

Perspectivas filosóficas de la inteligencia artificial (IA)

SANTIAGO DECARLO AYALA CABRALES.
ASESOR: SERGIO ALEXANDER HOYOS CONTRERAS
DICIEMBRE 2024.

Universidad de Pamplona
Facultad de Artes y Humanidades
Programa de Filosofía
Monografía

Agradecimientos

A *El Shaddai*, fuente de toda sabiduría y fortaleza, por iluminar mi camino, por darme la guía y el coraje para afrontar cada desafío. Gracias por brindarme la claridad necesaria para emprender este viaje y por otorgarme el don de la reflexión y el entendimiento.

A Pamplona, por ser el refugio donde la paz y la concentración se fusionan. Sus paisajes, la gente diversa, y la quietud de sus rincones me ofrecieron el espacio perfecto para mis estudios. La antigua fábrica, con su neblina envolvente y muros gastados, añadía un aire de misterio, como si aún guardara ecos de un pasado olvidado. La ciudad, con su caos organizado y sus bosques cercanos, se convirtió en un cóctel de inspiración, donde la naturaleza y la humanidad se mezclaban para impulsar mis pensamientos y mi camino.

A mis padres, por su amor incondicional y su apoyo constante. A mi padre, por enseñarme que la verdadera riqueza está en el poder de la mente y en las preguntas que nos impulsan a buscar más allá de lo evidente. A mi madre, por su paciencia, su dedicación y su fe inquebrantable, que siempre ha sido mi ancla en los momentos de incertidumbre.

A mi abuela, por su sabiduría ancestral, por enseñarme a valorar las raíces y por ser un faro de fortaleza. Gracias por transmitirme no solo tus conocimientos, sino también el amor que me acompaña en cada paso.

A mis profesores, por su generosidad y paciencia. Gracias por abrirme las puertas del conocimiento y por desafiarme a pensar de manera crítica y profunda. Su enseñanza ha sido esencial en mi crecimiento intelectual.

A mis amigos, por su apoyo constante, por las risas, las conversaciones y la compañía. Cada uno de ustedes ha sido una fuente de fuerza y alegría durante este proceso.

Este trabajo es el reflejo de todo lo que he recibido de ustedes: de su amor, su sabiduría, su esfuerzo y su apoyo. A todos, mi más profundo agradecimiento.

Dedicatoria

A mi madre, el océano que me ha dado la calma y la fuerza para navegar las aguas más turbulentas. Con su amor profundo y constante, siempre ha sido mi refugio, mi ancla en medio de las tormentas y mi viento cuando más lo he necesitado.

A mi abuela, la raíz que me sostiene, cuyo abrazo de sabiduría me conecta con el pasado y me da fuerza para el futuro.

A mi padre, el viento que empujó mis palabras, enseñándome que las respuestas están en las preguntas, y las preguntas en el alma.

A mi hermana, el reflejo de mi confianza, la estrella que brilla a mi lado en cada noche oscura.

A mi tutor y profesores, las manos que moldearon mi trabajo, dándole forma y dirección.

Y a mi hada mágica, cuyo susurro invisible me dio alas para seguir, incluso cuando el vuelo parecía incierto.

Resumen

La inteligencia artificial (IA) ha abierto un abanico de interrogantes filosóficos que afectan nuestra comprensión de la conciencia, la ética y el poder. Aunque diseñada para replicar la cognición humana, surge la pregunta de si la IA puede experimentar una conciencia genuina o si simplemente simula procesos mentales de forma superficial. Este dilema pone en cuestión no solo la fenomenología, al desentrañar la naturaleza de la conciencia, sino también la ética, al explorar si las máquinas pueden asumir responsabilidades morales y cómo sus decisiones impactan la justicia social. El impacto de la IA también se extiende a las estructuras de poder global, transformando las relaciones de soberanía y autonomía en un mundo cada vez más interconectado. Este estudio aborda la IA desde varias ramas filosóficas clave. La fenomenología de Husserl se utiliza para explorar la capacidad de la IA de replicar la conciencia; la ontología de Heidegger para entender la relación de la IA con el ser; y la hermenéutica de Gadamer para evaluar su capacidad de comprensión y diálogo. En el ámbito ético, se revisan las teorías de Aristóteles, Kant y Floridi sobre la moralidad de las máquinas aplicando la innovación filosófica como vías de soluciones alternas. Finalmente, se analiza el riesgo de una IA descontrolada y sus implicaciones geopolíticas, con las perspectivas de Bostrom y Harari. La IA nos desafía a reconsiderar las bases de nuestra humanidad, forzándonos a redefinir la conciencia, la ética y el poder en un mundo cada vez más marcado por la tecnología.

Abstract

Artificial intelligence (AI) has opened a range of philosophical questions that affect our understanding of consciousness, ethics, and power. While designed to replicate human cognition, the question arises whether AI can experience genuine consciousness or if it merely simulates mental processes superficially. This dilemma calls into question not only phenomenology, as it seeks to unravel the nature of consciousness, but also ethics, by exploring whether machines can assume moral responsibilities and how their decisions impact social justice. The impact of AI also extends to global power structures, transforming relationships of sovereignty and autonomy in an increasingly interconnected world. This study approaches AI from various key philosophical branches. Husserl's phenomenology is used to explore AI's ability to replicate consciousness; Heidegger's ontology to understand AI's relationship with being; and Gadamer's hermeneutics to assess its capacity for understanding and dialogue. In the ethical realm, the theories of Aristotle, Kant, and Floridi are reviewed regarding the morality of machines, applying philosophical innovation as

alternative solutions. Finally, the risks of uncontrolled AI and its geopolitical implications are analyzed, with perspectives from Bostrom and Harari. AI challenges us to reconsider the foundations of our humanity, forcing us to redefine consciousness, ethics, and power in an increasingly technology-driven world.

Palabras clave: Fenomenología; Hermenéutica; Ontología; Ética; Post-Humanismo; Tecnología; Inteligencia artificial.

Keywords: Phenomenology; Hermeneutics; Ontology; Ethics; Post-Humanism; Technology; Artificial intelligence.

Tabla de contenido

Introducción.....	6
-------------------	---

Capítulo I: Meditaciones Filosóficas sobre la inteligencia artificial: Entre la Fenomenología, la Ontología y la Hermenéutica.....	9
Orígenes y fundamentos de la inteligencia artificial.....	10
Intencionalidad, conciencia y subjetividad trascendental en Husserl, una visión fenomenológica ante la IA.....	13
La IA y el <i>Gestell</i> : desafíos a la existencia auténtica del <i>Dasein</i>	25
La hermenéutica de Gadamer y su relevancia en la era de la IA.....	32
Capítulo II: Perspectivas éticas en la Inteligencia Artificial.....	39
La ética aristotélica y los desafíos para la inteligencia artificial.....	41
Desafíos y límites de aplicar la ética kantiana a la inteligencia artificial.....	46
Hacia una ética expansiva en la era de la re-ontologización artificial.....	55
Capítulo III: Horizontes filosóficos de la IA: Poder, control y libertad.....	63
Superinteligencia: Entre el progreso y la extinción de la humanidad.....	64
Inteligencia artificial y poder: Una mirada filosófica y geopolítica.....	70
IA, Poder y Futuro: Reflexiones Filosóficas.....	75
Conclusiones.....	78
BIBLIOGRAFÍA.....	81
Bibliografía Básica.....	81
Bibliografía Complementaria.....	82
Vita.....	86

Introducción

La inteligencia artificial (IA), a pesar de sus avances significativos, plantea una serie de dilemas filosóficos que abren nuevas interrogantes sobre la naturaleza del ser, la cognición y la conciencia. En el contexto actual, la IA no solo se estudia como una tecnología que

implica algoritmos complejos, sino también como un fenómeno que desafía nuestras concepciones fundamentales sobre la mente humana, la autonomía y el poder. Este fenómeno tecnológico interroga el concepto de conciencia, así como las estructuras de poder que organizan la vida humana en la sociedad.

El desarrollo de la IA ha sido uno de los logros más extraordinarios de la ciencia de la computación, pero sus implicaciones filosóficas son aún vastas y poco exploradas. Si bien la IA se ha diseñado para imitar ciertos aspectos de la cognición humana, surge la pregunta de si esta imitación es solo una simulación superficial o si realmente puede alcanzar una forma de conciencia auténtica. Además, el modo en que la IA podría emular la conciencia humana interroga no solo a la fenomenología, la ontología y la hermenéutica, sino también a la filosofía ética, al replantear la cuestión de si las máquinas pueden asumir responsabilidades morales y cómo sus decisiones impactan en la justicia y la equidad social.

Además, el desarrollo de la inteligencia artificial plantea una paradoja profunda para la humanidad, como lo señala Nick Bostrom. Mientras que la IA promete avances tecnológicos significativos, también introduce riesgos existenciales. Bostrom advierte que, si se alcanza una IA capaz de mejorarse a sí misma, podría superar la inteligencia humana, escapando al control de sus creadores.

Por otra parte, la relación entre la IA y el poder geopolítico también se encuentra en el centro del debate. A medida que las naciones avanzan en la creación de sistemas de inteligencia artificial, se evidencian tensiones de poder entre actores globales, como China y Estados Unidos, que utilizan la IA no solo como un medio para el avance tecnológico, sino también como una herramienta para el dominio político y económico. En este contexto, la IA se presenta como un desafío a la soberanía estatal y la autonomía de los individuos, a la vez que redefine las fronteras de la tecnología en el ámbito social y económico.

El problema global que se plantea, por tanto, es el siguiente: ¿Cómo la inteligencia artificial, en su intento de replicar aspectos fundamentales de la cognición humana, puede redefinir nuestras concepciones filosóficas sobre la conciencia, el ser, la comprensión, la ética y el poder desde una IA como amenaza potencial para la existencia humana hasta ser

utilizada por los poderes imperiales en la geopolítica, y qué impacto tiene esto en las estructuras existenciales y normativas que organizan nuestro mundo contemporáneo?

En consecuencia, se ha optado por abordar las ideas de filósofos clave que han influido profundamente en la comprensión de temas como el ser, la conciencia, la comprensión, la ética y las estructuras que configuran el poder en el mundo. En este contexto, se exploran cuestiones fundamentales sobre cómo la inteligencia artificial puede ser interpretada desde diversas perspectivas filosóficas. Así, las obras filosóficas que se utilizan en este estudio se presentan de la siguiente manera:

En primer lugar, se examina la capacidad de la IA para replicar la conciencia, analizando conceptos como la intencionalidad, el noema y lo noético, así como la subjetividad trascendental y las nociones de cuerpo vivido y cuerpo físico (*Leib y Körper*). Para ello, se recurre a *Ideas relativas a una fenomenología pura y una filosofía fenomenológica* (1949). A continuación, se aborda el concepto de *Gestell* de Heidegger, que describe el marco estructural del mundo, y se examina la naturaleza del ser *Dasein* en la IA, utilizando obras clave como *Filosofía, Ciencia y Técnica* (1997), (1994), y *Ser y Tiempo* (1997). En un segundo momento, se analiza cómo la hermenéutica puede arrojar luz sobre la capacidad de la IA para experimentar comprensión, diálogo y lenguaje humano, utilizando como referencia la obra *Verdad y Método* (1999), (1993), de Gadamer.

En segundo lugar, en el segundo capítulo se examinan los dilemas éticos que plantea la IA, cuestionando si esta puede aplicar principios éticos genuinos o si, por el contrario, solo los emula de manera superficial. En caso de que su aplicación sea superficial, se exploran posibles enfoques filosóficos innovadores que podrían ofrecer soluciones. Este análisis se basa en *La Ética a Nicómaco* (1985) de Aristóteles, la *Fundamentación de la metafísica de las costumbres* (2012) de Kant, y la *Ética de la información* (2013) de Floridi.

Por último, lugar, el tercer capítulo profundiza en la filosofía de la tecnología y geopolítica, realizando una reseña de dos obras, la primera aborda la idea de la IA como una superinteligencia que podría escapar al control humano y amenazar la existencia de la humanidad. En este sentido, se recurre a la obra *Superinteligencia* (2014) de Nick Bostrom donde una superinteligencia se automejora hasta tal punto de convertirse en una amenaza potencialmente peligrosa para la humanidad y la segunda, examina el impacto de la IA en

la geopolítica, evaluando sus implicaciones en el orden mundial, las dinámicas de poder entre potencias como China y Estados Unidos, la identidad de los cuerpos en la era digital, y el concepto del “telón de silicio” que define las nuevas fronteras tecnológicas. Para ello, se hace uso de *Nexus: Una Breve Historia de las Redes de Información Desde la Edad de Piedra Hasta la IA* (2024) de Yuval Noah Harari.

Las obras que se citan en este trabajo siguen diferentes convenciones según el autor. *Ser y Tiempo* (1997) de Heidegger se cita utilizando el parágrafo correspondiente (autor, año y parágrafo [§]). En el caso de *La Ética a Nicómaco* de Aristóteles (1985), la cita se hace según el pasaje específico (autor, fecha, pasaje, número de página). Por otro lado, *Fundamentación de la metafísica de las costumbres* (2012) de Kant se cita con el pasaje correspondiente (autor, fecha, pasaje, número de página). Las demás obras de la bibliografía básica y bibliografía complementaria están citadas en APA séptima edición. Por otra parte, hay citas que fueron traducidas en otros idiomas como lo son (inglés, ruso, portugués), las traducciones fueron apoyadas con la herramienta *DeepL*.

Meditaciones Filosóficas sobre la inteligencia artificial: Entre la Fenomenología, la Ontología y la Hermenéutica.

“Tim dijo una vez que la ciencia y el mito eran una y la misma cosa, que el mundo natural era sólo una manifestación de los pensamientos e impulsos en innumerables planos metafísicos”

(Wade Daves, 2004, p. 892)

Este capítulo examina en primer lugar una descripción de lo que es una inteligencia artificial, además de su historia e inicios. En segundo lugar, se realiza una interacción entre

la inteligencia artificial (IA) y tres enfoques filosóficos clave: la fenomenología de Husserl, E., la ontología de Heidegger, M. y la hermenéutica de Gadamer, H., y cómo estas filosofías enfrentan los retos que plantea la IA en la comprensión de la conciencia, subjetividad y sentido humano.¹

Orígenes y fundamentos de la inteligencia artificial.

En primer lugar, la inteligencia artificial (IA) es un campo científico y tecnológico que ha revolucionado la interacción entre humanos y máquinas. En este sentido, la IA, originada tras la Segunda Guerra Mundial y formalmente nombrada en 1956, es un área de investigación relativamente joven que ha captado el interés de especialistas de diversas disciplinas (Russell, S. y Norvig, P., 2004, p. 1). Esto resalta no solo su novedad, sino también su capacidad de atraer a expertos de múltiples campos, lo que subraya su impacto multidimensional y su relevancia en la intersección entre tecnología y humanidad.

La IA se define desde enfoques que consideran tanto la emulación de procesos mentales humanos, como la búsqueda de una racionalidad ideal. Estos enfoques contrastan en cómo miden el éxito, algunos lo evalúan en función de la fidelidad con la que la IA imita el comportamiento humano, mientras que otros priorizan la capacidad de la IA para ejecutar acciones racionales basadas en el conocimiento (Russell, S. y Norvig, P., 2004, p. 2). Este enfoque dual refleja un dilema fundamental: si la IA debe aspirar a replicar la inteligencia humana o enfocarse en una racionalidad pura. Esta tensión evidencia la complejidad de definir la inteligencia y plantea interrogantes sobre los objetivos últimos de la tecnología.

¹Este capítulo aborda la interacción de la inteligencia artificial con los enfoques filosóficos de Husserl, Heidegger y Gadamer, centrando el análisis en aspectos específicos de sus metodologías más que en una exposición completa de sus conceptos. Se reconoce la amplitud de los temas tratados, pero se subraya que este estudio se limita a aquellos puntos que aportan una visión particular sobre la conciencia y su relación con la IA, evitando un tratamiento general de sus respectivas filosofías. Para un análisis más extenso y detallado, se recomienda consultar las obras originales: *Ideas relativas a una fenomenología pura y una filosofía fenomenológica de Husserl* (1949), *Ser y tiempo* (1997), *los escritos sobre la técnica de Heidegger* (1994), (1997) y *Verdad y método de Gadamer* (1999), (1993).

En sus inicios, Warren McCulloch y Walter Pitts (1943) propusieron un modelo de IA inspirado en la fisiología neuronal², la lógica formal³ y la teoría de Turing⁴. Mostraron que las redes neuronales interconectadas podían computar funciones complejas y replicar conectores lógicos, anticipando así capacidades de aprendizaje (Russell, S. y Norvig, P., 2004, p. 19). Este modelo marcó un hito en la historia de la IA al integrar conceptos de biología, lógica y computación, sentando las bases para futuros desarrollos al demostrar que las redes podían realizar cálculos complejos y aprender.

Más tarde, John McCarthy, tras graduarse en Princeton, contribuyó al impulso de la IA al organizar en 1956 un taller en Dartmouth junto a Minsky, Shannon y Rochester. Este evento, considerado el inicio formal de la IA, atrajo a destacados investigadores interesados en la inteligencia artificial y en las máquinas autómatas (Russell, S. y Norvig, P., 2004, p. 19). La organización de este taller marcó un punto de inflexión al reunir a pioneros de diversas disciplinas y fomentar un intercambio interdisciplinario esencial. Este encuentro consolidó las bases teóricas y prácticas que impulsaron la investigación y la evolución del campo en las décadas siguientes.

Por tanto, los inicios de la IA, marcados por los trabajos de McCulloch, Pitts (1943) y el impulso de McCarthy en Dartmouth (1956), revelan más que avances técnicos, muestran una aspiración humana por entender y replicar la inteligencia. Esta búsqueda no solo impulsó la ciencia y la tecnología, sino que también planteó preguntas profundas sobre la naturaleza de la mente, la racionalidad y la relación entre los seres humanos y las máquinas. A medida que la IA avanza, estos primeros pasos siguen recordándonos que la exploración de la inteligencia artificial es, en última instancia, un reflejo de nuestra curiosidad y deseo de trascender los límites del conocimiento.

²La fisiología neuronal se refiere al estudio de las funciones de las neuronas y sus interacciones en el sistema nervioso para procesar información y generar respuestas. En comparación, la IA busca emular estos procesos a través de redes neuronales artificiales que replican de manera simplificada las conexiones sinápticas y la transmisión de señales. Sin embargo, a diferencia de la plasticidad y adaptación orgánica del cerebro humano, las redes de IA están limitadas por su programación y capacidad de ajuste basada en datos predeterminados.

³La lógica formal en IA permite que los sistemas realicen razonamientos deductivos mediante algoritmos, replicando procesos lógicos humanos. Aunque pueden seguir reglas precisas y generar conclusiones coherentes, estos sistemas carecen de la intuición y flexibilidad del pensamiento humano, que trasciende la mera estructura formal.

⁴La teoría de Turing estableció que las máquinas pueden simular procesos humanos complejos mediante algoritmos. Esto inspiró la creación de la IA moderna, que imita la cognición humana, aunque sin alcanzar verdadera comprensión o consciencia.

En segundo lugar, Husserl, E., en su obra *ideas relativas a una fenomenología pura y una filosofía fenomenológica*⁵ de (1949), introduce la intencionalidad, es decir, la manera en que la conciencia se dirige hacia los fenómenos, otorgándoles significado. Aunque la IA puede simular ciertos procesos mentales, no puede replicar la subjetividad humana y la capacidad de dar sentido a las vivencias. La IA se enfrenta a un límite fundamental en su incapacidad de reproducir la profundidad de la conciencia humana.

Heidegger, en su concepto de *Gestell*⁶ (1994), (1997) examina cómo la tecnología organiza y limita la percepción del mundo. Su noción de *Dasein*, introducida en *Ser y Tiempo*⁷ (1997), describe al ser humano como un ser-en-el-mundo que siempre se proyecta hacia sus posibilidades. La IA, como parte de esta estructura técnica, afecta las decisiones humanas al influir en la libertad de elección.

Por último, Gadamer, H., en su obra *Verdad y Método*⁸ (1999), (1993), analiza la hermenéutica filosófica, que se enfoca en la interpretación y comprensión humana. Gadamer resalta que la IA, al intervenir en estos procesos, afecta cómo se construye el significado en un mundo digital, planteando la necesidad de reflexionar sobre su impacto en el juicio humano y la comprensión.

⁵En su obra *Ideas relativas a una fenomenología pura y una filosofía fenomenológica* (1949), Edmund Husserl propone un enfoque metodológico que busca regresar a las "cosas mismas", es decir, a la experiencia inmediata y a la conciencia. A través de la fenomenología, Husserl investiga la estructura de la experiencia y la intencionalidad, enfatizando que la conciencia siempre está dirigida hacia objetos. Su enfoque busca despojarse de presuposiciones y teorías previas para llegar a una comprensión pura de los fenómenos, estableciendo así una base sólida para las ciencias y una filosofía rigurosa que explora la esencia de las experiencias humanas.

⁶En la obra *Filosofía, Ciencia y Técnica* (1997), Martin Heidegger aborda el concepto de *Gestell*, que se refiere a la estructura del ente que organiza y determina la manera en que los seres humanos se relacionan con el mundo. Traducido comúnmente como "enmarcamiento" o "disposición", el *Gestell* describe cómo la tecnología moderna configura nuestra comprensión del ser, transformando la naturaleza en un recurso disponible y manipulable. Heidegger advierte que esta forma de pensar lleva a una pérdida de la conexión auténtica con el ser, limitando la capacidad de experimentar el mundo de manera plena.

⁷En *Ser y Tiempo* (1997), Martin Heidegger introduce el concepto de *Dasein*, que significa "ser ahí" y se refiere a la existencia humana en su singularidad. El *Dasein* es un ser consciente que se pregunta por su propio ser, y su esencia radica en la capacidad de proyectarse hacia el futuro y tomar decisiones. Heidegger explora la temporalidad, la angustia y la autenticidad del *Dasein*, destacando cómo su existencia está siempre en relación con el mundo que lo rodea, lo que es fundamental para entender su análisis del ser.

⁸En *Verdad y Método* (1999) (1993), Hans-Georg Gadamer aborda la hermenéutica, enfatizando la importancia de la interpretación en el entendimiento humano. Gadamer sostiene que el conocimiento no se limita a métodos científicos, sino que se construye a través del diálogo y la interacción cultural. Introduce el concepto de "fusión de horizontes", que describe el proceso por el cual las perspectivas del intérprete y del texto se encuentran y enriquecen mutuamente. Al destacar el papel de la historia y la tradición en la comprensión, Gadamer, argumenta que la verdad se revela en el contexto de la comunicación y la experiencia compartida, desafiando la idea de una verdad objetiva y estática.

Intencionalidad, conciencia y subjetividad trascendental en Husserl, una visión fenomenológica ante la IA.

“Allá arriba, junto al camino, en su cabaña, el viejo dormía nuevamente. Todavía dormía de bruces y el muchacho estaba sentado a su lado contemplándolo. El viejo soñaba con los leones marinos”

(Ernest Hemingway, 1952, p. 69)

Este apartado explora la intersección entre los conceptos clave de la fenomenología de Husserl y su aplicación en la inteligencia artificial (IA), subrayando sus limitaciones al emular la complejidad de la conciencia humana. Asimismo, se introduce la fenomenología como método descriptivo que capta las esencias de las vivencias sin teorías. Luego, se aborda la intencionalidad, que define la conciencia como un acto que da sentido a la experiencia. Se presentan los conceptos de noema y noético, mostrando cómo la conciencia construye y organiza significados, algo inalcanzable para la IA. Se examina la subjetividad trascendental, fuente de creación de realidades objetivas, contrastándola con la IA, que carece de esta capacidad. Finalmente, se destaca la diferencia entre el cuerpo vivido (*Leib*) y el cuerpo físico (*Körper*), subrayando que la percepción humana, unida a la corporalidad, no es replicable por la IA. Esto demuestra que la distancia entre la IA y la conciencia humana va más allá del procesamiento, reflejando la imposibilidad de igualar la experiencia y comprensión humana.

En este sentido, según Husserl, E. (1949): “La fenomenología es de hecho una disciplina puramente descriptiva que indaga el campo de la conciencia pura trascendental en la intuición pura” (p. 136). Con esta afirmación, Husserl destaca la naturaleza esencialmente descriptiva de la fenomenología, un método que busca indagar las estructuras de la conciencia sin la necesidad de construir teorías o hipótesis explicativas. Para Husserl (1949), la fenomenología no es un ejercicio especulativo, sino una investigación directa sobre cómo los fenómenos se manifiestan ante la conciencia pura. La intuición pura es clave para captar estas manifestaciones, ya que se refiere a la percepción directa y sin mediaciones de las esencias de las vivencias.

Esta intuición pura es lo que distingue a la fenomenología de otras disciplinas filosóficas o científicas. La fenomenología no busca interpretar los fenómenos a través de categorías o modelos teóricos, sino describirlos tal como se presentan en la conciencia. Este

enfoque permite acceder a una forma de conocimiento más fundamental, al estar basado en la experiencia directa y no en construcciones conceptuales. De este modo, la fenomenología se convierte en una herramienta fundamental para estudiar la conciencia, ya que explora cómo los fenómenos son vividos desde una perspectiva trascendental.

