

Comparación entre modelos de inteligencia artificial: USA-Empresarial vs. UE-Universitario*

Antonio Sánchez-Bayón

Prof. Contratado Doctor de Economía Aplicada.

Dpto. Economía Aplicada II, Universidad Rey Juan Carlos, España.

ORCID: <https://orcid.org/0000-0003-4855-8356>

e-mail: antonio.sbayon@urjc.es

Moshe Yanovskiy

Prof. Ingeniería Eléctrica y Electrónica

Dpto. Ingeniería Eléctrica y Electrónica, Jerusalem College of Technology-Israel.

ORCID: 0000-0002-7014-0522

e-mail: mosheya@jct.ac.il

Yehoshua Socol

Prof. Ingeniería Eléctrica y Electrónica

Dpto. Ingeniería Eléctrica y Electrónica, Jerusalem College of Technology-Israel.

ORCID: 0000-0003-4167-248X

e-mail: socol@jct.ac.il

Francisco Javier Sastre

Prof. Economía y Negocios

ESIC Business & Marketing School, España.

ORCID: <https://orcid.org/0000-0003-1751-9506>

e-mail: franciscojavier.sastre@esic.edu

RESUMEN

Estudio crítico e interdisciplinario de Ingeniería, Economía y Ética Aplicada sobre el auge 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 Austriaca 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.

Comparison of artificial intelligence models: Usa-Business vs. EU-University*

ABSTRACT

This is a critical and interdisciplinary study of Engineering, Economics, and Applied Ethics on the rise in the everyday implementation of artificial intelligence algorithms, the biases and risks for human life due to a transparency lack or the black box effect. Here, the focus is on the evaluation of a problem that has influenced the development of artificial intelligence: the ethical-economic dilemma of the black box, along with its paradox. Attention to this problem, about whether transparency prevails over algorithmic performance (and how it is valued, with its biases and risks), allows us to understand the paradox leading to the current dichotomy between the Anglo-Saxon world and continental Europe. Through a bibliometric-narrative and critical-hermeneutic study, along with the theoretical and methodological frameworks of the Austrian Economics and the Neoinstitutional Economics (given their experience in 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 whether a future convergence of positions on the matter can be expected.

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

Comparação de modelos de inteligência artificial: EUA-Empresas vs. UE-Universidade*

RESUMO

Um estudo crítico e interdisciplinar de Engenharia, Economia e Ética Aplicada sobre a ascensão da implementação quotidiana de algoritmos de inteligência artificial, os seus enviesamentos e riscos para a vida humana devido à falta de transparência ou ao efeito caixa negra. Este artigo centra-se na avaliação de um problema que influenciou o desenvolvimento da inteligência artificial: o dilema ético-económico da caixa negra, juntamente com o seu paradoxo. Concentrarmo-nos neste problema, em saber se a transparência prevalece sobre o desempenho algorítmico (e como é valorizada, com os seus enviesamentos e riscos), permite-nos compreender o paradoxo que leva à actual dicotomia entre o mundo anglo-saxónico e a Europa continental. Através de um estudo bibliométrico-narrativo e crítico-hermenêutico, juntamente com os quadros teóricos e metodológicos da Escola Austríaca e dos neoinstitucionalistas (dada a sua experiência na análise de outras caixas negras, como o Estado, o sector público ou a economia do bem-estar), este trabalho oferece uma exposição e explicação do problema, do seu alcance e se se pode esperar uma futura convergência de posições sobre o tema.

Palavras-chave: Inteligência artificial, transparência, sustentabilidade, dilema da caixa negra, enviesamentos e riscos, abordagens heterodoxas.

