



UNIVERSIDAD
SANTO TOMÁS
TUNJA
ESCUELA DE INGENIERÍA - FUNDADA 1712

Quaestiones Disputatae

ISSN: 2011-0472 - IBN Publindex Categoría C / ISSN 2422-2186 (En Línea)



FILOSOFÍA DE LA INTELIGENCIA ARTIFICIAL: LOS RETOS FILOSÓFICOS FRENTE A LAS MÁQUINAS ‘PENSANTES’ *

*PHILOSOPHY OF ARTIFICIAL INTELLIGENCE: THE PHILOSOPHICAL CHALLENGES FACING
“THINKING” MACHINES*

Antony Zapata-Flórez³

Información del artículo

* Este texto es un ensayo que se desprende de un análisis en curso basado en los estudios y prácticas docente que he realizado durante años en las asignaturas a mi cargo como *Filosofía de la inteligencia artificial* y *Ética de la inteligencia artificial en las Instituciones universitarias* en las que estoy afiliado como la Universidad de Antioquia y Universidad Pontificia Bolivariana, Medellín.

* Artículo de reflexión.

* Artículo Recibido 12/02/2025 – Aceptado 21/04/2025

Cómo citar este artículo: Zapata-Flórez, A. (2024-2). Filosofía de la inteligencia artificial: los retos filosóficos frente a las máquinas pensantes. *Quaestiones Disputatae*, 17 (35), 34-57

³ Filósofo. Profesor de Cátedra de la Universidad de Antioquia. Facultad de Ingeniería y Universidad Pontificia Bolivariana. Escuela de Teología, Filosofía y humanidades. Contacto: antony.zapata@udea.edu.co
ORCID: <https://orcid.org/0000-0003-0555-6406>

Resumen

Más que un proyecto tecnológico, la ‘inteligencia artificial’ (*IA*) constituye una tesis filosófica que pretende explicar el fenómeno de la inteligencia. El diseño de sistemas que procesan información para resolver problemas humanos, como lo hacen las computadoras, plantea cuestiones de fondo sobre qué es razonar, pensar, aprender y tener consciencia, es decir, la *IA* plantea discusiones que desbordan los problemas técnicos de la misma informática y pasa a ser un objeto de la discusión filosófica. En este escrito expondré en qué consistiría una Filosofía de la inteligencia artificial, aclarando, primero, como este fenómeno no se limita al campo de la ingeniería, sino que se descubre cómo filosofía; segundo, que el abordaje filosófico de la *IA* se divide entre un análisis cognitivo y una reflexión de carácter ético; y, tercero, que la filosofía debe acercarse a la *IA* desde la sospecha y la crítica en lugar de la especulación del optimismo futurista.

Palabras clave: Inteligencia artificial; Filosofía; Inteligencia; Ética; Cognición.

Abstract

More than a technological project, Artificial Intelligence (*AI*) constitutes a philosophical thesis that seeks to explain the phenomenon of intelligence. The design of systems that process information to solve human problems, as computers do, raises fundamental questions about what it means to reason, think, learn, and be conscious. In other words, AI raises discussions that go beyond the technical problems of computer science itself and becomes an object of philosophical discussion. In this article, I will explain what a Philosophy of Artificial Intelligence would consist of, clarifying, first, how this phenomenon is not limited to the field of engineering but is discovered as philosophy; second, that the philosophical approach to AI is divided between cognitive analysis and ethical reflection; and, third, that philosophy should approach AI from a place of suspicion and criticism rather than the speculation of futuristic optimism.

Keywords: Artificial intelligence; Philosophy; Intelligence; Ethic; Cognition.

Introducción

La Filosofía de la inteligencia artificial (*IA*) como término que se refiere al estudio y análisis filosófico sobre el fenómeno de la *IA* como tecnología en sus diversas implicaciones y alusiones en el mundo humano ha ganado una atención considerable. El análisis filosófico de la *IA* no es algo nuevo en los últimos años, de hecho, el abordaje filosófico de la *IA* lleva décadas, inclusive desde el origen de la *IA* con su nacimiento en los laboratorios del *MIT* (Instituto Tecnológico de Massachussets) (Martínez-Freire, 1995). De hecho, la *IA* nació como *filosofía*⁴ mucho antes de ser una tecnología, puesto que la idea de construir máquinas inteligentes o pensantes ya se consideraba en la primera mitad del siglo XX en las discusiones filosóficas sobre la naturaleza de la mente (Lowe, 2000). Luego en 1950, en un artículo filosófico - científico Alan Turing argumentó de manera seria tal posibilidad⁵ (Turing, 1950), siete años antes de que el termino Inteligencia artificial naciera con John McCarthy.

La filosofía sobre la *IA* era un subanálisis dentro del análisis amplio tanto en filosofía de la mente, de la técnica o de la ciencia (Rabossi, 1995). De ahí que no se considerara hablar de una Filosofía de la *IA* propiamente. No obstante, con la relevancia de la *IA* en los últimos años, la filosofía sobre la *IA* ha logrado su autonomía como un análisis propio al que varios autores se han venido apropiando y que han visto adecuado usar el termino con propiedad para titular sus investigaciones académicas y literarias bajo el rotulo de Filosofía de la inteligencia artificial (Boden, 1994) (Madrid, 2024). Considero que la *IA* es un tema de importancia particular que debe tener su propio foco de análisis como otros problemas de la filosofía, tales como la mente, el lenguaje, la ciencia, el arte, etc. Por lo que hablar de una Filosofía de la *IA* como disciplina y el análisis autónomo dentro de la investigación filosófica no sólo tienen sentido, sino que es necesaria para un tema tan acuciante para la contemporaneidad humana. En este escrito me dedicaré a exponer el sentido de una Filosofía

⁴ La consolidación del movimiento de la Cibernética, las investigaciones de Walter Pitts y McCurroll la lógica y los procesos funcionales, y los avances en computación de *Alan Turing*, motivó el pensamiento funcionalista a entender que los estados mentales son procesos funcionales que podrían ser formalizados y ejecutados por materiales que no fueran propiamente orgánicos.

⁵ A. Turing en su artículo “Maquinaria computacional e inteligencia” propone el juego de la imitación, luego conocido como *Prueba de Turing*, con el cual argumentó que en si una máquina podría engañar a un ser humano, la máquina se consideraría pensante, postulando dicha prueba como un criterio de probar la inteligencia de una computadora.

de la *IA*, y, sobre todo, de en qué debería consistir, como también, el papel que debe jugar para que sea un análisis filosófico central.

