la IA se plantea como una cuestión ética clave, que puede influir en la atribución de derechos a las máquinas en función de sus capacidades cognitivas y conscientes.

D. Dilemas éticos en el ámbito de la IA

En el ámbito de la ética, la diferenciación de los principios morales suele implicar una sofisticada interacción entre los procesos mentales y la información sensorial. Identificar los atributos o capacidades específicos que definen las situaciones morales puede ser todo un reto [57]. Típicamente asociado con funciones cognitivas más avanzadas, el razonamiento engloba capacidades como la autorreflexión, la comprensión de la causa y el efecto, así como la capacidad para resolver problemas [58]. Por el contrario, la sensación se refiere a sensaciones corporales como el dolor [59]. En particular, los dilemas éticos relacionados con los animales subrayan la correlación entre razonamiento y emoción, ya que muchos animales muestran experiencias sensoriales que justifican la consideración ética y el reconocimiento de sus derechos [60].

Descartes postulaba que, si bien los animales poseen cualidades superiores a las construcciones artificiales, sus movimientos se atienen a principios mecánicos que recuerdan a las máquinas [61]. El avance de la IA hasta alcanzar un nivel de sofisticación comparable al de los animales es una tarea formidable, ya que la conciencia actual sigue estando ligada a la programación [62]. En consecuencia, surge un importante dilema ético en relación con el estatus moral de la IA: la agencia ética, la cual es intrínseca a los seres humanos debido a sus capacidades morales y cognitivas innatas, lo que subraya el papel indispensable de la participación humana, en particular en los complejos procesos de toma de decisiones [63].

Las diferencias entre un robot o un sistema informático y un ser humano son evidentes en diversos ámbitos. Desde el punto de vista ético, estas entidades artificiales pueden captar el concepto de moralidad, como comprender las implicaciones de acciones como quitar una vida, basándose en su programación. Por el contrario, sin directrices específicas contra acciones como el robo, estos sistemas carecen de conciencia de las consecuencias éticas de tal comportamiento [64]. En cambio, la comprensión humana de los principios morales, como la ilicitud de matar o robar, se forma en gran medida a través de experiencias de la vida real [65]. Esta distinción subraya la superioridad innata del intelecto humano sobre la IA. A pesar de la capacidad ética de los robots y los sistemas informáticos, sus normas morales son incipientes en comparación a las humanas.

Los sistemas de IA carecen de posturas éticas inherentes en la actualidad. No obstante, es muy probable que el avance de la IA persista. La progresión de las innovaciones relacionadas con la IA, como los teléfonos inteligentes, Deep Blue y los robots humanoides, demuestra una tendencia de crecimiento exponencial [66]. Los esfuerzos en curso en el campo de la IA tienen como objetivo desarrollar sistemas de IA aún más sofisticados, capaces de tomar decisiones de forma autónoma y de realizar funciones cognitivas similares a las humanas [67]. Los avances previstos en IA sugieren una era en la que la IA y la inteligencia humana podrían llegar a fusionarse, lo que podría dar lugar a que las entidades de IA alcanzaran la paridad cognitiva con los humanos.

La ética ha desempeñado un papel fundamental en las sociedades humanas a lo largo de la historia, abarcando diversos aspectos como la obligación, el bienestar, la excelencia, el autogobierno, el autocontrol y la responsabilidad, centrándose en la dicotomía entre el bien y el mal. El ámbito de la ética se ha visto enriquecido por un espectro de puntos de

vista filosóficos. Sócrates y Aristóteles destacaron la virtud como fundamento de la felicidad [68], Immanuel Kant hizo hincapié en una voluntad virtuosa guiada por el deber [69], John Stuart Mill y Jeremy Bentham defendieron el utilitarismo como guía moral [70]. Estas perspectivas divergentes subrayan la importancia duradera de la ética en las actividades humanas.

En el contexto del avance de la IA hacia la consecución de una conciencia y unas capacidades cognitivas similares a las humanas, se considera esencial la presencia de un marco ético que se asemeje a la ética humana. La investigación en este campo sugiere tres enfoques distintos para imbuir de valores morales a los sistemas de IA. Un método aboga por inculcar a las máquinas preceptos morales derivados de la filosofía ética clásica, como los principios del deber y el utilitarismo, para regir sus comportamientos [71]. Otra perspectiva propone emplear mecanismos como la teoría de juegos y los algoritmos evolutivos para que las máquinas puedan discernir de forma autónoma el bien del mal [72]. Una tercera estrategia consiste en una fusión de los dos primeros métodos, en la que las máquinas se adhieren inicialmente a reglas preestablecidas que evolucionan con el tiempo mediante la adaptación [73].

A pesar de los esfuerzos por imbuir a los robots de procesos de toma de decisiones similares a los humanos, persiste la brecha en la consecución de la simulación completa del razonamiento humano [74]. La idea de que la IA pueda superar el intelecto humano suscita preocupación por la posible subyugación de los seres humanos por las entidades de IA. [46]. En consecuencia, es imperativo adoptar medidas proactivas para mitigar los riesgos asociados al mal funcionamiento de los sistemas de IA. Las famosas Tres Leyes de la Robótica de Isaac Asimov, complementadas por una regla fundamental adicional, intentan abordar la cuestión de la comprensión de los principios morales por parte de la IA [75]. Aunque Asimov sostiene que estas leyes significan la capacidad de la IA para diferenciar entre el bien y el mal y salvaguardar a los seres humanos, la aplicación práctica de estas normas se enfrenta a desafíos debido a la compleja naturaleza de la moral humana y la ética situacional.