INTRODUCCIÓN

¿Cuál es el modelo mejor para el desarrollo de la inteligencia artificial (IA)? ¿El eficiente modelo de negocio y financiación privada estadounidense o el transparente modelo universitario y de financiación pública europeo? La IA, tal como se conoce en la actualidad, tiene sus antecedentes en la colaboración entre profesores universitarios y el ejército, para labores tales como la encriptación (v.g. proyecto enigma: *Government Code & Cypher School*, con la participación de Turing). Tras la II Guerra Mundial, inició su desarrollo en las universidades con financiación del sector público (Huang et al, 2023; Gofman & Jin, 2024; Neumann et al, 2024): un ejemplo popular, porque fue el origen oficial de la propia denominación IA (en 1956, en la Universidad de Dartmouth, con fondos públicos), fue “Dartmouth Summer Research Project on Artificial Intelligence”, evento organizado por académicos como McCarthy, Minsky, Rochester y Shannon (Doroudi, 2023).

La denominación IA comprende varios campos (prestando atención a las relaciones entre Economía, Ingeniería y Ética aplicada). Suele emplearse para referir la capacidad de las máquinas como imitadoras de las funciones cognitivas humanas: aprender de la experiencia, adaptarse a nuevas tareas o realizar funciones como el reconocimiento de imágenes, voz o sonido; la traducción de idiomas; etc., 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 forma autónoma (Tan et al, 2024). Sin embargo, esto tiene varias implicaciones, que difieren según el modelo utilizado para su desarrollo. Como ya se ha señalado, inicialmente, después de la Segunda Guerra Mundial, la situación era similar en Occidente, con el estudio de la IA promovido a nivel universitario y centros de investigación estatales (v.g. máquina y test de Turing, 1950). Sin embargo, tras décadas de lento progreso (con eficacia, sí, pero sin eficiencia), con la llegada de la globalización y la intensificación de la digitalización (Sánchez-Bayón, 2020 y 2021a-b), las empresas comenzaron a interesarse por su desarrollo y aplicaciones (Sánchez-Bayón, 2025a). Así, surgió una división entre el modelo europeo (inspirado en el intervencionismo público, el código abierto y un enfoque en la transparencia y la ética) frente al modelo estadounidense (impulsado por la iniciativa privada, sin código abierto y orientado a resultados eficientes). Para comprender por qué el modelo estadounidense (con su emprendimiento privado y orientación hacia el mundo de los negocios) ha prevalecido desde noviembre de 2022 (Sánchez-Bayón et al., 2024a-b y 2025), es necesario analizar cómo la comunidad científica ha abordado el tema, cuyo gran debate se ha planteado en los siguientes términos: los algoritmos de IA se implementan progresivamente en el procesamiento de imágenes, el procesamiento del lenguaje natural (PLN), el

apoyo a la toma de decisiones clínicas, las fuerzas del orden y otros ámbitos. Al mismo tiempo, se plantearon muchas inquietudes con respecto a cuestiones éticas y riesgos potenciales de la IA aplicada para resolver problemas relacionados con la vida (Awad et al, 2018; Benkler, 2019; Biller-Andorno y Biller, 2019; Ngiam y Khor, 2019; Sánchez-Bayón, 2025b), con cientos de artículos publicados sobre este tema, especialmente durante los últimos años (antes de 2022 y el punto de ruptura para el modelo de EE.UU. con LLM: ChatGPT, Gemini, Grok, etc., Xie y Avila, 2025) . En este trabajo abordamos una de las inquietudes mencionadas con frecuencia: muchos algoritmos de IA son "cajas negras" para que el usuario no tenga idea de por qué la máquina elige una solución u otra (Benkler, 2019; Petkus et al, 2020; Wang et al, 2023; Marcus y Teuwen, 2024). La preocupación por ser una caja negra es completamente aplicable a las redes neuronales profundas (DNN) con una gran cantidad de parámetros (hasta 10^8 y más). Sin embargo, otros métodos de IA (regresión logística, árbol de decisión, máquinas de vectores de soporte, etc.) generalmente se consideran incomparablemente más transparentes (el primer ejemplo de IA probablemente se deba rastrear hasta las espoletas de proximidad basadas en radar durante la Segunda Guerra Mundial). Si bien las DNN se convirtieron en verdaderos cambiadores de juego en campos como el PLN y el procesamiento de imágenes, su superioridad en otros campos (es decir, en la investigación y las aplicaciones biomédicas, está siendo cuestionada, Wang et al, 2023). La ventaja de los modelos de IA que no son DNN como más transparentes es cuestionada por algunos autores. Por ejemplo, Lipton (2017) examina escrupulosamente varios aspectos de la comprensión humana del trabajo de un sistema de inteligencia artificial. El autor considera la transparencia del sistema de IA en diferentes niveles: el modelo completo (simulable, Teufel et al, 2023; Chen et al, 2023; Chaudhary, 2024), los componentes individuales (descomponible) y el algoritmo de entrenamiento (transparencia algorítmica, Grimmelikhuijsen, 2023; Cheong, 2024) y explica cómo los componentes clave del sistema de IA se relacionan con la comprensión humana de cómo el sistema obtiene sus resultados. También se deben considerar las oportunidades de interpretación de los resultados que podrían contribuir adicionalmente a la "informatividad" de todo el sistema (Romanova, 2025). Lipton argumenta que aunque los modelos no DNN más simples tienen algoritmos más comprensibles, toda la clase de estos modelos no demuestra una ventaja obvia. El autor también muestra que el análisis por expertos humanos se ajusta bastante bien a la definición de "caja negra". De todos modos, la percepción pública de los métodos de ML no DNN como más transparentes puede causar una ventaja significativa en su competencia con las DNN. Por lo tanto, la percepción misma puede convertirse en una gran ventaja, posiblemente sobreponiendo un tanto el deterioro del rendimiento, ya que los algoritmos "transparentes" son vistos como mucho más compatibles con las soluciones "con intervención humana".