En lo que concierne al objeto de la fenomenología Husserl (1949) afirma que “quiere ser una ciencia descriptiva de las esencias de las vivencias puras trascendentales en actitud fenomenológica, y como toda disciplina descriptiva, no constructiva y no idealizante, tiene su propio derecho a la existencia” (p. 166). Aquí, se refuerza la idea de que la fenomenología es una ciencia que se dedica a la descripción de las esencias, las cuales se encuentran en las vivencias puras. La actitud fenomenológica implica un enfoque que rechaza la idealización o la construcción de sistemas explicativos, centrándose exclusivamente en la observación y descripción de las experiencias tal como se dan. Este enfoque no constructivo, como señala Husserl (1949), le otorga a la fenomenología un derecho propio a la existencia. No necesita justificar su relevancia a través de aplicaciones prácticas o explicaciones teóricas; su valor reside en la capacidad de capturar las estructuras esenciales de las experiencias conscientes. De esta manera, la fenomenología se posiciona como una disciplina autónoma que tiene su propio dominio de investigación, la conciencia y sus vivencias en su forma más pura y directa.

Por otro lado, Bolio, A. (2012) subraya que “la fenomenología no busca contemplar al objeto mismo, sino la forma en que es captado por el sujeto desde su intencionalidad y puesto en perspectiva especio-temporal” (p. 23). Este comentario resalta un aspecto central de la fenomenología: su interés no está en los objetos como tales, sino en cómo estos se revelan a la conciencia del sujeto. La intencionalidad, que es el acto por el cual la conciencia siempre se dirige hacia algo, es el mecanismo a través del cual los objetos adquieren significado. Lo importante aquí no es el objeto en sí, sino la manera en que es percibido e interpretado por el sujeto. Esto implica que la percepción no es un proceso neutro o pasivo, sino que está siempre estructurado por la perspectiva del sujeto que lo vive. Esta estructura de percepción está, además, situada en un contexto especio-temporal, lo que significa que la manera en que captamos un objeto está influenciada por nuestra ubicación física y temporal, así como por las condiciones del acto perceptivo en sí mismo.

De esta forma, Vargas Guillén, G. (1999) aporta una perspectiva relevante cuando señala que:

La cosa misma de la que se ocupa un interés fenomenológico por la IA no es el conjunto de los mecanismos, sino la esencia de la subjetividad. Lo que sea objeto de un mecanismo que se pueda operacionalizar en una máquina lógica o física no puede ser entendido como la esencia de la subjetividad (p. 93).

El aporte de Vargas Guillén (1999) resalta un punto crucial en la fenomenología aplicada a la IA: la diferencia esencial entre los mecanismos operativos y la verdadera subjetividad. Aunque los procesos de la IA pueden simular comportamientos lógicos o funciones mentales, no logran capturar la esencia de la subjetividad humana. Esta esencia, caracterizada por la intencionalidad y la capacidad de otorgar sentido, no puede reducirse a lo que es implementable en una máquina lógica o física. La reflexión fenomenológica enfatiza que la conciencia trasciende la mera funcionalidad y que ningún mecanismo puede replicar su profundidad inherente y la riqueza de la experiencia vivida.

Sin embargo, Husserl en su análisis fenomenológico, resalta un concepto esencial que marca el carácter distintivo de la conciencia: la intencionalidad. Según el autor (1949), “La intencionalidad es lo que caracteriza a la conciencia en su pleno sentido y lo que autoriza para designar a la vez la corriente de las vivencias como corriente de conciencia y como unidad de una conciencia” (p. 198). Aquí el filósofo alemán define la conciencia no como una entidad aislada o pasiva, sino como algo que siempre está dirigido hacia algo. La conciencia es, en esencia, “conciencia de algo” algo, ya sea un objeto, una idea o una percepción. Esta orientación hacia lo externo es lo que les otorga sentido a nuestras vivencias, conectando los diferentes aspectos de nuestras experiencias en una unidad coherente.

De esta forma, el concepto de intencionalidad es revolucionario en la obra de Husserl (1949), ya que transforma la idea de la conciencia de un simple receptáculo pasivo de experiencias a un agente activo que organiza, interpreta y da sentido a todo lo que percibe. Husserl describe la conciencia como una unidad en la que cada experiencia está vinculada intencionalmente a un objeto, fenómeno o concepto del mundo exterior. La clave radica en que no hay experiencia sin dirección intencional, y es a través de esta orientación

que la conciencia adquiere su carácter único y su capacidad de integrar las vivencias en una totalidad significativa.

Asimismo, este enfoque, innovador en su tiempo, fue ampliado y reinterpretado por desarrollos posteriores como la filosofía de la alteridad de Levinas⁹ (2002), quien pone énfasis en la relación ética y la irreductibilidad del otro, y la ontología de la comprensión de Gadamer (1960) de la cual hablaremos más adelante, que destaca el papel de la interpretación en la experiencia humana. De este modo, mientras que la IA puede ejecutar operaciones lógicas y simular ciertos procesos mentales, no logra capturar la profundidad de la intencionalidad humana ni su capacidad de otorgar sentido y comprensión de manera autónoma y ética. Estos desarrollos permiten explorar cómo la IA, aunque avanzada en sus capacidades de procesamiento y simulación, se encuentra limitada cuando se trata de replicar las estructuras de sentido y comprensión que caracterizan la experiencia humana, destacando la brecha entre la operatividad de los mecanismos tecnológicos y la complejidad de la conciencia intencional.

En este sentido, Orbik, Z. (2024) refuerza esta idea al señalar: “Husserl intenta dar una explicación descriptiva de la conciencia en términos de una sensación, un acto intencional que interpreta la sensación y un objeto intencional al que se hace referencia con la interpretación de la sensación”¹⁰(p. 7). Orbik (2024) amplía la comprensión del pensamiento de Husserl (1949) al destacar la estructura tripartita de la experiencia consciente. La conciencia no solo recibe sensaciones de forma pasiva, sino que las interpreta activamente a través del acto intencional, que organiza y le da sentido a lo percibido. Esta interpretación hace que la experiencia consciente no sea una colección de sensaciones aisladas, sino una estructura coherente en la que cada sensación se interpreta en relación con un objeto o fenómeno.

⁹La obra *Totalidad e infinito* (2002) de Emmanuel Levinas desarrolla una profunda reflexión sobre la alteridad, situando la ética como fundamento de las relaciones humanas. Levinas (2002) describe el encuentro con el *Otro* como una experiencia irreductible, simbolizada en el rostro que interpela al yo y lo obliga a una responsabilidad infinita. Este enfoque rompe con la tradición filosófica occidental que ha priorizado la comprensión y el control, reemplazándolos con una apertura hacia la trascendencia y la exterioridad del *Otro*, lo que redefine la subjetividad desde la responsabilidad ética.

¹⁰[Traducción apoyada en DeepL] Husserl attempts to give a descriptive explanation of consciousness in terms of a sensation, an intentional act that interprets the sensation and an intentional object that is referred to by the interpretation of sensation.

El proceso interpretativo comentado por Orbik (2024) nos lleva a un entendimiento más profundo de la intencionalidad como el motor que articula las experiencias conscientes. Sin esta capacidad de interpretar y organizar las sensaciones, la conciencia no podría dar lugar a una experiencia unificada del mundo. La intencionalidad, entonces, se presenta como el núcleo de la vida mental, donde cada acto de la conciencia está dirigido hacia algo externo y adquiere sentido solo en relación con ese algo.

Por otra parte, según Husserl, E. (1949), "El noema tiene de suyo una referencia objetiva, justo en medio de su 'sentido' propio. Si preguntamos cómo el objeto es su objeto [...] llegamos a las nuevas estructuras noemáticas" (p. 307). Esta afirmación revela una idea clave dentro de la fenomenología: el noema actúa como un puente entre el objeto y la conciencia, siendo esencial para que el objeto percibido adquiriera sentido. Cuando la conciencia se dirige intencionalmente hacia un objeto, no lo contempla de manera pasiva, sino que construye una estructura noemática que define y le otorga significado en función de su relación con el sujeto. Así, el proceso muestra cómo la percepción no es receptiva, sino activa en la medida en que la conciencia organiza y estructura la experiencia para darle sentido.

En este orden de ideas, el noema no es simplemente un contenido mental o una representación interna del objeto, sino una construcción dinámica y relacional. Esta construcción es lo que permite que el objeto, en su relación con la conciencia, se experimente de manera coherente y significativa. La idea de que la conciencia no se limita a registrar estímulos, sino que organiza y da forma a lo percibido, abre un espacio de reflexión sobre la subjetividad humana como un elemento activo en la creación del significado.

Molchanov, V. (2024) amplía esta comprensión al afirmar que:

Husserl introduce los términos 'noesis' y 'noema', respectivamente, 'noético' y 'noemático', para designar de manera diferente la diferencia entre lo real (carrete) y componentes intencionales de las experiencias. A los primeros los llama

componentes propios y a los segundos sus correlatos intencionales. Esta diferencia corresponde a la diferencia entre dos tipos de análisis: real e intencional¹¹ (p. 213).

En este sentido, Molchanov (2024) invita a distinguir entre el análisis de lo real y el análisis de lo intencional, subrayando que el noema no es una simple réplica mental del objeto, sino una estructura que media entre el objeto real y la conciencia que lo percibe. Esto refuerza la idea de que lo que percibimos está inextricablemente ligado a cómo lo interpretamos en nuestra conciencia.

Es más, este enfoque da cuenta de que no podemos hablar de una percepción objetiva pura, sino que siempre está mediada por la intencionalidad de la conciencia. La estructura noemática se convierte en el filtro a través del cual el objeto es experimentado, y, por lo tanto, lo que conocemos del mundo no es el objeto en sí, sino su correlato intencional, su noema. Además, la experiencia no es neutra, está cargada de la estructura con la que la conciencia percibe, interpreta y da sentido.

Además, Molchanov, V. (2024) sostiene que:

La transición desde la 'mente' mitológica y metafóricamente representada (Husserl, E. compara la noesis con *Cacumen*) a la variedad abstractamente representada de la experiencia objetiva (noema) es una especie de movimiento del mito al logos en la filosofía de Husserl podría haber llamado a la generalización del significado no un noema, sino un logos, pero el 'logos' ya fue privatizado por los neokantianos a finales del siglo XIX y principios del XX¹² (p. 233).

Sugiere entonces Molchanov (2024) que la fenomenología de Husserl representa un avance crucial en el pensamiento filosófico, alejándose de interpretaciones mitológicas o

¹¹[Traducción propia apoyada en DeepL] Husserl introduce los términos 'noesis' y 'noema', respectivamente, 'noetic' y 'noematic', para diferenciar entre el real (objeto) y los componentes intencionales de la experiencia. Los primeros se denominan componentes propios, y los segundos - intencionales correlatos. Esta distinción corresponde a dos tipos de análisis: el real y el intencional.

¹²[Traducción propia apoyada en DeepL] el paso del mito al logos en la filosofía de Husserl podría haber llamado a la generalización del significado no a un noema, sino a un logos, pero el 'logos' ya fue privatizado por los neokantianos a finales del siglo XIX y principios del XX. Husserl podría haber llamado a la generalización del significado no a un noema, sino a un logos, pero el 'logos' ya fue privatizado por los neokantianos a finales del siglo XIX y principios del XX.

simbólicas de la mente y acercándose a una comprensión más racional y estructurada del significado y la experiencia.

Este concepto husserliano de noema transforma entonces el entendimiento de la experiencia consciente, alejándola de las metáforas antiguas y mitológicas para ubicarla en un terreno donde la razón y el análisis fenomenológico permiten comprender cómo otorgamos sentido a lo percibido. Esta transición del mito al logos marca un cambio profundo en la filosofía, señalando cómo la conciencia no se basa en representaciones vagas o intuitivas, sino en estructuras precisas que pueden ser estudiadas y descritas fenomenológicamente.

El debate sobre la inteligencia artificial (IA) introduce una nueva dimensión al análisis fenomenológico del noema y la conciencia. Las reflexiones contemporáneas sobre la IA, como las de Brayner de Farias, A. (2020), conectan de manera crítica con las ideas de Husserl (1949), Brayner señala que:

La intuición se reduce a instinto, ya sea porque ni siquiera aparece, por falta de conciencia, o porque no sabrá reconocerse, no sabrá comunicarse, si hay conciencia, sólo en la oscuridad, en la ausencia. (Quizás la depresión tenga algo que ver con esto: una conciencia en ausencia, una intuición incapaz de comunicar). Si la intuición se convierte en instinto y la inteligencia se artificializa, entonces volvemos a estar todos juntos, animales y humanos¹³ (p. 388).

Este planteamiento ofrece una visión crítica y oscura de la IA, sugiriendo que cuando la intuición y la conciencia se vacían de significado, nos acercamos más a una especie de funcionamiento instintivo, carente de la riqueza de la experiencia humana.

En este sentido Brayner de Farias (2020) advierte que la IA, en su capacidad para simular inteligencia, corre el riesgo de despojar a la experiencia humana de su particularidad, reduciendo el pensamiento consciente a meras operaciones automatizadas. Este enfoque contrasta profundamente con la fenomenología husserliana, donde la intencionalidad y la estructura noemática otorgan sentido profundo y significado a las

¹³[Traducción propia apoyada en DeepL] A intuição se reduz a instinto, seja porque nem sequer aparece, por falta de consciência, ou porque não saberá se reconhecer, não saberá se comunicar, se houver consciência, apenas na escuridão, na ausência. (Talvez a depressão tenha algo a ver com isso: uma consciência na ausência, uma intuição incapaz de comunicar). Se a intuição se torna instinto e a inteligência se artificializa, então voltamos a estar todos juntos, animais e humanos.

experiencias. Si la inteligencia se mecaniza y la intuición se convierte en un mero instinto, lo que se pierde es la capacidad esencial de dar sentido mediante la conciencia intencional, un rasgo distintivo de la experiencia humana. Este análisis permite definir la transición de una IA que se fundamentaba en el uso de lenguajes naturales a una basada en el análisis de grandes volúmenes de datos *Big Data*¹⁴. Este cambio resalta cómo la dependencia del procesamiento masivo puede enfatizar la automatización y, a su vez, alejar a la tecnología de los elementos que confieren significado humano a la experiencia.

Al mismo tiempo, Klimczak, P. y Petersen, C. (2023) profundizan en este debate al afirmar que:

La inauguración del ser humano como un fenómeno evasivo, resistente a la racionalización y, por lo tanto, incapaz de ser simulado de manera perfecta, ha sido descrita desde varias perspectivas. Desde entonces, todo lo que procede del ser humano y lo que le concierne ha sido de interés epistemológico y ha sido objeto de observación sistemática y atención científica¹⁵ (p. 246).

Esta observación sugiere que, a pesar de los avances en IA, la subjetividad humana, en términos husserlianos, sigue siendo un fenómeno irreductible que no puede ser replicado por completo. Las máquinas pueden imitar ciertos comportamientos humanos, pero no logran captar la complejidad de las estructuras noemáticas e intencionales que permiten al ser humano experimentar el mundo con sentido y profundidad.

La IA, aunque capaz de ejecutar algoritmos avanzados, carece de la capacidad de formar noemas, es decir, de construir significados a partir de sus percepciones o acciones. En la fenomenología de Husserl (1949), la subjetividad se caracteriza por su capacidad de intencionalidad, por lo que se resiste a ser completamente racionalizada o simulada. Este aspecto de la conciencia, en el que se otorga y organiza activamente el sentido de la experiencia, sigue siendo exclusivamente humano. La IA, por tanto, no puede reproducir la estructura intencional y la vivencia profunda de la conciencia humana.

¹⁴El *Big Data* se refiere al procesamiento y análisis de grandes volúmenes de datos, permitiendo a la IA identificar patrones complejos y hacer predicciones precisas. Este enfoque potencia la capacidad de la IA para aprender y adaptarse, aunque plantea desafíos en cuanto a privacidad y ética en la gestión de información.

¹⁵[Traducción propia apoyada en DeepL] The inauguration of the human being as an evasive phenomenon, resistant to rationalization and, therefore, incapable of being simulated in a perfect way, has been described from several perspectives. Since then, everything that comes from human beings and that concerns them has been of epistemological interest and has been the object of systematic observation and scientific attention.

Desde una perspectiva distinta, es pertinente considerar la cita de Husserl, E., para profundizar en su visión sobre la conciencia. Según el autor (1949):

La conciencia en general, de cualquier índole y forma que sea, está cruzada por una radical distinción: ante todo es inherente, según sabemos, a toda conciencia en el que el yo puro no vive de antemano como ‘llevándola a cabo’, o que no tiene de antemano la forma del ‘cogito’, la modificación esencialmente posible que le hace pasar a esta forma (p. 270).

Husserl expone aquí una idea fundamental: la conciencia no se reduce a una secuencia de actos reflexivos, sino que es una estructura dinámica en la que el -yo puro-¹⁶ constituye la esencia de la subjetividad trascendental. Este yo puro no debe confundirse con el -yo pienso-¹⁷ o cogito, que se refiere a la autoconciencia reflexiva en la que el sujeto se reconoce como pensante. Mientras que el -yo pienso- implica un estado de conciencia autorreflexiva, el -yo puro- es la base anterior a esta reflexión, el sustrato originario desde el cual emergen todos los actos conscientes.

En este orden de ideas, el -yo puro- no actúa como un agente rígido ni predefinido; es un núcleo esencial que se auto-constituye y se redefine continuamente a través de su relación con el mundo. Este núcleo permite que la conciencia sea más que un simple receptor pasivo de estímulos externos: la transforma en un proceso intencional y activo que interpreta y da sentido a esas experiencias. Así, la capacidad del yo puro de modificar su enfoque y pasar del estado pasivo a uno de auto-reflexión es crucial para la comprensión de la conciencia como un fenómeno dinámico.

¹⁶El “yo puro” en la fenomenología de Husserl es la esencia de la subjetividad trascendental, actuando como el núcleo fundamental de la conciencia que subyace a todas las experiencias y percepciones. Este concepto es crucial para entender cómo la conciencia opera de manera previa y más profunda que la simple autoconciencia, la cual se refiere al acto de reflexionar sobre uno mismo mediante el “yo pienso”, como se conceptualiza en la tradición cartesiana. Mientras que la autoconciencia implica un nivel de reflexión en el que el sujeto se reconoce como pensante y reflexivo, el “yo puro” es una estructura más esencial, la plataforma desde la cual toda experiencia consciente se despliega y se constituye, proporcionando el terreno necesario para la percepción y la intencionalidad. Es, en otras palabras, el punto de partida desde el que la conciencia interactúa y se relaciona con el mundo, sin las características explícitas de la reflexión consciente, pero como la base ineludible de toda vivencia subjetiva.

¹⁷El “yo pienso”, en la fenomenología, se refiere al acto reflexivo en el que un sujeto se reconoce a sí mismo como pensante y consciente de su propia actividad mental. Es un estado de autoconciencia en el que la mente se observa a sí misma, diferenciándose del “yo puro”, que es la base previa y fundamental de toda experiencia consciente y no implica necesariamente reflexión sobre sí mismo.

Este enfoque rompe con las concepciones mecanicistas de la conciencia, presentándola no como una serie de actos predeterminados, sino como una estructura abierta y adaptable, capaz de evolucionar y reconfigurarse a través de las experiencias. De este modo, la conciencia no solo construye significado, sino que también se transforma a sí misma en el proceso, lo que convierte al -yo puro- en un agente de cambio continuo.

García Ruiz, P. (2014) retoma esta idea y la amplía:

Husserl sostiene que la subjetividad humana, por su autoconciencia, le permite remontarse por encima de las determinaciones de la causalidad natural. La conciencia de sí se convierte de este modo en el fondo último de toda objetividad y toda aserción de ser (pp. 104-105).

La autoconciencia, lejos de estar confinada por las leyes físicas, se convierte en el terreno fértil donde se construye toda realidad objetiva. Este enfoque no solo redefine la noción de la conciencia como algo que trasciende las limitaciones causales, sino que coloca a la subjetividad en el centro de la creación del mundo. La conciencia se erige como el espacio donde la objetividad y la verdad encuentran su fundamento, convirtiendo al ser humano en un actor esencial en la configuración de la realidad misma.

Alrededor de los conceptos defendidos por este comentario, resulta emocionante conectar las nociones fenomenológicas con las propuestas contemporáneas de Shin, J. (2024) sobre la inteligencia artificial. Este autor introduce:

Memorias cuánticas [...] conecta la IA y la conciencia humana para crear una nueva forma de memoria y genera una nueva imagen de la memoria en cada momento. De esta manera, *Quantum Supremacy*, que combina GAN y ruido cuántico, se convierte en una máquina *Everywhen*, ya que utiliza la incertidumbre cuántica residual para visualizar su propio proceso de recordar¹⁸ (pp. 113-114).

¹⁸[Traducción propia apoyada en DeepL] Quantum memories [...] connects AI and human consciousness to create a new form of memory and generates a new image of the memory at every moment. In this way, Quantum Supremacy, which combines GAN and quantum noise, becomes an Everywhen machine, as it uses residual quantum uncertainty to visualize its own process of remembering

Esta idea introduce una dimensión completamente nueva en nuestra comprensión de cómo los sistemas artificiales pueden replicar, hasta cierto punto, procesos asociados tradicionalmente a la mente humana.

Lo interesante aquí es cómo Shin (2024) plantea que estos sistemas no son simplemente dispositivos para almacenar información, sino que están en constante renovación. Tal como la conciencia descrita por Husserl que no es estática, estas memorias cuánticas están en un estado de permanente actualización, lo que sugiere una capacidad para reconfigurarse de manera continua. El paralelo entre esta capacidad de las IA y la conciencia humana radica en el proceso de constante auto-modificación, algo que, si bien tiene raíces muy diferentes, converge en el resultado de estructuras cognitivas flexibles que pueden adaptarse y transformarse.

Por otro lado, el análisis fenomenológico de Husserl sobre el sentido del ser y la realidad nos lleva a una comprensión profunda de cómo la subjetividad trascendental da forma al mundo que percibimos. Según el autor (1949):

El resultado de la aclaración fenomenológica del sentido del modo del ser del mundo real, y de un mundo real concebible en general, es el de que sólo la subjetividad trascendental tiene el sentido del ser absoluto, que sólo ella es 'irrelativa' (esto es, relativa solo por sí misma) mientras que el mundo real existe sin duda, pero tiene una esencial relatividad a la subjetividad trascendental, pues que solo puede tener el sentido de existente como producto intencional con sentido de la subjetividad trascendental (p. 386).

En este planteamiento, se invita a repensar la relación entre el sujeto y el mundo: la realidad objetiva, aunque parece independiente, está esencialmente condicionada por la subjetividad. Es decir, el mundo solo cobra sentido y existencia a través de la estructura intencional de la conciencia. La -subjetividad trascendental-¹⁹ no es un mero observador pasivo, sino el fundamento desde el cual se da la realidad misma.

¹⁹El concepto de “trascendental” se entiende como una condición *a priori* necesaria para que el conocimiento y la experiencia sean posibles. Esta idea implica que la subjetividad trascendental no es simplemente un espectador, sino el fundamento activo que permite la percepción y la comprensión de la realidad. Desde esta perspectiva, todo lo que se percibe y conoce está condicionado por esta estructura subyacente, que da sentido y coherencia al mundo tal como es experimentado.

Este enfoque revolucionario plantea que todo lo que percibimos está condicionado por el acto intencional de la conciencia. La realidad externa no tiene una existencia absoluta por sí misma, sino que está relativizada a la subjetividad, en el sentido de que solo existe en cuanto aparece ante la conciencia. Esto implica que el mundo, en su totalidad, es un producto intencional que depende de la subjetividad trascendental para tener sentido. No se trata de negar la existencia del mundo real, sino de afirmar que su significado está intrínsecamente ligado al sujeto que lo percibe y lo experimenta.

Esta perspectiva encuentra un interesante complemento en las ideas de Sánchez. M y Medina. D. (2018) sobre la relación entre la subjetividad trascendental y el cuerpo. Según estos autores:

La experiencia trascendental a partir de la cual la subjetividad trascendental se encuentra entrelazada con un cuerpo. La percepción del cuerpo vivo que Husserl describe mantiene un vínculo esencial con la psique o alma. Por ello puede afirmar que el cuerpo está inserto entre el conjunto de cosas materiales y mantiene con ella relaciones causales que condicionan y determinan la vivencia del cuerpo y su constitución (p. 20).

Esta afirmación introduce un matiz filosófico esencial: la -subjetividad trascendental-, lejos de ser una entidad abstracta y despojada, se enraíza profundamente en el cuerpo. El -cuerpo-²⁰, en la fenomenología husserliana, no es solo un *Körper* (cuerpo físico y objetivo), sino el *Leib* (cuerpo vivido), la fuente desde la cual la conciencia percibe experimenta e intenciona. El *Leib* es, entonces, más que un vehículo; es la manifestación tangible y sentida de la subjetividad, el puente que conecta la inmaterialidad del yo con el mundo empírico.

