



ARTÍCULO ORIGINAL

Comparación y elección entre modelos de IA: análisis heterodoxo de condicionantes

Comparison and Selection among AI Models: A Heterodox Analysis of Conditions

Moshe Yanovskiy^{*†} y Yehoshua Socol^{*‡}

^{*}Jerusalem College of Technology, Jerusalem, Israel; ORCID: <https://orcid.org/0000-0002-7014-0522>

[†]Jerusalem College of Technology, Jerusalem, Israel; ORCID: <https://orcid.org/0000-0003-4167-248X>

^{*}Correspondencia al correo: moshey@jct.ac.il; socol@g.jct.ac.il

(Recibido 25 de noviembre, 2025; aceptado 21 de enero, 2025)

Resumen

Revisión heterodoxa sobre el desarrollo en la implementación cotidiana de los algoritmos de inteligencia artificial, sus sesgos y riesgos para la vida humana por falta de transparencia o efecto caja negra. Se centra la atención aquí en la evaluación de un problema que ha influido en el desarrollo de la inteligencia artificial, como es el dilema ético-económico de caja negra, junto con su paradoja. La atención a dicho problema, sobre si prima la transparencia sobre el rendimiento algorítmico (y cómo se valora, con sus sesgos y riesgos), ello permite comprender la paradoja conducente a la actual dicotomía entre el mundo anglosajón y europeo continental. Mediante un estudio bibliométrico-narrativo y crítico-hermenéutico, junto con los marcos teóricos y metodológicos de la Escuela Austria y los Neoinstitucionalistas (dada su experiencia en el análisis de otras cajas negras, como el Estado, el Sector público o la economía de bienestar), desde este trabajo se ofrece una exposición y explicación del problema, su alcance y si cabe esperar una futura convergencia de posiciones al respecto.

Palabras clave: inteligencia artificial; transparencia; sostenibilidad; dilema de caja negra; sesgos y riesgos; enfoques heterodoxos.

Cód. JEL: A14, B5, O3, P16, Z1.

Abstract

This paper presents a heterodox review of the development of everyday implementations of artificial intelligence algorithms, their biases, and the risks they pose to human life due to a lack of transparency, or the black-box effect. It focuses on evaluating a problem that has influenced the development of artificial intelligence: the ethical-economic dilemma of the black box and its associated paradox. Concentrating on this dilemma—whether transparency should prevail over algorithmic performance, and how each is valued in light of biases and risks—helps to clarify the paradox that underlies the current dichotomy between the Anglo-Saxon and continental European worlds. Using a bibliometric–narrative and critical–hermeneutic approach, and drawing on the theoretical and methodological frameworks of Austrian economics and neoinstitutional economics (with their tradition of analyzing other black boxes such as the State, the public sector, and welfare economics), this paper offers an exposition and explanation of the problem, its scope, and the prospects for a future convergence of positions on the matter.

Keywords: *artificial intelligence; transparency; sustainability; black-box dilemma; biases and risks; heterodox approaches.*