En este trabajo, gracias al análisis de bibliometría narrativa (Jahin et al, 2023) y la ilustración empírica de los teoremas de la economía austriaca (con su traslación al aula, mejorando el aprendizaje,

Alonso et al, 2024; Sánchez-Bayón, 2015), se propone verificar y cuantificar la hipótesis diferencial (entre el modelo de EE.UU. y el modelo de la UE), de que una reputación de "caja negra" es ampliamente considerada como un inconveniente importante de los algoritmos de IA, que influye directamente en las oportunidades de implementación (en especial en los sectores relacionados con la vida y la salud).

METODOLOGÍA

Este estudio se basa en enfoques heterodoxos (Sánchez-Bayón, 2020, 2022a-c y 2025c), que aplica elementos de análisis de: a) la bibliometría narrativa (Torres, 2023; Rivas et al, 2024), más allá de la tradicional revisión sistemática 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), como el teorema de la imposibilidad del cálculo económico en el socialismo (Mises, 2000[1922] y 1949) – actualmente revisado por Boettke, 2000; Huerta de Soto, 2010, etc.-, y alguno otro principal principios de economía política (Menger, 2007[1871]; Sánchez-Bayón, 2025c). El debate sobre el teorema de la imposibilidad del cálculo económico es un elemento definitorio del pensamiento de la economía austriaca, distinguiéndola de otras escuelas (Huerta de Soto, 2000 y 2008; Smith, 2024); además, este teorema ha vivido un resurgimiento bajo la gestión de la última crisis (Sánchez-Bayón et al, 2023 y 2024c). El teorema de cálculo económico o la imposibilidad de socialismo, es discutido y aplicado por los académicos de esta tradición heterodoxa en una amplia gama de contextos, además de inspirar futuras líneas de investigación (es decir, la evaluación de la gestión pública en la digitalización del sector turístico -Sánchez-Bayón et al., 2024c-d-, el sector sanitario -Tajodin y Sánchez-Bayón, 2024-, o la digitalización del empleo -Sastre et al, 2024). Este trabajo se centra en la comparación entre dos modelos principales: a) el modelo empresarial estadounidense (basado en el emprendimiento privado, con un código cerrado que prioriza la eficiencia de los resultados); 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 de Menger-Hayek (Menger, 2007[1871]; Hayek, 1988) sobre evolución institucional (modelo estadounidense) y constructivismo (modelo europeo), y el teorema de Huerta de Soto-Sánchez-Bayón (Huerta de Soto et al., 2021) sobre procesos dinámicos, emprendimiento y bienestar con ilustración empírica (Alonso et al., 2023 y 2024), además de abordar efectos secundarios como la caja negra. Según estos teoremas, es posible una lectura heterodoxa sobre el auge de la IA en 2022 (Floridi, 2024; Sánchez-Bayón et al, 2025), a favor del modelo estadounidense (dejando atrás el modelo académico público de la UE); este trabajo estudia si este evento fue casualidad o causalidad, basándose en dichos principios económicos (Sánchez-Bayón, 2025c).

