



MADRID CASADO, C.M.: *Filosofía de la Inteligencia Artificial*. Pentalfa Ediciones, Oviedo 2024, 216 pp. ISBN: 9788478486502

Manuel Castillo Moro

Académico Correspondiente de la Real Academia de Jurisprudencia y Legislación Española

lorigadecastilla@hotmail.com  0000-0003-3881-2285

Recibido: 14 de julio de 2024 | Aceptado: 08 de noviembre de 2024

Publicado en 2024 por Pentalfa Ediciones, *Filosofía de la Inteligencia Artificial* de Carlos Miguel Madrid Casado constituye una obra rigurosa y crítica acerca de la naturaleza y función de la Inteligencia Artificial (en adelante IA) en la sociedad contemporánea. Este libro llega en un momento donde el desarrollo de la IA suscita polémicos debates filosóficos, éticos y jurídicos. El autor parte de un enfoque que gravita alrededor de la filosofía materialista de Gustavo Bueno, lo que proporciona una base sólida y coherente para el análisis de la IA. Este marco filosófico le permite a Madrid Casado no solo examinar las capacidades técnicas de la IA, sino también sus implicaciones filosóficas y sociopolíticas. Al aplicar los principios del materialismo filosófico, enfatiza la importancia de entender la IA como un fenómeno concreto, enraizado en realidades materiales que operan dentro de contextos socioeconómicos específicos. Este enfoque contrasta la tendencia de abstraer la tecnología de sus aplicaciones y efectos reales, promoviendo una comprensión de la IA que es tanto práctica como profundamente reflexiva. La adopción de este marco permite una crítica rigurosa de las narrativas dominantes y ofrece herramientas para evaluar de manera razonada cómo la tecnología puede y debe integrarse en la sociedad para servir al bien común.

En el prefacio de *Filosofía de la Inteligencia Artificial*, el autor plantea la tesis de lo que podríamos denominar como «mito de la Inteligencia Artificial», argumentando que mucha de la retórica que rodea a la IA está afectada por visiones distópicas y utópicas que encuentran en los medios de comunicación su principal portal y escaparate. Una narrativa que arrastra a la opinión pública hacia una visión distorsionada del verdadero alcance de la IA. Una confusión que genera expectativas en la sociedad poco realistas sobre estos avances tecnológicos y sus repercusiones prácticas. El autor nos propone desvelar estos malentendidos y proporcionar una perspectiva más precisa y fundamentada de la IA, basada no solo en sus potenciales logros sino también en sus desafíos inherentes.

Madrid Casado explica como este tipo de narrativas se alimentan de la confluencia entre ciencia ficción y reportajes sensacionalistas que a menudo presentan a esta como una entidad casi omnisciente, capaz de superar las capacidades humanas en casi todos los aspectos. Un mito que se ve reforzado por los avances en tecnologías específicas, como el aprendizaje profundo (Deep Learning) y los grandes modelos de lenguaje (LLM, Large Language Models), pero que no son suficientes como para establecer una equivalencia con la inteligencia de los seres humanos. El autor critica lo que denomina como «filosofía espontánea» de la IA, adoptada por determinados ingenieros y desarrolladores que, consciente o inconscientemente, proyectan capacidades humanas a sistemas operativos que son solo herramientas complejas controladas y al servicio de los seres humanos.

El prefacio también llama la atención sobre la necesidad urgente de un análisis filosófico más riguroso y sistemático de la IA. Madrid Casado propone que, para entender verdaderamente la IA y sus impactos potenciales, es crucial dismantelar la narrativa citada y examinar críticamente los fundamentos gnoseológicos y ontológicos que sostienen estos sistemas operativos. Esto incluye discernir entre las capacidades reales de las tecnologías actuales y las especulaciones futuristas que a menudo se presentan como inevitables. Por lo tanto, este análisis no solo despeja malentendidos, sino que también establece un terreno más fértil para las discusiones éticas y jurídicas sobre cómo deberíamos integrar estas tecnologías en nuestra sociedad.