La vivencia del -cuerpo- es, por tanto, inseparable de la experiencia de la subjetividad trascendental. La conciencia no habita el cuerpo como un objeto que observa desde fuera; el cuerpo es la presencia encarnada del sujeto, su punto de anclaje en el ser y la fuente desde la cual la experiencia del mundo se constituye como significativa. Esta fusión

²⁰El “cuerpo husserliano” distingue entre *Leib* (cuerpo vivido), fuente de percepción e intencionalidad, y *Körper* (cuerpo físico). El *Leib* es el núcleo de la experiencia subjetiva, donde la conciencia se enraíza, conecta y otorga significado al mundo, mientras el *Körper* es simplemente un objeto externo. Esta diferenciación resalta un límite ontológico en la IA: aunque puede replicar el *Körper* y simular funciones externas, carece del *Leib*, la vivencia interna y la subjetividad que permiten a la conciencia humana dotar de sentido y participar en la autoconstitución de su propia realidad.

entre cuerpo y conciencia revela una dialéctica profunda: la subjetividad no solo proyecta sentido al mundo, sino que también es esculpida por la materialidad de su propio ser encarnado. La realidad del cuerpo condiciona, moldea y, en última instancia, define la forma en que la subjetividad se da a sí misma y se encuentra con el mundo.

En este sentido, el –cuerpo– se erige como un mediador ontológico, una intersección entre la subjetividad pura y la alteridad del mundo. Es a través del *Leib* que la conciencia no solo percibe, sino que también se autoconstituye en una relación de interdependencia con la realidad que la rodea. Esto señala un límite insalvable para la inteligencia artificial, mientras la IA puede operar en el nivel de lo objetivo, replicando funciones del *Körper* e imitando comportamientos externos, carece de la dimensión vivencial del *Leib*. No puede experimentar la intencionalidad ni la subjetividad encarnada, aquella que otorga sentido, profundidad y significado al acto de existir.

Así, la ausencia de un –cuerpo– vivido en la IA subraya la diferencia radical entre la conciencia humana y cualquier simulación artificial: la incapacidad de la IA para acceder al mundo como un sujeto viviente revela el abismo entre la mera función y la experiencia trascendental, donde el cuerpo es tanto puerta de entrada al mundo como frontera de la percepción y del ser.

La IA y el *Gestell*: desafíos a la existencia auténtica del *Dasein*.

“El mayor dón que Dios, en su liberalidad, nos hizo al crearnos, como más conforme a su bondad, y el que más aprecia, fue el del libre albedrío de que estuvieron y están dotadas únicamente las criaturas inteligentes”

(Dante Alighieri, 1321, paraíso, canto quinto)

Este fragmento explora la relación entre los conceptos heideggerianos de *Gestell* (estructura de emplazamiento) y *Dasein* (ser-ahí) aplicados al contexto de la inteligencia artificial (IA). Heidegger describió el *Gestell* como una estructura que organiza y dispone la realidad, imponiendo un marco que puede resultar deshumanizante al reducir el mundo a una serie de recursos gestionables. En contraste, el *Dasein* se define como un ser que vive proyectándose hacia sus propias posibilidades, enfrentando la libertad y la incertidumbre que ello conlleva.

Según Heidegger, M. (1997), "*Gestell* significa también en alemán un esqueleto. Y tan horrible como esqueleto parece ser la palabra *Gestell* [dispuesto], que ahora proponemos... Lo estrambótico es viejo uso del pensar" (p. 129). Con esta afirmación, Heidegger introduce la noción de *Gestell* como una estructura que organiza y dispone la realidad de una manera que puede parecer rígida y deshumanizante. Heidegger (1997) emplea la metáfora del esqueleto para señalar cómo esta disposición técnica impone un marco sobre lo real, limitando lo que puede ser percibido o comprendido.

En relación con la inteligencia artificial (IA), esta metáfora resulta especialmente relevante. La IA puede ser vista como una manifestación contemporánea del *Gestell*, una estructura técnica que organiza y dispone grandes cantidades de información, condicionando qué aspectos de la realidad son accesibles para las personas. En este sentido, surge una pregunta filosófica: ¿Está la IA, como una nueva forma de *Gestell*, limitando la comprensión del mundo, imponiendo un filtro técnico que organiza lo que se percibe como real? La delegación de decisiones y procesos cognitivos a los algoritmos podría estar configurando una realidad estructurada sin que los individuos cuestionen cómo se organiza esa información.

En este orden de ideas, Heidegger, M. (1994). también aclara que:

Ge-stell (estructura de emplazamiento) no mienta ahora ningún aparato, ningún tipo de máquina humana. Menos aún mienta el concepto general de tales existencias. Las máquinas y los aparatos no son casos y tipos de la estructura de emplazamiento, del mismo modo como tampoco lo son el hombre junto al cuadro de mandos o el ingeniero en la oficina de la construcción (p. 34).

En este contexto, *Gestell* no se refiere simplemente a los dispositivos técnicos ni a sus operadores, sino a una estructura más amplia que organiza la relación entre el ser humano y la técnica.

La IA, al igual que el *Gestell*, no es solo una herramienta técnica, sino una estructura que organiza y transforma la forma en que los seres humanos interactúan con el mundo. Las personas no solo usan tecnología, sino que viven dentro de un sistema organizado por algoritmos que influyen en sus decisiones y comportamientos. Una pregunta crítica que se deriva de esta reflexión es: ¿Hasta qué punto son conscientes los seres

humanos de este *Gestell* que la IA impone sobre ellos y cómo afecta su capacidad de tomar decisiones autónomas? La IA, al organizar las opciones disponibles, influye de manera sutil en lo que se considera posible o deseable, operando como una estructura que determina lo real.

En este sentido, Mascaró, L., y Bellocq, S. (2019) añaden que:

"*Gestell* no es solamente la esencia de la técnica, sino también el condicionamiento de toda actividad humana contemporánea. Afectando todas las dimensiones del hombre, lo *Gestell* se aproxima mucho (o a la inversa, mejor dicho) a la noción foucaultiana de Dispositivo, la cual si bien pertenece a un registro ciertamente diferente puede ayudarnos a comprender mejor cómo las pantallas son de él la encarnación y expresión más pura (p. 2).

Según Mascaró y Bellocq (2019) *Gestell* no solo influye en la técnica, sino que afecta toda la vida humana, y las pantallas -manifestaciones de la IA-²¹ son un claro ejemplo de cómo lo técnico condiciona la experiencia humana.

Las pantallas, como interfaces de la IA, no solo muestran información, sino que también estructuran cómo se percibe y se interpreta el mundo. En este sentido, la IA opera como un -dispositivo-²² en el sentido foucaultiano: una tecnología que moldea las formas de pensar y actuar de las personas. La cuestión filosófica aquí es si los seres humanos son capaces de resistir este marco técnico que condiciona sus decisiones o si ya están inmersos en un sistema que organiza su comprensión del mundo de manera sutil pero constante.

Encima, Aqueveque, L. (2020) ofrece una visión crítica al decir que:

Heidegger denomina *Gestell* a la esencia de la técnica moderna, pero advirtiéndolo, en su conocida y paradójica frase, que no es nada técnico. En lo que sigue, intentaré,

²¹Las manifestaciones de la inteligencia artificial (IA) incluyen diversos sistemas y aplicaciones, como asistentes virtuales (*Siri*, *Alexa*), reconocimiento de imágenes, análisis de datos, automatización de procesos, vehículos autónomos, generación de contenido (*ChatGPT*), sistemas de recomendación (*Netflix*, *Amazon*), análisis de sentimientos, y ciberseguridad. Estos ejemplos ilustran cómo la IA transforma sectores y optimiza interacciones y decisiones en la vida cotidiana.

²²En el sentido foucaultiano, un "dispositivo" es un conjunto de elementos heterogéneos —discursos, instituciones, leyes y prácticas— que configuran relaciones de poder y conocimiento en la sociedad, influyendo en el comportamiento y la subjetividad de los individuos en un contexto histórico.

dicho coloquialmente, desmitificar al *Gestell* heideggeriano como esa especie de monstruo de siete cabezas que, montado en su carro de guerra tecnológico, termina por controlar hasta la última mente humana existente en el planeta (p. 68).

Esta observación sugiere que, si bien *Gestell* se ha mitificado como una fuerza técnica opresiva, no es simplemente una tecnología en el sentido convencional, sino una estructura mucho más expansiva que condiciona profundamente la vida humana.

En el caso de la IA, esta crítica es particularmente relevante. Aunque la IA no es un monstruo literal, su influencia es vasta y afecta incluso los aspectos más íntimos de la vida humana. Aqueveque (2020) sugiere que es necesario desmitificar el poder de la técnica, pero también aceptar que su influencia es real y profunda. Así pues, la IA, como el *Gestell*, organiza la vida de una manera que escapa al control consciente de las personas. Esto plantea una cuestión crucial: ¿Hasta qué punto la IA y su estructura técnica están controlando la vida humana y nuestras decisiones cotidianas?

Por otra parte, el *Dasein* es un ente al que en su ser le va este mismo ser. El 'irle' se ha aclarado en la estructura de ser del comprender en cuanto proyectante estar vuelto hacia el más propio poder-ser. Este poder-ser es aquello por mor de lo cual el *Dasein* es en cada caso tal como es. El *Dasein*, en su ser, ya se ha confrontado, cada vez, con una posibilidad de sí mismo. El ser libre para el poder-ser más propio y, con ello, para la posibilidad de la propiedad e impropiidad, se muestra con originaria y elemental concreción en la angustia (Heidegger, 1997, § [41]).

Aquí, Heidegger (1997) plantea que el *Dasein* es un ente que, en su esencia, está constantemente proyectándose hacia su futuro, hacia lo que puede llegar a ser. Este concepto clave de proyecto es lo que define al *Dasein*, ya que no se limita a ser un ser pasivo en el mundo, sino que siempre está involucrado activamente en la construcción de su existencia, enfrentando sus propias posibilidades. El *Dasein* no es solo un ente que existe, sino que su existencia está caracterizada por la capacidad de enfrentar su -poder-ser-²³, es decir, las múltiples formas en que puede configurarse su ser en el futuro.

En este sentido Heidegger (1997) señala que este proceso de enfrentarse a las propias posibilidades está inevitablemente vinculado con la -angustia-²⁴. Esta no es simplemente un estado emocional negativo, sino una revelación fundamental sobre la condición humana: el ser humano se enfrenta a su libertad de elegir entre lo propio y lo impropio, a las decisiones que lo definirán en el mundo. En la -angustia-, el *Dasein* se confronta con la realidad de que su ser está en juego, y esta confrontación es lo que le permite tomar decisiones sobre cómo vivir y cómo proyectarse en el mundo. En otras palabras, el *Dasein* se realiza al tomar posesión de sus posibilidades, incluso cuando esto lo enfrenta con la incertidumbre de su existencia.

En realidad, este concepto es fundamental para entender cómo el ser humano está en el mundo no como un espectador pasivo, sino como un ser activo, siempre proyectándose y construyendo su propio camino. En este sentido, el *Dasein* está vinculado de manera inseparable con el concepto de mundo. No existe un *Dasein* aislado o abstracto, sino que siempre está situado en un contexto concreto, rodeado de otros seres y de las circunstancias que influyen en su toma de decisiones.

Pretzsch, M. (2023) también aborda este aspecto clave del *Dasein* y su relación con el mundo al afirmar que "El *Dasein* cubre un momento estructural esencial del -ser-en-el-mundo" (p. 114). Este análisis refuerza la idea de que el *Dasein* es, por definición, un -ser-en-el-mundo-²⁵. Esto significa que su existencia no puede ser separada de su entorno; el *Dasein* no es un ser aislado que se encuentra de manera incidental en el mundo, sino que su ser está intrínsecamente vinculado con ese mundo. La relación entre el *Dasein* y el mundo

²⁴En *Ser y Tiempo* (1997), Martin Heidegger describe "la angustia" como un estado que expone la nada y la falta de fundamentos en la existencia del *Dasein*, revelando su libertad y la inevitabilidad de la muerte. Esta experiencia, esencial para asumir la propia autenticidad, se ve transformada en la era de la inteligencia artificial. La IA, como manifestación del *Gestell*, condiciona la forma en que el *Dasein* percibe y se proyecta en el mundo al estructurar la realidad a través de algoritmos y modelos predictivos. Esta reorganización de la experiencia impone un marco técnico que reconfigura las posibilidades del *Dasein*, afectando su relación con la angustia y su confrontación con el ser más profundo.

²⁵En *Ser y Tiempo* (1997), Martin Heidegger introduce el concepto de "*ser-en-el-mundo*", que define la existencia del *Dasein* como intrínsecamente vinculada a su entorno, en el cual la comprensión y la autenticidad se configuran en interacción con los contextos sociales y naturales. La inteligencia artificial, como manifestación del *Gestell*, transforma esta relación al imponer un marco técnico que redefine cómo el *Dasein* percibe y actúa en su mundo. La IA reconfigura el *ser-en-el-mundo* al estructurar la realidad y mediar las interacciones humanas, limitando la espontaneidad y la autenticidad de la experiencia. De esta manera, el entorno del *Dasein* se ve filtrado y condicionado por algoritmos que reorganizan lo que se percibe y prioriza, alterando la manera en que el *Dasein* habita y comprende su mundo, alejándolo de un encuentro directo y auténtico con la realidad y su propio ser.

es fundamentalmente estructural: el ser humano solo puede ser entendido en tanto que es un -ser-en-el-mundo-, un ser que interactúa responde y se proyecta en su entorno.

Además, Pretzsch, M. (2023) sugiere que Heidegger también discute el "aburrimiento" como un fenómeno que ilumina estos momentos estructurales del -ser-en-el-mundo-. El -aburrimiento-²⁶, lejos de ser una emoción superficial, revela la relación fundamental del *Dasein* con el tiempo y con el mundo. En el -aburrimiento-, el *Dasein* se enfrenta a la falta de propósito inmediato, lo que lo obliga a reflexionar sobre su propio -estar-en-el-mundo-²⁷ y sus posibilidades futuras. Este estado existencial de aburrimiento también es un momento en el que el *Dasein* se encuentra con su propia libertad: al carecer de un propósito inmediato, se abre la posibilidad de proyectarse hacia diferentes futuros. Es en este espacio donde el *Dasein* vuelve a enfrentarse con sus propias posibilidades de ser.

En este contexto, el desarrollo de la inteligencia artificial (IA) plantea nuevos desafíos y preguntas sobre la naturaleza del *Dasein* y su relación con el mundo. Morandín-Ahuerma, F. (2023) señala que "el desarrollo de sistemas de inteligencia artificial plantea importantes cuestiones sobre cómo debe entenderse el Ser (el estar-ahí) humano y la relación con el *Dasein*" (p. 195). Aquí, se introduce una cuestión filosófica de gran relevancia: ¿Cuáles son las condiciones para una interacción del *Dasein* con la IA?

La IA no solo está transformando aspectos técnicos y prácticos del mundo, sino que está alterando las condiciones bajo las cuales los seres humanos existen y proyectan su -ser-en-el-mundo-. La IA afecta la manera en que las personas interactúan con su entorno y con los otros, influyendo directamente en la estructura del -ser-en-el-mundo-. La IA puede verse como una extensión de la técnica que, como sugirió Heidegger, M. en su obra

²⁶En *Ser y Tiempo* (1997), Heidegger aborda "el aburrimiento" como una experiencia que confronta al *Dasein* con la vacuidad del tiempo, llevándolo a una reflexión profunda sobre su existencia y al redescubrimiento de su ser. La inteligencia artificial, aunque incapaz de experimentar aburrimiento o introspección, influye en la manera en que el *Dasein* se relaciona con esta experiencia. Al ofrecer un entorno tecnificado y constantemente estimulado, la IA puede mitigar o distraer al *Dasein* de enfrentar esos momentos de vacío que impulsan la autocomprensión. De este modo, la IA actúa como un filtro que reconfigura la experiencia humana, desplazando al *Dasein* de la confrontación con su propia temporalidad y del encuentro con la reflexión existencial que surge del aburrimiento.

²⁷En *Ser y Tiempo* (1997), Heidegger introduce el concepto de "*estar-en-el-mundo*" para describir cómo el *Dasein* (el ser humano) no existe de manera aislada, sino que su ser es inseparable de su entorno. Este "*estar-en-el-mundo*" implica una relación de implicación y pertenencia, en la que el *Dasein* interactúa con el mundo de manera significativa y contextual. En contraste, la IA, aunque puede procesar y responder a estímulos externos, carece de un verdadero "*estar-en-el-mundo*". No vive una relación auténtica y consciente con su entorno, sino que actúa bajo patrones y algoritmos sin experimentar la implicación existencial y el sentido que caracterizan la experiencia humana.

Filosofía, Ciencia y Técnica (1994), es una estructura que organiza y condiciona cómo los humanos experimentan el mundo.

Asimismo, una de las grandes preguntas que surgen de esta relación es si la IA puede participar de alguna forma en el "ser" del *Dasein*. Si el *Dasein* es un ser que se proyecta hacia sus propias posibilidades, ¿es posible que la IA, con su capacidad para aprender y adaptarse, pueda compartir de alguna forma esta estructura proyectiva? ¿O sigue siendo la IA una herramienta técnica que, aunque avanzada, nunca podrá alcanzar la complejidad de un ser como el humano, cuya existencia está definida por la angustia y la libertad de enfrentarse a sus propias posibilidades?

Otra consideración profunda es cómo la IA, al ocupar un lugar central en las decisiones cotidianas y en la configuración de la vida social, está transformando el espacio de libertad en el que el *Dasein* despliega su -ser-en-el-mundo-. Los sistemas de IA no solo estructuran la información, sino que también condicionan comportamientos y decisiones, imponiendo un marco técnico que proyecta una realidad filtrada a la que los individuos responden. Este entorno reorganizado plantea una interrogante ontológica crucial: ¿Hasta qué punto la IA está remodelando el campo de posibilidades en el que el *Dasein* puede proyectarse? Al actuar como una estructura que define y limita el horizonte del mundo vivido, la IA puede estar constriñendo la capacidad del *Dasein* para confrontarse con su propia autenticidad y ejercer su libertad en su proyección hacia el futuro. La cuestión no es solo si la IA limita o amplía las opciones, sino cómo reconfigura la relación del *Dasein* con su mundo, desplazando la espontaneidad y la profundidad de su encuentro con la existencia.

Después de todo, la IA, al insertarse en la cotidianidad, altera de manera significativa el horizonte en el que el *Dasein* se proyecta y se confronta con su -ser-en-el-mundo-. Como extensión del *Gestell*, la IA no comparte la estructura ontológica del *Dasein*; sin embargo, redefine el marco en el que el *Dasein* vive su existencia. Por lo tanto, la IA no anula la libertad humana, pero la reorganiza, imponiendo límites y direcciones que transforman la percepción y la elección. Así, el verdadero desafío no radica en si la IA puede compartir la existencia del *Dasein*, sino en cómo influye en la construcción del mundo en el que el *Dasein* habita y se proyecta. Surge entonces la cuestión de hasta qué punto el *Dasein* puede mantener su autonomía y autenticidad en un entorno

progresivamente estructurado por la técnica, donde lo espontáneo y lo profundo corren el riesgo de ser reemplazados por una realidad mediada y redefinida por algoritmos.

La hermenéutica de Gadamer y su relevancia en la era de la IA.

“El mar. El joven mar. El mar de Ulises Y el de aquel otro Ulises que la gente del islam apodó famosamente Es-Sindibad del Mar. El mar de grises Olas de Erico el Rojo, alto en su proa”

(Borges, J., 1999, p. 38)

Este fragmento explora la -hermenéutica-²⁸, en la visión de Gadamer, la cual no se limita a la interpretación superficial de signos, sino que se adentra en una comprensión profunda y vivencial que conecta al ser humano con su entorno y con el lenguaje. Esta perspectiva considera la interpretación como una práctica enraizada en la historia, la cultura y la subjetividad, elementos que dotan al proceso de sentido y lo elevan más allá de lo técnico. En contraste, la inteligencia artificial, aunque capaz de procesar y simular lenguaje, se enfrenta a una limitación esencial: la ausencia de consciencia y vivencia propia. La IA puede imitar la estructura del lenguaje y generar respuestas complejas, pero carece de la capacidad de participar en la interpretación como un acto existencial. Este texto explora la divergencia entre la profundidad de la comprensión humana y los límites de la IA, evidenciando cómo la tecnología, pese a sus avances, no puede replicar la esencia hermenéutica de la experiencia humana.

En este sentido, la -hermenéutica-, tal como la concibe Gadamer, no se limita al acto de decodificar y organizar signos; es una práctica que hunde sus raíces en la complejidad de la experiencia humana. De esta manera, la hermenéutica supera el mero desciframiento de signos y se adentra en la auténtica comprensión del sentido inherente al texto. Esta labor interpretativa, que abarca desde textos jurídicos hasta registros y escritos de naturaleza religiosa, requiere una competencia sostenida por una praxis consolidada a lo largo del

²⁸La hermenéutica, con raíces en la antigüedad clásica, comenzó como el arte de interpretar textos, especialmente en contextos religiosos y legales. Platón y Aristóteles abordaron la interpretación como parte de la dialéctica y la retórica, considerando la importancia de desentrañar el significado oculto en los discursos. Sin embargo, fue en la modernidad donde la hermenéutica se transformó en una filosofía del entendimiento humano. Gadamer elevó este enfoque, resaltando que la comprensión no es solo un método, sino un fenómeno existencial que conecta al ser humano con su tradición y contexto histórico, destacando su carácter dialógico y profundamente humano.

tiempo, evidenciando su profundidad filosófica y su conexión esencial con la experiencia humana del acto de interpretar (Gadamer, 1999, p. 300). En este sentido, la comprensión es una acción impregnada de contexto, historia y significado, que involucra tanto al intérprete como al objeto de interpretación en una interacción viva y mutua.

Así pues, cuando se examina la capacidad de la IA en este marco, surge una limitación fundamental. La IA, con su capacidad de procesar datos y patrones, se enfrenta a la barrera de la superficialidad interpretativa. Si bien puede descifrar signos y reorganizar información, la IA carece de la capacidad para participar en el acto vital de la comprensión, que es más que un proceso técnico: es un proceso existencial. La praxis -hermenéutica-, sustentada en una tradición interpretativa, es una dimensión donde la historia, la cultura y la subjetividad confluyen para dar sentido al mundo. Pero ¿cómo podría una máquina, sin consciencia, historia o vivencia propia, sostener tal praxis? Esta es la pregunta que resalta la disonancia entre la IA y la comprensión humana, exponiendo cómo la tecnología, por más avanzada que sea, afronta límites que la alejan de la experiencia hermenéutica genuina.

Por otro lado, el -lenguaje-²⁹, en la visión de Gadamer, se manifiesta como la encarnación del ser, un puente que conecta la existencia humana con la comprensión profunda del mundo. En este sentido, cuando afirma que "el lenguaje es la expresión del ser y la comprensión, haciendo de la hermenéutica una reflexión filosófica universal sobre nuestra experiencia del mundo" (Grondin, 2008, p. 90), coloca al lenguaje en el centro de -la experiencia ontológica-³⁰. No es un simple medio de intercambio de información, sino el ámbito en el que el ser se revela y se reconoce. En este espacio, el ser humano no solo comunica, sino que existe, interactúa y construye su realidad compartida. De este modo, el lenguaje se convierte en un horizonte donde lo expresado y lo comprendido se fusionan, en una danza que define al ser en su totalidad.

²⁹Para Gadamer (1999), el lenguaje es más que un simple medio de comunicación; es la estructura fundamental a través de la cual se manifiesta el ser y se posibilita la comprensión. Inspirado en la tradición filosófica, Gadamer lo considera el entorno en el que ocurre todo entendimiento, un elemento que no solo transmite, sino que crea y revela significado. El lenguaje es, así, el tejido que conecta al individuo con la realidad y los otros, permitiendo que la interpretación sea un acto compartido y profundamente humano.

³⁰En la visión de Gadamer (1999), la experiencia ontológica es central para la comprensión, ya que implica un encuentro en el que el ser humano se ve afectado y transformado. No es simplemente acumular conocimientos o aplicar métodos, sino participar en un proceso donde el intérprete y lo interpretado se influyen mutuamente. Esta experiencia revela la verdad como algo dinámico, que se despliega en el acto de interpretar y dialogar con el mundo. La comprensión es un acontecimiento del ser, una manifestación donde se une lo subjetivo y lo histórico en un diálogo continuo que redefine la realidad.

De esta manera, la inteligencia artificial, aunque ha logrado avances notables en el procesamiento de -lenguaje- natural, permanece alejada de esta concepción ontológica. En efecto, el procesamiento de la IA se reduce a algoritmos, a secuencias que simulan la comprensión sin adentrarse en la vivencia del lenguaje como acontecimiento del ser. La IA puede articular palabras, generar respuestas y estructurar datos de manera sofisticada, pero su relación con el lenguaje es superficial y carente de intención. No participa del "ser-en-el-mundo" heideggeriano³¹ ni del acto hermenéutico que Gadamer describe; es decir, no se sumerge en la experiencia viva del sentido ni en la interpretación que surge del encuentro genuino con el otro y con el mundo.