Una búsqueda en la *Web of Science Core Collection* por la secuencia clave “ética de la inteligencia artificial” (hasta 2022, con el auge del modelo estadounidense, Floridi, 2024; Sánchez-Bayón et al, 2025),

se encontró por encima de 600 resultados con uno o dos artículos por año en 1990-2000, hasta 117 artículos en 2018 y 189 artículos en 2019 (posteriormente fue el auge de la COVID-19 y en 2022 el auge de la IA). Dado que nuestro objetivo es sugerir recomendaciones para el desarrollo actual de la IA (entre el modelo universitario de la UE frente al modelo empresarial estadounidense), las tendencias modernas son de primordial importancia. Por lo tanto, decidimos limitar nuestra consideración a los artículos publicados desde principios de 2017 hasta 2022, cuando se realizó la búsqueda (entre las recuperaciones de la crisis y el auge de la IA, Challoumis, 2024; Noncheva y Baykin, 2025). Se identificaron un total de 400 artículos, de los cuales 267 se encontraban en revistas con revisión por pares, incluyendo las de mayor prestigio con factores de impacto de 30 a 70 (i.e., 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 una selección primaria de los artículos (a criterio personal) con base en los resúmenes. De un total de 400, se seleccionaron 198 artículos (49 %) de interés para el tema de la transparencia del algoritmo. Se accedió a estos artículos y se calificaron manualmente según la siguiente escala de cuatro grados:

1. El documento no menciona la cuestión de la transparencia del algoritmo.
2. Se menciona el problema de la transparencia del algoritmo, pero no se aborda.
3. Se presta especial atención al problema de la transparencia de la IA.
4. El problema de la transparencia de la IA, principio para el artículo

No logramos formular una puntuación formal, por lo que debemos admitir que fue algo subjetiva. Presentamos un ejemplo de puntuación, considerando cuatro artículos publicados en las revistas más influyentes (factor de impacto 30-70) y con puntuaciones de 1 a 4, respectivamente.

1. Awad et al. (2018) presentan en *Nature* los resultados de una investigación sociológica impactante sobre la elección de personas de diversas culturas en situaciones como el problema del tranvía (en el contexto de los vehículos autónomos). No se menciona la elección del algoritmo (transparente o de "caja negra").
2. Un análisis detallado de la IA (aprendizaje automático) que se aplica actualmente en oncología clínica (Ngiam y Khor, 2019) en *Lancet Oncology* solo menciona 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 especial ('Moralidad, Transparencia, Humanidad') a la cuestión de la transparencia.
4. Benkler (2019) en *Nature* percibe la creciente aplicación de sistemas de IA como una amenaza grave para la sociedad y designa los algoritmos no transparentes ("caja negra") como un problema clave.

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

Los factores de impacto (FI) de 2019 de las revistas revisadas por pares se registraron de acuerdo con el Journal Citation Reports™ (JCR, 2019). Para actas de congresos y otras publicaciones no revisadas por pares, se estableció un FI=0. También registramos el número de veces que se citó cada publicación. Sin embargo, teniendo en cuenta que la mayoría de las publicaciones son muy recientes, no consideramos el número de citas como 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 datos sobre autores, título del artículo, detalles de la publicación (revista, volumen, etc.), veces citado, FI de la revista, la puntuación de transparencia 1-4 y DOI (identificación de objeto digital).