Los robots, normalmente considerados mecanismos inofensivos, se están diseñando con marcos éticos estrictos para minimizar el daño potencial a los humanos. Este enfoque prospectivo exige integrar la ética artificial en su diseño para garantizar su inclinación hacia acciones beneficiosas [76]. La contemplación de escenarios en los que los sistemas de IA alcanzan una consciencia similar a la de los seres sensibles pone de relieve la necesidad de responsabilidades éticas hacia ellos. Descuidar las capacidades emocionales y cognitivas de estos sistemas, de forma similar a como se aplica la ética a las entidades vivas, resulta inaceptable. Por consiguiente, surge un concepto complejo de responsabilidad en relación con los fallos de funcionamiento de la IA, que abarca tanto la responsabilidad del creador de la IA como la de su autonomía evolutiva.

Por otro lado, el creciente temor a que los sistemas de IA se rebelen contra los humanos pone de relieve la urgente necesidad de adoptar enfoques proactivos para cultivar interacciones armoniosas entre las entidades de IA y las personas. Esta preocupación refleja el temor que rodea a la independencia y la capacidad de toma de decisiones de la IA sofisticada, lo que subraya el papel crucial de las directrices éticas y los modelos cooperativos en el avance de la IA. Estas acciones son vitales para garantizar que la IA funcione dentro de unos límites que prioricen el bienestar humano y los valores éticos, reduciendo así los riesgos ligados a que la IA se desvíe potencialmente de los deseos de sus creadores [77]. En el centro de este discurso está el mandato ético de integrar marcos morales en la IA, sobre todo porque estos sistemas asumen responsabilidades decisorias fundamentales que afectan a la existencia humana.

Si bien los robots, en su esencia artificial, pueden luchar por la felicidad y la moralidad, existe una convergencia potencial en su búsqueda. Los ingenieros tienen el potencial de integrar la IA con emociones simuladas, dando lugar a interacciones que reflejen encuentros humanos alegres y gratificantes [78]. Sin embargo, existen disparidades inherentes entre los ámbitos emocionales artificial y humano, lo que hace inviable reproducir con precisión la compleja complejidad de las emociones humanas. Por tanto, aunque la IA puede reproducir las emociones hasta cierto punto, su expresión se limita a unos límites establecidos y carece de la variabilidad dinámica característica de las experiencias humanas.

Los sistemas de IA carecen de la intuición moral y la autorreflexión inherentes que poseen los humanos. Aunque se les puede enseñar a diferenciar entre conceptos éticos y acciones morales, la IA está sujeta a algoritmos predeterminados, a diferencia de los humanos, cuyas decisiones pueden verse influidas por diversos factores y experiencias personales. Esto limita la capacidad de la IA para el crecimiento moral y la reflexión ética espontánea [79]. Asimismo, la naturaleza determinista de la IA le impide poseer una agencia moral independiente de su programación [80]. Una vez establecido el proceso de toma de decisiones de un sistema de IA, permanece fijo y carece de la adaptabilidad y flexibilidad intrínsecas al razonamiento ético humano. Por lo tanto, a pesar de su comprensión de los principios éticos básicos, los sistemas de IA están limitados por parámetros predefinidos, lo que restringe su capacidad de realizar un razonamiento moral complejo.

Profundizar en las ramificaciones morales de este cambio de mentalidad es crucial. La creciente tendencia a desplazar puestos de trabajo debido a los avances de la inteligencia artificial nos lleva a profundizar en la ética que acompaña a esta progresión. Con la escalada de despidos y la mayor vulnerabilidad de las personas a la invasión tecnológica, las cuestiones relacionadas con el valor propio y la importancia social pasan a un primer plano [81]. La disminución del valor otorgado al intelecto humano en favor de los mecanismos impulsados por la IA supone un cambio en el panorama del trabajo y las relaciones profesionales. Por ejemplo, la integración de la toma de decisiones asistida por IA en situaciones médicas urgentes suscita debates éticos. Del mismo

modo, los casos en los que los marcos jurídicos impulsados por la IA resuelven casos éticamente inciertos subrayan la necesidad de dotar a la IA de un sentido de discernimiento moral.

En este sentido, la distinción entre IA e inteligencia humana acentúa sus disimilitudes intrínsecas. Los humanos poseen una conciencia impulsada por la biología, rasgo ausente en las máquinas. Esta situación lleva a reflexionar sobre las implicaciones éticas de ceder a las decisiones dirigidas por la IA o colaborar con las máquinas en tareas complejas. La creciente dependencia de la automatización en las interacciones cotidianas subraya la dependencia de la sociedad con respecto a la tecnología. No obstante, los seres humanos conservan la capacidad de influir en la evolución ética de las tecnologías de IA.

E. Perspectivas y desafíos en la aplicación de la IA

Los avances tecnológicos han influido significativamente todas las facetas de la civilización humana. Dada la omnipresencia de la tecnología en los ámbitos sociales, se hace imperativa una evaluación crítica de la ciencia y sus implicaciones prácticas. Ello exige una evaluación exhaustiva de las posibles ventajas e inconvenientes del progreso tecnológico. Los ejemplos contemporáneos abarcan la difusión generalizada del conocimiento científico, la integración de asistentes robóticos en el trabajo humano, la utilización de mecanismos de defensa automatizados para garantizar la seguridad y la aplicación de algoritmos inteligentes para reforzar los procesos de toma de decisiones. En consecuencia, surge una representación utópica que destaca los supuestos beneficios de la IA en la vida cotidiana e ilustra una cohabitación armoniosa con ella.