En la primera parte de su obra, al examinar la IA desde una óptica gnoseológica, el autor nos plantea cómo es percibida y aplicada en el mundo real. Al considerar la IA más como una herramienta tecnológica que como una disciplina científica, propone una visión pragmática que concibe a la IA como un conjunto de instrumentos diseñados para cumplir con funciones específicas, en vez de ser un ámbito de investigación enfocado en desentrañar verdades universales. Esta perspectiva invita a los desarrolladores y usuarios de IA a reflexionar sobre las limitaciones y potencialidades de estas herramientas, reconociendo que, a pesar de su posibilidades, están concebidas para alcanzar metas determinadas y solucionar problemas concretos.

Adoptar esta visión pragmática conlleva también a que la evaluación de la IA se oriente menos hacia la comparación con la inteligencia humana y más hacia la eficiencia con que estas herramientas alcanzan los propósitos para los cuales fueron diseñadas. Se enfatiza la necesidad de una evaluación constante sobre la interacción de estas herramientas tecnológicas con su entorno y el impacto que generan en los ámbitos sociales y económicos donde se desarrollan. Al definir la IA como una aplicación tecnológica, se subraya la importancia de una regulación y una gestión meticulosa de su desarrollo y uso. Este enfoque pragmático garantiza que las discusiones sobre IA se mantengan centradas en su utilidad práctica y en las consecuencias sociales que conlleva, evitando así desviarse hacia terrenos especulativos.

En la segunda parte, centrada en la ontología de la IA, el autor propone una evaluación crítica de la creencia extendida de que la IA podría, en algún momento, emular completamente la inteligencia humana. Mediante un escrutinio meticuloso, se desmonta la idea de que estos sistemas operativos, en su estado actual de diseño, puedan alcanzar

una autonomía y facultades cognitivas equiparables a las humanas. Esta afirmación se basa en un análisis de las restricciones presentes en la tecnología, enfatizando que, si bien la IA puede sobresalir en funciones específicas, continúa siendo dependiente de la intervención y control humanos para su funcionamiento y efectividad.

Uno de los debates filosóficos más significativos que Madrid Casado examina es el conocido argumento de la «habitación china» de John Searle, que cuestiona la idea de que los programas informáticos puedan poseer una comprensión semántica auténtica. Searle sostiene que, aunque una máquina pueda procesar símbolos lingüísticos de tal manera que parezca entender el lenguaje, en realidad está operando con símbolos sin ninguna comprensión verdadera de su significado. Este razonamiento destaca la diferencia entre la simulación superficial de la IA y la verdadera posesión de una mente consciente y comprensiva como la humana. Mediante esta crítica, el autor sugiere que muchas expectativas sobre el futuro de la IA se fundamentan más en suposiciones optimistas que rozan la ciencia ficción que en una apreciación realista de las capacidades tecnológicas actuales.

Otra cuestión importante que se analiza en esta segunda parte del libro es el choque entre la inteligencia emocional humana y la artificial. La inteligencia emocional implica la habilidad de reconocer y reaccionar ante las emociones propias y de los demás, una capacidad profundamente enraizada en la biología y la experiencia humana. La IA, por más sofisticados que sean sus algoritmos, carece de la habilidad para experimentar emociones genuinamente. Esta discusión es vital para comprender las barreras de la IA en escenarios que demandan empatía y juicio moral, como en el ámbito de la medicina o la abogacía, donde las decisiones deben tener en cuenta complejos factores emocionales y éticos. Unas limitaciones que nos llevan a mantener expectativas realistas sobre el papel que este tipo de tecnologías pueden desempeñar en nuestra sociedad, en el estado actual de conocimiento y desarrollo tecnológico.