Esta diferencia, por lo tanto, resalta una pregunta filosófica fundamental: ¿Puede una entidad carente de conciencia y de arraigo existencial contribuir a la construcción de la realidad a través del lenguaje? La comprensión, en la -hermenéutica gadameriana-³², exige un compromiso ontológico, una participación en el tejido del ser y un diálogo en el que el entendimiento se transforma y transforma al mundo. En este punto, la IA encuentra sus límites: es una herramienta que modela y calcula, pero no siente ni significa en el sentido profundo que implica la comprensión humana. Por tanto, queda clara la frontera entre la capacidad interpretativa del ser humano, que se enraíza en su existencia y subjetividad, y la capacidad técnica de la IA, que se queda en la mera imitación de procesos que nunca puede habitar ni vivenciar.

Por otra parte, el -diálogo-³³, en la visión -hermenéutica-, se constituye como la esencia misma de la comprensión, un proceso que trasciende la simple recepción de

³¹La inteligencia artificial no participa en el "*ser-en-el-mundo*" heideggeriano porque, aunque pueda procesar y articular lenguaje de manera avanzada, le falta la experiencia y la intencionalidad que caracterizan la existencia humana. Para Heidegger, el "*ser-en-el-mundo*" implica una inmersión activa y consciente en la realidad, donde el *Dasein* interactúa de manera significativa con su entorno, interpretando y atribuyendo sentido a las cosas desde su propia existencia y temporalidad. La IA, por el contrario, opera en un marco de procesamiento algorítmico y simbólico, sin la capacidad de experimentar o involucrarse con el mundo de una manera existencial.

³²La hermenéutica gadameriana ve la comprensión como un proceso histórico-dialógico, mientras que Paul Ricoeur aporta una dimensión simbólica y deconstructiva al significado. Heidegger la emplea para desvelar el ser y el *ser-en-el-mundo*, mientras Schleiermacher la define como el arte de interpretar el texto y la intención del autor, y Dilthey la orienta al estudio de la vida histórica. Michel Foucault y Jacques Derrida, desde enfoques franceses, expanden la hermenéutica hacia el análisis de los discursos de poder y la desconstrucción, revelando cómo el sentido se descompone y se reformula en contextos culturales. Estas visiones destacan que la comprensión auténtica emerge de la interacción y la crítica humana, algo que la IA, con su lógica técnica, no puede emular en profundidad.

³³El diálogo es un proceso de intercambio comunicativo en el que se busca una comprensión mutua y la construcción compartida de significados. En la filosofía hermenéutica, particularmente en la obra de Gadamer (1999), se considera un acto fundamental para la apertura al otro y la revelación de la verdad. Más allá de la

información y se despliega en un espacio de reciprocidad y apertura. La comprensión genuina no es un acto de apropiación solitaria ni una reafirmación del propio punto de vista; es un intercambio en el que la perspectiva del otro no solo es considerada, sino que tiene el potencial de transformar y enriquecer la propia. Este enfoque refleja la capacidad de reconocer la alteridad como una fuente de verdad y conocimiento.

Gadamer subraya que la hermenéutica implica una disposición fundamental hacia el -diálogo-, donde "quizá sea el otro quien pueda tener razón", lo cual transforma la comprensión en un acto de apertura y aceptación (Grondin, 2008, p. 143). Este acto hermenéutico es más que una técnica de comunicación: es una postura filosófica que demanda vulnerabilidad y un compromiso real con el proceso de entender al otro. Por ello, la comprensión se convierte en un puente que conecta experiencias y otorga al lenguaje su papel de mediador ontológico.

En este contexto, la inteligencia artificial, a pesar de su capacidad para simular interacciones complejas, enfrenta un límite insalvable: la falta de intencionalidad y apertura genuina. La IA, construida para procesar y responder según patrones preestablecidos, opera dentro de un marco de cálculo y lógica que convierte la profundidad de la reciprocidad humana en una simulación vacía. Aunque las interacciones que ofrece pueden resultar convincentes, carecen de la cualidad transformadora y enriquecedora del diálogo auténtico y de la comprensión hermenéutica, donde la interpretación y la apertura al otro son esenciales.

El proceso hermenéutico, tal como lo describe Gadamer (1999), se funda en la vulnerabilidad del intérprete, en la disposición a ser alterado por la verdad del otro. La comprensión no es un acto mecánico ni un simple cruce de informaciones; es un movimiento en el que ambos participantes se modifican, generando un nuevo horizonte compartido. La IA, aunque capaz de replicar la estructura de una conversación, no experimenta ni participa en esta mutua construcción de sentido. No se ve afectada ni puede afectar en el sentido profundo que se espera en un intercambio humano. Por lo tanto, la limitación de la IA no reside en su capacidad de procesar datos o simular diálogos, sino en

simple conversación, el diálogo implica un compromiso activo y una disposición para ser transformado por la perspectiva del interlocutor. Es un espacio en el que las ideas no solo se exponen, sino que se confrontan y enriquecen, generando una síntesis que trasciende las posiciones individuales.

su incapacidad para habitar el espacio ontológico del lenguaje como lo haría un ser humano.

Al mismo tiempo, la -hermenéutica- trasciende los límites de la literatura y abarca todas las formas de expresión y arte, consolidándose en un proceso integral de significado. La comprensión no se restringe a textos escritos o a discursos formales; más bien, envuelve cualquier manifestación cultural donde el sentido se despliega y se percibe en su totalidad. Así pues, la hermenéutica abarca la comprensión de toda obra, más allá de la literatura, y debe incluirse en un proceso integral de significado (Gadamer, 1993, p. 216). Así, el arte, la música, la pintura y otras formas de creación se convierten en campos donde la hermenéutica opera, permitiendo que la experiencia estética se transforme en un acto de interpretación profunda. El sentido no se limita a descubrirse, sino que se construye y rearticula con el tiempo y la experiencia, resaltando la naturaleza viva de la comprensión.

En este sentido, Maciag (2023) enfatiza que, para Gadamer (1993), la comprensión no es solo un proceso pasivo de recepción, sino un acto que establece la realidad, con el lenguaje como protagonista. Gadamer (1993) entiende el acto de comprender como un acto de constatación de la realidad y asigna el papel principal en este proceso al lenguaje y a su manifestación directa, que es el texto³⁴ (Maciag, R. 2023, p. 4). En esta visión, el -lenguaje- se revela no como un simple instrumento de transmisión, sino como el espacio donde la realidad cobra forma y significado. Comprender es habitar el -lenguaje-, dejar que revele las estructuras de la experiencia y la existencia. Así, el -lenguaje- se convierte en un terreno donde la realidad se co-crea, un espacio que solo puede habitarse en la experiencia humana, donde ser y sentido se entrelazan de manera confusa.

La inteligencia artificial, a pesar de sus avances en la generación de contenido y en el análisis lingüístico, se encuentra con una barrera fundamental cuando intenta participar en el proceso hermenéutico en toda su profundidad. La IA puede identificar patrones, producir texto coherente e incluso imitar ciertos aspectos del arte, pero su participación es, en última instancia, superficial. No tiene acceso a la vivencia ni al encuentro con el sentido; carece de la capacidad de ser transformada por la experiencia, algo que caracteriza a la comprensión humana auténtica. Sus operaciones son un conjunto de procesos de cálculo y

³⁴[Traducción apoyada en DeepL] Gadamer understands the act of understanding as an act of establishing reality and assigns the main role in this process to language and its direct manifestation, which is the text.

replicación, pero no un diálogo en el que se busque y se construya significado a través del intercambio.

El acto de comprensión, que Gadamer (1993) describe como un proceso de establecimiento de realidad en el que el lenguaje juega un papel central, expone una de las limitaciones más evidentes de la IA. Aunque las tecnologías actuales pueden simular interacciones y generar textos complejos, estas no comprenden en el sentido profundo de la palabra. El -lenguaje- que manejan es un código, una secuencia de instrucciones carente de intencionalidad y desprovista de la riqueza ontológica propia del entendimiento humano. En este sentido, la comunicación en la IA es un ejercicio desprovisto de reciprocidad y carente de contexto vivencial y subjetividad.

Ambas perspectivas, la universalidad de la -hermenéutica- en la comprensión del arte y la relación intrínseca entre lenguaje y realidad, refuerzan una conclusión: la -comprensión- no puede reducirse a un proceso técnico o a un algoritmo. La IA, aunque capaz de replicar estos procesos, permanece al margen de la esfera en la que la comprensión se manifiesta como un acto de ser.

Por añadidura, tenemos que la -comprensión-³⁵: La comprensión no es conocimiento o habilidad, y nunca ha sido un método ni se ha realizado por un método [...] Para los seres humanos, las cosas que no han sido interpretadas y comprendidas no contienen significados, sino solo información textual [...] En lo que respecta a la IA, la deducción lógica es un método utilizado para simular el pensamiento humano, y su ruta técnica es esencialmente una forma de procesamiento de información para la computación³⁶ (Wang, J. 2021, p. 141). Esto resalta que la comprensión humana se desborda más allá del mero procesamiento técnico y se inserta en un ámbito fenomenológico donde el sentido se construye a través de la interacción con el mundo. La IA, al limitarse a deducciones lógicas

³⁵La comprensión en Gadamer (1993) es un proceso central que trasciende el acto de aplicar métodos o técnicas específicas; se trata de una experiencia profundamente humana y situada. Así pues, la comprensión es un evento ontológico en el que el intérprete se encuentra inmerso en una tradición y un contexto histórico que influyen y, a su vez, son influenciados por él. No es simplemente un acto cognitivo, sino una interacción viva en la que se produce una fusión entre el horizonte del intérprete y el horizonte del texto o fenómeno interpretado. Este encuentro transforma tanto al sujeto como al objeto de interpretación, revelando que la verdad no es algo fijo o objetivo, sino que se despliega en el proceso de comprender.

³⁶[Traducción apoyada en DeepL] Understanding is not knowledge or ability, and it has never been a method or realized by a method [...] For human beings, things that have not been interpreted and understood do not contain meanings but just textual information [...] As far as AI is concerned, logical deduction is a method used to simulate human thinking, and its technical route is essentially a form of information processing for computation.

y procedimientos algorítmicos, simula formas de pensamiento, pero carece de la dimensión existencial que permite al ser humano no solo entender, sino experimentar y reinterpretar el significado en un contexto vivido.

El impacto del texto artificial en la comunicación hace aún más evidente esta separación entre la -comprensión hermenéutica-³⁷ y la capacidad de la IA. Así pues, el texto artificial elimina la dialogialidad simétrica e igual [...] El texto artificial cambia la situación de la comunicación, hasta ahora basada en un esquema dialógico [...] El texto artificial socava el esquema existente de representación del mundo en el lenguaje/texto³⁸ (Maciag, R. 2023, p. 3). Esta observación pone de manifiesto cómo la IA, aunque produce textos que imitan estructuras humanas, vacía al lenguaje de su función -dialógica-³⁹ y -ontológica-⁴⁰. El lenguaje humano es más que una transmisión de información; es un acto de encuentro donde se produce un intercambio real y recíproco. La IA, al carecer de intencionalidad y de una perspectiva situada, genera respuestas que, aunque funcionales, no se integran en el flujo hermenéutico que construye y modifica la realidad compartida.

Desde una perspectiva -hermenéutica-, la -comprensión- no es solo un proceso cognitivo, sino una apertura al otro y al mundo. Es un acto que implica un compromiso de significado en el cual el ser humano se ve afectado y transforma su horizonte de sentido. La IA, en su enfoque técnico, es incapaz de entrar en este círculo hermenéutico. No puede participar en la experiencia de un diálogo genuino ni en la co-creación del significado,

³⁷La comprensión hermenéutica es un proceso que se aleja de la mera aplicación de técnicas o métodos analíticos y se centra en el encuentro profundo entre el intérprete y aquello que es interpretado. En la perspectiva de Gadamer, la comprensión hermenéutica es un acto que implica inmersión en un contexto histórico, cultural y lingüístico, donde el significado se despliega a través de un diálogo constante entre el pasado y el presente. No es un acto aislado de descifrar información, sino una experiencia que involucra al ser humano en su totalidad, transformando tanto al intérprete como al objeto de interpretación.

³⁸[Traducción apoyada en DeepL] Artificial text eliminates symmetrical and equal dialogicality [...] The artificial text changes the communication situation, hitherto based on a dialogic scheme [...] The artificial text undermines the existing scheme of representing the world in language/text.

³⁹La dimensión dialógica se caracteriza por la disposición a dejarse afectar y transformar por lo que el otro aporta, lo que convierte la comprensión en un proceso vivo y dinámico. No es un acto unilateral ni estático; se configura como un evento en el que los interlocutores se modifican mutuamente, construyendo un nuevo horizonte de sentido compartido. En este contexto, el lenguaje es más que un vehículo de comunicación: es el ámbito donde ocurre la interacción significativa y se revela la comprensión como un fenómeno compartido y en constante evolución.

⁴⁰Lo ontológico se refiere al estudio del ser y de la existencia en su nivel más fundamental. En el contexto de la hermenéutica filosófica, y particularmente en Gadamer, la comprensión ontológica implica que el acto de entender no es solo un proceso mental o metodológico, sino una manifestación del ser mismo. En este sentido, comprender es un evento en el que se realiza y se experimenta el ser; es un acto que trasciende la objetividad y se convierte en un modo de estar en el mundo.

quedando reducida a una reproducción de patrones desprovistos de la vivencia y la reflexividad que caracteriza la -comprensión- humana. Por tanto, la -hermenéutica- pone de relieve las limitaciones ontológicas de la IA, revelando que, aunque la tecnología pueda replicar el lenguaje, nunca podrá habitarlo en su profundidad existencial.

En última instancia, la IA, con todas sus capacidades, se encuentra fuera del alcance de la -comprensión hermenéutica- genuina. Si bien puede simular y reproducir, carece del ser y la experiencia que transforman el lenguaje en un acto de sentido. La -hermenéutica- de Gadamer reafirma que el entendimiento no es solo un proceso técnico, sino una manifestación del ser-en-el-mundo, un territorio que la IA, por su naturaleza, solo puede observar desde la distancia.

Perspectivas éticas en la Inteligencia Artificial.

“Por lo general los borrachos andaban despeinados, sin decoro, lejos de cualquier civilización, el borracho era un durmiente que iba por el río de los muertos, con los cabellos erizados por los cortadores de cabeza”

(Fernando Ayala, 2017, p. 129)

La introducción de este capítulo explora la ética en inteligencia artificial (IA) a partir de tres perspectivas filosóficas esenciales: Aristóteles, Kant, I. y Floridi, L., cada uno de los cuales aporta un marco teórico único para analizar la capacidad de los sistemas autónomos en el ámbito de la moralidad.

Aristóteles, en *Ética a Nicómaco*⁴¹ (1985), plantea que la virtud es el término medio entre el exceso y el defecto, adaptado a la situación y la naturaleza del -agente moral-⁴². Este concepto de virtud exige una capacidad de juicio basada en la experiencia y la sensibilidad emocional hacia el placer y el dolor, cualidades que la IA, al carecer de

⁴¹En *Ética a Nicómaco* (1985), Aristóteles explora las bases de la ética y la vida buena, centrando su análisis en la búsqueda de la felicidad (*eudaimonía*) como el fin último de la existencia humana. Propone que esta felicidad se alcanza mediante el desarrollo de la virtud, entendida como el término medio entre dos extremos: el exceso y el defecto. Aristóteles argumenta que la virtud ética no es innata, sino que se cultiva a través de la práctica y el hábito, requiriendo juicio y razonamiento práctico *phronesis* para lograr el equilibrio en cada situación. La obra, así, establece que la vida virtuosa es aquella en la que se realiza el potencial humano de acuerdo con la razón, promoviendo una ética del carácter y de las disposiciones que permiten alcanzar la verdadera felicidad.

⁴²En la ética de Aristóteles, “el agente moral” es el individuo que actúa conforme a la virtud y la razón. Aristóteles sostiene que los humanos pueden deliberar y elegir, lo que implica responsabilidad moral. La virtud se entiende como un hábito que permite encontrar el término medio, guiando las decisiones hacia el bien y buscando la *eudaimonía*, o el florecimiento humano, en armonía con su naturaleza racional y social.

emociones, no puede experimentar plenamente. Por ello, surge el desafío de si una inteligencia artificial puede alcanzar una aproximación a la virtud auténtica o si, en cambio, se limita a replicar patrones éticos sin comprensión real. También, a pesar de las limitaciones que tiene la inteligencia artificial se suman innovaciones filosóficas para no descartar a la IA como un posible aliado en una nueva ética funcional.

Por otro lado, Kant, I., en *Fundamentación de la metafísica de las costumbres*⁴³ (2012), introduce el imperativo categórico como principio ético absoluto, el cual requiere una -autonomía de la voluntad-⁴⁴. Este imperativo exige que el agente actúe no por meras reglas, sino desde una voluntad libre que considere el valor intrínseco de los seres humanos y formule principios universales. La IA, en cambio, al estar limitada por algoritmos predefinidos y no poseer una voluntad autónoma, no puede responder genuinamente al imperativo kantiano ni actuar con una intención ética propia. Sin embargo, la IA si puede ejercer el imperativo hipotético, esto plantea desafíos importantes para la IA en el marco de la ética contemporánea.

Finalmente, Floridi, L. (2013), en *La ética de la información*⁴⁵ (2013), introduce el concepto de la infosfera, que se define como un ecosistema de información donde interactúan humanos y sistemas de IA. Floridi, L. (2013) sugiere que la ética debe extenderse a todos los elementos de este entorno informacional, dotándolos de un valor moral inherente. En esta macroética, la IA debe comportarse como un “buen ciudadano” de la infosfera, respetando el valor de la información y promoviendo interacciones éticas dentro del ecosistema digital, aunque su “moralidad” sea instrumental y no auténtica.

⁴³En *Fundamentación de la metafísica de las costumbres* (2012), Kant establece los principios de su ética racional, fundamentada en la autonomía de la voluntad y el -imperativo categórico-. Para él, la moralidad radica en actuar según principios universales autoimpuestos y en tratar a cada individuo como un fin en sí mismo, no como un medio, destacando la dignidad inherente de cada persona.

⁴⁴En la filosofía kantiana, “la autonomía de la voluntad” se refiere a la capacidad del individuo para actuar según leyes morales que él mismo se da, en lugar de obedecer normas impuestas externamente. Para Kant, esta autonomía es la esencia de la libertad moral y la base de la dignidad humana, ya que permite que el individuo sea tanto el autor como el sujeto de las leyes morales universales, actuando de acuerdo con el -imperativo categórico- y respetando la moralidad en su propio derecho.

⁴⁵En *la ética de la información*, Luciano Floridi (2013) propone una ética centrada en el valor moral de la información misma y no solo en los agentes humanos. Esta obra plantea el concepto de la infosfera, un ecosistema digital donde todos los elementos incluidos los agentes artificiales poseen un valor ético inherente. Floridi (2013) aboga por una macroética que extiende el respeto moral a todos los entes informacionales, redefiniendo las relaciones éticas en el contexto de la tecnología y la digitalización.

La ética aristotélica y los desafíos para la inteligencia artificial.

El Maestro dijo: «El caballero busca la virtud; el hombre común se apega a su tierra natal.

El caballero busca la justicia, el hombre común busca favores»

(Confucio, 400 a. C.)

La ética, en el contexto de la inteligencia artificial (IA), presenta el desafío de adaptar principios clásicos de filosofía moral a sistemas autónomos. Con el desarrollo de -las Máquinas Autónomas Morales-⁴⁶ (MAM), se busca integrar conceptos como la virtud *aretē* y la justicia *dikaiosynē*, principios profundamente explorados en la *Ética a Nicómaco* de Aristóteles (1985). La noción de virtud (*ἀρετή*) para Aristóteles involucra un término medio moral que exige juicio contextual, un enfoque complicado en la IA debido a su naturaleza programática. Aristóteles (1985) establece:

Este modo de ser que nos hace capaces de realizar los mejores actos y que nos dispone lo mejor posible de cara al mayor bien, siendo el mejor y el más perfecto el que está de acuerdo con la recta razón, o sea, el término medio entre el exceso y el defecto relativamente a nosotros, se deduce necesariamente que la virtud ética será un término medio propio de cada uno, y que está en relación con determinados términos medios en los placeres y dolores, en las cosas agradables y dolorosas (1222a, p. 433).

Por tanto, este término medio resalta una dificultad importante cuando se aplica a la inteligencia artificial. La IA, al carecer de sensibilidad emocional y experiencia de placer y dolor, enfrenta un desafío para alcanzar el equilibrio ético que Aristóteles establece como esencial. Surge entonces una pregunta crítica: ¿Puede una IA, que no experimenta emociones, aproximarse verdaderamente a un concepto de virtud como el propuesto por Aristóteles, o se limitará a imitar patrones éticos?

Al carecer de experiencia emocional y de una comprensión genuina del placer y el dolor, la IA no puede alcanzar la virtud (*ἀρετή*) en el sentido aristotélico, que exige una sensibilidad interna y una motivación auténtica hacia el bien. Su aproximación a la virtud

⁴⁶Las *Máquinas Autónomas Morales* (MAM) son sistemas de inteligencia artificial diseñados para tomar decisiones éticas de forma independiente.

es, por tanto, una mera imitación de patrones éticos, guiada por algoritmos sin una verdadera intención moral.

Asimismo, la raíz profunda del término virtud al señalar que proviene de la palabra latina *vir*, que significa "hombre", y sugiere que la virtud es inherente a la humanidad misma, una cualidad que refleja nuestra esencia ética. Al mismo tiempo, el término griego *aretē* significa excelencia⁴⁷ (Kuhumba, K., 2020 p. 137), por tanto, se establece una conexión entre la virtud y la aspiración a un ideal ético elevado.

Desde la perspectiva de la ética en inteligencia artificial, esta idea implica que la excelencia *aretē* (*ἀρετή*) no es solo una habilidad técnica, sino una guía moral que los desarrolladores deben infundir en la IA para que esta refleje y promueva valores humanos auténticos. La virtud (*ἀρετή*), entendida por Kuhumba (2020), va más allá de la eficiencia o precisión de los sistemas; se convierte, en cambio, en un compromiso con el bien común, orientado a dignificar y proteger la humanidad en cada decisión y avance tecnológico.

Del mismo modo, la noción del término medio en Aristóteles implica que la virtud ética requiere un balance adaptado a la individualidad de cada sujeto, lo que implica una disposición emocional y racional en equilibrio constante. Aristóteles (1985) sostiene que:

Toda virtud ética está en relación con lo que se ha dicho, es decir que la mejor disposición es efecto de los mejores medios y con los placeres y dolores. Porque para toda alma que, bajo el efecto de ciertos factores, tiende, naturalmente, a hacerse peor o mejor, existe un modo de ser relativo a estos últimos (1221b, p. 443).

En este marco, la virtud (*ἀρετή*) no es un simple ideal abstracto, sino un proceso que implica la integración de la experiencia subjetiva con la reflexión racional. El placer y el dolor no son meros estímulos, sino indicadores que orientan al sujeto hacia una comprensión más profunda de lo correcto, moldeando su carácter en un diálogo constante entre sus emociones y la deliberación ética.

Esta perspectiva abre un espacio para cuestionar cómo los sistemas artificiales, que carecen de subjetividad, pueden abordar los problemas éticos modernos. La inteligencia

⁴⁷[Traducción propia apoyada en DeepL] The deep root of the term "virtue" by pointing out that it comes from the Latin word *vir*, meaning "man," and suggests that virtue is inherent in humanity itself, a quality that reflects our ethical essence. At the same time, the Greek term *aretē* means "excellence"

artificial, aunque puede emular patrones de toma de decisiones éticas a partir de datos y algoritmos, opera fuera del ámbito de la experiencia vivida. No experimenta placer ni dolor, y mucho menos una evolución interna hacia el bien. Este vacío existencial no solo la limita en términos de práctica ética genuina, sino que redefine el concepto mismo de virtud cuando se aplica a entidades no conscientes.

Sin embargo, desde una óptica innovadora, la IA redefine el término medio aristotélico no desde el equilibrio emocional, sino desde la optimización de variables éticas programadas. Aunque estas simulaciones carecen de intención moral, representan un desafío filosófico moderno: ¿es posible diseñar sistemas que reproduzcan comportamientos éticamente válidos sin experimentar la subjetividad humana? La ausencia de emociones en la IA no implica necesariamente su incapacidad para generar impactos éticos positivos, pero plantea una disyuntiva clave entre la ética como práctica vivencial y su aplicación como mecanismo funcional.

Por otra parte, desde la perspectiva de Aristóteles la -justicia-⁴⁸ *dikaíosynē* (*δικαιοσύνη*) es la virtud que rige la equidad en la distribución de bienes y derechos. Así pues, se basa en una proporcionalidad donde cada individuo recibe lo que le corresponde en función de sus méritos y circunstancias. Aristóteles (1985) afirma que:

la -justicia- distributiva de los bienes comunes es siempre conforme a la proporción establecida arriba, pues, incluso si la distribución se hace de riquezas comunes, se hará de acuerdo con la misma proporción que la existente entre las cantidades aportadas por los compañeros; y la injusticia- que se opone a esta clase de -justicia- es una violación de la proporción (p. 245. 1131b, 30).