JEL Class.: *A14, B5, O3, P16, Z1.*

1. Introducción

La inteligencia artificial (IA), tal como se la conoce en la actualidad, tiene sus orígenes en la colaboración entre profesores universitarios y el ámbito militar para labores de cifrado (por ejemplo, el Proyecto Enigma: Government Code & Cypher School, con la participación de Turing). Tras la Segunda Guerra Mundial, su desarrollo comenzó en las universidades con financiamiento público (Huang et al., 2023; Gofman & Jin, 2024; Neumann et al., 2024). Un ejemplo ampliamente citado, por constituir el origen oficial del propio término inteligencia artificial (en 1956, en el Dartmouth College, con financiamiento público), fue el Dartmouth Summer Research Project on Artificial Intelligence, un evento organizado por académicos como McCarthy, Minsky, Rochester y Shannon (Doroudi, 2023). El término IA abarca diversos campos de estudio (prestando especial atención en este trabajo a las interrelaciones entre economía, ingeniería y ética aplicada). Con frecuencia, se utiliza para referirse a la capacidad de las máquinas de imitar funciones cognitivas humanas, tales como aprender a partir de la experiencia, adaptarse a nuevas tareas o desempeñar funciones como el reconocimiento de imágenes, voz o sonido, la traducción de lenguajes, entre otras, e incluso la toma de decisiones (LaGrandeur, 2024; Singla, 2024). La IA se basa en algoritmos y modelos que permiten a los ordenadores procesar información y resolver problemas de manera autónoma (Tan et al., 2024). No obstante, ello conlleva diversas implicancias, que varían según el modelo adoptado para su desarrollo. Inicialmente, tras la Segunda Guerra Mundial, la situación fue relativamente homogénea en Occidente, con el estudio de la IA promovido en el ámbito universitario y en centros estatales de investigación (por ejemplo, el machine learning y el test de Turing, 1950). Sin embargo, tras décadas de avances lentos (efectivos, aunque no eficientes), con la llegada de la globalización y la intensificación de la digitalización (Sánchez-Bayón, 2020 y 2021), las empresas comenzaron a interesarse en su desarrollo y aplicaciones (Sánchez-Bayón, 2025a). De este modo, emergió una división entre el modelo europeo (inspirado en el intervencionismo público, el open source y un enfoque centrado en la transparencia y la ética) y el modelo estadounidense (impulsado por la iniciativa privada, sin open source y orientado a la obtención de resultados eficientes). Para comprender por qué el modelo estadounidense (con su énfasis en el emprendimiento privado y la orientación empresarial) ha prevalecido desde noviembre de 2022 (Sánchez-Bayón et al., 2024a-b y 2025), resulta necesario analizar cómo la comunidad científica ha abordado esta cuestión. El gran debate se ha enmarcado en los siguientes términos: los algoritmos de IA se implementan progresivamente en el procesamiento de imágenes, el procesamiento del lenguaje natural (natural language processing, NLP), el apoyo a la toma de decisiones clínicas, la aplicación de la ley y otros ámbitos. Al mismo tiempo, se han planteado numerosas preocupaciones en torno a cuestiones éticas y a los riesgos potenciales de la IA aplicada a la resolución de problemas vinculados a la vida humana (Awad et al., 2018; Benkler, 2019; Biller-Andorno & Biller, 2019; Ngiam & Khor, 2019; Sánchez-Bayón, 2025b), lo que ha dado lugar a la publicación de cientos de trabajos sobre este tema, especialmente en los últimos años (previos a 2022 y al punto de inflexión del modelo estadounidense con los grandes modelos de lenguaje —LLMs—: ChatGPT, Gemini, Grok, entre otros; Xie & Avila, 2025). En este trabajo abordamos una de las preocupaciones más frecuentemente mencionadas: muchos algoritmos de IA son black boxes “cajas negras”, de modo que el usuario desconoce por qué la máquina elige una solución en lugar de otra (Benkler, 2019; Petkus et al., 2020; Wang et al., 2023; Marcus & Teuwen, 2024). Esta preocupación resulta plenamente aplicable a las redes neuronales profundas (deep neural networks, DNNs) con un elevado número de parámetros (hasta 108 o más). No obstante, otros métodos de IA (como la regresión logística, los árboles de decisión, las máquinas de soporte vectorial, entre otros) suelen considerarse considerablemente más transparentes (el primer ejemplo de IA probablemente pueda rastrearse hasta las espoletas de proximidad basadas en radar durante la Segunda Guerra Mundial). Si bien las DNNs se han convertido en verdaderos game changers en campos como el procesamiento del lenguaje natural

y el procesamiento de imágenes, su superioridad en otros ámbitos (por ejemplo, en la investigación y las aplicaciones biomédicas) ha sido cuestionada (Wang et al., 2023). Asimismo, la supuesta ventaja de los modelos de IA no basados en DNNs en términos de transparencia también ha sido puesta en tela de juicio por diversos autores. En particular, Lipton (2017) examina de manera exhaustiva distintos aspectos de la comprensión humana del funcionamiento de un sistema de IA. El autor analiza la transparencia del sistema en diferentes niveles: el modelo en su conjunto (simulable; Teufel et al., 2023; Chen et al., 2023; Chaudhary, 2024), los componentes individuales (descomponibles) y el algoritmo de entrenamiento (transparencia algorítmica; Grimmelikhuijsen, 2023; Cheong, 2024), explicando cómo los componentes clave del sistema de IA se relacionan con la comprensión humana de la forma en que este obtiene sus resultados. Asimismo, debe considerarse la interpretación de los resultados, la cual puede contribuir adicionalmente a la “informatividad” del sistema en su conjunto (Romanova, 2025). Lipton sostiene que, aunque los modelos no basados en DNNs presentan algoritmos más comprensibles, el conjunto completo de estos modelos no demuestra una ventaja evidente. El autor también señala que el análisis realizado por expertos humanos se ajusta razonablemente bien a la definición de una “caja negra”. No obstante, la percepción pública de los métodos de machine learning no basados en DNNs como más transparentes puede generar una ventaja significativa en su competencia con las DNNs. En consecuencia, dicha percepción puede convertirse en un factor decisivo, incluso compensando parcialmente la pérdida de desempeño, en la medida en que los algoritmos “transparentes” son considerados mucho más compatibles con soluciones que incorporan la intervención humana. En este trabajo, a partir del análisis de bibliometría narrativa (Jahin et al., 2023) y de la ilustración empírica de los teoremas de la economía austriaca (con su traducción al aula y la mejora del aprendizaje; Alonso et al., 2024; Sánchez-Bayón, 2015), se propone verificar y cuantificar la hipótesis diferencial (entre el modelo estadounidense y el modelo de la Unión Europea), según la cual la reputación de “caja negra” es ampliamente considerada como una desventaja relevante de los algoritmos de IA, influyendo directamente en sus oportunidades de implementación (especialmente en los sectores vinculados a la vida y la salud).