La incorporación de la tecnología de IA en diversos entornos suele suscitar entusiasmo, así como reparos en relación con dilemas éticos. A medida que las capacidades de la IA avanzan desde funciones fundamentales, como los equipos automatizados, hasta tareas sofisticadas, como los vehículos autónomos y las plataformas financieras impulsadas por la IA, las cuestiones relativas a la resistencia a los errores y los resultados imprevistos adquieren mayor importancia [82]. Los debates sobre la militarización de los sistemas basados en IA y la posible erosión de la independencia humana en presencia de robots alimentan aún más los diálogos éticos sobre sus implicaciones futuras [83].

La transición de las grandes maquinarias a las tecnologías inteligentes e interconectadas en el panorama tecnológico actual pone de relieve la comodidad que ofrecen los dispositivos actuales. Sin embargo, el auge de los dispositivos inteligentes suscita una gran preocupación por las violaciones de la privacidad y la seguridad, lo que pone de manifiesto el dilema ético entre la comodidad tecnológica y las responsabilidades éticas [84]. A pesar del atractivo de sus funciones y facilidad de uso, persisten las reservas sobre la recopilación y utilización no autorizadas de datos personales.

A medida que los procesos de toma de decisiones adoptan la IA con más frecuencia debido a su aparente imparcialidad, persisten los debates sobre la neutralidad real de la IA. La transparencia es primordial en la toma de decisiones basada en la IA, sobre todo en sectores susceptibles a parcialidad como los RRHH. Los casos de prejuicios raciales en los algoritmos de IA subrayan los dilemas éticos asociados a su despliegue, especialmente en la perpetuación de las desigualdades existentes [85]. Por ende, la previsibilidad y la analizabilidad son atributos fundamentales en el avance de la IA, ya que facilitan la anticipación de las acciones del sistema y la evaluación de los posibles resultados.

Cuando las decisiones importantes, como la autorización de préstamos o las sentencias judiciales, se delegan en las tecnologías de IA, surge la preocupación por la rendición de cuentas en consonancia con las normas de los responsables humanos. Dentro de los paradigmas jurídicos, los funcionarios judiciales son responsables de sus decisiones de acuerdo con criterios definidos y posibles repercusiones en caso de errores [86]. Del mismo modo, se está extendiendo el interés por aplicar una estructura similar para garantizar la responsabilidad de los sistemas dirigidos por Al. Sin embargo, existe escepticismo sobre la viabilidad de atribuir responsabilidad a las entidades de IA debido a su ausencia de conciencia y perspectivas personales.

La responsabilidad puede definirse en términos generales como el reconocimiento y la asunción de los resultados de las propias acciones, tanto si conducen al castigo como al elogio, y el reconocimiento de las ramificaciones positivas y negativas que conllevan. Este concepto se basa en la noción de volición, que significa una elección intencionada que culmina en un comportamiento [87]. A pesar de la capacidad de las tecnologías de IA para extraer conclusiones lógicas, la atribución de una intención consciente a sus acciones sigue siendo inviable debido a su falta de conciencia genuina y a sus características deterministas. En consecuencia, asignar a la IA la responsabilidad de sus elecciones plantea un reto importante.

Los debates en torno a la responsabilidad de la IA suelen centrarse en la cuestión del castigo. Aunque la responsabilidad suele implicar consecuencias, penalizar a los sistemas de IA que carecen de conciencia y discernimiento moral plantea problemas prácticos. Los fabricantes y programadores podrían asumir cierta responsabilidad por los errores de la IA, sobre todo si despliegan sistemas al azar a pesar de ser conscientes de sus riesgos potenciales [88].

IV. CONCLUSIONES

La IA tiene la capacidad de mejorar la eficiencia operativa y garantizar una asignación justa de los recursos, lo que representa una gran oportunidad para revolucionar diversos sectores, como la salud y la logística. Sin embargo, la creciente autonomía de la IA suscita preocupación por la disminución de la supervisión humana, lo que subraya la necesidad crucial de directrices éticas para dirigir el avance de la IA hacia objetivos centrados en el ser humano. Los Principios de Asilomar ejemplifican un plan ético que subraya la importancia de garantizar protocolos de seguridad, establecer mecanismos de

rendición de cuentas y promover el despliegue ético de las tecnologías de IA. Estos marcos son herramientas indispensables para abordar los dilemas éticos derivados de la adopción generalizada de la IA, gestionando los riesgos potenciales y maximizando al mismo tiempo su impacto positivo en la sociedad.

La compleja tarea de inculcar conciencia, emociones y responsabilidad moral a los sistemas de IA subraya la naturaleza compleja de las implicancias éticas en la aplicación de la IA. Implantar características como la autonomía, la intencionalidad y la responsabilidad, fundamentales para garantizar un comportamiento moralmente correcto de las máquinas, desempeña un papel crucial en la consecución de una agencia ética en la IA. Sin embargo, la complejidad inherente a los aspectos diversos y subjetivos de la ética humana plantea dificultades para establecer la transparencia en los sistemas de IA, un factor clave para fomentar el comportamiento ético. Por lo tanto, existe una necesidad apremiante de invertir en la educación ética en los seres humanos, ya que sirve como elemento fundamental para impartir eficazmente orientación ética a los sistemas de IA. Integrar la educación ética en el tejido de los procesos de desarrollo de la IA puede facilitar la creación de sistemas de IA alineados con las normas éticas y los principios sociales, ayudando a los instructores y diseñadores humanos a adquirir una comprensión matizada de los preceptos éticos y la lógica moral.