Desarrollaré mi exposición en tres apartados. Primero, hablaré del sentido de una *Filosofía de la IA* basándome en que la relación que existe entre filosofía e *IA* es natural, ya que la *IA* nace como filosofía dado a que su tesis central alude a discusiones filosóficas sobre la inteligencia y el ser humano, por lo que la filosofía es inherente a todo debate referente a la *IA*. Segundo, en relación con el apartado anterior, expongo de qué tipo es la relación analítica de la filosofía con la *IA* interponiendo la discusión cognitiva, como, por ejemplo, el problema de la inteligencia, por encima de la reflexión ética y sus implicaciones sociales. Con ello no quiero decir que el abordaje ético de la *IA* no corresponda a una Filosofía de la *IA* o que no sea necesaria, al contrario, es una de las reflexiones centrales a realizar, por ello debe hacerse de la mejor manera posible, y una manera de hacerlo bien es mediante un estudio cognitivo pre-ético. Esto se debe a que los impactos y miedos éticos de la *IA* dependen de las posibilidades técnicas que podrían cumplir. Argumentaré que la investigación filosófica puede determinar lo que es técnicamente posible realizar en materia de *IA*, pues muchas de las promesas de la *IA* están inspiradas en ideas y supuestos filosóficos. Por último, expondré cuál es la función que debe tener una Filosofía de la *IA* para ser eficiente y autónoma adoptando una aptitud crítica más que un abordaje especulativo frente a las tesis y promesas de la *IA*. En resumen, mostraré porque tiene sentido hablar de una Filosofía de la *IA*, en qué consiste y su función aludiendo a las prioridades que la caracterizan y la definen.

Prioridad filosófica frente a la prioridad tecnológica

Issac Asimov dentro de sus cuentos de ciencia ficción acerca de robots inventa el término ‘Robótica’⁶, el cual definió como “construcción, mantenimiento y uso de los robots”, convirtiéndose luego en un término científico que denomina a una ciencia ya instalada y con mucho futuro por delante. A la par de inventar un término que en su momento solo era ficticio, pero que con el tiempo se usó en ciencia real como lo es la *Robótica* hoy en día,

⁶ La primera vez que aparece la palabra *Robótica* es en un relato de Asimov en 1942 titulado “Círculo vicioso”, exponiendo también, las tres leyes de la robótica, las cuales definirán la concepción de los robots por parte del autor y un rasgo característicos de sus cuentos de ciencia ficción.

Asimov acuña, dentro de sus relatos de ficción, el término '*Robopsicología*' como la disciplina científica que estudia la psicología de los robots, de las inteligencias artificiales o de las máquinas humanoides (Asimov, 2023). A diferencia del término robótica que tuvo éxito de pasar del mundo ficticio de la literatura al mundo real de la ingeniería avanzada, el término *Robopsicología* aún permanece sólo en las páginas de la literatura de entretenimiento, al menos, en los relatos de Asimov. La Robopsicología o el *robopsicólogo*, dentro del mundo ficticio de Asimov, es el encargado de estudiar el comportamiento errático o aleatorio de un *robot* cuando las tres leyes de la robótica⁷ entran en conflicto dentro del cerebro positrónico de la máquina. El robopsicólogo entra a estudiar el resultado errático de un *robot* cuando frente a una orden humana insólita, el silogismo lógico que coordina las tres leyes de la robótica del artefacto se ven conflictuadas, provocando impulsos imprevistos en el *robot* como 'sentimientos' e incluso sueños. Todo este repertorio imaginativo del mundo ficticio de Asimov para con la tecnología y sobre todo con los *robots*, dice entre líneas algo que va más allá del simple juego imaginativo del futurismo de entretenimiento, pues no deja de ser latente un sentimiento, una creencia, una intención, o más bien, una filosofía que motiva a estructurar estas ideas que no es otra que la de humanizar a las máquinas (López Corredoira, 2019).

El común denominador de los relatos de *robots* de Asimov fue el de transmitir cierta idea de reivindicación de las máquinas y del poder computacional de la tecnología para igualar o superar las capacidades humanas. Pues es curiosos que el término '*Robot*' inventado por el escritor checo Karel Capek en 1920⁸, signifique precisamente 'esclavo'. Para Asimov, el robot es un ser 'humano' artificial, que en tanto mecánico y artificial, era sometidos a trabajos forzosos y bajo la servidumbre de los seres humanos biológicos. Asimov, entre sus motivos de dotar de capacidades humanas a las máquinas robóticas pareciera tener la meta de reivindicar a estos seres artificiales e incluso liberarlos del yugo humano, como pasa en

⁷ Las tres leyes de la *Robótica* descritas por Asimov son: 1) Un robot no puede hacer daño a un ser humano, por medio de la inacción o permitir que un ser humano sea lesionado; 2) Un robot debe obedecer las órdenes recibidas por los humanos, excepto si estás órdenes entrasen en conflicto con la primera ley; 3) Un robot debe proteger su propia existencia en la medida en que está protección no sea incompatible con la primera y segunda ley.

⁸ El término *Robot*, aparece en el relato de ficción R.U.R. de Karel Capek, cuyas siglas quieren decir: Robots Universales de Rossum. Este *Rossum* es el personaje del cuento que crea humanos artificiales para servir a los humanos naturales en sus labores más forzosas por su naturaleza mecánica.

su célebre relato de “El hombre bicentenario”⁹ (Asimov, 2023): “Asimov pretendía que sus narrativas de *robots* funcionaran como una metáfora de la discriminación racial.” (Inaba, 2023, p. 7). Y parece ser que tal sentimiento de liberación de los *robots* frente a los humanos estuviera motivado a la vez por cierta ideología o, mejor dicho, por cierta filosofía acerca de qué es el ser humano y en qué consiste la inteligencia. Así, según la tesis filosófica que se lee entre líneas en los relatos de Asimov, nos señala que los *robots* y los seres humanos en esencia no son diferentes y unos no tienen que estar por encima de otros, porque los dos seres, tanto artificiales como biológicos, gozan de la misma inteligencia, una tesis que se promueve desde la filosofía computacional de la mente o desde el movimiento transhumanista (Fodor, 2002) (Dieguéz, 2020).

Issac Asimov, pues, fue uno de los pioneros de esta forma de pensar acerca de que el alma, la inteligencia o lo humano mismo podría entenderse desde la mecánica de la lógica formal o de las matemáticas, más precisamente, desde la mecánica computacional. Dicha idea con el pasar de los años a mediados del siglo XX serían tomadas en serio, como, por ejemplo, Alan Turing¹⁰ (1950), que, a diferencia de Asimov, empezó a hablar de inteligencia real en las máquinas, no como ficción para entretener, sino como ciencia estricta para investigar (Hofstadter & Dennett, 2000). Es así que la *IA* siendo la tecnología que hoy en día conocemos como los sistemas computacionales que procesan información implementados tanto en máquinas virtuales como en máquinas físicas, representan una tecnología muy particular y especial en sí misma, porque al hacer referencia directa al comportamiento humano, a sus capacidades e inteligencia, desborda el cuestionar meramente tecnológico, y entra inherentemente en el cuestionar filosófico de las discusiones cognitivas acerca de qué es el ser humano y en qué consiste su inteligencia (Masís, 2011).

John McCarthy acuñó la expresión ‘Inteligencia artificial’ entre 1956-1957, para referirse a computadoras con la capacidad de ejecutar procesos racionales similares a los que realiza un ser humano para resolver un problema complejo, tanto en la Conferencia de

⁹ En este relato, el *robot* que protagoniza la historia, *Andrew*, consigue vía legal el reconocimiento como un ser jurídico con derechos de propiedad e intelectual, independizándose de cualquier dominio humano alcanzando su libertad.

¹⁰ *A. Turing*, como criterio de probar inteligencia real en las computadoras, planteó que si una computadora al interactuar con un ser humano, éste no puede distinguir a la máquina de una persona real, demostraría que la máquina piensa ...