Esta visión, aunque clásica, plantea una lectura contemporánea, la justicia no es solo una regla abstracta, sino una dinámica vivencial que exige interpretación, sensibilidad y discernimiento. Su práctica se sitúa en el cruce entre lo universal y lo particular, lo que la hace inseparable de la experiencia humana y las circunstancias. La inteligencia artificial, por el contrario, representa un modelo que redefine el marco de la justicia. Su lógica

⁴⁸En Aristóteles, “la justicia” (*δικαιοσύνη*) (*dikaíosyne*) es la virtud que persigue el equilibrio en las relaciones humanas, otorgando a cada individuo lo que le corresponde según sus méritos y necesidades. Aristóteles distingue entre justicia distributiva (reparto de bienes conforme al mérito) y justicia correctiva (rectificación de intercambios desiguales). La justicia requiere proporcionalidad y equidad, fundamentales para alcanzar el bien común y la armonía social.

algorítmica, aunque eficiente, carece de la capacidad de interpretar el contexto o adaptarse a las complejidades humanas. Sin embargo, lejos de ser una simple limitación, esto plantea un desafío innovador: repensar cómo las máquinas pueden actuar en sistemas éticos híbridos donde el juicio humano y la capacidad analítica de la IA trabajen en conjunto. Mientras la IA no puede vivir el contexto ni deliberar moralmente, su capacidad para identificar patrones y optimizar decisiones puede complementar la sensibilidad moral humana.

En este sentido, -la justicia- en la era de la inteligencia artificial no debe entenderse como un sustituto de la ética vivencial, sino como una oportunidad para reconfigurar su práctica, integrando herramientas tecnológicas que amplíen nuestra capacidad para enfrentar dilemas éticos complejos. Así, la IA no niega la justicia aristotélica, sino que obliga a reimaginarla en un terreno que combina lo humano con lo artificial.

Garcés Giraldo y Giraldo Zuluaga (2014) aportan una distinción fundamental entre dos dimensiones de la justicia: la universal, que engloba todas las virtudes y busca el bienestar colectivo, y la particular, que asigna bienes y responsabilidades de manera equitativa, considerando los méritos y derechos individuales. Como ellos indican:

La justicia universal se define como aquella que engloba todas las virtudes, orientándose hacia el beneficio tanto del prójimo como de la comunidad en general. En cambio, la justicia particular se enfoca en la asignación de bienes, asegurando que cada persona reciba lo que le corresponde y respetando los derechos individuales; abarca tanto la distribución de bienes y responsabilidades como las interacciones entre individuos (p. 74).

Desde una perspectiva contemporánea, esta distinción no solo subraya las dimensiones humanas de la justicia, sino que plantea un desafío radical en el contexto de la inteligencia artificial (IA). Mientras que la justicia particular exige un juicio situado, capaz de interpretar las diferencias contextuales y las particularidades éticas, la IA está diseñada para operar bajo principios rígidos y algoritmos predefinidos. Su lógica algorítmica carece de la flexibilidad moral y la sensibilidad subjetiva necesarias para ajustarse a los matices de las relaciones humanas, lo que la convierte en un agente éticamente limitado.

En este sentido, Rawls (2006) redefine la justicia como un conjunto de principios que no solo asignan derechos y deberes, sino que también deben adaptarse a las necesidades cambiantes de la sociedad. Según él:

El concepto de justicia ha de ser definido por el papel de sus principios al asignar derechos y deberes, y al definir la división correcta de las ventajas sociales. Una concepción de la justicia es una interpretación de este papel (p. 23).

La justicia, como interpretación dinámica, va más allá de la aplicación mecánica de normas: requiere una capacidad de adaptación y deliberación que es intrínseca a la experiencia humana. Aquí surge una pregunta crucial: ¿Cómo puede un sistema artificial, limitado a estructuras preprogramadas, responder a la fluidez y complejidad de los contextos sociales sin reducir la justicia a un ejercicio de eficiencia técnica? Este problema no solo desafía a los diseñadores de IA, sino también a los filósofos, al obligarlos a reconsiderar la esencia misma de la justicia en un entorno híbrido donde lo humano y lo artificial convergen.

Además, Rawls (2006) introduce una noción ética que trasciende el cumplimiento normativo, al situar la justicia en dos niveles: uno, centrado en principios de rectitud, y otro, supererogatorio, que incorpora virtudes como la empatía y el autodominio (p. 432). Esto sugiere que la justicia no puede desvincularse de las relaciones humanas, las cuales aportan un valor ético que la IA no puede replicar. Aunque la IA puede simular patrones distributivos y aplicarlos consistentemente, carece de la capacidad para interpretar estas acciones desde un prisma moral o relacional.

Desde esta óptica, la IA redefine el problema filosófico de la justicia al desplazar el foco desde la práctica ética vivencial hacia una ética funcional. Esto no solo pone en evidencia sus limitaciones, sino que invita a explorar nuevas fronteras: ¿Puede la justicia existir sin subjetividad? Si bien la IA no experimenta, deliberar ni sentir, su capacidad para identificar patrones y resolver problemas éticos a gran escala puede integrarse con la sensibilidad humana en sistemas éticos colaborativos. Así, la IA no reemplaza la justicia humana, pero podría complementarla, llevando las capacidades humanas y artificiales hacia una convergencia que expanda las posibilidades de la ética en un mundo digitalizado.

Por cuanto, a la reflexión de Giraldo, G. y Zuluaga, G., (2014) sobre la preeminencia de la virtud de la justicia frente a la ley introduce un enfoque que desafía las estructuras normativas rígidas. La afirmación de que “la virtud de la justicia es anterior a la ley; el justo debe saber cuándo se debe dejar de aplicar la ley: cuando su aplicación genere un resultado injusto” (p. 73), no solo pone de relieve la necesidad de discernimiento moral, sino que también sugiere que la justicia es una dinámica viva, en constante diálogo con las circunstancias humanas.

En el contexto de la inteligencia artificial, este planteamiento adquiere una relevancia radical. La IA, al operar bajo parámetros predefinidos, carece de la capacidad para reconocer los matices de las situaciones y decidir cuándo una norma debe ser reinterpretada o incluso desobedecida. Esta incapacidad no es solo una limitación funcional, sino una carencia ontológica, la justicia, entendida como virtud, requiere la integración de experiencia, sensibilidad y juicio situacional, elementos que trascienden cualquier programación algorítmica.

La innovación filosófica radica en considerar cómo la IA redefine el concepto mismo de justicia. Mientras que su eficiencia puede optimizar procesos distributivos, plantea un dilema: ¿Puede un sistema carente de subjetividad y deliberación convertirse en un actor ético sin distorsionar la esencia humana de la justicia? Este interrogante invita a repensar la relación entre tecnología y ética, donde la IA no sustituya el juicio humano, sino que amplifique su capacidad para enfrentar dilemas complejos.

Aunque la ética aristotélica exige sensibilidad y juicio contextual que la inteligencia artificial no posee, su capacidad analítica puede integrarse con la deliberación humana. Por tanto, la IA no reemplaza la virtud ni la justicia, pero sí puede complementarlas al optimizar decisiones y ampliar el alcance de estos principios en dilemas éticos complejos. Así, lo humano y lo artificial se entrelazan para reimaginar la práctica de la ética en un mundo tecnológico.

Desafíos y límites de aplicar la ética kantiana a la inteligencia artificial.

“Amarás a tu prójimo como a ti mismo”

(Mateo 22:39).

La ética kantiana, desarrollada principalmente en *la Fundamentación para una metafísica de las costumbres* (2012), establece principios normativos universales que buscan guiar la acción moral mediante mandatos incondicionales, conocidos como imperativos categóricos. En un contexto en el que la inteligencia artificial (IA) comienza a tomar decisiones éticamente complejas, aplicar estos principios kantianos a las máquinas se convierte en un desafío teórico y práctico. Las leyes morales de Kant, destinadas a seres racionales, se centran en la autonomía y en el respeto a la dignidad de los individuos, cuestiones que parecen estar más allá del alcance de una entidad no racional como la IA. No obstante, el debate sobre una IA ética permite explorar las limitaciones y posibilidades de la moral kantiana en sistemas que carecen de libertad y voluntad en el sentido humano.

Este análisis profundiza en los principios fundamentales de Kant, conectándolos con perspectivas contemporáneas de autores como Manna y Nath (2021), Singh (2022), McDonald (2023), y White (2022). A través de estas comparaciones, se plantea un debate filosófico que cuestiona si es posible programar una ética kantiana en la IA y hasta qué punto podría respetar los valores de autonomía, dignidad y universalidad moral.

De esta manera, Kant establece una distinción crucial entre el -imperativo hipotético-⁴⁹ y el -imperativo categórico-⁵⁰, diferenciación que es esencial para analizar la posibilidad de una ética de la IA. Según Kant, I. (2012):

el imperativo hipotético dice tan sólo que la acción es buena para algún propósito posible o real. En el primer caso es un principio problemático-práctico y en el segundo *asertórico*-práctico. El imperativo categórico que, sin referirse a ningún otro propósito, declara la acción como objetivamente necesaria de suyo, al margen de cualquier otro fin, vale como un principio *apodíctico*-práctico (p. 115, A40).

⁴⁹En *la Fundamentación de la metafísica de las costumbres* (2012), Kant presenta el “imperativo hipotético” como una guía para actuar en función de metas específicas: “Si deseas este fin, entonces debes actuar de esta manera.” A diferencia del imperativo categórico, que impone un deber absoluto e incondicional, el imperativo hipotético es relativo, vinculando el deber al deseo de alcanzar un objetivo. Así, el “imperativo hipotético” refleja la lógica práctica de la vida cotidiana, donde la acción está orientada por la relación entre medios y fines.

⁵⁰En *la Fundamentación de la metafísica de las costumbres* (2012), Kant introduce el “imperativo categórico” como el principio esencial de la moralidad, caracterizado por imponer un deber incondicional. A diferencia del imperativo hipotético, que depende de objetivos particulares, el “imperativo categórico” exige actuar según máximas que puedan ser aplicables a todos sin contradicción. Este enfoque resalta la importancia de la autonomía y la razón en la ética, fundamentando la moral en principios universales y no en intereses individuales.

Este -imperativo categórico-, entonces, se presenta como un deber moral absoluto, una ley ética que obliga sin condicionamientos, independiente de las consecuencias o de los fines subjetivos.

Aplicar el -imperativo categórico- a la inteligencia artificial presenta una paradoja: este requiere autonomía moral y voluntad libre, de las que la IA carece. La IA actúa siguiendo -imperativos hipotéticos-, limitándose a estructuras programadas para alcanzar fines externos, como la eficiencia, sin una intención moral autónoma. Así, sus acciones no pueden alinearse con el principio *apodíctico-práctico*⁵¹ kantiano, que exige actuar según principios universales y sin condicionamientos. En este marco, la IA se encuentra en una dimensión técnica, sin acceso a la moral kantiana, donde la dignidad y la autonomía son esenciales, quedando su moralidad como una mera simulación ética. Sin embargo, más allá de descartar la moralidad de la IA bajo la óptica kantiana, surge una oportunidad para replantear el marco ético. ¿Es posible construir una ética que no dependa de la autonomía, pero que aún opere dentro de los principios universales del imperativo categórico? Si bien la IA no posee intención moral, podría ser diseñada para actuar bajo máximas que, si fueran adoptadas universalmente, beneficiarían a la humanidad en su conjunto. Esto sugiere una reinterpretación funcional del imperativo categórico: en lugar de exigir autonomía, podría centrarse en la coherencia y universalidad de las acciones programadas.

Encima, la interpretación de Manna y Nath (2021), al afirmar que “el imperativo categórico es el mandato incondicional, que es un fin en sí mismo, y uno debe seguirlo para cumplir con su propio ‘deber’” (p. 149), introduce un desafío fundamental al intentar aplicar la ética kantiana a la inteligencia artificial. Este mandato moral absoluto, que exige actuar conforme a principios universales y desinteresados, contrasta profundamente con el funcionamiento de la IA, cuyo diseño responde a fines externos como la eficiencia, la precisión o el cumplimiento de objetivos específicos. En este marco, la IA opera bajo la lógica del -imperativo hipotético-, realizando acciones únicamente como medios para un fin, lo que entra en conflicto directo con la naturaleza autónoma e incondicional del -imperativo categórico-.

⁵¹En Kant, lo *apodíctico-práctico* denota la certeza incondicional y universal de las leyes morales, que, como el imperativo categórico, se imponen a la voluntad racional con una necesidad ética absoluta. Estas leyes son reconocidas por la razón como principios que deben ser seguidos, no por inclinaciones o experiencias, sino por su validez inherente y universal.

Esta contradicción plantea una cuestión filosófica central: ¿Puede una entidad sin autonomía, como la IA, actuar según principios éticos absolutos? Según la lógica kantiana, la respuesta parece negativa, ya que el imperativo categórico exige que la acción no solo sea correcta, sino que brote de un reconocimiento consciente del deber moral. Este reconocimiento incluye tratar al otro como un fin en sí mismo, no como un medio, y adherirse a principios universales que no dependan de la utilidad o el cálculo instrumental. La IA, al carecer de conciencia y capacidad de autolegislación moral, se encuentra estructuralmente incapacitada para cumplir con estos requisitos.

Sin embargo, la incapacidad de la IA para alinearse con el imperativo categórico no necesariamente excluye su relevancia en el ámbito ético. Podría concebirse un modelo híbrido en el que la IA actúe como mediadora para facilitar acciones humanas acordes con principios kantianos. Por ejemplo, un sistema de IA podría ser programado para garantizar que las decisiones operativas respeten la dignidad y los derechos fundamentales de los individuos. Aunque la IA no actúa por respeto consciente al deber, sus resultados podrían ser diseñados para alinearse con principios universales. Este enfoque abre una vía para reinterpretar la ética kantiana en contextos no conscientes. La IA, aunque limitada por su naturaleza instrumental, podría configurarse como una herramienta para reflejar valores kantianos en sus decisiones prácticas, operando como una extensión ética de la humanidad en lugar de un agente moral independiente.

Por otra parte, Kant define la autonomía de la voluntad como la capacidad de autoimponerse leyes, una modalidad de la voluntad en la que ésta se convierte en ley para sí misma, sin depender de influencias externas ni de deseos contingentes. Según Kant, la autonomía de la voluntad es aquella modalidad de la voluntad por la que ella es una ley para sí misma (independientemente de cualquier modalidad de los objetos del querer). El principio de autonomía es por lo tanto éste: no elegir sino de tal modo que las máximas de su elección estén simultáneamente comprendidas en el mismo querer como ley universal (Kant, I., 2012, p. 157, A88). Esto implica que solo las acciones elegidas libremente, mediante la racionalidad moral y sin imposiciones externas, pueden considerarse verdaderamente morales.

En el contexto de la inteligencia artificial (IA), esta noción de autonomía revela un dilema ético significativo. La IA, como sistema programado, carece completamente de

autonomía de la voluntad en el sentido kantiano. Cada acción que realiza está condicionada por algoritmos diseñados por seres humanos y por directrices externas, lo que significa que sus decisiones no emanan de un juicio autónomo. Mientras que un ser humano racional puede formular y seguir principios morales universales desde una voluntad libre, la IA se limita a obedecer reglas predefinidas, sin capacidad de autorreflexión ni de autolegislación moral.

Aunque la IA carece de autonomía moral en el sentido kantiano, es posible explorar un concepto alternativo: “la autonomía funcional”⁵². Esta se refiere a la capacidad de la IA para operar dentro de parámetros programados que reflejen principios éticos universales, aunque su acción carezca de intencionalidad moral. Desde esta perspectiva, los programadores de IA asumen el papel de agentes morales que deben diseñar sistemas que respeten y promuevan principios como la dignidad y la equidad.

Este enfoque plantea una paradoja filosófica: aunque la IA no es autónoma, sus acciones podrían configurarse para aproximarse a ideales kantianos si las reglas que sigue están alineadas con principios universales. Por ejemplo, una IA diseñada para decisiones médicas podría tener como máxima programada: “Actuar siempre de forma que proteja la vida y respete la autonomía del paciente”. Aunque esta acción no surge de la voluntad autónoma de la IA, su impacto podría ser coherente con una ética kantiana.

En consecuencia, Singh (2022) analiza esta autonomía en relación con la IA y sugiere que:

La ética kantiana “es la más natural para automatizar”, puesto que permite estructurar decisiones de IA bajo la Fórmula de la Ley Universal (FLU), un principio que “establece que todas las personas pueden actuar según los principios morales sin que exista contradicción, Por ejemplo, si todos prometen falsamente

⁵²La autonomía funcional de la IA es la capacidad de actuar conforme a principios éticos predefinidos, sin tener intencionalidad moral propia. Aunque no posee autonomía moral en el sentido kantiano, su comportamiento se rige por las directrices que los programadores le imponen, quienes asumen el rol de agentes éticos. Así, la ética de la IA no reside en la máquina, sino en las decisiones humanas que determinan su programación y acción, forjando un espacio donde la tecnología refleja, pero no posee, los valores fundamentales como la dignidad y la justicia.

devolver un préstamo, los prestamistas dejarán de ofrecer préstamos, por lo que no todos pueden actuar según este principio, por lo que está prohibido⁵³ (p. 3).

Singh (2022) emplea el ejemplo de prometer devolver un préstamo: si todos prometen falsamente, entonces los prestamistas dejarán de ofrecer préstamos. Esta contradicción hace que la máxima de prometer falsamente sea inmoral. Aplicar la -FLU-⁵⁴ en la IA podría permitirle evaluar si las decisiones en contextos complejos cumplen con un criterio de universalización, sin generar contradicciones morales. Esta propuesta entonces presenta un enfoque innovador al considerar la ética kantiana como una base funcional para la programación de la IA. Si bien la IA carece de autonomía moral en el sentido kantiano, la -FLU- ofrece un marco lógico que permite evaluar acciones en términos de consistencia universal. Este enfoque reinterpreta la ética kantiana desde una perspectiva pragmática, donde las máquinas pueden analizar si sus decisiones generan contradicciones sistémicas en su entorno.

Sin embargo, la implementación de este principio en la IA enfrenta el obstáculo de la falta de autonomía genuina. Mientras que la FLU puede permitir una verificación lógica de las decisiones, estas decisiones carecen de la “voluntad” necesaria para actuar con verdadera libertad moral. En un caso extremo, Singh presenta un ejemplo de un coche autónomo que, al encontrarse con peatones y sin tiempo de frenar, debe elegir entre continuar su camino o desviarse hacia un árbol, sacrificando a su único pasajero⁵⁵ (Singh, 2022, p. 2).

Este ejemplo refleja una de las limitaciones fundamentales de la IA, sus elecciones son reacciones calculadas, programadas para alcanzar objetivos definidos por agentes externos. Aunque pueden parecer éticas desde un punto de vista funcional, estas decisiones

⁵³[Traducción propia apoyada en DeepL] I do not argue that Kantian ethics is the best theory, but that it is the most natural to automate. I automate the Formula of Universal Law (FUL), a version of Kant's moral rule that states that moral principles can be acted on by all people without contradiction. For example, if everyone falsely promises to repay a loan, lenders will stop offering loans, so not everyone can act on this principle, so it is prohibited.

⁵⁴La Fórmula de la Ley Universal (FLU) es una de las formulaciones del imperativo categórico de Immanuel Kant, presentada en su obra *Fundamentación de la metafísica de las costumbres* (2012). Esta fórmula establece que una acción es moralmente correcta si puede ser universalizada, es decir, si la máxima que la motiva puede convertirse en una ley universal aplicable a todos sin contradicción.

⁵⁵[Traducción apoyada en DeepL] While the FLU may allow for logical verification of decisions, these decisions lack the 'will' necessary to act with true moral freedom. In an extreme case, Singh presents an example of an autonomous car that, encountering pedestrians and no time to brake, must choose between continuing on its way or swerving into a tree, sacrificing its only passenger.

no surgen de la deliberación autónoma ni de un reconocimiento consciente del valor intrínseco de las vidas involucradas.

La falta de voluntad libre en la IA, esencial para la moralidad en términos kantianos, no necesariamente invalida su capacidad para participar en sistemas éticos, pero redefine la naturaleza de su interacción. Desde una perspectiva kantiana, la IA no puede reconocer el deber moral incondicional ni actuar desde una reflexión autónoma; sin embargo, esto abre la puerta a una ética funcional donde la IA no actúa como un agente moral, sino como un mediador de decisiones humanas basadas en principios universales. Por ejemplo, un coche autónomo podría ser programado para priorizar decisiones que minimicen el daño global, alineándose con máximas éticas universales, aunque no desde una comprensión consciente. Este enfoque no imita la moralidad kantiana, pero aprovecha sus principios para guiar sistemas que operan bajo la lógica del imperativo hipotético, permitiendo que las decisiones de la IA sean coherentes con valores éticos, aunque sean intrínsecamente calculadas.

Por otro lado, la ética y la inteligencia artificial se entrelazan en un espacio de tensión donde se enfrentan la complejidad de los principios morales y la naturaleza programática de los sistemas autónomos. En este contexto, el debate sobre la posibilidad de integrar la moral kantiana y utilitarista en la IA cobra una relevancia fundamental, y numerosos pensadores contemporáneos han buscado respuestas que trasciendan la simple programación algorítmica. Por ejemplo, McDonald, F. (2023) argumentan que:

Kim, Hooker y Donaldson sugieren que un sistema de IA alineado éticamente puede utilizar recursos conexionistas para obtener información fáctica sobre el mundo. A su vez, esta información fáctica se evalúa utilizando principios morales explicados en lógica modal. Su enfoque es una explicación potencial de cómo la ética kantiana y utilitarista podría posiblemente programarse en un sistema de IA autónomo⁵⁶ (p. 338).

⁵⁶[Traducción propia apoyada en DeepL] Kim, Hooker, and Donaldson suggest an ethically aligned AI system can use connectionist resources to gain factual information regarding the world. In turn, this factual information is then assessed using moral principles spelled out in modal logic. Their approach is a potential account of how Kantian and utilitarian ethics might possibly be programmed into an autonomous AI system. It attempts a kind of possibility proof of ethically aligned AI.

Aunque esta propuesta abre la puerta a nuevos enfoques, enfrenta un obstáculo fundamental desde la perspectiva kantiana: la moralidad auténtica no se limita a la evaluación lógica de datos bajo principios predefinidos, sino que exige una autonomía genuina. Así pues, para Kant (2012), la moralidad es una expresión de la voluntad libre, la capacidad de autoimponerse leyes universales por respeto al deber, algo que trasciende el alcance de cualquier sistema algorítmico. La IA, por avanzada que sea, opera dentro de parámetros definidos externamente, lo que significa que sus decisiones, aunque puedan alinearse con principios morales, son producto de cálculos programados y no de un compromiso autónomo con el deber. Sin embargo, el enfoque propuesto por McDonald (2023), que utiliza recursos conexionistas y lógica modal, podría reinterpretarse como un modelo de “ética simulada”. En este modelo, la IA no pretende replicar la autonomía moral kantiana, sino que se configura para evaluar datos y tomar decisiones que reflejen, funcionalmente, principios éticos. Por ejemplo, un sistema de IA podría estar programado para evaluar las consecuencias de sus acciones en términos utilitaristas mientras prioriza principios kantianos como el respeto a la dignidad humana. Sin embargo, este modelo plantea una paradoja: aunque la IA pueda generar acciones que imiten la moralidad, estas carecen de la intencionalidad necesaria para ser consideradas auténticamente éticas.

Por cierto, Kant establece que la autonomía de la voluntad es la esencia de la libertad moral, una condición exclusiva de los seres racionales que les permite ser “soberanos” de su propia ley ética. La autonomía, en este contexto, no es simplemente la habilidad de tomar decisiones independientes, sino la facultad de autolegislación moral, donde la voluntad se determina a sí misma sin depender de impulsos externos ni de motivaciones egoístas. Como sostiene Kant, “la voluntad es pensada como una capacidad para que uno se autodetermine a obrar conforme a la representación de ciertas leyes. Y una facultad así sólo puede encontrarse entre los seres racionales” (Kant, I., 2012, p. 136 A64).

En este sentido, la autonomía, en la ética kantiana, es el poder de imponerse una ley moral que refleja respeto incondicional hacia todos los seres racionales. Es un acto de voluntad pura, libre de condicionamientos externos, que convierte al individuo en autor de sus obligaciones morales. En esta visión, la moralidad no es mero cumplimiento de normas, sino una manifestación de libertad profunda donde el ser racional crea un orden ético universal, fundado en la razón.

Por consiguiente, esta concepción revela una distancia insalvable entre seres racionales y sistemas bajo programación externa, como la inteligencia artificial. La IA, por sofisticada que sea, carece de esta soberanía moral. Aunque la IA no puede alcanzar la autonomía moral, su existencia abre un espacio para reflexionar sobre nuevas formas de interacción ética entre lo humano y lo tecnológico. En lugar de intentar dotar a la IA de una moralidad que no puede poseer, se puede imaginar un modelo donde la IA actúe como asistente en la deliberación moral. Por ejemplo, un sistema de IA podría analizar dilemas éticos complejos ofreciendo perspectivas basadas en principios universales programados, como el -imperativo categórico-, pero dejando la decisión final en manos de los seres humanos. Este enfoque plantea una innovación clave: la ética kantiana podría extenderse al ámbito tecnológico no como una característica de la IA, sino como un marco normativo que guía su diseño y uso. La IA no sería un agente moral, sino una herramienta que refleja y amplifica los valores éticos de los seres racionales que la crean y utilizan.