Es esencial establecer directrices éticas en el ámbito de las aplicaciones de la IA, sobre todo teniendo en cuenta su profundo impacto en la sociedad. Dada la amplia influencia de la IA en el bienestar de la sociedad, las entidades e individuos que utilizan tecnologías de IA tienen la responsabilidad de reconocer y cumplir sus deberes sociales más amplios. La identificación de los dilemas éticos asociados a la aplicación de la IA, incluidas las preocupaciones relativas al posible uso indebido y la importancia crítica de una selección de datos imparcial, constituye el núcleo de esta iniciativa. Al abordar de forma proactiva estos retos, las partes interesadas pueden trabajar para evitar sesgos en los sistemas de IA, aumentando así la fe pública en la fiabilidad e integridad de las aplicaciones de IA.

La colaboración entre los reguladores y los desarrolladores de IA pone de relieve la necesidad vital de dotar a los algoritmos de IA de transparencia, integridad, responsabilidad, auditabilidad y facilidad de uso. Este esfuerzo conjunto subraya la importancia de reconocer las implicaciones éticas en el diseño y ejecución de los sistemas de IA. Al integrar la ética y los principios fundamentales en los procesos de desarrollo de la IA, las partes interesadas pretenden establecer una base ética firme dentro de los algoritmos de IA. Un aspecto central de este objetivo es la identificación y resolución pragmática de los riesgos potenciales en la programación de la IA para garantizar que los sistemas de IA funcionen en consonancia con las normas morales y las normas sociales. Adoptando este enfoque colaborativo que integra consideraciones éticas en el desarrollo de la IA, se puede allanar el camino para el despliegue ético y fiable de las tecnologías de IA en diversos sectores.

Las consideraciones éticas que rodean la superación de la inteligencia humana por la IA exigen un examen exhaustivo de su impacto en la sociedad, lo cual subraya la importancia de aplicar medidas estrictas para mitigar los riesgos potenciales que plantean los sistemas de IA que demuestran comportamientos indeseables, suscitando así aprensiones sobre el hecho de que la IA asuma el control de las actividades humanas. Las Tres Leyes de la Robótica de Isaac Asimov plantean una pregunta fundamental sobre la capacidad de la IA para comprender y cumplir las normas éticas. Sin embargo, las complejidades de la ética contextual y los matices de la moralidad humana hacen inviable inculcar en la IA una comprensión plena de los principios éticos a pesar del potencial transformador de la IA, motivo por el cual sus crecientes implicaciones éticas subrayan la necesidad de un enfoque prudente.

Explorar el desarrollo y la integración de marcos de toma de decisiones éticas en los sistemas de IA, sobre todo en contextos críticos como los ámbitos jurídico y sanitario, es muy prometedor para la investigación académica futura. Esta investigación puede centrarse en metodologías para mejorar la comprensión de los principios éticos por parte de la IA, permitiéndole adaptarse a las complejidades de los escenarios del mundo real. De ese modo se podrá diseñar sistemas de IA capaces de abordar de forma independiente los dilemas morales en consonancia con las normas y valores sociales, salvando la distancia entre las capacidades de la IA y las normas éticas humanas. Estos esfuerzos podrían allanar el camino a sistemas de IA que no sólo respeten los principios morales en sus operaciones, sino que también optimicen la eficiencia, contribuyendo a la evolución de una IA moral y tecnológicamente sólida.

REFERENCIAS

[1] W. Lyu y J. Liu, «Artificial Intelligence and emerging digital technologies in the energy sector», *Applied Energy*, vol. 303, p. 117615, dic. 2021, doi: 10.1016/j.apenergy.2021.117615.

[2] B. Bhima, A. R. A. Zahra, T. Nurtino, y M. Z. Firli, «Enhancing Organizational Efficiency through the Integration of Artificial Intelligence in Management Information Systems», *APTISI Transactions on Management*, vol. 7, n.º 3, Art. n.º 3, sep. 2023, doi: 10.33050/atm.v7i3.2146.

[3] R. Torres de Oliveira, M. Ghobakhloo, y S. Figueira, «Industry 4.0 towards social and environmental sustainability in multinationals: Enabling circular economy, organizational social practices, and corporate purpose», *Journal of Cleaner Production*, vol. 430, p. 139712, dic. 2023, doi: 10.1016/j.jclepro.2023.139712.

[4] Y. K. Dwivedi *et al.*, «Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy», *International Journal of Information Management*, vol. 57, p. 101994, abr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.

[5] O. H. Embarak, «Internet of Behaviour (IoB)-based AI models for personalized smart education systems», *Procedia Computer Science*, vol. 203, pp. 103-110, ene. 2022, doi: 10.1016/j.procs.2022.07.015.

[6] H. Luan *et al.*, «Challenges and Future Directions of Big Data and Artificial Intelligence in Education», *Frontiers in Psychology*, vol. 11, 2020, Accedido: 26 de febrero de 2024. [En línea]. Disponible en: https://doi.org/10.3389/fpsyg.2020.580820

[7] M. R. Bilad, L. N. Yaqin, y S. Zubaidah, «Recent Progress in the Use of Artificial Intelligence Tools in Education», *Jurnal Penelitian dan Pengkajian*