Darmouth de 1956, como cuando fundó el primer laboratorio de *IA* en el *MIT* (Instituto Tecnológico de Massachussets) en 1957 (Martinez-Freire, 1995) (Varela, Thompson, & Rosch, 1992). Desde entonces la *inteligencia artificial* se conoció como la ciencia de la construcción, mantenimiento y uso de sistemas computacionales que procesan información con el fin de resolver problemas complejos de una manera parecida a como lo haría la inteligencia humana. La *IA* o las máquinas inteligentes en apariencia sería desde entonces un tema reservado solo para el campo tecnológico de las computadoras y la ciencia informática. Sin embargo, su origen más profundo y su campo de juego de base fue, y nunca deja de serlo, la filosofía misma (Masis, 2023). Más precisamente, la disciplina que da sentido a la discusión sobre la posibilidad de las máquinas inteligentes es la filosofía¹¹, ya que toma como foco de reflexión al comportamiento humano y su despliegue multidimensional en ámbitos como el cognitivo, social, ético, científico, técnico, etc. De ahí que la *IA* siempre devenga en discusiones de corte filosófico y conceptual acerca de qué es el pensamiento, la memoria, el aprendizaje, las emociones, la consciencia, etc., desbordando la mera actividad ingenieril de diseñar y construir computadoras (Agre, 2005).

¿Cuál es la filosofía que se oculta detrás de la idea de que las máquinas pueden pensar? Se trata de una idea implícita incluso en los relatos clásicos de Asimov y que fue el motivante de sus ideas literarias con los robots, y que, incluso, dicha tesis filosófica está presente en el mismo término de Inteligencia artificial cuya idea central es: la inteligencia también puede ser artificial (Minsky, 1986) (Searle, *Mentes, cerebros y ciencia*, 1994) (Putnam; Nilsson, 2001). Esta tesis es muy interesante como poderosa, lo cual, sin embargo, no la hace necesariamente verdadera. Pero en su centro dicha tesis ya plantea un cuestionamiento profundo de la concepción del ser humano, a saber, que lo humano, nosotros, tú y yo, no tenemos nada especial (Prades, 2006; Crane, 2008). Pensamos normalmente que nuestra razón, pensamiento e inteligencia nos hace diferentes o superiores a plantas, animales o máquinas. Intentar afirmar en un sentido filosófico que la inteligencia también puede ser artificial es afirmar que las facultades mentales, como pensar, razonar y ser conscientes también puede darse en sistemas no naturales o biológicos, sino que también

¹¹ Ha sido común que el problema de la *Inteligencia artificial* como tema filosófico se haya limitado a las discusiones de nicho dentro de la *Filosofía de la Mente*, como pasaje de temas referentes a teorías funcionalistas de la mente, los qualia, la intencionalidad, la consciencia, etc.

pueden ser reproducidos por la ciencia humana, es decir, que el humano puede replicarse así mismo dentro de un ser mecánico mediante métodos lógico-matemáticos <Varela, Thompson & Rosch, 1992; Suay Belengue, 2012; Padial, 2019>. En consecuencia, la inteligencia de las máquinas no es una inteligencia mecánica, diferente o extraña a la nuestra, ya que la tesis en cuestión señala que las máquinas también tendrían nuestra inteligencia, puesto que no existen otras inteligencias diferentes entre sí, sino que sólo existe una, la cual puede darse tanto en seres biológicos como en seres mecánicos (Searle, 2000). Por lo tanto, la expresión inteligencia artificial en un sentido filosófico defiende que la mente, el pensamiento y la conciencia puede originarse en cualquier sistema que pueda desarrollarla, sostenerla o ejecutarla, ya sea en un cerebro biológico o no¹² (Dennett, 2006).

La idea de descentralizar al ser humano como el único ser pensante y llegar a igualarlo por medio de las máquinas, ya que la inteligencia es una facultad que puede compartirse con las computadoras, tiene un antecedente en un poco conocido cuento de Asimov que se titula precisamente “Segregacionista” (Asimov, 2023), en el cual se narra que en una hipotética sociedad humana futurista, el avance cognitivo de los *robots* ha sido tal, que dichos seres ya podían gozar de ciudadanía, y ante el avance de la tecnología prostética, tanto humanos como robots podían transformar su corporalidad, tanto de biológicos a metálicos como de metálicos a biosintéticos. Los dialogantes del relato se cuestionan si ello no era intrigante, a lo que uno de los interlocutores responde que al final no hay diferencia entre humanos y *robots*, dado que los dos son inteligentes, las diferencias que alguna vez tuvieron se limitaban a sus corporalidades, pero si éstas se pueden cambiar, el límite entre lo humano y lo robótico, o más bien, entre lo natural y artificial desaparecería (Inaba, 2023). Dice uno de los personajes que alega por la unicidad de la inteligencia antes mencionada:

¿Por qué no puede llegarse a ello? Y también seres humanos metalizados. En este momento tenemos en la Tierra dos variedades de inteligencias, pero por que tener dos. Dejemos que se acerquen la una a la otra, al final no seremos capaces de ver la diferencia. ¿Por qué íbamos a querer que se notase la diferencia?

¹² Existe una postura filosófica muy influyente a lo largo del siglo XX y XXI que se conoce como *Funcionalismo*, en la que se defiende que los estados mentales de los seres humanos podrían entenderse como procesos funcionales, con relaciones causales entre sí y con metas procedurales muy concretas, los cuales podrían formalizarse y luego aplicarse a cualquier soporte físico que pueda ejecutar dichos procesos así no sean biológicos. Allí se sugiera la posibilidad de cerebros artificiales que puedan poseer mente e inteligencia real.

La filosofía que motiva la creencia de que las máquinas pueden tener mente o pensar parte de la idea fisicalista de que el ser humano puede comprenderse como una cosa común y dentro de la naturaleza (Acosta, 2019). La mente o la conciencia serían objetos naturales sometidos a las mecánicas causales de la materia, por ende, pueden comprenderse mediante una lógica mecánica y matemática (Varela, Thompson & Rosch, 1992). De manera que el ser humano puede ser entendido como una máquina más... *¿De qué tipo de máquina sería el ser humano?* Gracias al surgimiento de la ciencia informática y a la creación de las computadoras se inició a pensar a mediados del siglo XX de que lo más probable es que el tipo de máquina que sería el ser humano es de tipo computacional (Masís, 2011). De allí el título de la obra fundadora de lo que años después se conocería como Inteligencia artificial en ese artículo ya clásico de Alan Turing: “Maquinaria computacional e inteligencia”. Turing al ver el funcionamiento de las computadoras que él mismo construía, encontraba muchas similitudes con como supuestamente operaba un ser humano al pensar. Por lo tanto, no es descabellado -proponía Turing-, que las máquinas también llegaran a ser mentales, pues el ser humano reconocido como un ser mental, sólo estaría ejecutando operaciones lógicas mediante su control central de información que no es otra cosa que su cerebro. Sí replicamos el mismo funcionamiento en un cerebro artificial, por tanto, la máquina también producirá lo que produce el cerebro biológico, a saber, mente pensante.