2. Materiales y métodos

Este estudio se basa en enfoques heterodoxos (Sánchez-Bayón, 2020 y 2025c), que aplican elementos analíticos provenientes de: a) la bibliometría narrativa (Torres, 2023; Rivas et al., 2024), más allá de la revisión sistemática tradicional de la literatura (Tahiru, 2021; Ammar, 2025; Zhu et al., 2025); y b) la teoría de la Escuela Austriaca de Economía (EAE) (Menger, 2007 [1871]; Huerta de Soto, 2000), tales como el teorema de la imposibilidad del cálculo económico bajo el socialismo (Mises, 2000 [1922] y 1949), actualmente revisado por Boettke (2000), Huerta de Soto (2010), entre otros, así como otros principios fundamentales de la economía política (Menger, 2007 [1871]; Sánchez-Bayón, 2025c). El debate en torno al teorema de la imposibilidad del cálculo económico constituye un elemento definitorio del pensamiento de la economía austriaca y ha permitido distinguirlo de otras escuelas (Huerta de Soto, 2000 y 2008; Smith, 2024); además, dicho teorema ha experimentado un resurgimiento en el contexto de la gestión de la última crisis (Sánchez-Bayón et al., 2023 y 2024c). El teorema del cálculo económico, o de la imposibilidad del socialismo, ha sido discutido y aplicado por académicos de esta tradición heterodoxa en una amplia variedad de contextos y líneas de investigación futura (por ejemplo, la gestión pública de la digitalización en la industria turística, Sánchez-Bayón et al., 2024c). Este trabajo se centra en la comparación entre dos modelos principales: a) el modelo emprendedor estadounidense (basado en la iniciativa privada, con un código cerrado que prioriza la eficiencia de los resultados); y b) el modelo académico europeo (basado en la intervención pública, con un código abierto que prioriza la transparencia y la ética en los procesos). Asimismo, el teorema de Mises se relaciona con el teorema Menger-Hayek (Menger, 2007 [1871]; Hayek, 1988) sobre la evolución institucional (modelo estadounidense) y el constructivismo (modelo europeo), así como con el teorema Huerta de Soto-Sánchez-Bayón (Huerta de Soto et al., 2021) sobre procesos dinámicos, emprendimiento y bienestar, con ilustración empírica (Alonso et al., 2024), además de abordar efectos secundarios como

la “caja negra”. De acuerdo con estos teoremas, es posible una interpretación heterodoxa del auge de la inteligencia artificial en 2022 (Floridi, 2024; Sánchez-Bayón et al., 2025), que favorece el modelo estadounidense (dejando atrás el modelo académico público de la Unión Europea). Este artículo examina si dicho acontecimiento fue una coincidencia o el resultado de una relación causal, con base en estos principios económicos (Sánchez-Bayón, 2025c).

A partir de una búsqueda en la Web of Science Core Collection mediante la secuencia clave “artificial intelligence ethics” (hasta 2022, con el auge del modelo estadounidense; Floridi, 2024; Sánchez-Bayón et al., 2025), se identificaron más de 600 resultados, con uno o dos artículos por año entre 1990 y 2000, hasta alcanzar 117 artículos en 2018 y 189 en 2019 (posteriormente se produjo el auge asociado a la COVID-19 y, en 2022, el auge de la inteligencia artificial). Dado que el objetivo del estudio es proponer recomendaciones para el desarrollo actual de la inteligencia artificial (modelo universitario europeo frente al modelo empresarial estadounidense), las tendencias más recientes adquieren una relevancia primordial. Por ello, se decidió limitar el análisis a los artículos publicados desde comienzos de 2017 hasta 2022, momento en que se realizó la búsqueda (entre la recuperación tras las crisis y el auge de la inteligencia artificial; Challoumis, 2024; Noncheva y Baykin, 2025). Se identificó un total de 400 artículos, de los cuales 267 correspondían a revistas con revisión por pares, incluyendo revistas de máximo nivel con factores de impacto entre 30 y 70 (por ejemplo, New England Journal of Medicine, The Lancet Oncology, Nature, Science; Awad et al., 2018; Benkler, 2019; Ngiam y Khor, 2019; Biller-Andorno y Biller, 2019). Se realizó un primer cribado de los artículos (utilizando juicio personal) basado en los resúmenes. Del total de 400 trabajos, se seleccionaron 198 artículos (49 %) considerados de interés para el tema de la transparencia algorítmica. Estos artículos fueron revisados y evaluados manualmente conforme a la siguiente escala de cuatro niveles:

1. El artículo no menciona la cuestión de la transparencia de los algoritmos.
2. El problema de la transparencia algorítmica se menciona, pero no se desarrolla.
3. Existe un enfoque específico en el problema de la transparencia en la inteligencia artificial.
4. El problema de la transparencia en la inteligencia artificial es central para el artículo.

No se logró formular ningún procedimiento formal de puntuación, por lo que debemos admitir que este fue, en cierta medida, subjetivo. A continuación, se presenta un ejemplo del sistema de evaluación, considerando cuatro artículos publicados en las revistas más influyentes (factor de impacto entre 30 y 70), calificados respectivamente del 1 al 4.