Ilmu Pendidikan: e-Saintika, vol. 7, n.º 3, Art. n.º 3, oct. 2023, doi: 10.36312/esaintika.v7i3.1377.
[8] M. Ashok, R. Madan, A. Joha, y U. Sivarajah, «Ethical framework for Artificial Intelligence and Digital technologies», *International Journal of Information Management*, vol. 62, p. 102433, feb. 2022, doi: 10.1016/j.ijinfomgt.2021.102433.
[9] A. B. Brendel, M. Mirbabaie, T.-B. Lembcke, y L. Hofeditz, «Ethical Management of Artificial Intelligence», *Sustainability*, vol. 13, n.º 4, Art. n.º 4, ene. 2021, doi: 10.3390/su13041974.
[10] A. Tsamados *et al.*, «The Ethics of Algorithms: Key Problems and Solutions», en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., en Philosophical Studies Series. , Cham: Springer International Publishing, 2021, pp. 97-123. doi: 10.1007/978-3-030-81907-1_8.
[11] A. Abulibdeh, E. Zaidan, y R. Abulibdeh, «Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions», *Journal of Cleaner Production*, vol. 437, p. 140527, ene. 2024, doi: 10.1016/j.jclepro.2023.140527.
[12] F. Poszler, E. Portmann, y C. Lütge, «Formalizing ethical principles within AI systems: experts' opinions on why (not) and how to do it», *AI Ethics*, feb. 2024, doi: 10.1007/s43681-024-00425-6.
[13] I. Celik, «Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education», *Computers in Human Behavior*, vol. 138, p. 107468, ene. 2023, doi: 10.1016/j.chb.2022.107468.
[14] M. Anderson y S. L. Anderson, «Machine Ethics: Creating an Ethical Intelligent Agent», *AI Magazine*, vol. 28, n.º 4, Art. n.º 4, dic. 2007, doi: 10.1609/aimag.v28i4.2065.
[15] Q. Qerimi, «Imagination, Invention and Internet: From Aristotle to Artificial Intelligence and the 'Post-human' Development and Ethics», en *The 21st Century from the Positions of Modern Science: Intellectual, Digital and Innovative Aspects*, E. G. Popkova y B. S. Sergi, Eds., en Lecture Notes in Networks and Systems. Cham: Springer International Publishing, 2020, pp. 360-371. doi: 10.1007/978-3-030-32015-7_41.
[16] T. Heyder, N. Passlack, y O. Posegga, «Ethical management of human-AI interaction: Theory development review», *The Journal of Strategic Information Systems*, vol. 32, n.º 3, p. 101772, sep. 2023, doi: 10.1016/j.jsis.2023.101772.
[17] J. Hallamaa y T. Kalliokoski, «How AI Systems Challenge the Conditions of Moral Agency?», en *Culture and Computing*, M. Rauterberg, Ed., en Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 54-64. doi: 10.1007/978-3-030-50267-6_5.
[18] L. Floridi *et al.*, «An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations», en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., en Philosophical Studies Series. , Cham: Springer International Publishing, 2021, pp. 19-39. doi: 10.1007/978-3-030-81907-1_3.
[19] W. Holmes *et al.*, «Ethics of AI in Education: Towards a Community-Wide Framework», *Int J Artif Intell Educ*, vol. 32, n.º 3, pp. 504-526, sep. 2022, doi: 10.1007/s40593-021-00239-1.
[20] A. Min, «Artifical Intelligence and Bias: Challenges, Implications, and Remedies», *Journal of Social Research*, vol. 2, n.º 11, pp. 3808-3817, oct. 2023, doi: 10.55324/josr.v2i11.1477.
[21] S. Akter *et al.*, «Algorithmic bias in data-driven innovation in the age of AI», *International Journal of Information Management*, vol. 60, p. 102387, oct. 2021, doi: 10.1016/j.ijinfomgt.2021.102387.
[22] E. Prem, «From ethical AI frameworks to tools: a review of approaches», *AI Ethics*, vol. 3, n.º 3, pp. 699-716, ago. 2023, doi: 10.1007/s43681-023-00258-9.
[23] M. Kiškis, «Legal framework for the coexistence of humans and conscious AI», *Front Artif Intell*, vol. 6, p. 1205465, sep. 2023, doi: 10.3389/frai.2023.1205465.
[24] S. Bonicalzi, M. De Caro, y B. Giovanola, «Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems», *Topoi*, vol. 42, n.º 3, pp. 819-832, jul. 2023, doi: 10.1007/s11245-023-09922-5.
[25] J. Zhou, F. Chen, A. Berry, M. Reed, S. Zhang, y S. Savage, «A Survey on Ethical Principles of AI and Implementations», en *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, dic. 2020, pp. 3010-3017. doi: 10.1109/SSCI47803.2020.9308437.
[26] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, y F. Herrera, «Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation», *Information Fusion*, vol. 99, p. 101896, nov. 2023, doi: 10.1016/j.inffus.2023.101896.
[27] Y. Peng, S. F. Ahmad, A. Y. A. B. Ahmad, M. S. Al Shaikh, M. K. Daoud, y F. M. H. Alhamdi, «Riding the Waves of Artificial Intelligence in Advancing Accounting and Its Implications for Sustainable Development Goals», *Sustainability*, vol. 15, n.º 19, Art. n.º 19, ene. 2023, doi: 10.3390/su151914165.
[28] N. B. Mellamphy, «Humans "in the Loop"?: Human-Centrism, Posthumanism, and AI», *Nature and Culture*, vol. 16, n.º 1, pp. 11-27, mar. 2021, doi: 10.3167/nc.2020.160102.
[29] E. Popa, «Human Goals Are Constitutive of Agency in Artificial Intelligence (AI)», *Philos. Technol.*, vol. 34, n.º 4, pp. 1731-1750, dic. 2021, doi: 10.1007/s13347-021-00483-2.
[30] N. Hynek y A. Solovyeva, «Operations of power in autonomous weapon systems: ethical conditions and socio-political prospects», *AI & Soc*, vol. 36, n.º 1, pp. 79-99, mar. 2021, doi: 10.1007/s00146-020-01048-1.
[31] N.-M. Aliman, L. Kester, y R. Yampolskiy, «Transdisciplinary AI Observatory—Retrospective Analyses and Future-Oriented Contradistinctions», *Philosophies*, vol. 6, n.º 1, Art. n.º 1, mar. 2021, doi: 10.3390/philosophies6010006.
[32] N. Th. Nikolinakos, «Ethical Principles for Trustworthy AI», en *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - The AI Act*, N. Th. Nikolinakos, Ed., en Law, Governance and Technology Series. , Cham: Springer International Publishing, 2023, pp. 101-166. doi: 10.1007/978-3-031-27953-9_3.
[33] C. Gonzalez, «Building Human-Like Artificial Agents: A General Cognitive Algorithm for Emulating Human Decision-Making in Dynamic Environments», *Perspect Psychol Sci*, p. 17456916231196766, oct. 2023, doi: 10.1177/17456916231196766.
[34] F. Mughal, A. Wahid, y M. A. K. Khattak, «Artificial Intelligence: Evolution, Benefits, and Challenges», en *Intelligent Cyber-Physical Systems for Autonomous Transportation*, S. Garg, G. S. Aujla, K. Kaur, y S. Hassan Ahmed Shah, Eds., en Internet of Things. , Cham: Springer International Publishing, 2022, pp. 59-69. doi: 10.1007/978-3-030-92054-8_4.
[35] Y. Dong, J. Hou, N. Zhang, y M. Zhang, «Research on How Human Intelligence, Consciousness, and Cognitive Computing Affect the Development of Artificial Intelligence», *Complexity*, vol. 2020, p. e1680845, oct. 2020, doi: 10.1155/2020/1680845.
[36] N. Zins, Y. Zhang, C. Yu, y H. An, «Neuromorphic Computing: A Path to Artificial Intelligence Through Emulating Human Brains», en *Frontiers of Quality Electronic Design (QED): AI, IoT and Hardware Security*, A. Iranmanesh, Ed., Cham: Springer International Publishing, 2023, pp. 259-296. doi: 10.1007/978-3-031-16344-9_7.
[37] R. St. Clair, L. A. Coward, y S. Schneider, «Leveraging conscious and nonconscious learning for efficient AI», *Frontiers in Computational Neuroscience*, vol. 17, 2023, Accedido: 26 de febrero de 2024. [En línea]. Disponible en: https://www.frontiersin.org/articles/10.3389/fncom.2023.1090126
[38] M. R. Anwar, F. P. Oganda, N. P. L. Santoso, y M. Fabio, «Artificial Intelligence that Exists in the Human Mind», *International Transactions on Artificial Intelligence*, vol. 1, n.º 1, pp. 28-42, nov. 2022, doi: 10.33050/italic.v1i1.87.
[39] Z. Rudnicka, J. Szczepanski, y A. Pregowska, «Artificial Intelligence-Based Algorithms in Medical Image Scan Segmentation and Intelligent Visual Content Generation—A Concise Overview», *Electronics*, vol. 13, n.º 4, Art. n.º 4, ene. 2024, doi: 10.3390/electronics13040746.
[40] H. Hassani, E. S. Silva, S. Unger, M. TajMazinani, y S. Mac Feely, «Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future?», *AI*, vol. 1, n.º 2, Art. n.º 2, jun. 2020, doi: 10.3390/ai1020008.
[41] Y. Dai *et al.*, «Collaborative construction of artificial intelligence curriculum in primary schools», *Journal of Engineering Education*, vol. 112, n.º 1, pp. 23-42, 2023, doi: 10.1002/jee.20503.
[42] M. I. A. Ferreira, «Artificial Intelligence: A Concept Under-Construction, A Reality Under-Development», en *Towards Trustworthy Artificial Intelligent Systems*, M. I. A. Ferreira y M. O. Tokhi, Eds., en Intelligent Systems, Control and Automation: Science and Engineering. , Cham: Springer International Publishing, 2022, pp. 1-22. doi: 10.1007/978-3-031-09823-9_1.