Según este orden de ideas, la tesis filosófica que refleja el término mismo de Inteligencia artificial al proponer la descentralización del ser humano como único ser pensante expone un movimiento de ideas que se retroalimentan entre sí, que por lo menos resulta curioso, a saber: se mecaniza al ser humano para humanizar a las máquinas. El ser humano, según dicha filosofía computacional, no tiene nada especial en comparación a la mecánica de las cosas naturales, por ende, las máquinas con un funcionamiento muy similar al ser humano compartirían eso especial de lo humano (Sadin, 2020). Mecanizar a lo humano permite al mismo tiempo atribuir facultades cognitivas y estados mentales a la máquina¹³.

¹³ Por su parte, John Searle menciona el porqué tendemos a atribuir psicología a las máquinas, señalando dos causas de esta acción humana: primero, aplicamos un cierto conductismo psicológico a todo aquello que se comporte como nosotros; y, segundo, un dualismo filosófico al pensar que, sin importar que aquello que se comporte como lo humano no sea materialmente igual, debe tener un alma o una mente igual a la nuestra.

Resulta curioso entonces que, dentro de la discusión, tanto tradicional como actual de la *IA*, para referirnos al comportamiento humano usemos términos mecanizantes o informáticos, como, por ejemplo, que la mente humana procesa y almacena información; y para referirnos a las tareas que ejecutan las computadoras lo hacemos mediante términos psicologizantes, al decir que la máquina piensa, entiende, aprende y habla (Searle, 1994). Al informatizar al ser humano buscamos curiosamente psicologizar a las máquinas, lo que significa es que se pone en práctica la tesis filosófica de la *IA* antes aludida y que Asimov reflejaba en su literatura, a saber: no hay diferencias entre el ser humano y el robot porque solo hay una sola inteligencia y ésta puede habitar tanto en cerebros biológicos como en cerebros positrónicos.

Esta filosofía antisegregacionista respecto a la mente humana frente a la supuesta mente mecánica se viralizó durante toda la segunda mitad del siglo XX y hasta muy adentrado el siglo XXI en nuestros días, llevándose a cabo un giro curioso con repercusiones filosóficas por lo menos interesantes dentro de ese otro giro antes mencionado acerca de la mecanización del ser humano para psicologizar a las máquinas. Y es que si el ser humano es un sistema mecánico cuyo funcionamiento es computacional, como en su momento lo defendió Turing y los demás que le siguieron, entonces los nuevos filósofos serían los ingenieros informáticos (Dreyfus, 1992). De manera que, imperceptiblemente, la ingeniería se transformó en psicología, pues ahora las respuestas acerca de las nociones más profundas del comportamiento humano como la razón, el pensamiento, la consciencia y el aprendizaje vendrían por parte de los programadores expertos en computadoras. Una práctica muy común hoy en día en la que se confía ingenuamente en este giro implícito desde muchos sectores y sobre todo desde los medios de comunicación: para obtener respuestas sobre alguna cuestión cognitiva recurrimos a los expertos informáticos o a los tecnólogos de las computadoras. Sería entonces como si la tecnociencia triunfara epistemológicamente frente a la reflexión filosófica acerca de qué es lo humano y lo mental (Agre P. J., 2005).

Ahora bien, frente a esta apropiación ilegítima de la filosofía por la ingeniería, Hubert Dreyfus respondió frente a los ingenieros del *MIT* -cuando éstos le expresaron que la informática había desplazado a la filosofía, pero, más bien, hubiera sucedido realmente lo contrario: la filosofía había infiltrado a la ingeniería computacional, y los programadores de

IA habían transformado cierto tipo de filosofía (la cartesiana, racionalista) en un proyecto tecnológico de investigación. Nos relata Dreyfus (2007):

Pero en 1963, cuando me invitaron a evaluar el trabajo de Alan Newell y Herbert Simon sobre los sistemas de símbolos físicos, encontré para mi sorpresa que, lejos de reemplazar la filosofía, estos investigadores pioneros habían aprendido mucho, directa o indirectamente, de nosotros los filósofos: por ejemplo, la afirmación de Hobbes de que el razonamiento era calcular, de las representaciones mentales de Descartes, de la idea de Leibniz de una característica ‘universal’ (un conjunto de primitivos en los que todo conocimiento podría expresarse), la teoría de Kant al afirmar que los conceptos eran reglas, la formalización de tales reglas por parte de Frege y la postulación de Wittgenstein de átomos lógicos en su *Tractatus Logico-philosophicus*. En definitiva, sin darnos cuenta, los investigadores de la IA estaban trabajando arduamente para convertir la filosofía racionalista en un programa de investigación. (Dreyfus, 2007, p. 247)

Así, la IA no representa el triunfo de la ciencia empírica sobre la filosofía o la psicología, sino que más bien, la computación se transformó en un campo de prueba que ha intentado materializar una tesis filosófica que dio vida a dicha idea que hoy en día vemos como un gran avance técnico. La IA es, ante todo, una tesis filosófica que se materializa y se desarrolla mediante la ciencia informática, pues la metafísica que existe de base es la que nutre sus más profundas metas y permea sus más arriesgada y radicales investigaciones tecnológicas, a saber: replicar al ser humano por completo. De ahí que, en un sentido más básico y esencial, la IA puede ser abordada desde la priorización filosófica que desde la prioridad tecnológica.

Prioridad cognitiva frente a la prioridad ética

Lo argumentado en el apartado anterior muestra la fuerte relación entre Filosofía e IA, de hecho, como expuse, la IA ya es de por sí filosófica. La aproximación legítima de la filosofía sobre la IA ya se ha venido defendiendo por muchos años y es una relación que se ha aceptado mayormente en todos los ámbitos (Agre P. E., 2005). Sin embargo, prevalece una idea que se debe precisar sobre esta relación entre filosofía e IA, la cual consiste en pensar que el abordaje filosófico de la IA sólo se limita a la reflexión ética y a la regulación humanista en un momento a posteriori de la aplicación de la IA. Estoy en desacuerdo en parte de este supuesto, pues considero que el abordaje ético de la IA por parte de la filosofía es fundamental; sin embargo, el análisis filosófico no se limita a ello, sino que la *ética de la IA*,

es una parte de la filosofía de la *IA*, la cual se compone de otro análisis que no es ético, y que incluso es anterior y más fundamental que lo propiamente ético en el ámbito reflexivo¹⁴ de la *IA*. Me refiero a que el análisis filosófico de la tesis cognitiva de la *IA* es más fundamental y guarda prioridad frente a los abordajes éticos (Agustí-Cullell, 2022; Monasterio Astobiza, 2021).

La *IA* ha significado un logro científico y tecnológico a considerar, tanto, que está cambiando la forma de vida del ser humano respecto a cómo tradicionalmente era vivida. Los avances en la automatización de tareas humanas complejas tanto en el mundo físico como digital gracias a la optimización de modelo computacionales como las Redes neuronales artificiales, supone una reconfiguración de las actividades humanas en todos los ámbitos (Benhamou, 2022). Por lo que es muy común esperar propuestas tanto discretas como pomposas desde la publicidad tecnológica. Los avances de la tecnología y de la *IA* en especial, motiva a plantear promesas y metas *problémicas* que desafían o estimulan la imaginación, como, por ejemplo, lo mencionado en el apartado anterior: *¿Se podría crear máquinas con conciencia propia?* La *IA*, por lo tanto, deambula en dos ámbitos que van desde las tesis filosóficas más radicales a las posibilidades realistas de la computación, y en esa confusión que no logra delimitar una claridad de campos, se genera los sueños más estrafalarios de las promesas futuristas, las cuales suponen que la tecnología puede realizar todo lo que podamos imaginar. De hecho, pareciera ser que la imaginación es el único instrumento a disposición para determinar el rumbo de la historia e incluso el futuro de la *IA* y sus implicaciones, lo cual es un error (López Corredoira, 2019).