La tercera parte del libro, enfocada en la ética de la IA, el autor profundiza en las cuestiones morales y políticas que surgen del desarrollo de este tipo de tecnologías. Se plantea una pregunta fundamental: ¿pueden las máquinas, como los algoritmos y los robots, ostentar un estatus ético, o son meramente herramientas en manos de sus creadores humanos? Esta distinción es crucial para comprender cómo se asigna la responsabilidad en escenarios que involucran decisiones autónomas tomadas por sistemas operativos de IA, particularmente en situaciones donde dichas decisiones podrían tener consecuencias perjudiciales o éticamente ambiguas. Al explorar estos dilemas, el autor sugiere que la capacidad de una máquina para ejecutar acciones complejas no implica necesariamente la posesión de una moralidad inherente o la capacidad para tomar decisiones éticas de manera independiente. Aparece así una clara distinción entre las IA como inteligencia débil y específica (limitadas a cuestiones concretas) y las humanas caracterizadas por una inteligencia genérica y multitarea.

El debate se amplía hacia las políticas de privacidad y los sesgos de datos, identificados como dos de los desafíos más críticos asociados con la IA en la sociedad contemporánea. La preocupación por la privacidad emerge de la habilidad de los sistemas de IA para procesar y almacenar ingentes cantidades de información personal, en ocasiones

sin el consentimiento explícito del individuo, lo que representa riesgos significativos en términos de violaciones de privacidad y un uso interesado de dicha información. Por otra parte, el sesgo en los algoritmos puede perpetuar y amplificar discriminaciones preexistentes en la sociedad, dado que los sistemas de IA a menudo se entrenan con datos que pueden reflejar prejuicios culturales o intereses económicos. El autor enfatiza que una regulación efectiva y una ética sólida son fundamentales para mitigar estos riesgos, resaltando la importancia de establecer marcos normativos que aseguren que la tecnología de IA se desarrolle y utilice de manera que respete los derechos humanos e impida una nueva tipología de desigualdad social. Difícil labor en un mundo globalizado donde cada Estado es soberano.

Ante el impacto transfronterizo de la IA, es vital desarrollar un entendimiento y un enfoque coherentes a nivel internacional para asegurar que su desarrollo beneficie a la sociedad en general, sin profundizar desigualdades o exacerbar conflictos. El autor propone la cooperación internacional para el desarrollo de estándares éticos y legales que guíen tanto la investigación como la implementación de la IA, argumentando que una colaboración global podría optimizar los beneficios de la IA mientras se minimizan sus riesgos potenciales. Esta sección del libro hace un llamado a un diálogo continuo entre tecnólogos, legisladores, filósofos y el público en general para construir un futuro donde la tecnología de IA se maneje de forma ética y responsable.

El autor no solo desafía una narrativa que enaltece la tecnología sin un adecuado escrutinio, sino que también proporciona un marco riguroso a través del cual podemos evaluar los beneficios y los riesgos de la IA. Su crítica al «mito de la inteligencia artificial» es particularmente pertinente en un momento en que el debate público a menudo se ve inundado por visiones extremas, ya sea de utopías tecnológicas o de distopías apocalípticas. Al abogar por un enfoque equilibrado y fundamentado en la filosofía materialista, ofrece un contrapunto necesario a las discusiones que frecuentemente se desvían hacia el «tecno-optimismo» sin crítica o el pesimismo paralizante ante la denominada «singularidad tecnológica», que nos pondría a los pies de nuestra creación.

La obra de Madrid Casado es un llamamiento hacia la responsabilidad. Argumenta convincentemente que el desarrollo de la IA no debe ser guiada únicamente por la búsqueda de innovación o beneficio económico, sino que debe estar profundamente arraigada en principios éticos que prioricen el bienestar humano y la justicia social. En última instancia, *Filosofía de la Inteligencia Artificial* es más que un texto académico, es una guía esencial para legisladores, profesionales de la tecnología, filósofos y el público en general para poner rumbo con sabiduría y prudencia la nave tecnológica de las IA. Su lectura invita a una reflexión crítica sobre cómo podemos, como sociedad, utilizar la IA de manera que refuerce nuestros valores más preciados y fomente un futuro sostenible y equitativo.