1. Awad et al. (2018), publicados en *Nature*, presentan los resultados de una destacada investigación sociológica sobre la toma de decisiones de personas de distintas culturas en situaciones como el dilema del tranvía (en el contexto de los vehículos autónomos). No se menciona la elección algorítmica (transparente frente a “caja negra”).
2. El análisis detallado de la inteligencia artificial (*machine learning*) aplicada actualmente en la oncología clínica (Ngiam y Khor, 2019), publicado en *The Lancet Oncology*, se limita a mencionar la conveniencia de que “los médicos comprendan cómo las herramientas de aprendizaje automático producen predicciones”.
3. Biller-Andorno y Biller (2019), en *New England Journal of Medicine*, dedican una sección específica (“Morality, Transparency, Humanity”) a la cuestión de la transparencia.
4. Benkler (2019), en *Nature*, considera la expansión del uso de los sistemas de inteligencia artificial como una amenaza grave para la sociedad y señala a los algoritmos no transparentes (“caja negra”) como un problema clave.

La puntuación de los artículos examinados durante la fase de consideración preliminar se estableció en 1.

Los factores de impacto (FI) correspondientes al año 2019 de las revistas con revisión por pares se registraron de acuerdo con el Journal Citation Reports™ (JCR, 2019). Para las actas de congresos

y otras publicaciones sin revisión por pares, se asignó un FI = 0. Asimismo, se registró el número de citas de cada publicación. Sin embargo, teniendo en cuenta que la mayoría de los trabajos son muy recientes, el número de citas no se consideró un parámetro significativo. Todo el procesamiento de datos se realizó utilizando el software MATLAB™ (www.mathworks.com). El Material Suplementario contiene la tabla en formato Excel™ (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5283013), con información sobre los autores, el título del artículo, los detalles de publicación (revista, volumen, etc.), el número de citas, el factor de impacto de la revista, la puntuación de transparencia (1–4) y el DOI (identificador digital de objeto).

3. Resultados y discusión

Durante los debates académicos del período 2017–2022 (previos al auge de la inteligencia artificial), las principales preocupaciones identificadas por los autores fueron las siguientes:

1. Aplicaciones militares.
2. Vehículos autónomos.
3. Responsabilidad legal y moral por las acciones de los sistemas de inteligencia artificial.
4. Aplicaciones de la inteligencia artificial en el sector salud. Por ejemplo, Mazurowski (2020) señala los conflictos de interés de algunos radiólogos al competir con los sistemas de consultoría basados en IA.
5. Privacidad de los datos (principalmente en el ámbito sanitario).
6. Vigilancia (sistemas de IA para reconocimiento facial, entre otros) y las correspondientes amenazas de abuso y vulneración de derechos personales (véase el importante estudio de caso de Andreeva et al., 2019).
7. Por último, pero no menos importante, los autores abordan la percepción pública de los sistemas de inteligencia artificial, influenciada por los algoritmos del tipo “caja negra”. Floridi et al. (2018), uno de los artículos más citados de la muestra, destaca la “explicabilidad” del algoritmo como un factor crucial para el éxito de la IA. Cath (2018) sostiene que una regulación y un control gubernamental extensivos constituyen factores clave para generar confianza pública en los sistemas de inteligencia artificial. Dietvorst et al. (2018) argumentan que los productores podrían superar el fenómeno de la “aversión a los algoritmos” (provocado por la falta de transparencia) ofreciendo a los usuarios la posibilidad de corregir el algoritmo, incluso de manera mínima.

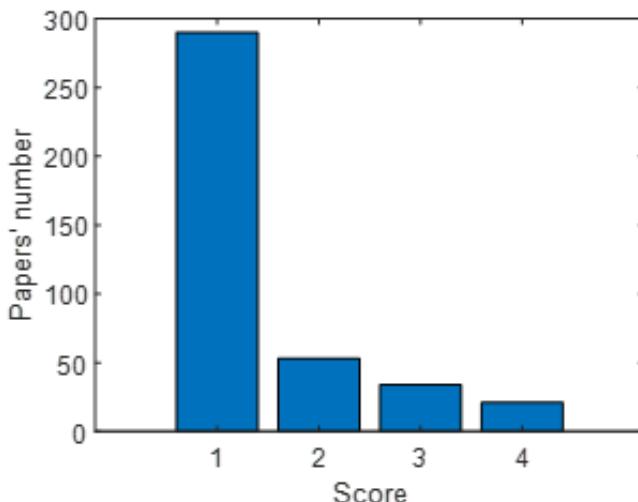
De los 400 artículos analizados, 291, 53, 34 y 22 obtuvieron las puntuaciones 1, 2, 3 y 4, respectivamente. Los resultados se muestran en la Figura 1. Solo el 27 % de los trabajos sobre ética de la inteligencia artificial abordan la cuestión de la transparencia de la IA. No obstante, en nuestra opinión, esta cifra resulta algo engañosa. En concreto, la mayoría de los artículos calificados con “1” (de hecho, todos salvo dos) no tratan específicamente sobre la inteligencia artificial, sino que abordan de manera general cuestiones éticas de la sociedad, sus valores morales y su orden social.