Ante el avance técnico de la computación y las campañas mediáticas que inflan las novedades de la tecnología con el fin de maravillar la percepción de las personas, se instala exitosamente la idea de una exponencialidad tecnológica, que, a su vez, alimenta una noción de incertidumbre futurista sobre el destino de la sociedad. (Gayozzo Huamanchumo, 2021) En esa incertidumbre indeterminada de la tecnología, la imaginación se apodera del escenario de discusión para proponer promesas sobre las posibilidades de la ciencia y la *IA*, llenando

¹⁴ Es normal que se considere la reflexión filosófica como un pensar sobre las implicaciones futuras del uso de la tecnología o sobre la evaluación de lo ya impactado por cómo se aplica la *IA*, explorando normativas, regulaciones o políticas de gobernanza del diseño y uso de la tecnología, que sí, pero esto es una parte. De hecho, la filosofía podría tener la facultad de invalidar futuros riesgos e impactos simplemente refutando los principios teóricos de un proyecto tecnológico deslegitimando sus posibilidades de éxito técnico y práctico.

los vacíos con especulaciones, hipótesis, ficciones y fantasías de las que los más entusiastas *tecnófilos* encuentran su material para sus más osadas profecías (Valles-Peris, 2021)¹⁵. Como buen relato literario de ficción, el futuro que se promete en estas profecías tecnológicas inicia con una relación armónica y progresiva de las personas con los dispositivos, *robots e IA* hasta que inevitablemente estalla el conflicto. Tomando de ejemplo los relatos de ficción tanto en novelas como en películas, el ser humano se ve amenazado cuando la *IA* toma consciencia de su situación dado a que tecnológicamente se pudo dotar de inteligencia genuina al algoritmo computacional. Es así que, para otros tanto futuristas, el futuro puede ser apocalíptico cuando las máquinas sean inteligentes (Bales, 2024). De manera que la inteligencia es un concepto que tiene unas implicaciones prácticas muy fuertes en estas suposiciones filosóficas que los futuristas envuelven en sus promesas tecnológicas (López Corredoira, 2019). Es así que vemos dos tendencias perceptivas ante las novedades de la *IA*: los entusiastas y los temerosos. La *IA* genera unos miedos sobre los futuros impactos que puede traer en un sentido negativo, incluso apocalíptico, no sólo para personas particulares sino para la especie humana en general, y uno de esos conceptos que catalizan esos miedos y preocupaciones es el de la inteligencia, inclusive, el de la *superinteligencia* (Dieguez, 2020).

Es en ese momento en el que aparece la ética como un abordaje necesario frente a la *IA* cuando se revela un lado oscuro de esta tecnología. Nos damos cuenta, entonces, de que la *IA* puede provocar miedos, preocupaciones y posibles riesgos devastadores a los cuales hay que hacer frente, tanto para mitigar o prevenir para que el daño que podamos padecer como especie pueda ser evitado. Sea atendiendo o no al relato de las máquinas inteligentes o superinteligentes, la ética cada vez toma más importancia, ya que se considera el instrumento para encausar las acciones para el bien social y evitar los efectos negativos (Monasterio Astobiza, 2021). Pero para que esta empresa preventiva o reparatoria de los impactos negativos de la *IA* sobre la sociedad sea apropiada y efectiva primero se debe determinar qué tipo de reflexión ética se debe emplear: *¿Cuál es la mejor ética frente a la IA? ¿Cuáles son*

¹⁵ Junto a las promesas tecnológicas de la filosofía computacional de crear IAs pensantes, se adhiere la hipótesis de la Singularidad tecnológica en la que se postula el advenimiento de una realidad tan disruptiva para los seres humanos que solo será gobernada para las Super IAs, dejando al ser humano obsoleto. Una profecía basada en la percepción del progreso tecnológico alegando un supuesto exponencialidad de las máquinas cuyo avance de cálculo superará cualquier habilidad intelectual humana o acción para detenerlas.

los verdaderos riesgos que conlleva la IA? ¿a qué problemas hay que darle prioridad? ¿hay preocupaciones y riesgos válidos y con sustento? Estas son preguntas válidas y problemáticas antes de afrontar éticamente los usos y efectos de la IA.

Ahora bien, ¿cómo determinar estos temas *pre-éticos* antes de formular un programa ético en concreto frente a los riesgos de la IA? Precisamente con un análisis filosófico no ético que determine el origen de los miedos de la IA que son los que luego plantea las preguntas éticas. De manera que la filosofía de la IA no se limita a una reflexión meramente ética, sino que tiene un análisis previo a esta problemática. Este análisis filosófico *noético* o pre-ético es de tipo cognitivo, por el mero hecho que como vimos más arriba, el concepto de inteligencia, y, en consecuencia, el de superinteligencia (Gayozzo Huamanchumo, 2021), es una de las fuentes de los miedos hipotéticos más hablados en la opinión formada o no sobre IA.

Para evitar que las máquinas actúen en contra del bienestar humano se debe diseñar una ética robusta que encause el diseño, uso y el comportamiento de las máquinas. Sin embargo, para determinar la efectividad del diseño ético que se aplique como orientación de la IA, primero hay que determinar los impactos de la IA, es decir, primero se debe determinar de qué es capaz la IA y de qué no. En otras palabras, la ética de la IA depende de los impactos que provoque esta tecnología, y a la vez, dichos impactos dependen de las posibilidades tecnológicas de la IA. Pero como hemos visto, las posibilidades tecnológicas se han entremezclado filosóficamente con posibilidades cognitivas, como es el caso de proponerse dotar de inteligencia a las máquinas para que sean igual o más inteligente que los seres humanos (Bostrom, 2021). Pero, cabría preguntarse: *¿La cognición es un tema de la ingeniería?* Este ya es un tema polémico y de por sí filosófico. En consecuencia, para determinar una ética de la IA, primero hay que determinar las posibilidades de la IA, las cuales deben ser analizadas y discutidas por análisis filosóficos y disciplinas afines sobre la naturaleza de la cognición y la inteligencia (Agre P. E., 1997).

Al día de hoy se discuten en congresos a nivel global los riesgos de la IA y cómo asumirlos para evitar daños, en cuyas preocupaciones se encuentran problemas tan acuciantes y reales que vulneran la dignidad humana, que van desde los sesgos algorítmicos, pasando por la privacidad y la autonomía de las personas frente al mal manejo de los datos personales

(Grigore, 2022), hasta el desempleo a causa de la automatización de los procesos productivos (Benhamou, 2022). Como también, miedos tan hipotéticos, tales como la futura extinción humana a manos de una máquina super consciente que nos vería como una especie inferior (Bales, 2024). Pero esta indeterminación de tipos de riesgos de la *IA*, entre reales e hipotéticos, ya es problema que la reflexión ética debería abordar, antes que nada. Y una forma de determinar los verdaderos problemas éticos de la *IA*, es determinar sus posibilidades cognitivas mediante las posibilidades computacionales de sus técnicas. Antes de preguntarse: ¿Por qué deberíamos hacer frente a una super *IA*?, mejor deberíamos preguntarnos: ¿Es posible la super *IA*?; ¿qué entendemos con superinteligencia?; ¿la inteligencia se reduce a mera computación? Como vemos, las preguntas indeterminadas y confusas en el contexto ético, remite de inmediato al contexto cognitivo.