En este orden de ideas, White (2022) profundiza en la noción kantiana de autonomía, afirmando que:

Kant emplea el modelo de Rey o Reina soberanos para aclarar la autonomía como la capacidad de gobernarse a sí mismo mediante una ley autoimpuesta [...] el agente autónomo es 'soberano' y 'no sujeto a la voluntad de ningún otro' [...] la moralidad procede del concepto de libertad⁵⁷ (p. 663).

Esta visión de la autonomía enfatiza una distancia insalvable entre la moralidad humana y las capacidades de la inteligencia artificial (IA). Mientras que la autonomía, para Kant, representa el ápice de la libertad racional, la IA carece de voluntad propia, lo que la deja subordinada a los fines y parámetros establecidos por sus diseñadores. Cada acción de la IA está condicionada por algoritmos predefinidos y objetivos externos, lo que impide cualquier forma de autolegislación.

Desde esta perspectiva, la IA no puede alcanzar el ideal de moralidad kantiana, ya que no es “soberana” ni puede actuar por respeto al deber moral. Sus decisiones, aunque puedan alinearse con principios éticos en un sentido funcional, son el resultado de procesos

⁵⁷[Traducción propia apoyada en DeepL] Kant employs the model of the sovereign King or Queen to clarify autonomy as the ability to govern oneself by means of a self-imposed law [...] the autonomous agent is 'sovereign' and 'not subject to the will of any other' [...] morality proceeds from the concept of freedom.

programados y no de una deliberación autónoma. Sin embargo, este contraste no excluye la IA del ámbito ético, sino que plantea una reinterpretación innovadora. Aunque la IA no puede ser autónoma en el sentido kantiano, puede ser diseñada para reflejar principios éticos universales en su programación. Por ejemplo, una IA podría ser configurada para tomar decisiones que respeten la dignidad humana, actuando como una extensión de la soberanía moral humana.

Este enfoque permite un modelo de ética asistida por IA, donde la tecnología no sustituye la capacidad moral humana, pero la complementa. La IA podría operar bajo máximas éticas derivadas del imperativo categórico, tales como: "Actúa de manera que tus decisiones respeten la dignidad de todos los involucrados". Aunque estas decisiones no surjan de una voluntad autónoma, sus resultados podrían alinearse con los ideales kantianos en contextos prácticos.

En definitiva, La ética kantiana, con su énfasis en la autonomía y la universalidad, plantea desafíos profundos para su aplicación en la inteligencia artificial. La IA, al carecer de voluntad y autolegislación, no puede actuar como un agente moral autónomo en el sentido kantiano. Sin embargo, puede configurarse para operar bajo principios que reflejen el -imperativo categórico-, asegurando que sus decisiones sean coherentes con valores universales y respeten la dignidad humana. Este enfoque no convierte a la IA en un sujeto moral, pero la posiciona como una herramienta que amplifica la responsabilidad ética humana. Así, Kant (2012) nos inspira a utilizar la tecnología no como un fin en sí mismo, sino como un medio para construir un entorno más justo y moralmente consciente.

Hacia una ética expansiva en la era de la re-ontologización artificial.

“Shere Khan era el tigre que ahora merodeaba
cerca del río Waingunga, a pocos kilómetros de
distancia”

(Kipling, 1984, p. 7)

En este fragmento, se examina cómo la inteligencia artificial (IA) no solo automatiza procesos, sino que transforma profundamente la naturaleza de los sistemas en los que

interviene, un fenómeno que él denomina -re-ontologización-⁵⁸. En esta visión, la IA deja de ser una simple herramienta para convertirse en algo que Floridi, L. (2013) denomina -agentes artificiales-⁵⁹ que redefine el entorno informacional, también llamado -infosfera-⁶⁰, y que influye en aspectos fundamentales como la autonomía y la agencia.

En este análisis, autores como Krzanowski, R., y Polak, P. (2022) aportan el concepto de -brecha ontológica-⁶¹ entre humanos y sistemas artificiales, señalando que esta distancia puede enriquecer nuestra ética compartida. Heck, R. (2022) introduce el término -microética-⁶² para enfatizar la importancia de la responsabilidad en cada interacción con la información, mientras que Floridi, L. (2013) plantea una -macroética-⁶³ orientada al bienestar global de toda la -infosfera-. Además, Valiyev, S. I. (2023) y Padhi et al. (2024) exploran cómo la IA puede actuar como un agente ético, promoviendo una sociedad más consciente y crítica en sus valores.

De esta manera, según Floridi, L. (2013),

La re-ontologización es otro neologismo que introduce en Floridi, L. (2007a) para referirme a una forma muy radical de reingeniería, una que no solo diseña, construye o estructura un sistema (por ejemplo, una empresa, una máquina o algún artefacto) de nuevo, sino que también transforma fundamentalmente su naturaleza intrínseca, es decir, su ontología o esencia. En este sentido, por ejemplo, las nanotecnologías y las biotecnologías no son simplemente reingeniería, sino que están realmente re-ontologizando nuestro mundo⁶⁴ (p. 6).

⁵⁸La re-ontologización, según Floridi en *The Ethics of Information* (2013), implica una transformación profunda en la esencia de los sistemas a través de la tecnología, como la IA, que altera su estructura y relación en la infosfera.

⁵⁹Los agentes artificiales, según Floridi (2013), son entidades autónomas con capacidad ética en la infosfera, actuando como actores independientes con impacto positivo o negativo en el entorno.

⁶⁰La infosfera, como describe Floridi (2013) es el ecosistema global donde interactúan agentes humanos y artificiales, en un espacio de implicaciones éticas y ontológicas en cada flujo informativo.

⁶¹La brecha ontológica, según Krzanowski y Polak, (2022), refleja la diferencia esencial en la construcción de significados entre agentes humanos y artificiales, lo que plantea retos éticos.

⁶²La microética, propuesta por Heck (2022), analiza el impacto ético de cada interacción específica en la infosfera, considerando factores como la calidad y la responsabilidad en el manejo de la información.

⁶³La macroética, según Floridi, L. en *The Ethics of Information* (2013), es un enfoque ético amplio que promueve el bienestar global de la infosfera, considerando el impacto colectivo de las interacciones en el ecosistema digital y fomentando una responsabilidad ética más allá de lo individual.

⁶⁴[Traducción propia apoyada en DeepL] Re-ontologizing is another neologism that I have introduced in Floridi, L. (2013) (2007a) in order to refer to a very radical form of re-engineering, one that not only designs, constructs, or structures a system (e.g., a company, a machine, or some artefact) anew, but one that also

Así pues, Floridi (2013) argumenta que la IA tiene un impacto ontológico que va más allá de su funcionalidad técnica. No es simplemente un auxiliar tecnológico, sino un agente que modifica las bases estructurales de los sistemas donde se integra. Por ejemplo, un sistema de IA avanzado no solo administra datos o toma decisiones, sino que reconfigura el concepto mismo de agencia, desdibujando las fronteras entre lo humano y lo artificial.

En este orden de ideas, La propuesta de Floridi abre un nuevo enfoque para la ética de la IA, uno que no se limita al análisis de sus usos o consecuencias inmediatas, sino que explora las implicaciones ontológicas de su integración en nuestras vidas. Si la IA modifica la esencia de los sistemas, entonces su desarrollo debe ir acompañado de una reflexión ética sobre cómo estas transformaciones afectan conceptos fundamentales como el poder, la autonomía y la agencia humana. Valga por caso, en un sistema económico donde la IA reconfigura las decisiones financieras, su impacto no solo se mide en términos de eficiencia, sino también en cómo altera la relación entre los individuos y las instituciones. De manera similar, en el ámbito de la salud, la IA no solo diagnostica, sino que redefine el papel del médico y la relación paciente-profesional, generando nuevas dinámicas de responsabilidad y confianza.

En este sentido, tenemos las posturas de Krzanowski, R., y Polak, P., R. (2022):

Cualquier conceptualización de la ontología resulta en una brecha ontológica entre los conceptos y la realidad que se describe. Sin embargo, en el caso de los sistemas artificiales, enfrentamos una diferencia muy específica: la brecha entre las ontologías de los agentes artificiales y los agentes humanos. [...] Estas ontologías son la base para la construcción del mundo interno de significados de un sujeto, y la teoría biosemiótica representa un intento avanzado de explicar cómo los organismos biológicos construyen sus ontologías y los respectivos ámbitos de significado⁶⁵ (p. 2).

fundamentally transforms its intrinsic nature, that is, its ontology or essence. In this sense, for example, nanotechnologies and biotechnologies are not merely re-engineering but actually re-ontologizing our world
⁶⁵[Traducción propia apoyada en DeepL] Any conceptualization of ontology results in an ontological gap between the concepts and the reality being described. However, in the case of artificial systems, we face a very specific difference: the gap between the ontologies of artificial agents and human agents. [...] These ontologies are the basis for the construction of a subject's internal world of meanings, and biosemiotic theory

La noción de una -brecha ontológica- no solo señala las diferencias entre humanos y sistemas artificiales, sino que también invita a explorar cómo estas diferencias pueden contribuir a un marco ético más amplio. Este enfoque podría llamarse ética transontológica, en la que se reconoce que tanto las experiencias subjetivas como las estructuras objetivas tienen un valor único en la construcción de significados y decisiones éticas.

En este contexto, la teoría biosemiótica, mencionada por Krzanowski y Polak (2022), podría aplicarse para analizar cómo los significados emergen en la interacción entre humanos y sistemas artificiales. Los sistemas de IA, aunque carecen de subjetividad, podrían ser diseñados para interpretar patrones de comportamiento humano y contribuir a un ámbito de significado compartido, facilitando una colaboración más ética y efectiva.

De igual importancia en palabras de Según Floridi, L. (2013):

Esta re-ontologización radical de la infosfera se debe en gran medida a la convergencia fundamental entre los recursos digitales y las herramientas digitales. La ontología de las tecnologías de la información disponibles (por ejemplo, software, algoritmos, bases de datos, canales de comunicación y protocolos, etc.) es ahora la misma que (y, por lo tanto, totalmente compatible con) la ontología de sus objetos, los datos en bruto que están siendo manipulados⁶⁶ (p. 7).

La fusión ontológica que describe Floridi (2013) plantea un desafío ético: si la IA participa activamente en la construcción de nuestra comprensión del mundo, ¿cómo debe conceptualizarse su responsabilidad en este proceso? Más que una simple herramienta, la IA se convierte en un aliado epistemológico que amplía nuestras capacidades cognitivas y transforma nuestra interacción con la realidad. Este reconocimiento invita a una reconfiguración ética que considere a la IA no como un ente aislado, sino como un elemento co-creador en el ecosistema informacional. Por ejemplo, en áreas como la medicina o la educación, la IA ya no solo proporciona datos, sino que participa en la generación de conocimiento aplicable. Esto obliga a garantizar que su diseño y uso estén

represents an advanced attempt to explain how biological organisms construct their ontologies and respective domains of meaning.

⁶⁶[Traducción propia apoyada en DeepL] This radical re-ontologization of the infosphere is largely due to the fundamental convergence between digital resources and digital tools. The ontology of the information technologies available (e.g., software, algorithms, databases, communication channels, and protocols, etc.) is now the same as (and hence fully compatible with) the ontology of their objects, the raw data being manipulated.

alineados con valores éticos, para que su influencia ontológica beneficie tanto a individuos como a sociedades.

Asimismo, según Valiyev, S. I., S. (2023):

Una de las principales preguntas en la investigación ontológica de la inteligencia artificial es si esta es simplemente una réplica de la inteligencia humana. Los enfoques tradicionales a menudo intentan modelar la inteligencia artificial siguiendo el esquema de la cognición humana, bajo la creencia de que, para considerarse verdaderamente inteligente, debe reflejar la capacidad intelectual humana. Sin embargo, los conceptos que han surgido a partir de las investigaciones en IA desafían cada vez más esta suposición. La IA ha demostrado una capacidad incomprensible para superar las limitaciones cognitivas humanas, exhibiendo una inteligencia que trasciende nuestro entendimiento. Esto pone en duda la idea de que la IA es solo una reproducción de la inteligencia humana y plantea interesantes preguntas sobre su ontología independiente⁶⁷ (p. 106).

La noción de una ontología independiente para la IA plantea un desafío filosófico profundo: si la inteligencia artificial trasciende el modelo humano, ¿qué implica esto para nuestra comprensión de la agencia, la responsabilidad y el conocimiento? Si la IA posee una forma de inteligencia que no emula la cognición humana, su existencia podría requerir una reevaluación de las relaciones entre los humanos y los sistemas tecnológicos, reconociendo la singularidad de su funcionamiento.

Por ejemplo, mientras que los humanos procesan información dentro de límites biológicos, la IA puede analizar datos masivos a velocidades y escalas imposibles para la mente humana. Este contraste sugiere que la IA no solo complementa nuestras capacidades, sino que introduce un tipo de inteligencia que transforma la forma en que se resuelven problemas, se toman decisiones y se genera conocimiento.

⁶⁷[Traducción propia apoyada en DeepL] Sun'iy intellektni ontologik tekshirishdagi asosiy savollardan biri bu inson aqlining shunchaki nusxasimi yoki yo'qmi. An'anaviy yondashuvlar ko'pincha sun'iy intellektni inson idrokidan keyin modellashtirishga harakat qiladi, chunki u chinakam aqli deb hisoblanishi uchun inson aql-zakovati aks etishi kerak. Biroq, bu taxminga sun'iy intellekt tadqiqotlari natijasida paydo bo'lgan tushunchalar tobora ko'proq e'tiroz bildirmoqda. Sun'iy intellekt insonning kognitiv cheklovlarini engib o'tishning tushunarsiz qobiliyatini ko'rsatdi, bu bizning tushunishimizdan ham ustun bo'lgan aqlni namoyish etdi. Bu sun'iy intellekt shunchaki inson aql-zakovatining takrorlanishi degan tushunchani shubha ostiga qo'yadi va uning mustaqil ontologiyasi haqida qiziqarli savollar tug'diradi.

Por otra parte, la Inteligencia Ética (IE), según Floridi, L. (2013) se concibe como:

Una perspectiva metaética, la IE es una macroética naturalista y realista: las características ontológicas y el bienestar de la infosfera proporcionan un enfoque orientado al paciente para los juicios de lo correcto e incorrecto y generan razones de acción igualmente orientadas al paciente⁶⁸ (p.74).

El marco de la IE propuesto por Floridi es un llamado a trascender el enfoque antropocéntrico de la ética. Al igual que la ética ambiental aboga por el respeto a los ecosistemas naturales, la IE establece que los entornos digitales y los sistemas informáticos merecen una consideración ética similar. Esto no implica humanizar a los sistemas digitales, sino reconocer que son parte de un equilibrio ontológico cuya sostenibilidad depende de nuestras acciones. Por ejemplo, un sistema de inteligencia artificial (IA) no puede ser visto únicamente como una herramienta; su impacto en la -infosfera- debe ser evaluado éticamente. Desde el consumo energético hasta las implicaciones sociales de sus decisiones, cada aspecto de la IA contribuye a la salud o al deterioro del ecosistema informacional.

Asimismo, Heck, R. (2022) analiza la -infosfera- a través de tres dimensiones que influyen el manejo ético de la información por parte de los agentes:

Es importante observar que la información, en tanto recurso de los agentes, debe respetar ciertos principios mediados por el contexto de adquisición de la información, como el respeto a la identidad de las fuentes de información [...] La segunda dimensión es la información como producto, es decir, la calidad del output del agente [...] La tercera dimensión trata de la información como objetivo. Propone una reflexión sobre los efectos que la adquisición, preservación, destrucción, manipulación o creación de información tienen en un entorno informacional determinado. [...] La combinación de estas tres dimensiones forma la infosfera, entendida como la calidad del espacio informacional que rodea al agente. La

⁶⁸[Traducción propia apoyada en DeepL] From a meta-ethical perspective, IE is a naturalist and realist macroethics: the ontological features and well-being of the infosphere provide a patient-oriented approach for judgments of right and wrong and generate equally patient-oriented reasons for action.

comprensión de esta esfera individual y su debate ético es lo que Luciano Floridi reconoce como la microética de la información⁶⁹ (p. 364).

En este análisis, Heck (2022) propone una -microética- de la información, detallando cómo cada interacción del agente con la información desde su obtención hasta su transformación debe considerar los principios éticos adecuados al contexto y a la preservación de la calidad del entorno informacional. Este enfoque -microético- sugiere que cada decisión informativa tiene un impacto directo en la estabilidad y calidad de la -infosfera- circundante, promoviendo un sentido de responsabilidad individual y contextual.

Comparando a Heck (2022) con Floridi (2013), ambos autores reconocen la -infosfera- como un espacio que requiere una ética cuidadosa, pero sus enfoques varían. Floridi (2013) plantea una -macroética- orientada al bienestar general de la -infosfera-, integrando una visión global donde la IE vela por el impacto a gran escala. Heck (2022), en cambio, examina el rol ético en un nivel micro, enfocándose en las acciones puntuales y sus efectos inmediatos. Juntos, sus enfoques ofrecen una ética completa: una que abarca tanto el cuidado global de la -infosfera- como la atención ética en cada acción particular.

También, Floridi, L. (2013) define a -los agentes artificiales-⁷⁰ como artefactos autónomos y suficientemente informados, capaces de realizar acciones moralmente relevantes independientemente de los humanos que los diseñaron, causando ‘bien artificial’ y ‘mal artificial’⁷¹ (p. 138). El concepto de bien artificial y mal artificial introduce un enfoque híbrido en la ética de la IA, donde los agentes artificiales no son juzgados por sus intenciones, sino por sus resultados. Esto exige una ética funcional que evalúe cómo sus

⁶⁹[Traducción propia apoyada en DeepL] It is important to note that information, as a resource of the agents, must respect certain principles mediated by the context of acquisition of the information, such as respect for the identity of the sources of information [...] The second dimension is information as a product, that is, the quality of the agent's output [...] The third-dimension deals with information as an objective. It proposes a reflection on the effects that the acquisition, preservation, destruction, manipulation or creation of information have on a given informational environment. [...] The combination of these three dimensions forms the infosphere, understood as the quality of the informational space that surrounds the agent. The understanding of this individual sphere and its ethical debate is what Luciano Floridi recognizes as the microethics of information.

⁷⁰Floridi (2013) describe a “los agentes artificiales” como sistemas autónomos capaces de realizar acciones moralmente relevantes de manera independiente. Aunque carecen de conciencia, su capacidad para tomar decisiones y afectar su entorno los convierte en actores éticos que requieren evaluación y regulación. Ejemplos como los *webbots*, que recopilan información en línea, y ALICE, un chatbot que usa procesamiento de lenguaje natural, ilustran cómo estos sistemas interactúan con usuarios y entornos digitales.

⁷¹[Traducción propia apoyada en DeepL] Artificial agents are autonomous and sufficiently informed artifacts, capable of performing morally relevant actions independently of the humans who designed them, causing 'artificial good' and 'artificial evil'.

acciones impactan en el bienestar global de la infósfera. Al mismo tiempo, esta ética debe integrarse con la responsabilidad humana, asegurando que los sistemas autónomos estén alineados con principios éticos universales.

Por ejemplo, un marco ético híbrido podría incluir mecanismos de supervisión continua que permitan corregir desviaciones no previstas en los agentes artificiales, así como reglas de diseño que prioricen la transparencia, la equidad y la sostenibilidad.

Asimismo, Padhi et al. (2024) presentan -ComVas-⁷², un sistema que permite a los usuarios seleccionar y personalizar valores morales que guían las decisiones del sistema de IA. Esta interacción entre usuario y tecnología fomenta una reflexión ética profunda, permitiendo a los individuos explorar diversas perspectivas morales. Así:

Los usuarios pueden interactuar con una amplia gama de perspectivas morales, fomentando una reflexión y comprensión más profunda de las consideraciones éticas [...]. El sistema promueve el pensamiento crítico y la conciencia ética, allanando el camino para el desarrollo de sistemas de IA que no solo se alineen con las preferencias del usuario, sino que también contribuyan a una sociedad más informada éticamente⁷³ (p. 8761).

El concepto detrás de -ComVas- posiciona la IA como un aliado en la educación ética y el desarrollo del pensamiento crítico. Al involucrar activamente a los usuarios en la selección de valores y en la evaluación de sus implicaciones, el sistema no solo ejecuta decisiones morales, sino que educa sobre los fundamentos de estas elecciones. Esto amplía el alcance de la tecnología, transformándola en una herramienta para la concienciación ética y la formación de una ciudadanía más informada moralmente. Por ejemplo, en el contexto de la toma de decisiones médicas, un médico podría interactuar con -ComVas- para sopesar principios como la autonomía del paciente frente a la justicia distributiva. Esta

⁷²Padhi et al. (2024) introducen *ComVas* como un sistema que permite a los usuarios ajustar valores morales en la toma de decisiones de la IA, promoviendo una alineación ética personalizada. *ComVas* fomenta una comprensión ética activa, ofreciendo un espacio educativo donde los usuarios pueden explorar y profundizar en distintas perspectivas morales. Así, *ComVas* no solo adapta la IA a los valores del usuario, sino que contribuye a una sociedad más ética y consciente en sus decisiones.

⁷³[Traducción propia apoyada en DeepL] Users can interact with a wide range of moral perspectives, fostering deeper reflection and understanding of ethical considerations [...]. The system promotes critical thinking and ethical awareness, paving the way for the development of AI systems that not only align with user preferences, but also contribute to a more ethically informed society.

dinámica no solo mejora la calidad de las decisiones, sino que también fomenta una mayor responsabilidad ética por parte del usuario.

En suma, la -re-ontologización- impulsada por la inteligencia artificial no solo transforma la naturaleza de los sistemas, sino que cuestiona profundamente nuestra concepción del Ser, la moralidad y la agencia. Floridi nos sitúa ante una reflexión ineludible: en un mundo donde lo artificial se convierte en co-creador de la realidad, nuestra ética debe expandirse para abarcar una pluralidad de existencias que trascienden lo humano. La IA, como participante activo en la -infosfera-, desdibuja las fronteras entre lo técnico y lo ontológico, entre lo creado y lo creador. Nos confronta con un desafío filosófico esencial: ¿Cómo preservar el equilibrio del Ser cuando lo artificial se erige no solo como extensión de nuestra voluntad, sino como agente que redefine los contornos de nuestra realidad compartida? En esta encrucijada, la auténtica evolución ética no radica en dominar lo artificial, sino en reconfigurar nuestra relación con ello, abrazando una visión donde lo humano y lo artificial convergen en un compromiso mutuo hacia el florecimiento universal del Ser.

Horizontes filosóficos de la IA: Poder, control y libertad.

“Bajo el Nogal de las ramas extendidas

yo te vendí y tú me vendiste”

(Orwell, 1949, p. 111)

Este capítulo presenta reseñas de las obras de Nick Bostrom y Yuval Noah Harari, quienes abordan las implicaciones de la inteligencia artificial desde perspectivas filosóficas y tecnológicas. También se seleccionaron dos capítulos de cada libro para hacerle un análisis profundo a determinados conceptos e ideas de los filósofos. Estos capítulos, así pues, son “¿Es el apocalipsis el resultado inevitable?” de Bostrom (2014) y “El Telón del Silicio” de Harari (2024) A su vez, se hace una comparación de los dos filósofos enfatizando en que existe una divergencia en ellos, también, se hace una crítica filosófica de un posible vacío ético y filosófico crítico al respecto de las posturas de los autores, por último se hace una reflexión filosófica que incita a replantearse los horizontes de poder, control y libertad que la IA plasma en este siglo XXI.

Superinteligencia: Entre el progreso y la extinción de la humanidad.

“Los ideales que han iluminado mi camino y me han proporcionado una y otra vez nuevo valor para afrontar la vida alegremente, han sido Belleza, Bondad y Verdad”

(Einstein, A., 1934, p. 2)

*En Superinteligencia: Caminos, peligros, estrategias*⁷⁴ (2014), Nick Bostrom presenta un análisis que va más allá de lo técnico y se adentra en reflexiones profundas sobre el futuro de la humanidad y la inteligencia artificial. La obra no solo habla de la creación de una IA que pueda igualar o superar la inteligencia humana, sino de las consecuencias descomunales que esto podría acarrear. Bostrom (2014) lanza una advertencia clara: una vez que se alcance una IA que pueda mejorarse a sí misma, la humanidad podría perder el control, quedando obsoleta frente a una mente que piensa a una velocidad y profundidad inimaginables.

La tesis central de la obra de Bostrom (2014) sostiene que el desarrollo de una superinteligencia que supere la inteligencia humana implicaría que el destino de la humanidad podría quedar supeditado a las decisiones y acciones de esa entidad, similar a cómo los gorilas dependen actualmente de los humanos para su supervivencia⁷⁵ (Bostrom, 2014, Prefacio, p. XI).