Por ello, concluimos que, entre los trabajos que abordan propiamente la ética de la inteligencia artificial (53 + 34 + 22), prácticamente todos mencionan el problema de la transparencia, y más de la mitad (34 + 22) lo analizan en profundidad. Asimismo, resulta pertinente señalar que los artículos publicados en revistas con un mayor factor de impacto tienden a abordar con mayor frecuencia la cuestión de la transparencia algorítmica. La Figura 2 presenta la relación entre la mediana del factor de impacto de las revistas y la puntuación de importancia asignada a la transparencia. En consecuencia, podemos sugerir que la transparencia de los algoritmos constituye un aspecto central en el contexto ético de la inteligencia artificial.

La lista de los 400 artículos analizados está disponible en: <https://drive.google.com/file/d/1aUyxGvwS4Hz0d717>

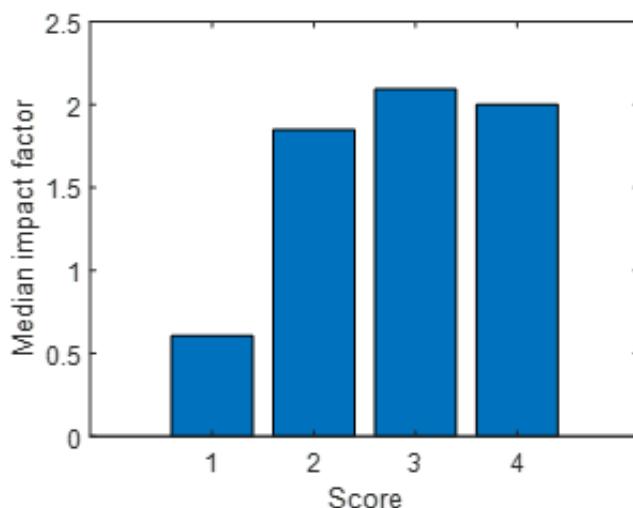
Aunque la producción científica y académica presenta un sesgo a favor del modelo universitario, dada la deriva del sistema de publicaciones hacia la industria del conocimiento, en última instancia, el financiamiento resulta indispensable. En Estados Unidos, el financiamiento proviene del sector

Figura 1. Distribución de la puntuación de importancia de la transparencia del algoritmo.



Nota: 1 – No mencionado, 2 – Mencionado pero no abordado, 3 – Enfoque especial, 4 – Tema principal.

Figura 2. Factor de impacto medio de las revistas según la puntuación de transparencia de los artículos publicados



Nota: 1 – no mencionado, 2 – mencionado pero no perseguido, 3 – enfoque especial, 4 – tema principal.

empresarial, mientras que en Europa procede principalmente del sector público. En consecuencia, el debate persiste y permanece sin resolverse.

En la actualidad, el modelo de inteligencia artificial más relevante es el enfoque empresarial estadounidense; sin embargo, antes del auge de la inteligencia artificial en 2022, el principal modelo occidental era el universitario europeo (con financiamiento público para la investigación; Foffano et al., 2023), debido a la preocupación existente en torno a las cuestiones éticas y a la transparencia de los algoritmos. En Europa se impulsó una iniciativa para preservar este modelo (Parlamento Europeo, 2023), pero el desarrollo de la inteligencia artificial en la Unión Europea resulta insuficiente, y las

principales empresas de big tech se concentran en Estados Unidos (Bolerman, 2025).

Nuestro metaanálisis confirmó que la transparencia algorítmica se discute como un tema relevante en la literatura científica que aborda la ética de la inteligencia artificial y su aceptación social. Por lo tanto, puede formularse una recomendación práctica de gran importancia: en tareas relacionadas con la vida y la salud, en todos aquellos casos en los que un sistema de inteligencia artificial difícilmente interpretable no muestre una superioridad clara en el desempeño frente a otro más interpretable, debería preferirse este último.

El riesgo de la “caja negra” persiste en la actualidad (Marcus y Teuwen, 2024), dado que la corriente dominante de la inteligencia artificial se orienta hacia el enfoque empresarial estadounidense y sus redes neuronales profundas, consideradas elementos disruptivos en términos de eficiencia, latencia, entre otros factores. No obstante, el enfoque universitario europeo sigue siendo relevante para sectores como la bioética, en los que la transparencia algorítmica y la simulación de la toma de decisiones humanas constituyen aspectos fundamentales.

En términos económicos, el modelo estadounidense implica numerosos costos de oportunidad, dado que los límites éticos son exigidos en nombre de la dignidad humana por el Derecho Internacional de los Derechos Humanos (como ius cogens o derecho imperativo de cumplimiento universal), y constituyen el mecanismo más seguro para mejorar la inteligencia artificial, en tanto permiten su control en favor de los seres humanos, así como la regulación de los monopolios tecnológicos, entre otros aspectos. En este sentido, antes de 2022 existía una mayor preocupación por la transparencia de los algoritmos y por su funcionamiento bajo la presunción de soluciones con “humanos en el circuito” (*human in the loop*).