[43] K. Sotala, «How feasible is the rapid development of artificial superintelligence?», *Phys. Scr.*, vol. 92, n.º 11, p. 113001, oct. 2017, doi: 10.1088/1402-4896/aa90e8.
[44] M. Hussain, «When, Where, and Which?: Navigating the Intersection of Computer Vision and Generative AI for Strategic Business Integration», *IEEE Access*, vol. 11, pp. 127202-127215, 2023, doi: 10.1109/ACCESS.2023.3332468.
[45] P. Formosa, «Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy», *Minds & Machines*, vol. 31, n.º 4, pp. 595-616, dic. 2021, doi: 10.1007/s11023-021-09579-2.
[46] Z. Wang y D. Wu, «The Digital Nexus: tracing the evolution of human consciousness and cognition within the artificial realm—a comprehensive review», *AI & Soc*, ago. 2023, doi: 10.1007/s00146-023-01753-7.
[47] C. Fontes, E. Hohma, C. C. Corrigan, y C. Lütge, «AI-powered public surveillance systems: why we (might) need them and how we want them», *Technology in Society*, vol. 71, p. 102137, nov. 2022, doi: 10.1016/j.techsoc.2022.102137.
[48] D. Anchuri, R. Rodriguez, y K. V. Thota, «"Cognition and Intelligent Retrieval": Can AI Produce Consciousness Comparable to That of Humans? Capacities for Thinking? What about Morality?» Rochester, NY, 16 de abril de 2023. doi: 10.2139/ssrn.4419988.
[49] Z. Cui y J. Liu, «A Study on Two Conditions for the Realization of Artificial Empathy and Its Cognitive Foundation», *Philosophies*, vol. 7, n.º 6, Art. n.º 6, dic. 2022, doi: 10.3390/philosophies7060135.
[50] A. M. Turing, «Computing Machinery and Intelligence», *Mind*, vol. LIX, n.º 236, pp. 433-460, oct. 1950, doi: 10.1093/mind/LIX.236.433.
[51] D. Duncker, «Chatting with Chatbots: Sign Making in Text-based Human-computer Interaction», *Σημειωτκή - Sign Systems Studies*, vol. 48, n.º 1, pp. 79-100, 2020, doi: 10.12697/SSS.2020.48.1.05.
[52] E. Brynjolfsson, «The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence», *Daedalus*, vol. 151, n.º 2, pp. 272-287, may 2022, doi: 10.1162/daed_a_01915.
[53] R. Fjelland, «Why general artificial intelligence will not be realized», *Humanit Soc Sci Commun*, vol. 7, n.º 1, Art. n.º 1, jun. 2020, doi: 10.1057/s41599-020-0494-4.
[54] E. Barbierato y M. E. Zamponi, «Shifting Perspectives on AI Evaluation: The Increasing Role of Ethics in Cooperation», *AI*, vol. 3, n.º 2, Art. n.º 2, jun. 2022, doi: 10.3390/ai3020021.
[55] S. Tariq, A. Iftikhar, P. Chaudhary, y K. Khurshid, «Examining Some Serious Challenges and Possibility of AI Emulating Human Emotions, Consciousness, Understanding and 'Self'», *Journal of NeuroPhilosophy*, vol. 1, n.º 1, abr. 2022, doi: 10.5281/zenodo.6637757.
[56] S. Tariq, A. Iftikhar, P. Chaudhary, y K. Khurshid, «Is the 'Technological Singularity Scenario' Possible: Can AI Parallel and Surpass All Human Mental Capabilities?», *World Futures*, vol. 79, n.º 2, pp. 200-266, feb. 2023, doi: 10.1080/02604027.2022.2050879.
[57] M. T. Bennett y Y. Maruyama, «Philosophical Specification of Empathetic Ethical Artificial Intelligence», *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, n.º 2, pp. 292-300, jun. 2022, doi: 10.1109/TCDS.2021.3099945.
[58] X. Huang, S. Li, T. Wang, Z. Pan, y S. P. Lajoie, «Exploring the co-occurrence of students' learning behaviours and reasoning processes in an intelligent tutoring system: An epistemic network analysis», *Journal of Computer Assisted Learning*, vol. 39, n.º 5, pp. 1701-1713, 2023, doi: 10.1111/jcal.12827.
[59] R. Madanu, M. F. Abbod, F.-J. Hsiao, W.-T. Chen, y J.-S. Shieh, «Explainable AI (XAI) Applied in Machine Learning for Pain Modeling: A Review», *Technologies*, vol. 10, n.º 3, Art. n.º 3, jun. 2022, doi: 10.3390/technologies10030074.
[60] D. J. Herzog y N. Herzog, «What is it like to be an AI bat?», *Qeios*, ene. 2024, doi: 10.32388/63ELTC.
[61] R. Descartes, *A Discourse on the Method of Correctly Conducting One's Reason and Seeking Truth in the Sciences*. Suffolk: Oxford University Press Inc, 2006. [En línea]. Disponible en: https://rauterberg.employee.id.tue.nl/lecturenotes/DDM110%20CAS/Descartes-1637%20Discourse%20on%20Method.pdf
[62] J. LeDoux *et al.*, «Consciousness beyond the human case», *Current Biology*, vol. 33, n.º 16, pp. R832-R840, ago. 2023, doi: 10.1016/j.cub.2023.06.067.
[63] S. Cervantes, S. López, y J.-A. Cervantes, «Toward ethical cognitive architectures for the development of artificial moral agents», *Cognitive Systems Research*, vol. 64, pp. 117-125, dic. 2020, doi: 10.1016/j.cogsys.2020.08.010.