De hecho, la *IA* genera muchos miedos, pero dentro de estos es preciso distinguir miedos reales de miedos hipotéticos o falaces (Valles-Peris, 2021). Los miedos reales son aquellos que se originan a partir del impacto real que vemos en el día a día, o que el día de mañana un *algoritmo* puede afectar mi vida al negarme el acceso a un servicio por un vicio algorítmico (Ferrante, 2021). Pero también están los miedos hipotéticos que se alimentan por las fantasías de la industria cultural o ideas de sustento filosófico dudoso que defienden la tesis de la superinteligencia o el de las máquinas pensantes. La posibilidad de la consciencia artificial es objeto de mucha crítica puesto que ha padecidos discusiones, objeciones y refutaciones (Searle, 1980; Dreyfus, 1992). De manera que la reflexión filosófica en un sentido cognitivo que aclare la posibilidad de darle consciencia o superinteligencia a una máquina es de mucha ayuda y permite despejar supuestos riesgos éticos que economizan el análisis en torno a ello, y da calma a la opinión general en un panorama tan absorto en la desinformación.

Así, John Searle ya aludía a ello en su refutación de la tesis computacional de la inteligencia, cuando refutó la posibilidad de que las computadoras pudieran pensar con su ya clásico argumento de *El Cuarto Chino*¹⁶. Searle sentenciaba que el problema de que las

¹⁶ El argumento de *El cuarto chino* consiste en demostrar la diferencia esencial entre la sintaxis y la semántica dentro del lenguaje. Con dicha diferencia se sentencia que toda conducta reglamentada como el de las computadoras (*algoritmos*) cae dentro de la sintaxis, mientras que la semántica refiere a todo aquello que contiene un significado, lo que para John Searle son los estados mentales o conscientes. Así, si una computadora no puede ser más que una máquina sintáctica, nunca puede pasar a ser una entidad semántica o con estados

computadoras pudieran tener conciencia no es un problema técnico de computación, sino un problema de principio, es decir, es un problema de lógica y razonamiento de aquellos que defiende tal idea de una *IA fuerte o general* (Searle, 1994). Si el problema radica en los principios de la estructura teórica de la *IA Fuerte*, que no es otra que identificar mente con *software*, y si al final nos damos cuenta que esta analogía es falsa, nunca la tecnología podrá cumplir o construir lo que teóricamente ya está refutado:

La argumentación no tiene nada que ver con los próximos y pasmosos avances de la ciencia de la computación. No tiene nada que ver con la distinción entre procesos en serie o en paralelo, o con el tamaño de los programas, o la velocidad de las operaciones del computador, o con computadores que puedan interaccionar casualmente con su entorno, o incluso con la invención de robots. Sin duda seremos capaces de simular mucho mejor la conducta humana en los computadores de lo que podemos hacer en la actualidad, y ciertamente mucho mejor de lo que hemos podido hacer en el pasado. Lo que quiero decir esencialmente es que, si estamos hablando sobre tener una mente, todas esas simulaciones son simplemente irrelevantes. (Searle, 1994, p. 43)

Objeciones como las de Searle generan una tranquilidad ante los futuros y apocalípticos miedos de una máquina con superinteligencia, puesto que podemos comprender que ello no es posible tal como nos lo cuenta desde la publicidad cinematográfica de la *IA*. Y tales objeciones han sido tan contundentes que a más de cuarenta años siguen teniendo validez, a las cuales muchos autores de hoy en día avalan su vigencia. (Boden M. E., 2016; Madrid, 2024; Masís, 2011)

Así, la filosofía de la *IA* no se limita a la reflexión ética, sino que se extiende a un análisis arduo sobre la naturaleza general de la cognición del que ya se ha venido desarrollando a lo largo de las décadas con discusiones de variados filósofos y autores, incluso antes y desde el nacimiento de la misma *IA*. Este análisis cognitivo no es un complemento a lo ético, sino un estudio más fundamental y previo desde el cual el estudio ético se sustentaría. El análisis cognitivo sobre la inteligencia o la conciencia permite determinar las posibilidades de éstas mismas, las posibilidades de las tecnologías para simularlas o replicarlas, y así mismo determinar los impactos de la *IA* dentro de sus posibilidades de diseño y uso. Determinando en consecuencia, los verdaderos miedos éticos

mentales. Una cosa es organizar la formalidad de la información, y otra comprender el contenido de la información, ésta última es imposible para la máquina.

ante la *IA*, y la mejor forma de afrontarlos para evitar daños. En definitiva, la prioridad cognitiva de la *IA* posibilita la prioridad de aquellos problemas reales e inmediatos que ya genera la aplicación de la *IA* en nuestro mundo real y en nuestro tiempo actual.

Prioridad crítica frente a la prioridad especulativa

Hasta este momento he venido mostrando cómo la filosofía de la *IA* no se limita a ser la encargada de reflexionar las implicaciones éticas de la *IA*, sino que esta es una de sus labores, pues precisamente para realizar de manera correcta la reflexión ética debe llevar a cabo un análisis previo más fundamental que es de tipo cognitivo, ya que el tema central frente a la teoría de la *IA* es la inteligencia. De manera que, a estas alturas, se podría afirmar que la Filosofía de la *IA* se compone de dos tipos de estudios, a saber, el cognitivo y el ético, y éste último se basa en el primero. La *IA* más que una ciencia o una tecnología, es una teoría, una tesis, una metáfora, una hipótesis sobre la inteligencia¹⁷, y como tal, dicha tesis se basa en unos principios conceptuales articulados dentro de un cierto marco teórico que fundamenta todo pensamiento espontáneo sobre el tema, a la vez que fundamenta consecuentes promesas tecnológicas y posteriores investigaciones tecnológicas. En suma, la *IA* es una filosofía que se construye con una serie de ideas y argumentos que la filosofía misma entra a revisar. La misión filosófica es atender ciertos argumentos y poner a prueba su sustento lógico y su claridad conceptual, entendiendo sus contextos, marco teórico que regula dicha tesis y comprendiendo el paradigma investigativo que la nutre (Agre P. E., 1997). La misión filosófica sobre la *IA* es de tipo *arqueológica y genealógica*, ya que debe comprender el origen de las ideas y metáforas que avivan ciertas tesis, y, sobre todo, la tesis particular de la *IA* que es de tipo cognitiva, puesto que la *IA* más filosófica constituye una tesis que busca responder a la pregunta milenaria: *¿Qué es la inteligencia?* O, en definitiva: *¿Qué es el ser humano?*

¹⁷ Para Phill Agre -investigador en *IA*-, las promesas y pretensiones tecnológicas en *IA*, están motivadas por metáforas que son comunes en las investigaciones empíricas. Entre más sólida la metáfora más prometedora será el proyecto tecnológico, o si no, pasaría lo contrario. En *IA*, metáforas filosóficas han permeado la investigación ingenieril. De manera que el estudio histórico y conceptual de los principios teóricos en los laboratorios de *IA* es fundamental para rescatar las metáforas más acordes a la realidad de la inteligencia y guíe su implementación en la maquinaria.