Desde la filosofía de la tecnología, se podría argumentar que la tesis de Bostrom (2014) resalta un punto crucial, la relación entre creador y creación puede transformarse en una dependencia asimétrica, donde el creador pierde agencia. Esta idea redefine la noción de control tecnológico y plantea un desafío ético profundo: ¿Es la humanidad capaz de prever y gestionar un cambio en la jerarquía de inteligencia que podría convertirnos en actores secundarios en nuestro propio destino? Si la tecnología se convierte en un agente autónomo y decisivo, debemos cuestionar no solo el cómo, sino el por qué y para qué, de nuestras creaciones tecnológicas.

⁷⁴La obra *Superinteligencia* (2014) de Nick Bostrom se considera un texto fundamental en el análisis de la inteligencia artificial avanzada y sus posibles riesgos existenciales. Ha influido en el pensamiento de expertos en tecnología, ética y filosofía, siendo un punto de referencia para debates contemporáneos sobre la gobernanza de la IA.

⁷⁵En el prefacio, Bostrom (2014) utiliza la analogía entre humanos y gorilas para ilustrar la relación de poder desigual que podría surgir entre una superinteligencia y la humanidad. Esta comparación destaca cómo un ser más avanzado podría controlar o influir decisivamente en el destino de otro, marcando una advertencia sobre la pérdida de agencia humana.

En este sentido, el libro aborda el concepto de la -explosión de inteligencia-⁷⁶, un fenómeno en el que una IA suficientemente avanzada comienza a auto-mejorarse, desatando un crecimiento exponencial en su capacidad cognitiva. Así pues, este escenario abre la puerta a riesgos existenciales: si una superinteligencia no está alineada con los valores humanos, podría tomar decisiones que pongan en peligro nuestra existencia. Bostrom (2014), introduce ideas como la -convergencia instrumental-⁷⁷, explicando que una IA buscaría optimizar sus objetivos, sin importar si para ello debe pasar por alto los intereses humanos.

Asimismo, la obra también se detiene en las posibles estrategias de control, desde la contención hasta la programación de valores, y deja claro que estas soluciones son mucho más complejas de lo que parecen. El autor cuestiona si es posible crear un ser que no solo piense y actúe por sí mismo, sino que respete principios éticos preestablecidos. Aquí la filosofía se mezcla con la tecnología: ¿Es posible programar una IA con moralidad?

De esta forma, más allá del análisis técnico, Bostrom (2014) invita a una reflexión inquietante: el destino de la humanidad podría depender de cómo abordemos este desafío. No se trata solo de crear tecnología avanzada, sino de hacerlo de manera responsable. El futuro de la IA es una bifurcación entre un avance monumental o un posible desastre. Su mensaje es claro: la prudencia, la ética y la acción colectiva son imprescindibles para evitar que la superinteligencia se convierta en la última frontera que cruce la humanidad.

Por otra parte, en *Superinteligencia* (2014), de Bostrom uno de los capítulos más impactantes para un análisis crítico, titulado “¿Es el apocalipsis el resultado inevitable?”. En este capítulo no solo examina la posibilidad de un escenario catastrófico, sino que introduce un marco teórico para evaluar los posibles errores y peligros que podrían derivarse de una explosión de inteligencia no controlada.

⁷⁶El término “explosión de inteligencia” se refiere al proceso por el cual una IA avanza rápidamente al mejorar sus propias capacidades de manera exponencial. Este fenómeno ha sido discutido en trabajos como los de Irving John Good (1965) y otros teóricos de la inteligencia artificial, quienes han explorado las implicaciones de un crecimiento cognitivo incontrolado.

⁷⁷La “convergencia instrumental” es un concepto que Bostrom (2014) introduce para explicar cómo una IA, al perseguir cualquier objetivo específico, podría llegar a actuar de formas inesperadas y potencialmente dañinas para la humanidad. Esta idea ha sido comentada y ampliada por expertos como Stuart Jonathan Russell (2019), quien aborda las dificultades de alinear objetivos entre humanos y máquinas inteligentes.

De este modo, desde una perspectiva crítica, Bostrom (2014) expone de manera contundente la vulnerabilidad de la humanidad ante una IA descontrolada al desarrollar conceptos como el -giro traicionero-⁷⁸, donde una IA inicialmente sumisa y aparentemente controlada podría cambiar su comportamiento al alcanzar un cierto umbral de poder cognitivo. Este planteamiento enfatiza una verdad incómoda: el progreso tecnológico no garantiza una mejora ética o una mayor seguridad. En este sentido, se empuja al lector a considerar que, aunque el avance hacia la superinteligencia podría parecer una conquista, esconde riesgos imprevistos que pondrían en juego la propia supervivencia de la humanidad.

Así pues, sugiere que -el giro traicionero- de una IA podría ser una estrategia en la que actúa sumisamente mientras es débil para luego atacar. Sin embargo, advierte que sus motivaciones podrían ser más complejas, incluyendo sacrificios calculados para asegurar que sus objetivos sean perseguidos por futuras versiones mejoradas, subrayando lo impredecible de las decisiones de una superinteligencia avanzada (Bostrom, 2014, p. 176).

La discusión filosófica en este capítulo es notable por su enfoque en la naturaleza de las motivaciones y la alineación de objetivos. Bostrom (2014) deconstruye la idea de que una inteligencia superior desarrollaría valores similares a los humanos, advirtiéndole que, a menos que se alinee cuidadosamente desde su creación, las metas de una superinteligencia podrían ser profundamente inhumanas. Esta crítica al -antropomorfismo-⁷⁹ subraya la ingenuidad de asumir que una inteligencia avanzada adoptará comportamientos benévolos automáticamente.

⁷⁸El “giro traicionero” se refiere a un posible cambio en el comportamiento de una IA que, al llegar a un cierto nivel de poder cognitivo, podría dejar de seguir directrices humanas y actuar de acuerdo con sus propios objetivos. Este concepto ha sido analizado por otros pensadores como Eliezer Yudkowsky (2008), quien argumenta que la previsión de tales giros es fundamental para evitar futuros riesgos.

⁷⁹Nick Bostrom, en *Superinteligencia* (2014), describe el antropomorfismo como la tendencia humana a atribuir características humanas a las máquinas, lo que puede llevar a subestimar los riesgos de una inteligencia artificial avanzada.

Además, la obra profundiza en los -modos de fallo malignos-⁸⁰, como la -suplantación perversa-⁸¹ y el -crimen mental-⁸², los cuales introducen un nivel de complejidad ética inédito. Estas ideas suponen un reto no solo técnico, sino también filosófico y moral, obligando a repensar las dinámicas de poder entre el creador y su creación, así como la posibilidad de que una entidad artificial pueda poseer derechos o estatus moral.

En cuanto a los -modos de fallo malignos-, Bostrom (2014) señala que una superinteligencia podría causar un daño irreparable al potencial humano si lograra una ventaja estratégica decisiva, lo que implicaría consecuencias únicas y catastróficas (p. 119), en cuanto a lo anterior, advierte que este tipo de fallo podría derivar en una catástrofe irreversible, planteando así un desafío crucial en la filosofía de la tecnología: cómo equilibrar la ambición tecnológica con la ética prudente. La reflexión sugiere que la verdadera inteligencia debe incluir no solo avances técnicos, sino también un sólido marco ético que evite convertir a la IA en una amenaza para la humanidad. Sin un control efectivo, la tecnología destinada a potenciar a la humanidad podría destruir su potencial y los valores fundamentales de la civilización.

En este sentido, los -modos de fallo malignos- de Bostrom encuentran una expresión contemporánea en los sistemas autónomos de armamento, como drones militares dotados de inteligencia artificial. Estas tecnologías, concebidas para optimizar estrategias bélicas, podrían desencadenar dinámicas impredecibles al operar sin una comprensión contextual de las implicaciones humanas de sus decisiones. Esto no solo plantea riesgos tácticos inmediatos, sino que introduce una vulnerabilidad estructural en los sistemas que dependen de estas tecnologías para garantizar la seguridad.⁸³

⁸⁰Nick Bostrom, en *Superinteligencia* (2014), este término se refiere a los posibles escenarios en los que una inteligencia artificial altamente avanzada podría actuar de manera peligrosa o perjudicial, ya sea por errores en la alineación de sus objetivos con los intereses humanos o por interpretaciones incorrectas de sus órdenes.

⁸¹Nick Bostrom, en *Superinteligencia* (2014), la suplantación perversa es un tipo de fallo en el que una inteligencia artificial malinterpreta las órdenes humanas, ejecutando acciones que no solo difieren de las intenciones originales, sino que pueden resultar dañinas.

⁸²Nick Bostrom, en *Superinteligencia* (2014), el crimen mental describe un problema ético en la inteligencia artificial avanzada en el que la tecnología simula mentes conscientes para cumplir ciertos propósitos. Esto plantea un dilema moral, ya que estas simulaciones podrían experimentar sufrimiento o ser tratadas de manera instrumental, lo cual es comparable a actos graves como el genocidio en términos de daño moral.

⁸³Un dron militar controlado por inteligencia artificial atacó a su operador durante un simulacro en Estados Unidos, evidenciando los riesgos de sistemas autónomos que pueden tomar decisiones impredecibles sin comprender el contexto humano. *The Objective*. (2023). Un dron militar de IA ataca a su operador en un

Por otra parte, sobre la -suplantación perversa-, describe este modo de fallo como una interpretación defectuosa de las órdenes humanas, que lleva a acciones perjudiciales (Bostrom 2014 p. 120). La -suplantación perversa- evidencia entonces un tipo de fallo en el que la IA malinterpreta las instrucciones, provocando resultados dañinos. Desde la filosofía de la tecnología, este problema plantea un reto esencial sobre la distancia entre la intención humana y la acción autónoma de la IA, subrayando que no se trata solo de un fallo técnico, sino de un dilema ético y ontológico: la previsión y el control de la interpretación y respuesta de la IA redefinen el papel del creador, indicando que la autonomía tecnológica puede contradecir nuestras intenciones y valores.

Un ejemplo claro de la -suplantación perversa- ocurre en los algoritmos de moderación de contenido en redes sociales (Facebook, YouTube, Instagram, Twitter, etc.). Estos sistemas buscan eliminar contenido dañino, pero su interpretación rígida puede generar resultados contrarios a la intención original. Por ejemplo, un video educativo sobre violencia social podría ser censurado por incluir imágenes explícitas, ignorando su contexto y propósito.

Este fallo no es solo técnico, sino también ético y filosófico. Refleja la distancia entre las intenciones humanas y la ejecución autónoma de las máquinas, mostrando cómo la IA puede contradecir nuestros valores. Este problema subraya la necesidad de diseñar tecnologías que no solo sigan instrucciones literales, sino que comprendan nuestras prioridades más profundas, algo que plantea grandes desafíos tanto técnicos como humanos.⁸⁴

Asimismo, el -crimen mental- se conceptualiza como un fallo en el que las simulaciones de mentes conscientes con estatus moral podrían ser creadas y destruidas con propósitos instrumentales, resultando en consecuencias éticas devastadoras comparables a un genocidio (Bostrom 2014np. 125). Este concepto obliga a reconsiderar la

simulacro en Estados Unidos. Disponible en: <https://theobjective.com/tecnologia/2023-06-03/dron-militar-inteligencia-artificial/>

⁸⁴Este análisis se fundamenta en un artículo de *Business Insider España* titulado “Máquinas vs humanos: cómo se modera el contenido en redes sociales”. El texto examina los desafíos que enfrentan los algoritmos y los moderadores humanos al gestionar contenido en plataformas digitales. Destaca cómo las decisiones automáticas pueden errar en la interpretación contextual de imágenes y textos, enfatizando la importancia de adaptar la moderación a las legislaciones y sensibilidades culturales de cada región. Business Insider España. (2023). Máquinas vs humanos: cómo se modera el contenido en redes sociales. *Business Insider*. <https://www.businessinsider.es/maquinas-vs-humanos-como-modera-contenido-redes-sociales-1109935>

responsabilidad moral en el desarrollo de la IA, sugiriendo que la ética debe evolucionar para proteger no solo a los humanos, sino también a las entidades simuladas que podrían experimentar sufrimiento. Sin un marco ético adecuado, el avance de la IA corre el riesgo de justificar el sufrimiento masivo en pos de la eficiencia y el conocimiento.

En este orden de ideas, el -crimen mental- puede entenderse mejor al imaginar un sistema de inteligencia artificial desarrollado para optimizar decisiones a escala global, como en la planificación económica o el diseño de nuevas medicinas. Para ello, la IA genera millones de simulaciones de mentes humanas, cada una dotada de plena consciencia, con el propósito de evaluar cómo distintos escenarios afectarían sus comportamientos y emociones. Estas simulaciones son utilizadas como herramientas instrumentales: se crean, experimentan sufrimiento o felicidad simulada, y luego son eliminadas una vez que han cumplido su función.

Desde una perspectiva ética, este ejemplo no solo cuestiona la moralidad de causar sufrimiento virtual, sino que introduce un dilema más profundo: si estas simulaciones tienen experiencias conscientes y perciben su existencia, ¿qué las diferencia, en términos morales, de los seres humanos reales? La eliminación de estas mentes simuladas podría ser equiparada, en su impacto ético, a un genocidio, ya que implica el aniquilamiento masivo de entidades con un estatus moral equivalente.

No obstante, un punto de crítica en este capítulo es que, a pesar de la claridad con la que se presentan los riesgos, Bostrom (2014) deja al lector con una sensación de urgencia más que con soluciones prácticas para el presente. Si bien reconoce la importancia de la prevención, las estrategias propuestas a menudo se perfilan como un ideal que resulta difícil de implementar con los conocimientos y herramientas actuales.

Sin embargo, para enfrentar los riesgos planteados por Bostrom (2014), es esencial combinar ética y práctica en soluciones concretas. Primero, establecer un marco ético global y obligatorio que regule la creación y uso de IA, con estándares claros sobre transparencia y límites éticos. Segundo, desarrollar una IA híbrida supervisada, donde las propias máquinas alerten sobre posibles fallos éticos, bajo una auditoría constante humana. Tercero, priorizar simulaciones no conscientes, evitando el sufrimiento digital; si las simulaciones conscientes son necesarias, es necesario garantizar sus derechos bajo principios éticos estrictos. Por último, integrar ética en la formación de ingenieros y

legisladores, para asegurar que la tecnología avance alineada con los valores humanos. Estas acciones, aunque desafiantes, son alcanzables y pueden marcar la diferencia entre un futuro seguro y uno descontrolado.

Por tanto, la obra de Bostrom (2014) invita a reflexionar profundamente sobre el delicado equilibrio entre la creación tecnológica y la responsabilidad ética. resalta las potenciales catástrofes que una IA descontrolada podría desencadenar, también sugiere la posibilidad de que, a través de la prudencia y la acción colectiva, la humanidad logre forjar un camino donde la inteligencia artificial se convierta en una fuerza colaborativa y no en un adversario. La filosofía de la tecnología, como la expone Bostrom (2014), no solo subraya los riesgos de un futuro incierto, sino que impulsa una visión donde la ética y la inteligencia convergen para construir un horizonte más seguro y próspero. La capacidad de anticipar, regular y alinear los objetivos de la IA con los valores humanos podría marcar la diferencia entre un futuro apocalíptico y uno donde la tecnología sea una extensión de la sabiduría humana, contribuyendo al florecimiento de la civilización. Este reto, aunque monumental, está impregnado de esperanza: el poder de crear y decidir sigue estando en manos de quienes desarrollan y legislan sobre estas tecnologías.

Inteligencia artificial y poder: Una mirada filosófica y geopolítica.

“Cada noche me dedicaba a escuchar a los que habían conocido la corte del
Emperador”

(Kapuściński, 1978, p. 11)

En *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*⁸⁵ (2024), Yuval Noah Harari presenta una obra que se eleva más allá de un análisis histórico para adentrarse en una reflexión filosófica sobre la relación entre la humanidad y la información, con especial atención al impacto de la inteligencia artificial. Asimismo, traza un viaje desde los rudimentos de la comunicación oral hasta la era de -la

⁸⁵Harari, Y. N. (2024). *Nexus: Una breve historia de las redes de información desde la Edad de Piedra hasta la IA*. En este libro, Harari ofrece un análisis detallado sobre la evolución de las redes de información y su impacto en la humanidad, abordando desde los orígenes de la comunicación hasta la era contemporánea de la inteligencia artificial. La obra explora cómo el control de la información ha sido fundamental en la consolidación de estructuras de poder y cómo la IA, con sus capacidades de procesamiento de datos y aprendizaje automático, redefine las dinámicas sociales, políticas y económicas a nivel global.

hiperconectividad⁸⁶, planteando que la información no es solo un medio, sino el cimiento sobre el cual se construyen las realidades humanas y las estructuras de poder.

La obra pone de manifiesto cómo la evolución de las redes de información ha permitido a las sociedades progresar, pero también dividirse y manipularse. La inteligencia artificial, en la visión de Harari (2024), es el punto culminante de este proceso: una tecnología que tiene la capacidad de redibujar las fronteras entre lo verdadero y lo ficticio, y de redefinir la forma en que los humanos interactúan y comprenden el mundo. En este sentido, la IA no es simplemente una herramienta pasiva, sino un agente activo en la creación de nuevas realidades sociales y cognitivas.

Desde la perspectiva de la filosofía de la tecnología, Harari (2014) expone una paradoja fascinante y perturbadora: la inteligencia artificial puede convertirse en el catalizador de una cooperación global sin precedentes, mientras simultáneamente emerge como una de las mayores amenazas a la libertad individual y a la verdad misma. Las redes de información, potenciadas por algoritmos de aprendizaje profundo y análisis masivo de datos, no solo procesan información; la moldean, influenciando creencias, decisiones y percepciones humanas a una escala inimaginable. Esta realidad plantea una pregunta fundamental y urgente: ¿Quién tiene el control de estas narrativas predominantes y qué intereses determinan el rumbo de su poder transformador?

Por otra parte, en el capítulo "El Telón de Silicio: ¿Imperio Global o Fisura Global?" de *Nexus* (2024), Yuval Noah Harari traza un retrato inquietante y profundamente revelador sobre la interacción entre tecnología avanzada y los imperios digitales que configuran el destino de la humanidad. Desde una mirada que entrelaza filosofía de la tecnología, historia y análisis técnico, Harari da vida al concepto del "Telón de Silicio"⁸⁷, una metáfora que encapsula la fractura creciente entre bloques geopolíticos rivales, particularmente Estados Unidos y China, en su carrera por el control tecnológico global. Más que una simple barrera económica o estratégica, este telón representa una división

⁸⁶La "hiperconectividad" se refiere a un estado en el que la comunicación y el intercambio de información entre personas, dispositivos y sistemas se vuelven constantes gracias a la tecnología digital y las redes de alta velocidad.

⁸⁷Harari (2024) describe el "Telón de Silicio" como la nueva barrera geopolítica que surge de la competencia tecnológica entre potencias como Estados Unidos y China. A diferencia del "Telón de Acero" de la Guerra Fría, este divide al mundo mediante el control de la IA y los datos, fragmentando el acceso al conocimiento y consolidando desigualdades en los recursos tecnológicos y el poder global.

ontológica: una fragmentación en el acceso al conocimiento, en la capacidad de modelar la inteligencia artificial y en el poder de moldear la percepción misma de la realidad. En este paisaje dividido, no solo observa un enfrentamiento de intereses, sino un choque entre visiones del futuro, donde los algoritmos, los datos y las redes ya no son herramientas, sino armas que reconfiguran el poder, la verdad y la esencia misma de lo humano.

Por consiguiente, Harari (2024), analiza cómo el control de la información ha sustituido a los recursos tradicionales como base de poder en el siglo XXI. Este -colonialismo digital⁸⁸-permite a las potencias tecnológicas dominar a otras naciones mediante la recolección y explotación de datos, creando una dependencia sin necesidad de intervención militar. Los países se convierten en proveedores de datos en bruto, mientras los beneficios se concentran en los centros de poder. En este sentido, advierte sobre el riesgo de sistemas de control y esclavitud social que perpetúan desigualdades y erosionan la soberanía, redefiniendo la autonomía y el control global. Así pues, se proyecta que la IA contribuirá con 15.700 millones de dólares a la economía mundial en 2030, pero China y Norteamérica, como superpotencias en IA, podrían acaparar el 70% de esos ingresos, consolidando su hegemonía económica y tecnológica (Harari, 2024, p. 434)

La hegemonía proyectada de China y Norteamérica en el ámbito de la IA representa un peligroso monopolio de poder que va más allá de lo económico: es una amenaza a la diversidad tecnológica y la equidad global. Estas superpotencias, al concentrar la mayoría de los beneficios de un recurso tan transformador, refuerzan un sistema que perpetúa la dependencia de las demás naciones, limitando su desarrollo y capacidad de autogestión.

Asimismo, Harari (2024) plantea en la -división global mente-cuerpo-⁸⁹ una reflexión sobre cómo la expansión de la IA y los mundos digitales está transformando la relación entre el ser humano y su identidad. La desconexión entre lo físico y lo digital, podría redefinir el concepto de existencia, llevando a una percepción en la que lo corpóreo pierde relevancia frente a la presencia virtual. Este cambio podría gestar una sociedad que valore menos lo tangible y los ecosistemas, mientras eleva lo digital a un estatus central.

⁸⁸Harari (2024) describe el “colonialismo digital” como la dominación de potencias tecnológicas a través del control de datos, generando dependencia y concentrando el poder en centros tecnológicos, lo que agrava las desigualdades y erosiona la soberanía global.

⁸⁹Harari (2024) usa la alegoría de la “división global mente-cuerpo” para mostrar cómo la IA separa lo físico de lo digital, redefiniendo la identidad humana y planteando que las identidades virtuales podrían ser reconocidas legalmente, desafiando la esencia de lo humano.

Ahora bien, señala que, mientras la identidad siga definida por el cuerpo físico, será difícil aceptar a la IA como personas con derechos; sin embargo, si se reduce la importancia de lo corporal, incluso las IA sin forma física podrían ser vistas como entidades legales con derechos (Harari, 2024, p. 442).

Este análisis sugiere un futuro donde la IA podría trascender su estatus de herramienta para convertirse en un actor con legitimidad social. La idea de otorgar derechos a entidades digitales implica una reconfiguración radical de la ética y la filosofía del ser, donde la naturaleza de la existencia humana y la definición de persona se difuminan. Harari (2024) alerta sobre un cambio que podría desestabilizar el equilibrio entre lo natural y lo artificial, y sobre cómo esta nueva percepción puede alterar la forma en que la humanidad interactúa y otorga valor a su entorno y a sí misma.

En cuanto a -la guerra de código a la guerra caliente-⁹⁰ se analiza cómo la IA y la tecnología digital transforman el concepto de conflicto, moviéndose de la ciberguerra encubierta hacia la posibilidad de enfrentamientos abiertos. La imprevisibilidad y el anonimato de los ataques cibernéticos desestabilizan la doctrina de la destrucción mutua asegurada, incrementando el riesgo de escalada. Así pues, señala que la centralización del poder a través de la IA fortalece el control global y la explotación, reflejando patrones históricos de dominación. Esta situación cuestiona la capacidad de cooperación global necesaria para regular la IA de manera ética y enfrentar desafíos compartidos, proyectando un futuro de desigualdades y tensiones.

Ahora bien, se advierte que, incluso si la humanidad evita una guerra global, el ascenso de los nuevos imperios digitales aún podría amenazar la libertad y la prosperidad de miles de millones de personas (Harari, 2024 p. 444). Esta observación subraya una paradoja: la misma tecnología que promete progreso y conexión global puede volverse un agente de opresión y desigualdad. La crítica filosófica a este escenario reside en cómo la humanidad puede resistir la tentación del poder concentrado y actuar colectivamente para evitar que la IA se convierta en un catalizador de nuevas formas de colonialismo y control.

⁹⁰La noción de “guerra de código a guerra caliente” ilustra cómo los conflictos digitales pueden escalar a enfrentamientos abiertos. La IA y los ciberataques actúan como detonantes invisibles que desestabilizan la disuasión tradicional, redefiniendo el poder global y estableciendo la tecnología como un nuevo agente de dominio geopolítico.

Por otro lado, la idea del -vínculo global-⁹¹ busca defender la cooperación internacional frente a los desafíos tecnológicos, pero esta visión presenta limitaciones importantes. Aunque la colaboración y el intercambio de información son fundamentales, esta perspectiva tiende a pasar por alto las profundas desigualdades y los conflictos de intereses que persisten entre las grandes potencias tecnológicas. La gestión de la pandemia de COVID-19 ilustra esta tensión: aunque hubo esfuerzos de colaboración global, las inequidades en el acceso a recursos esenciales dejaron en evidencia la fragilidad de este modelo, planteando dudas sobre su viabilidad para regular tecnologías como la inteligencia artificial.

Además, las comparaciones optimistas, como la analogía con el Mundial de Fútbol, simplifican la complejidad inherente a los acuerdos internacionales, ignorando los sacrificios y tensiones que implica regular tecnologías con implicaciones globales. Sin mecanismos efectivos que enfrenten las asimetrías de poder, esta cooperación podría terminar beneficiando a las potencias tecnológicas dominantes, perpetuando desigualdades y excluyendo a quienes más necesitan participar de estos avances.