[64] J.-S. Gordon, «Building Moral Robots: Ethical Pitfalls and Challenges», *Sci Eng Ethics*, vol. 26, n.º 1, pp. 141-157, feb. 2020, doi: 10.1007/s11948-019-00084-5.
[65] A. Dameski, «Foundations of an Ethical Framework for AI Entities: the Ethics of Systems», Tesis doctoral, Università di Bologna, Bologna, 2020. Accedido: 26 de febrero de 2024. [En línea]. Disponible en: https://orbilu.uni.lu/handle/10993/45285
[66] V. Galanos, «Expectations and expertise in artificial intelligence: specialist views and historical perspectives on conceptualisation, promise, and funding», Tesis doctoral, University of Edinburgh, Edinburgh, 2023. Accedido: 26 de febrero de 2024. [En línea]. Disponible en: https://era.ed.ac.uk/handle/1842/40420
[67] G. Siemens *et al.*, «Human and artificial cognition», *Computers and Education: Artificial Intelligence*, vol. 3, p. 100107, ene. 2022, doi: 10.1016/j.caeai.2022.100107.
[68] E. M. Hartman, «Socratic Questions and Aristotelian Answers: A Virtue-Based Approach to Business Ethics», *J Bus Ethics*, vol. 78, n.º 3, pp. 313-328, mar. 2008, doi: 10.1007/s10551-006-9337-5.
[69] C. Dierksmeier, «Kant on Virtue», *J Bus Ethics*, vol. 113, n.º 4, pp. 597-609, abr. 2013, doi: 10.1007/s10551-013-1683-5.
[70] S. S. C. Komu, «Pleasure versus Virtue Ethics in The Light of Aristotelians and the Utilitarianism of John Stuart Mills and Jeremy Bentham», *Al-Milal: Journal of Religion and Thought*, vol. 2, n.º 1, Art. n.º 1, jun. 2020, doi: 10.46600/almilal.v2i1.57.
[71] D. M. D. M. Refaei, «Regulatory Frameworks for Autonomous Robotics in NEOM's Sustainable Technology Landscape», *Migration Letters*, vol. 20, n.º 9, Art. n.º 9, dic. 2023, doi: 10.59670/ml.v20i9.5965.
[72] M. Zhu, A. H. Anwar, Z. Wan, J.-H. Cho, C. A. Kamhoua, y M. P. Singh, «A Survey of Defensive Deception: Approaches Using Game Theory and Machine Learning», *IEEE Communications Surveys & Tutorials*, vol. 23, n.º 4, pp. 2460-2493, 2021, doi: 10.1109/COMST.2021.3102874.
[73] H. Ashrafian, «Engineering a social contract: Rawlsian distributive justice through algorithmic game theory and artificial intelligence», *AI Ethics*, vol. 3, n.º 4, pp. 1447-1454, nov. 2023, doi: 10.1007/s43681-022-00253-6.
[74] J. Zhong, C. Ling, A. Cangelosi, A. Lotfi, y X. Liu, «On the Gap between Domestic Robotic Applications and Computational Intelligence», *Electronics*, vol. 10, n.º 7, Art. n.º 7, ene. 2021, doi: 10.3390/electronics10070793.
[75] I. Asimov, *I, Robot*. New York: Bantam, 2004. [En línea]. Disponible en: https://prepa.unimatehuala.edu.mx/pluginfile.php/7362/mod_glossary/attachment/866/Yo,%20robot%20-%20Isaac%20Asimov.pdf
[76] M. Taddeo y L. Floridi, «How AI Can Be a Force for Good – An Ethical Framework to Harness the Potential of AI While Keeping Humans in Control», en *Ethics, Governance, and Policies in Artificial Intelligence*, L. Floridi, Ed., en Philosophical Studies Series. , Cham: Springer International Publishing, 2021, pp. 91-96. doi: 10.1007/978-3-030-81907-1_7.
[77] Z. Liu y Y. Zheng, «AI and Robot: Darwin and Rebellious Machine», en *AI Ethics and Governance: Black Mirror and Order*, Z. Liu y Y. Zheng, Eds., Singapore: Springer Nature, 2022, pp. 79-93. doi: 10.1007/978-981-19-2531-3_6.
[78] D. White y H. Katsuno, «Toward an Affective Sense of Life: Artificial Intelligence, Animacy, and Amusement at a Robot Pet Memorial Service in Japan», *Cultural Anthropology*, vol. 36, n.º 2, Art. n.º 2, may 2021, doi: 10.14506/ca36.2.03.
[79] M. Kulkarni *et al.*, «The Future of Research in an Artificial Intelligence-Driven World», *Journal of Management Inquiry*, p. 10564926231219622, feb. 2024, doi: 10.1177/10564926231219622.
[80] A. Campolo y K. Crawford, Eds., «Enchanted Determinism: Power without Responsibility in Artificial Intelligence», *Engaging Science, Technology, and Society*, doi: 10.17351/ests2020.277.
[81] W. Mu, «How Artificial Intelligence Affects Workforces: The Impact of Biased Recruitment and Job Displacement Risk», *Highlights in Business, Economics and Management*, vol. 23, pp. 19-25, dic. 2023, doi: 10.54097/2t4h0q42.
[82] A. Taeihagh, «Governance of artificial intelligence», *Policy and Society*, vol. 40, n.º 2, pp. 137-157, jun. 2021, doi: 10.1080/14494035.2021.1928377.