Enfatizo esto para señalar el tipo de filosofía que debería ser la Filosofía de la *IA*. Pretendo señalar que esta posee una importancia central en estas discusiones en torno a la *IA*, dado que en su análisis crítico podrían revelarse las posibilidades e imposibilidades de la *IA* en tanto tecnología, otorgando necesidad a la crítica filosófica. La opinión generalizada de la *IA* se ha formado a partir de promesas futuristas de la tecnología (López Corredoira, 2019), las cuales se basan, a su vez, en metáforas que, como mínimo, deben verse con sospecha y conducir las al análisis con el fin de verificar sus fundamentos conceptuales y su solidez argumentativa. La filosofía de la *IA* debe tener un fuerte sentido crítico frente a las metáforas que permean su investigación tecnológica, ya que podría dar orientaciones teóricas sobre las posibilidades cognitivas de la técnica computacional.

La crítica filosófica le da independencia a la filosofía misma frente a la ciencia y la tecnología de la *IA*. La filosofía a veces se ve relegada a un simple ejercicio especulativo sobre las posibilidades futuristas de la *IA*, como, por ejemplo, en actividades de pensamiento que estimulan a la imaginación al concebir máquinas inteligentes en un futuro utópico o distópico (Kurzweil, 2005); o a imaginar si las máquinas gobiernan el planeta o si podemos convivir con *robots* cuando obtengan ciudadanía, nacionalidad o personería jurídica (Bostrom, 2016). A veces se percibe la filosofía como un apéndice de la tecnología que le ayuda a los científicos o ingenieros a aclarar este o aquel concepto o a reflexionar sobre este o aquel problema ético. De hecho, la tarea de la filosofía es ejercer la crítica, más no prestarse a la lluvia de ideas especulativas. Pues a partir de un análisis crítico de los fundamentos conceptuales de las metáforas que nutren las especulaciones de la *IA*, precisamente se podrían ahorrar dichas especulaciones y redirigir el esfuerzo reflexivo en lo verdaderamente importante. Para ello, la filosofía, los filósofos, no deben jugar al juego de la ingenuidad impresionable de muchos tecnólogos entusiastas frente al avance de ciertas tecnologías. La duda es más adecuada que el asombro frente a la *IA*, ya que como filosofía puede estar permeada por muchas metáforas falaces, y la *IA* cada vez más se vuelve un aspecto importante para la vida de cada uno y cada una de nosotros para dejarla viciar por falacias y narrativas nocivas.

El sentido crítico de una filosofía de la *IA* es fundamental porque la crítica le da sentido a la filosofía que aborda a la *IA*. La filosofía que especula, imagina y celebra las

promesas futuristas de la *IA* se convierte en una disciplina asistente de la tecnología, que se le hace útil cuando los tecnólogos quieren proponer y afirmar promesas sobre el futuro en sus arriesgadas profecías. Más aún, la filosofía no puede ser campo de juego de ideas imaginativas como caldo de cultivo para relatos de ciencia ficción que se hagan pasar por metáforas serias sobre la *IA*. Si la filosofía se limita a asistir y celebrar las profecías *tecnofuturistas*, pierde su autonomía, e incluso su sentido mismo de ser. Pues al celebrar las promesas y las tesis cognitivas sobre la *IA*, como, por ejemplo, que las máquinas podrán pensar cuando codifiquemos en un algoritmo el mecanismo del pensamiento; o cuando decodifiquemos el sentido de la consciencia y la podamos descargar en una computadora; o cuando las computadoras mismas construyan mejores computadoras y lleguen a ser *superinteligentes*; la filosofía estaría admitiendo que la solución y explicación final de la cognición y la inteligencia siempre había estado en la computación y en la pericia de los ingenieros y programadores, por lo que la filosofía no sería necesaria y no tendría sentido de existir porque, finalmente, la ingeniería habría reemplazado a la filosofía, como en su momento los estudiantes del *MIT* se lo expresaron a Dreyfus ingenuamente, pero en realidad la ingeniería no había reemplazado a la filosofía, sino que la ingeniería se había vuelto filosofía sin que ellos se hubieran dado cuenta (Dreyfus, 2007).

En cambio, si la filosofía demuestra -por medio de la crítica- que las metáforas de la *IA* presentan errores estructurales de principio en los que se basan sus metáforas, las posibilidades ingenieriles tendrían límites cognitivos considerables y se corroborarían ciertas imposibilidades de la computación frente a la inteligencia. De manera que la cognición o la inteligencia seguiría siendo un tema de la filosofía y de otras disciplinas sin ser agotada por la mera programación informática, recuperando con ello a la filosofía como disciplina autónoma. En efecto, si hablamos de una *Filosofía de la IA*, el análisis filosófico debe tener un sentido de ser, y para ello debe ser independiente de la ciencia y la tecnología, y la manera para mantener esa autonomía se constituye a partir de la función crítica frente a las bases y promesas filosóficas y tecnologías de la *IA*. La crítica filosófica salva a la filosofía misma, y -¿por qué no?-, también a la ciencia misma, pues la crítica aleja a la ciencia tecnológica de falacias y fantasías que muy comúnmente contaminan las opiniones y miedos espontáneos sobre la *IA* en general.

Conclusión

A modo de conclusión, sintetizo lo antes expuesto en los tres apartados para resaltar los puntos claves sobre en qué consiste una *Filosofía de la IA*. He venido argumentando que esta debe partir de un carácter crítico frente a las tesis y metáforas más filosóficas que sustentan las promesas de la *IA*, como lo es la *metáfora mente / computadora* que fundamenta la idea de que las computadoras puedan pensar. Desde su carácter crítico, la *Filosofía de la IA* asume una función analítica de dichas metáforas, lo cual tiene mucho de genealógico, ya que el análisis crítico de las promesas cognitivas de las máquinas se debe a una génesis de tesis que se originan en ciertas discusiones filosóficas de la mente. Desde luego, este análisis es necesario, ya que la Filosofía asume como objetivo superar dos retos, a saber: primero, desmitificar muchas tesis cognitivas de la *IA*, como de si la mente es una computadora, para luego -segundo-, filtrar los verdaderos miedos que genera la *IA* cuando se aplique en la vida social. La gran utilidad de una Filosofía de la *IA* es esta capacidad de filtrar los verdaderos impactos sociales de la *IA*, lo cual ayuda a determinar los verdaderos problemas y orientar a los programas de gobernanza ética y legal frente a los peligros de la *IA*.

No obstante, si la *Filosofía de la IA* no se comprende como análisis crítico de las posibilidades cognitivas de la tecnología, muchas metáforas sin sustento pasarán sin filtro al escenario selectivo de cuáles serían los problemas que traen consigo la *IA*. Un ejemplo de ello es la *metáfora de la superinteligencia* que, sin mucho análisis crítico, se postula como una de las principales fuentes de miedos de la *IA*. Si tomamos una metáfora sin analizar como sustento de una reflexión ética -el de las máquinas superinteligentes-, estaríamos diseñando una ética sin sustento y nada eficaz para los verdaderos impactos de la *IA*.