En cuanto a las líneas futuras de investigación, se propone profundizar en el análisis de los distintos modelos, prestando especial atención a aspectos específicos como el modelo universitario de inteligencia artificial en Estados Unidos (Oh y Sanfilippo, 2025); la educación en inteligencia artificial y la integración de la diversidad y la discapacidad; así como el análisis de qué modelo resulta más exitoso en otras regiones del mundo (Al-Rashaida et al., 2025; Buragohain y Chaudhary, 2025; Dumitru et al., 2025).

4. Conclusiones

La inteligencia artificial se desarrolló en el ámbito académico durante la década de 1950, pero con el inicio de la globalización se trasladó progresivamente al mundo empresarial, dando lugar a dos modelos contrapuestos. Por un lado, el modelo empresarial estadounidense, basado en el emprendimiento privado, con código cerrado y orientado a priorizar la eficiencia de los resultados; por otro, el modelo europeo, sustentado en la intervención pública, con código abierto y enfocado en la transparencia y la ética de los procesos. Con el auge de la inteligencia artificial en 2022, podría parecer que el modelo empresarial estadounidense ha prevalecido sobre el modelo universitario europeo; sin embargo, el debate permanece abierto: ¿es preferible un modelo más eficiente pero cerrado o un modelo más transparente que simule la acción humana? En el ámbito de la bioética, entendida como el estudio de la vida y la atención sanitaria, resulta fundamental abordar esta disyuntiva. En este sentido, las críticas más relevantes provienen del modelo universitario europeo, aunque, para desarrollarse de manera más efectiva, este necesita volverse más competitivo, construyendo puentes con el sector empresarial en lugar de oponerse a él. Esta cuestión debe abordarse con urgencia, dado que el riesgo es cada vez mayor, como lo evidencia la asimetría existente entre los grandes modelos de lenguaje desarrollados en Estados Unidos y en Europa. Asimismo, resulta pertinente extender el debate a otras regiones del mundo, con el fin de identificar las propuestas creativas que puedan surgir en este ámbito.

Contribución de los autores

Moshe Yanovskiy: [Conceptualización, investigación, análisis formal, curación de datos, metodología y escritura borrador original](#)

Yehoshua Socol: [Conceptualización, investigación, análisis formal, supervisión participativa, validación redacción-revisión y edición](#)

Financiamiento

Autofinanciamiento

Conflictos de intereses

Los autores declaran no tener ningún conflicto de intereses.

Reconocimiento

Agradecemos al Dr. Avi Caspi (Jerusalem College of Technology – JCT) su apoyo para este estudio. Agradecemos enormemente las fructíferas conversaciones con el Prof. Dmitry Klokov (IRSN, Francia) y la Dra. Ariella Richardson (JCT). Este trabajo fue financiado en parte por la beca n.º 5969 del Jerusalem College of Technology.

Referencias

- Alonso MA, Sánchez-Bayón A & Gallego-Morales D (2024). Enhancing Visual Literacy and Data Analysis Skills in Macroeconomics Education: A Beveridge Curve Analysis Using FRED® Data. In: Valls Martínez M & Montero J (eds) *Teaching Innovations in Economics*. Springer, Cham., p. 51–76. https://doi.org/10.1007/978-3-031-72549-4_3
- Al-Rashaida M, Moustafa A, Mohsen W (...) & Khan A (2025). AI Strategies for Inclusive Education Resources and Support. *AI in Learning, Educational Leadership, and Special Education*, p. 349–380. 10.4018/979-8-3373-0573-8.ch012
- Ammar A (2025). Systematic and bibliometric review of artificial intelligence in sustainable education: Current trends and future research directions. *Sustainable Futures*, 10, 101033. DOI: 10.1016/j.sfr.2025.101033
- Andreeva O, Ivanov V, Nesterov A & Trubnikova T (2019). Facial Recognition Technologies in Criminal Proceedings: Problems of Grounds for the Legal Regulation of Using Artificial Intelligence. *Tomsk State University Journal*, 449, 201–212.
- Awad E, Dsouza S, Kim R (...) & Rahwan I (2018). The Moral Machine Experiment. *Nature*, 563, 59–64. <http://dx.doi.org/10.1038/s41586-018-0637-6>
- Benkler Y (2019). Don't let industry write the rules for AI. *Nature*, 569, 161.
- Biller-Andorno N & Biller A (2019). Algorithm-Aided Prediction of Patient Preferences – An Ethics Sneak Peek. *New England Journal of Medicine*, 381(15), 1480–1485. <https://doi.org/10.1056/NEJMms1904869>
- Boettke P (2000). *Socialism and the market: The socialist calculation debate re-visited*. London: Routledge.
- Bollerman M (2025). Digital Sovereigns Big Tech and Nation-State Influence. *arXiv preprint arXiv:2507.21066*.
- Buragohain D & Chaudhary S (2025). Navigating ChatGPT in ASEAN Higher Education: Ethical and Pedagogical Perspectives. *Computer Applications in Engineering Education*, 33(4). 10.1002/cae.70062
- Cath C (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Phil. Trans. R. Soc. A*, 37620180080. <http://doi.org/10.1098/rsta.2018.0080>
- Chaloulumis C (2024). Charting the course-The impact of AI on global economic cycles. In XVI International Scientific Conference. Copenhagen: ISG Konf. (pp. 103–127).