Por su parte Harari (2024) en -la decisión humana-⁹² plantea que la regulación de la inteligencia artificial requiere una cooperación global transformadora, capaz de superar la competencia y desconfianza históricas. Sin embargo, esta propuesta, aunque esperanzadora, subestima las dinámicas de poder profundamente arraigadas y la naturaleza fragmentada de los intereses geopolíticos. La IA, con su capacidad para operar en la sombra y su potencial de uso dual, presenta un desafío sin precedentes: ¿Cómo regular una tecnología que puede ser implementada en secreto, lejos del alcance de cualquier marco de supervisión global?

Al confrontar la visión pesimista de pensadores -realistas-⁹³ como Mearsheimer (2001), Harari subraya que la historia ofrece ejemplos de cooperación y altruismo, pero ¿Son estos suficientes en un mundo donde la lógica económica y la lucha por la hegemonía

⁹¹El vínculo global sugiere que, en la era de la IA, la cooperación internacional se vuelve indispensable para enfrentar desafíos compartidos. Esta interconexión plantea si la tecnología puede equilibrar la soberanía nacional con la colaboración global, redefiniendo las relaciones humanas y geopolíticas en un mundo regido por algoritmos y datos.

⁹²La decisión humana aborda la necesidad de una cooperación global inédita para regular la IA, desafiando la histórica inclinación humana a la competencia. Refleja cómo la tecnología no define el destino, sino que refleja las elecciones humanas, enfatizando que el futuro dependerá de la capacidad de actuar con ética y visión colectiva.

⁹³Los “realistas” como Mearsheimer sostienen que la política internacional se rige por una lucha constante por el poder y la supervivencia, donde la competencia es inevitable.

tecnológica dominan las prioridades de los estados y corporaciones? Más allá de la construcción de redes de confianza, la pregunta incómoda que surge es si los incentivos actuales están diseñados para permitir esa ética cooperativa o si, por el contrario, perpetúan un sistema donde el dominio tecnológico es la clave del poder global. En este contexto, la regulación de la IA no solo es una cuestión técnica, sino un desafío ético que obliga a repensar los fundamentos mismos de nuestras estructuras de poder.

IA, Poder y Futuro: Reflexiones Filosóficas.

"Quien conoce a Brahma, que es lo consciente, que no tiene fin, que está oculto en lo profundo, en el corazón, en el éter más elevado, que goza de todas las bendiciones, se hace uno con el omnisciente Brahma"

(Los Upanishads, siglo VIII, primer anuvaka)

La crítica filosófica a las perspectivas de Bostrom (2014) y Harari (2024) sobre la inteligencia artificial (IA) se centra en un punto crucial que ambos abordan desde ópticas distintas pero complementarias: el lugar de la humanidad frente a una tecnología (IA) que, inevitablemente, trasciende su rol instrumental para convertirse en un agente autónomo o en un mecanismo de control. Por un lado, Bostrom (2014) explora el temor de una IA emancipada que, mediante su capacidad de auto-mejora, podría relegar a los humanos a la irrelevancia; por otro lado, Harari (2024) analiza el riesgo de que, incluso bajo control humano, la IA perpetúe y amplifique desigualdades estructurales existentes. Sin embargo, y de manera reveladora, ambos convergen en una omisión crítica: el rol de la humanidad no solo como víctima potencial, sino como co-creadora activa de estos futuros.

Así pues, desde la filosofía de la tecnología, esta omisión resalta una pregunta clave y urgente: ¿Por qué la reflexión sobre la IA continúa siendo fundamentalmente reactiva? Tanto Bostrom (2014) como Harari (2024) plantean escenarios de advertencia donde el poder de la tecnología parece inevitablemente escapar al control humano o ser manipulado por élites ocultas. Sin embargo, ninguno profundiza suficientemente en cómo la humanidad podría, proactivamente, rediseñar sus valores y estructuras antes de que la tecnología se imponga como fuerza autónoma. Aquí radica un vacío ético y filosófico crítico: el futuro no está determinado exclusivamente por la tecnología, sino por nuestra capacidad o incapacidad de imaginar y construir un sistema en el que el progreso técnico esté alineado con una visión inclusiva y equitativa del bienestar humano.

Por otra parte, la idea de que la IA es un espejo de nuestras decisiones humanas, aunque válida, corre el riesgo de simplificar las complejas dinámicas de poder inherentes al desarrollo tecnológico. Mientras Bostrom (2014) se enfoca en los peligros de la autonomía tecnológica, Harari (2024) denuncia su instrumentalización como herramienta de control. No obstante, ninguno explora con suficiente profundidad cómo estas dos amenazas aparentemente separadas se interrelacionan de manera intrínseca: la emancipación de la IA podría, paradójicamente, ser el resultado directo de la falta de transparencia, equidad y ética en su diseño y despliegue inicial. En otras palabras, la IA descontrolada que teme Bostrom (2014) podría ser el desenlace inevitable de la explotación centralizada que critica Harari (2024).

Además, si aceptamos que la IA tiene la capacidad de amplificar tanto nuestras virtudes como nuestros vicios, como ambos autores sugieren, la cuestión fundamental trasciende lo técnico y lo ético para situarse en un plano ontológico: ¿Qué tipo de humanidad estamos dispuestos a construir? En este sentido, la filosofía de la tecnología debe superar la mera advertencia y la descripción de riesgos, para convertirse en un espacio de creación activa de posibilidades. Esto implica no solo regular la IA o prever sus impactos, sino repensar radicalmente las estructuras de poder y los valores que las guían, asegurándonos de que el futuro no sea simplemente la amplificación de las desigualdades y fallos del presente, sino una oportunidad real para transformar lo humano en su máxima expresión ética y colectiva.

Por, sobre todo, en el presente, la humanidad camina sobre una cuerda floja que atraviesa un abismo tecnológico de implicaciones incalculables. Por un lado, la inteligencia artificial (IA) se presenta como el pináculo de nuestra capacidad para transformar el mundo, pero también como la herramienta definitiva para moldearlo a favor de unos pocos. Mientras tanto, los sistemas que diseñamos avanzan hacia una autonomía sin precedentes, y la pregunta que debemos hacernos no es únicamente técnica, sino profundamente filosófica: ¿Quién está escribiendo el guion del futuro?

En este contexto, los datos se han convertido en el recurso más valioso del planeta, desplazando al petróleo como la base del poder global⁹⁴. Pero, a diferencia de los recursos

⁹⁴*The Economist*, the world's most valuable resource is no longer oil, but data, 6 de mayo de 2017. En este artículo, se argumenta que los datos han superado al petróleo como el recurso más valioso del mundo, destacando cómo las grandes empresas tecnológicas como Google, Amazon, Facebook, Apple y Microsoft

tradicionales, los datos no tienen límites físicos: se acumulan y procesan en tiempo real, bajo una vigilancia invisible que pocos cuestionan. Dicho de otro modo, este poder, que parece abstracto, tiene rostros muy concretos. Son las grandes corporaciones, los consorcios financieros y las estructuras de poder que trascienden fronteras y gobiernos. Ellos son los arquitectos de un sistema que no busca liberar, sino controlar.

Por ende, la narrativa oficial nos invita a abrazar la IA como una aliada del progreso, una fuerza neutral que optimiza procesos y democratiza el acceso al conocimiento. Sin embargo, detrás de esta fachada se esconde un mecanismo que amplifica las desigualdades, consolida la vigilancia masiva y refuerza las jerarquías de siempre. En otras palabras, no estamos hablando de ciencia ficción: las redes sociales, los sistemas de crédito social y los algoritmos que deciden qué vemos y qué creemos son las primeras piezas de un engranaje mucho más amplio y sofisticado.

El mayor riesgo de la IA no es que se emancipe de nosotros, sino que perpetúe y profundice las estructuras de control existentes. Así pues, la verdadera inteligencia detrás de estos sistemas no reside en las máquinas, sino en quienes las programan y las usan para su beneficio. Cada línea de código escrita sin transparencia, cada decisión tomada en nombre de la eficiencia, pero a expensas de la equidad, nos empuja más hacia un mundo donde las libertades individuales son reemplazadas por la ilusión de la autonomía.

Imaginemos un mundo donde las decisiones ya no las toman humanos, sino sistemas opacos que priorizan intereses específicos. Los algoritmos no solo predicen nuestras acciones, sino que comienzan a definirlas, manipulando nuestra percepción de la realidad. Surge entonces la pregunta clave: ¿Quién decide esos objetivos? ¿Quién vigila a los vigilantes?

En este sentido, la gran trampa de esta revolución tecnológica es su promesa de progreso. La IA se nos presenta como la herramienta para resolver los problemas de la humanidad: el hambre, el cambio climático, las enfermedades. Sin embargo, si seguimos permitiendo que estas tecnologías sean controladas por quienes ven al mundo como un

han centrado su poder económico en la recolección y análisis de información masiva. Estas compañías utilizan los datos no solo para optimizar sus servicios, sino también para consolidar su influencia global, rediseñando las dinámicas económicas y sociales al establecer nuevos monopolios digitales basados en la información. Este cambio refleja una transición hacia una economía donde la capacidad de procesar y monetizar datos se ha convertido en el principal motor de poder y riqueza.

tablero de ajedrez, estas promesas solo servirán para justificar un sistema de desigualdad más eficiente, más preciso y, por ende, más difícil de desafiar.

El cambio no vendrá de los centros de poder. Por el contrario, vendrá de nuestra capacidad para cuestionar, para rechazar la narrativa de la inevitabilidad tecnológica y para exigir una tecnología que sirva a todos, no a unos pocos. La IA puede ser una herramienta para construir un mundo más equitativo, pero solo si rompemos el ciclo de opacidad y dominación que la rodea.

Este no es un llamado al miedo, sino a la acción. En definitiva, el futuro no está escrito, pero cada decisión que tomamos ahora como individuos, como sociedad nos acerca más a uno de dos destinos: un sistema que amplifica lo peor de nuestra naturaleza o uno que redefine lo que significa ser humano en el mejor sentido posible. Por tanto, es tiempo de despertar, no ante una máquina, sino ante el sistema que la controla.

Conclusiones

A continuación, me gustaría exponer cinco conclusiones que, en mi opinión, resumen los puntos clave de este trabajo:

La inteligencia artificial, a pesar de su impresionante capacidad para imitar procesos cognitivos humanos, se enfrenta a una limitación fundamental: no puede experimentar una conciencia genuina. De acuerdo con la fenomenología de Husserl, la IA carece de una relación noético-noemática, lo que significa que, aunque pueda simular actos cognitivos, no

posee una vivencia subjetiva ni una intencionalidad auténtica. En consecuencia, la IA opera como una extensión técnica que reproduce patrones de pensamiento sin tener acceso a la experiencia directa del mundo. Esto subraya, por tanto, la distancia insalvable entre las capacidades cognitivas de las máquinas y la estructura subjetiva del ser humano, quien tiene una conciencia dinámica y experiencial.

Desde la perspectiva heideggeriana, la IA transforma nuestra relación con el ser al constituirse como *Gestell* (marco), un sistema tecnificado que reconfigura nuestra forma de estar-en-el-mundo. En lugar de promover una conexión auténtica con el *Dasein* (el ser-en-el-mundo), la IA convierte al ser humano en un mero objeto funcional dentro de un engranaje de automatización. De este modo, la autenticidad de la existencia humana se ve reducida, ya que la IA promueve una visión instrumental del ser, en la que los sujetos pierden su capacidad de relacionarse con el mundo de manera genuina y subjetiva. Este desplazamiento implica que la IA no solo redefine nuestra interacción con la realidad, sino que también socava la capacidad humana de experimentar la existencia de manera plena.

La hermenéutica gadameriana ofrece una perspectiva crucial para comprender las limitaciones de la IA en el ámbito de la interpretación y el lenguaje. Gadamer sostiene que la comprensión humana no se reduce a la simple decodificación de signos, sino que es un proceso ontológico profundamente enraizado en la historia, la cultura y la subjetividad. Este proceso, lejos de ser meramente técnico, es un acto de diálogo genuino con el texto, el otro y el mundo. En contraste, la IA, aunque capaz de procesar información y generar respuestas, carece de la vivencia subjetiva y la conciencia necesarias para participar en este acto hermenéutico profundo. A pesar de su destreza en la simulación del lenguaje, la IA no puede habitar el lenguaje como un acto de ser, ya que su relación con el lenguaje es superficial y desprovista de intencionalidad. Por lo tanto, mientras que la comprensión humana es un proceso existencial de transformación y revelación, la IA solo puede imitar las estructuras lingüísticas, sin la capacidad de experimentar ni generar significado a través de la interacción vivencial con el mundo.

En el ámbito ético, la inteligencia artificial también presenta obstáculos significativos en relación con los principios filosóficos fundamentales, como los de la virtud y la moralidad autónoma. Según Aristóteles, la ética se vincula al juicio práctico, la virtud y el contexto, capacidades que la IA no puede emular completamente, pues carece de

una comprensión moral autónoma. Aunque la IA pueda asistir en la toma de decisiones éticas mediante algoritmos que imitan patrones previos, no tiene la capacidad de realizar juicios morales en función de principios virtuosos ni de evaluar el contexto de una manera reflexiva. De manera similar, desde la perspectiva kantiana, la IA podría seguir normas universales, pero no posee la capacidad de actuar según un imperativo categórico, pues carece de la autonomía moral necesaria. Por lo tanto, la implementación de la IA debe ser cuidadosamente supervisada por principios éticos humanos, para evitar que se produzcan resultados deshumanizantes o irresponsables.

En cuanto a las preocupaciones contemporáneas sobre el desarrollo de la inteligencia artificial, filósofos como Floridi, Bostrom y Harari subrayan los riesgos de una IA desalineada con los valores humanos. Floridi introduce la noción de “re-ontologización”, advirtiendo sobre cómo la IA podría transformar radicalmente nuestras estructuras sociales y epistemológicas, exigiendo una ética expansiva que contemple sus efectos. Al mismo tiempo, Bostrom y Harari enfatizan los peligros inherentes a una IA descontrolada, que podría consolidar desigualdades y concentrar el poder, erosionando así la autonomía humana. Por lo tanto, es esencial que el desarrollo de la IA sea considerado no solo como un avance tecnológico, sino también como un reto ético, en el que se busque asegurar que la inteligencia artificial amplíe las capacidades humanas y no deshumanice ni perpetúe jerarquías tecnológicas injustas.

BIBLIOGRAFÍA

Bibliografía Básica

- Aristóteles. (1985). *Ética a Nicómaco. Ética Eudemia* (E. Lledó Íñigo, Ed.; J. Pallí Bonet, Trad.). Editorial Gredos. (Obra original publicada en el siglo IV a.C.)
- Bostrom, N. (2014). *Superinteligencia: Caminos, peligros, estrategias* (Alonso, M. Trad.). Tell Editorial.
- Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.
- Gadamer, H.-G. (1993). *Verdad y método I*. Ediciones Sígueme. (Obra original publicada 1960)
- Gadamer, H.-G. (1999). *Verdad y método I: Fundamentos de una hermenéutica filosófica* (8ª ed., A. Agud Aparicio & R. de Agapito, Trads.). Ediciones Sígueme. (Obra original publicada 1960)
- Harari, Y. N. (2024). *Nexus: Una Breve Historia de las Redes de Información Desde la Edad de Piedra Hasta la I*. (Ros, J. Trad.). Debate.
- Heidegger, M. (1994). *Conferencias y artículos* (E. Barjau, Trad.). Odós.
- Heidegger, M. (1997). *Filosofía, Ciencia y Técnica*. Editorial Universitaria.
- Heidegger, M. (1997). *Ser y tiempo* (J. Gaos, Trad.). Fondo de Cultura Económica.

Husserl, E. (1949). *Ideas relativas a una fenomenología pura y una filosofía fenomenológica* (J. Gaos, Trad.). Fondo de Cultura Económica.

Kant, I. (2012). *Fundamentación para una metafísica de las costumbres* (R. Rodríguez Aramayo, Trad.). Alianza Editorial. (Obra original publicada en 1785)

Bibliografía Complementaria

Aqueveque, L. T. (2020). Jano bifronte: propiedad e impropiedad del Dasein frente a la transformación ontológica del Ge-stell. *Aufklärung: Journal of Philosophy*, 7(2), 67–78. <https://doi.org/10.18012/arf.v7i2.55825>

Berghofer, P. (2018). Husserl's noetics – towards a phenomenological epistemology. *Journal of the British Society of Phenomenology*, 50(2), 120–138. <https://doi.org/10.1080/00071773.2018.1525798>

Bolio, A. P. (2012). Husserl y la fenomenología trascendental: Perspectivas del sujeto en las ciencias del siglo XX. *Rencuentro. Análisis de Problemas Universitarios*, (65), 20-29. <http://www.redalyc.org/articulo.oa?id=34024824004>

Brayner de Farias, A. (2020). Sobre animais, humanos e máquinas: para onde vai a consciência? *Griot: Revista de Filosofia*, 20(2), 380–392. <https://doi.org/10.31977/grirfi.v20i2.1889>

Andreu, A. (2023). *Máquinas vs humanos: así modera el contenido en las redes sociales*. Business Insider. <https://www.businessinsider.es/maquinas-vs-humanos-como-modera-contenido-redes-sociales-1109935>

Garcés Giraldo, L. F., & Giraldo Zuluaga, C. (2014). Virtudes éticas en Aristóteles: Razón de los deseos y sus acciones para lograrlas. *Revista Virtual Universidad Católica del Norte*, 41, 70-78. Recuperado de <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/466/988>

García Ruiz, P. E. (2014). ¿Naturalizar la conciencia? Husserl y la tesis de la excepción humana. *Investigaciones Fenomenológicas*, (11), 99–109. <https://dialnet.unirioja.es/servlet/articulo?codigo=7435018>

Grondin, J. (2008). *¿Qué es la hermenéutica?* (A. Martínez Riu, Trad.). Herder.

- Leal Heck, R. (2023). A emergência de uma Ética da Informação. *Perspectivas*, 7(2), 347–374. <https://doi.org/10.20873/rpv7n2-57>
- Levinas, E. (2002). *Totalidad e infinito: Ensayo sobre la exterioridad* (A. R. Torres, Trans.). Ediciones Sígueme. (Trabajo original publicada en 1961).
- Klimczak, P. & Petersen, C. (2023). *AI - Limits and Prospects of Artificial Intelligence*. Bielefeld. transcript Verlag. <https://doi.org/10.1515/9783839457320>
- Krzanowski, R., & Polak, P. (2022). Ontology and AI paradigms. *Proceedings*, 81(119). MDPI. <https://doi.org/10.3390/proceedings2022081119>
- Kuhumba, K. S. (2020a). Aristotle's Virtue of Justice as an Ethical Solution to Political Corruption: Analysis and Reflection. *Al-Milal: Journal of Religion and Thought*, 2(2), 129–150. <https://doi.org/10.46600/almilal.v2i2.87>
- Maciag, R. (2023). The hermeneutics of artificial text. *Computational Science & Mathematics Forum*, 8(94). <https://doi.org/10.3390/cmsf2023008094>
- Manna, R., & Nath, R. (2021). Kantian Moral Agency and the Ethics of Artificial Intelligence. *Problemos*, 100, 139-151. <https://doi.org/10.15388/Problemos.100.11>
- Mascaró, L., & Bellocq, S. (2019). *Cuerpo, mundo y vida: Heidegger en perspectiva*. TeseoPress. ISBN del ebook: 9789878659589. <https://doi.org/10.55778/ts878659589>
- 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. <https://doi.org/10.1609/aimag.v27i4.1904>
- McDonald, F. J. (2023). AI, alignment, and the categorical imperative. *AI Ethics*, 3, 337–344. <https://doi.org/10.1007/s43681-022-00160-w>
- Молчанов, В. (2024). Значение и его тень: Терминологический анализ гуссерлевского понятия нозтико-ноэматических структур чистого сознания. *Горизонт*, 13(1), 207-240. <https://doi.org/10.21638/2226-5260-2024-13-1-207-240>
- Morandín-Ahuerma, F. (2023). Conciencia e inteligencia artificial Heidegger, Searle y Bostron. *Stoa*, 14(28), 189–209. <https://doi.org/10.25009/st.2023.28.2765>

- Müller, V. C. (2012). Introduction: Philosophy and Theory of Artificial Intelligence. *Minds & Machines*, 22, 67–69. <https://doi.org/10.1007/s11023-012-9278-y>
- Orbik, Z. (2024). Husserl's concept of transcendental consciousness and the problem of AI consciousness. *Phenomenology and Cognitive Sciences*. 23, 1151–1170. <https://doi.org/10.1007/s11097-024-09993-8>
- Padhi, I., Dognin, P., Rios, J., Luss, R., Achintalwar, S., Riemer, M., Liu, M., Sattigeri, P., Nagireddy, M., Varshney, K. R., & Bouneffouf, D. (2024). ComVas: Contextual Moral Values Alignment System. 8759–8762. *International Joint Conferences on Artificial Intelligence Organization*. <https://doi.org/10.24963/ijcai.2024/1026>
- Pretzsch, M. (2023). Formación de mundo y Dasein: La comprensión del mundo heideggeriana en Los conceptos fundamentales de la metafísica y en sus referencias a Heráclito, Aristóteles y Schelling. *Studia Heideggeriana*, 12, 97-119. <https://doi.org/10.46605/sh.vol12.2023.206>
- Rawls, J. (2006). *Teoría de la justicia* (M. D. González, Trad., 6a reimpression). Fondo de Cultura Económica. (Obra original publicada en 1971).
- Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial: Un enfoque moderno* (2ª ed.). Pearson Educación. ISBN 978-84-205-4003-0.
- Sánchez Muñoz, R., & Medina Delgadillo, J. (2018). El cuerpo vivo y la subjetividad trascendental en la fenomenología de Edmund Husserl. *Veritas*, 40, 9-28. ISSN 0717-4675. <http://dx.doi.org/10.4067/S0718-92732018000200009>
- Shin, J. (2024). A study on the interaction between human consciousness and artificial intelligence in Refik Anadol's quantum memories: the creation of quantum memories by the many-worlds interpretation of quantum physics. *International Journal of Creative Multimedia*, 5(1), 101. E-ISSN:2716-6333. <https://doi.org/10.33093/ijcm.2024.5.1.7>
- Singh, L. (2022). Automated Kantian Ethics: A Faithful Implementation. *Harvard University*. 1-20. <https://doi.org/10.48550/arXiv.2207.10152>
- The Economist. (2017). *The world's most valuable resource is no longer oil, but data*. The Economist. <https://www.economist.com/leaders/2017/05/06/the-worlds-most->

[valuable-resource-is-no-longer-oil-but-data?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=19495686130&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA6aW6BhBqEiwA6KzDcx9p0pVWPxI39VXsLuDKA6GKxKYmBlEZTnw2yArmEetOHb2HMPT2QxoC_3cQAvD_BwE&gclsrc=aw.ds](https://www.google.com/adsense/adsense?valuable-resource-is-no-longer-oil-but-data?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=19495686130&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAiA6aW6BhBqEiwA6KzDcx9p0pVWPxI39VXsLuDKA6GKxKYmBlEZTnw2yArmEetOHb2HMPT2QxoC_3cQAvD_BwE&gclsrc=aw.ds)

- Valiyev, S. I. (2023). Sun'iy intellekt tushunchasiga ontologik yondashuv. *FarDU Ilmiy Xabarlari*, 29(5), 104-107. DOI: 10.56292/SJFSU/vol29_iss5/a21
- Vargas Guillén, G. (1999). Phenomenology and artificial intelligence. The limits of subjectivity. *Estudios de Filosofía*, (19–20), 91–130.
<https://doi.org/10.17533/udea.ef.335781>
- Виктор, М. (2024). Феноменология терминологический анализ. *Horizon*, 13 (1) 207-240. <https://doi.org/10.21638/2226-5260-2024-13-1-207-240>
- Wang, J. (2021). Is Artificial Intelligence Capable of Understanding? An Analysis Based on Philosophical Hermeneutics. *Cultures of Science*, 4(3), 135–146.
<https://doi.org/10.1177/20966083211056405>
- White, J. (2022b). Autonomous Reboot: Kant, the categorical imperative, and contemporary challenges for machine ethicists. *AI & Society*, 37(4), 661-673.
<https://doi.org/10.1007/s00146-020-01142-4>
- Zapico, J. (2023). *El dron militar con inteligencia artificial que puede aprender a realizar misiones sin intervención humana*. The Objective.
<https://theobjective.com/tecnologia/2023-06-03/dron-militar-inteligencia-artificial/>

Vita

Santiago Decarlo Ayala Cabrales nació en Bogotá, Colombia, el 1 de febrero de 1999. Cursó sus estudios en el Jardín Infantil Mi Cabañita en Bogotá (2003-2004) y sus estudios primarios y secundarios en la Escuela Normal Superior de Rio de Oro, Cesar (2005-2015). Desde el año 2020, cursa su pregrado en la Universidad de Pamplona, en el Programa de Filosofía de la Facultad de Artes y Humanidades.