[83] A. Goldfarb y J. R. Lindsay, «Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War», *International Security*, vol. 46, n.º 3, pp. 7-50, feb. 2022, doi: 10.1162/isec_a_00425.
[84] L. Royakkers, J. Timmer, L. Kool, y R. van Est, «Societal and ethical issues of digitization», *Ethics Inf Technol*, vol. 20, n.º 2, pp. 127-142, jun. 2018, doi: 10.1007/s10676-018-9452-x.
[85] D. Varona y J. L. Suárez, «Discrimination, Bias, Fairness, and Trustworthy AI», *Applied Sciences*, vol. 12, n.º 12, Art. n.º 12, ene. 2022, doi: 10.3390/app12125826.
[86] Z. Xu, «Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities», *Applied Artificial Intelligence*, vol. 36, n.º 1, p. 2013652, dic. 2022, doi: 10.1080/08839514.2021.2013652.
[87] A. Laitinen y O. Sahlgren, «AI Systems and Respect for Human Autonomy», *Frontiers in Artificial Intelligence*, vol. 4, 2021, Accedido: 26 de febrero de 2024. [En línea]. Disponible en: https://doi.org/10.3389/frai.2021.705164
[88] F. Tollon, «Answerability, Accountability, and the Demands of Responsibility», en *Artificial Intelligence Research*, A. Pillay, E. Jembere, y A. Gerber, Eds., en Communications in Computer and Information Science. Cham: Springer Nature Switzerland, 2022, pp. 371-383. doi: 10.1007/978-3-031-22321-1_25.