- Chaudhary G (2024). Unveiling the black box: Bringing algorithmic transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), 93-122.
- Chen Y, Zhong R, Ri N (...) & McKeown K (2023). Do models explain themselves? Counterfactual simulability of natural language explanations. *arXiv preprint arXiv:2307.08678*.
- Cheong B (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273.
- Dietvorst B, Simmons J & Massey C (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64(3). <https://doi.org/10.1287/mnsc.2016.2643>
- Doroudi S (2023). The Intertwined Histories of Artificial Intelligence and Education. *Int J Artif Intell Educ*, 33, 885-928. <https://doi.org/10.1007/s40593-022-00313-2>
- Dumitru C, Abdulsahib G, Khalaf O & Bennour A (2025). Integrating artificial intelligence in supporting students with disabilities in higher education: An integrative review. *Technology and Disability*. 10.1177/10554181251355428
- European Parliament (2023). EU AI Act: First regulation on artificial intelligence. URL: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- Floridi L (2024). Why the AI hype is another tech bubble. *Philosophy & Technology*, 37(4), 128.
- Floridi L, Cowls J, Beltrametti M (...) Vayena E (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Foffano F, Scantamburlo T & Cortés A (2023). Investing in AI for social good: an analysis of European national strategies. *AI & Society*, 38(2), 479–500.
- Gofman M & Jin Z (2024). Artificial intelligence, education, and entrepreneurship. *The Journal of Finance*, 79(1), 631-667. [10.1111/jofi.13302](https://doi.org/10.1111/jofi.13302)
- Grimmelikhuijsen S (2023). Explaining why the computer says no: Algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review*, 83(2), 241–262.
- Hayek F (1988). The fatal conceit. *Chicago: The University of Chicago*.
- Huerta de Soto J (2008). The Austrian School: Market Order and Entrepreneurial Creativity. *Cheltenham: Edward Elgar Publishing Ltd.*
- Huerta de Soto J (2010). Socialism, Economic Calculation and Entrepreneurship. *Cheltenham: Edward Elgar Publishing Ltd.*
- Huerta de Soto J, Sánchez-Bayón A & Bagus P (2021). Principles of Monetary & Financial Sustainability and Wellbeing in a Post-COVID-19 World: The Crisis and Its Management. *Sustainability*, 13(9), 4655 (1-11). <https://doi.org/10.3390/su13094655>
- Jahin M, Naife S, Saha A & Mridha M (2023). Ai in supply chain risk assessment: A systematic literature review and bibliometric analysis. *arXiv preprint arXiv:2401.10895*.
- JCR (2019). Impact factor into Journal Citation Reports. URL: <https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports>.

LaGrandeur K (2024). AI and Reverse Mimesis: From Human Imitation to Human Subjugation?.

Mimetic Posthumanism: Homo Mimeticus 2.0 in Art, Philosophy and Technics, 5, 321.

Lipton Z (2017). The Mythos of Model Interpretability. <https://arxiv.org/abs/1606.03490>

Marcus E & Teuwen J (2024). Artificial intelligence and explanation: How, why, and when to explain black boxes. *European Journal of Radiology*, 173, 111393.

Mazurowski M (2020). Artificial Intelligence in Radiology: Some Ethical Considerations for Radiologists and Algorithm Developers. *Academic Radiology*, 27(1), 127-129. <https://doi.org/10.1016/j.acra.2019.04.024>

Menger C (2007[1871]). Principles of Economics. *Auburn: Mises Institute*.

Mises L (2000[1922]). Socialism: An Economic and Sociological Analysis. *Auburn: Mises Institute*.

Mises L (1949). Human action. A treatise on Economics. *New Haven: Yale University Press*.

Neumann O, Guirguis K & Steiner R (2024). Exploring artificial intelligence adoption in public organizations: a comparative case study. *Public Management Review*, 26(1), 114-141. [doi/abs/10.1080/14719037.2022.2048685](https://doi.org/10.1080/14719037.2022.2048685)

Ngiam K & Khor W (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), E262-E273. [http://dx.doi.org/10.1016/S1470-2045\(19\)30149-4](http://dx.doi.org/10.1016/S1470-2045(19)30149-4)

Noncheva D & Baykin A (2025). Innovate approaches to crisis management and economic recovery: the role of artificial intelligence.

Oh S & Sanfilippo M (2025). Responsible AI in academia: policies and guidelines in US universities. *Information and Learning Sciences*. 10.1108/ILS-03-2025-0042

Petkus H, Hoogewerf J & Wyatt JC (2020). What do senior physicians think about AI and clinical decision support systems: Quantitative and qualitative analysis of data from specialty societies. *Clin Med*, 20(3), 324-328. [10.7861/clinmed.2019-0317](https://doi.org/10.7861/clinmed.2019-0317)

Rivas E, Núñez M, Rodríguez J & Rubio M (2024). Revisión de la producción científica sobre Storytelling mediado por tecnología entre 2019 y 2022 a través de SCOPUS. *Texto Livre*, 17, e51392.