En definitiva, la *Filosofía de la IA* asume un carácter crítico, una función analítica y un reto de filtro muy importante en la actualidad, que la puede posicionar como una disciplina filosófica muy relevante para una realidad cada vez más tecnificada.

Referencias

- Acosta, M. (2019). Inteligencia artificial: la cibernética del ser vivo y la máquina. *Naturaleza y verdad*. 12, 12-30. <https://doi.org/10.24310/natylib.2019.v0i12>
- Agre, P. E. (1997). Toward a Critical Technical Practice: lessons Learned in Trying to Reform AI. En G. G. Bowker, *Bridging the great divide: Social Science, Technical Systems, and Cooperative Work*, (pp. 2-22). Los Ángeles, California: Erlbaum.
- Agre, P. J. (2005). The Soul Gained and Lost: Artificial Intelligence as a Philosophical project. En S. F. (eds.), *Mechanical Bodies, Computational Minds. Artificial Intelligence* (pp. 153-173). Cambridge, MA/London: The MIT Press.
- Agustí-Cullell, J. (2022). Beyond the AI Conundrum: the future of intelligence lies in its social flourishing. *Philosophy International Journal*, 1-8.
- Asimov, I. (2023). *Visiones de robot*. Barcelona: Penguin Random House.
- Bales, a. D.-G. (2024). Artificial intelligence: arguments for catastrophic risk. *Wiley*, 1-13. <https://doi.org/10.1002/wcs.1684>
- Benhamou, S. (2022). *La transformación del trabajo en la era de la inteligencia artificial: Análisis, ejemplos e interrogantes*. Santiago: Documentos de proyectos (LC/TS. 2022/85).
- Boden, M. (1994). *Filosofía de la inteligencia artificial*. México, D. F.: Fondo de Cultura económica.
- Boden, M. E. (2016). *IA: Its nature and future*. New York: Oxford University Press.
- Bostrom, N. (2016). *Superinteligencia: caminos, peligros y estrategias*. New York: Teell.
- _____ (2021). ¿Qué es el transhumanismo? *Futuro hoy*, Vol 2(2), 7-12. <https://doi.org/10.52749/fh.v2i2.1>
- Crane, T. (2008). *La mente mecánica*. México, D. F.: Fondo de Cultura económica.

- Dennett, D. (2006). *Dulces sueños: Obstáculos filosóficos para una ciencia de la conciencia*. Buenos Aires: Katz.
- Dieguéz, A. (2020). La función ideológica del transhumanismo y algunos de sus presupuestos. *Isegoría* 63, 367-386. <https://doi.org/10.3989/isegoria.2020.063.05>
- Dreyfus, H. (1992). *What computers still can't do: A critique of artificial reason*. Cambridge, Massachussetts: The MIT Press.
- _____ (2007). Why Heideggerian AI Failed and how Fixing it would require making more Heideggerian. *Philosophical psychology*, 20(2), 247-268.
- <http://dx.doi.org/10.1080/09515080701239510>
- Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos: ¿por qué debería importarnos? *Nueva Sociedad*, 294, 27-36.
- Fodor, J. (2002). *La mente no funciona así: Alcance y límite de la psicología computacional*. Madrid: Siglo XXI.
- Gayozzo Huamanchumo, P. A. (2021). Singularidad tecnológica y transhumanismo. *Teknokultura. Revista de Cultura Digitales y Movimientos Sociales*, 18(2), 195-200.
- <http://dx.doi.org/10.5209/Tekn.74056>
- Grigore, A. E. (2022). Derechos humanos e inteligencia artificial. *Ius et Scientia* 8 (1), 164-175.
- <http://doi.org/10.12795/iestscientia.2022.i01.10>
- Hofstadter, D. R., & Dennett, D. (2000). *The mind's I: Fantasies and reflections on self and soul*. New York: Basics Books.
- Inaba, S.-I. (2023). Isaac Asimov and the Current State of Space Science Fiction. *Philosophy of life*. 13(1), 5-28.
- Kurzweil, R. (2005). *La singularidad está cerca*. Londres: Penguin Books.
- López Corredoira, M. (2019). Del hombre-máquina a la máquina-hombre: materialismo, mecanicismo y transhumanismo. *Naturaleza y libertad*. Número 12, 179-190.

<https://doi.org/10.24310/natylib.2019.v0i12>

Lowe, E. J. (2000). *Filosofía de la mente*. Barcelona: Ideas Books.

Madrid, C. (2024). *Filosofía de la inteligencia artificial*. Oviedo: Pentalfa.

Martínez-Freire, P. (1995). *La nueva filosofía de la mente*. Barcelona: Gedisa.

Masís, J. (2011). El mito de lo mental: La transformación hermenéutica de la fenomenología y la inteligencia artificial. *Eikasia*, 113-132.

_____ (2023). Apostillas fenomenológicas a la filosofía de la inteligencia artificial de Phillip E, Agre. *Otrosiglo* 7(2), 73-100. Doi: 10.5281/zenodo.10442462

Minsky, M. (1986). *La sociedad de la mente*. Buenos Aires: Galápagos.

Monasterio Astobiza, A. (2021). Inteligencia artificial para el bien común (AI4SG): IA y los Objetivos del Desarrollo Sostenible. *Arbor*, 197(802), 1-19.
<https://doi.org/10.3989/arbor.2021.802007>

Nilsson, N. (2001). *Inteligencia artificial: Una nueva síntesis*. Madrid: McGraw-Hill.

Padial, J. (2019). ¿Deep Learning simulación o realización artificial de la inteligencia? *Naturaleza y Libertad* 12, 191-210. <https://doi.org/10.24310/natylib.2019.v0i12>

Prades, J. (2006). Filosofía de la mente: El estado de la cuestión. *Theoria; Revista de Teoría, Historia y Fundamentos de la Ciencia*, 315-332.

Putnam, H. (2001). *La trenza de tres cabos*. Madrid: Siglo veintiuno.

Rabossi, E. (1995). *Filosofía de la mente y la ciencia cognitiva*. Barcelona: Paidós.

Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo: Anatomía de un antihumanismo radical*. Buenos Aires: Caja Negra.

Searle, J. (1980). Minds, brains and programs. *The Behavioral and Brain Sciences*, 3, 417-424.

_____ (1994). *Mentes, cerebros y ciencia*. Madrid: Cátedra.

_____ (2000). *El misterio de la consciencia*. Barcelona: Paidós.

Suay Belengue, J. M. (2012). La mente mecánica. *Naturaleza y Libertad 1*, 139-169.
<https://doi.org/10.24310/nyl.v1i1>

Turing, A. (1950). Computing machinery and intelligence. *Mind*, 433-460.

Valles-Peris, N. (2021). Repensar la robótica y la inteligencia artificial desde la ética de los cuidados. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 18(2), 137-146. <http://dx.doi.org/10.5209/tekn.73983>

Varela, F., Thompson, E. & Rosch, E. (1992). *De cuerpo presente: La ciencia cognitiva y la experiencia humana*. Barcelona: Gedisa.