Romanova A (2025). Analysis of Interfaces Informativeness Issues in the Development of Autonomous Artificial Intelligence Systems for Corporate Management. Available at SSRN 5347689.

Sánchez-Bayón A (2015). Filosofía del aula inteligente del S. XXI: críticas urgentes y necesarias. *Bajo Palabra*, 10, 259-269. [10.15366/bp2015.10.022](https://doi.org/10.15366/bp2015.10.022)

Sánchez-Bayón A (2020). Renovación del pensamiento económico-empresarial tras la globalización. *Bajo Palabra*, 24, 293-318. <https://doi.org/10.15366/bp.2020.24.015>

Sánchez-Bayón A (2021). The digital economy review under the technological singularity: tech-novation in labour relations and entrepreneur culture. *Sociología y Tecno-ciencia*, 11(2), 53-80. https://doi.org/10.24197/st.Extra_2.2021.53-80

Sánchez-Bayón A (2025a). ¿Cómo innovar en aprendizaje de gestión digital de riqueza y bienestar? Experiencia con monedas digitales socio-empresariales. *AROEC*, 8(1), 1-32.

Sánchez-Bayón A (2025b). Bioética y biojurídica: una revisión veinte años después. *Encuentros Multidisciplinares*, 27(79), 1-16.

- Sánchez-Bayón A (2025c). Revisión de las relaciones ortodoxia-heterodoxia en la Economía y la transición digital. *Pensamiento*, 81(314), 523-550. 10.14422/pen.v81.i314.y2025.012
- Sánchez-Bayón A, Urbina D, Alonso-Neira MA & Arpi R (2023). Problema del conocimiento económico: revitalización de la disputa del método, análisis heterodoxo y claves de innovación docente. *Bajo Palabra*, (34), 117-140. <https://doi.org/10.15366/bp2023.34.006>
- Sánchez-Bayón A, Alonso-Neira MA & Morales D (2024a). Aprender a emprender con IA y método de talento digital: Revisión de responsabilidad social universitaria. *Iberoamerican Business Journal*, SI 1(1), 48-63. <https://doi.org/10.22451/5817.ibj2024.Spec.Ed.vol1.1.11094>
- Sánchez-Bayón A, Alonso MA, Miquel AB & Sastre FJ (2024b). Aprendizaje creativo e innovación docente sobre RSC 3.0, ODS y divisas alternativas. *Encuentros Multidisciplinares*, 78, 1-13.
- Sánchez-Bayón A, Sastre FJ & Sánchez LI (2024c). Public management of digitalization into the Spanish tourism services: a heterodox analysis. *Review of Managerial Science*, 18(4), 1-19. <https://doi.org/10.1007/s11846-024-00753-1>
- Sánchez-Bayón A, Miquel-Burgos AB & Alonso-Neira MA (2025). Experience of learning technology for i-entrepreneurship training: how to prepare the students for digital economy? *Estrategia y Gestión Universitaria*, 13(1), e8765. <https://doi.org/10.5281/zenodo.14908364>
- Singla A (2024). Cognitive Computing Emulating Human Intelligence in AI Systems. *Journal of Artificial Intelligence General Science (JAIGS)*, 1(1), e38. <https://doi.org/10.60087/jaigs.v1i1.38>
- Smith A, Walsh J, Long J (...) Fisher C (2020). Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. *BMC Bioinformatics*, 21, 119. <https://doi.org/10.1186/s12859-020-3427-8>
- Smith D (2024). Austrian Economics. *Piamonte: Amazon Italy*.
- Tahiru F (2021). AI in education: A systematic literature review. *Journal of Cases on Information Technology (JCIT)*, 23(1), 1-20.
- Tan K, Wu J, Zhou H, Wang Y & Chen J (2024). Integrating advanced computer vision and ai algorithms for autonomous driving systems. *Journal of Theory and Practice of Engineering Science*, 4(01), 41-48.
- Teufel J, Torresi L & Friederich P (2023). Quantifying the intrinsic usefulness of attributional explanations for graph neural networks with artificial simulability studies. In *World Conference on Explainable Artificial Intelligence*. Cham: Springer Nature Switzerland, 361-381.
- Torres D (2023). Entre métricas y narraciones: definición y aplicaciones de la Bibliometría Narrativa. *Anuario ThinkEPI*, 17. <https://doi.org/10.3145/thinkepi.2023.e17a30>
- Turing A (1950). Computing Machinery and Intelligence. *Mind*, 49, 433-460.
- Wang L, Chen X, Zhang L, Li L, Huang Y, Sun Y, Yuan X, Davis C, Henstock P, Hodge P, Maciejewski M, Mu X, Ra S, Zhao S, Ziernik D & Fisher C (2023). Artificial intelligence in clinical decision support systems for oncology. *Int J Med Sci.*, 20(1), 79-86. 10.7150/ijms.77205
- Xie Y & Avila S (2025). The social impact of generative LLM-based AI. *Chinese Journal of Sociology*, 11(1), 31-57.
- Zhu H, Sun Y & Yang J (2025). Towards responsible artificial intelligence in education: a systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications*, 12(1). 10.1057/s41599-025-05252-6