RESULTADOS Y DISCUSIÓN

Según los estudios analizados, las principales preocupaciones de los autores han venido siendo:

1. Aplicaciones militares
2. Vehículos autónomos
3. Responsabilidad legal y moral por las acciones de los sistemas de IA
4. Aplicaciones de IA en el ámbito sanitario. Por ejemplo, Mazurowski (2020) señala el conflicto de intereses de algunos radiólogos al competir con los sistemas de consultoría de IA.
5. Privacidad de datos (ante todo, en el ámbito sanitario)
6. Vigilancia (sistemas de IA para reconocimiento facial, etc.) y las correspondientes amenazas de abusos y violaciones 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 IA, afectados por algoritmos de tipo "caja negra". Floridi et al., 2018 (uno de los artículos más citados de la muestra) señala la "explicabilidad" del algoritmo como un factor crucial para el éxito de la IA. Cath (2018) insiste en que una amplia regulación y control gubernamental son factores clave para la confianza pública en los sistemas de IA. Dietvorst et al. (2018) argumentan que los fabricantes podrían superar la "aversión a los algoritmos" (causada por la falta de transparencia) brindando al usuario la oportunidad de corregir el algoritmo, incluso mínimamente.

De 400 artículos, 291, 53, 34 y 22 obtuvieron puntuaciones de 1, 2, 3 y 4, respectivamente. Los resultados se muestran en la Fig. 1. Solo el 27% de los artículos sobre ética de la IA abordan el tema de la transparencia de la IA. Sin embargo, en nuestra opinión, este número es algo engañoso. Es decir, la mayoría de los artículos puntuados con "1" (en realidad, todos menos dos) no tratan específicamente de la IA, sino esencialmente de cuestiones éticas de la sociedad, sus valores morales y el orden social. Por lo tanto, concluimos que de los artículos que tratan adecuadamente la ética de la inteligencia artificial

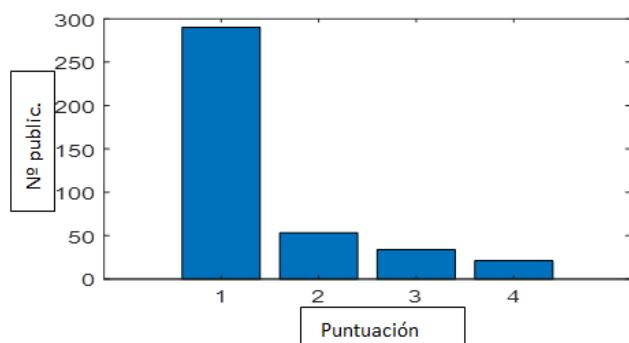
(2+53+34+22), casi todos mencionan el tema de la transparencia, y más de la mitad (34+22) lo abordan. También puede ser significativo observar que los artículos publicados en revistas con mayor FI tienden a abordar el tema de la transparencia de la IA con mucha más frecuencia. La Fig. 2 presenta la mediana de la puntuación de importancia del FI de la revista frente a la de la transparencia. Por lo tanto, podemos sugerir que la transparencia de los algoritmos es un problema importante en el contexto ético de la IA.

La lista de los 400 artículos analizados está disponible en:

<https://drive.google.com/file/d/1aUyxGvwS4Hz0d717LVmRk2clnabIIn2V/view?usp=sharing>

Figura 1:

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

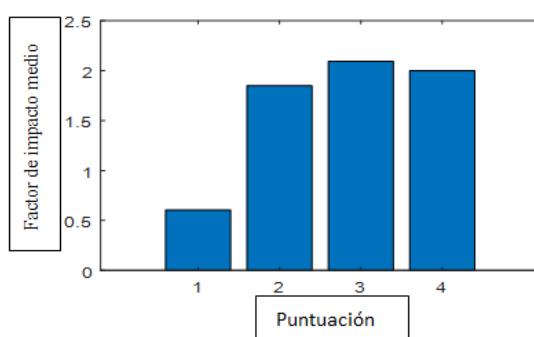


1 – no mencionados, 2 – mencionados no atendidos, 3 – enfoque especial, 4 – tema principal.

Fuente: elaboración propia

Figura 2:

Factor de impacto medio de revistas según puntuación de transparencia de artículos publicados

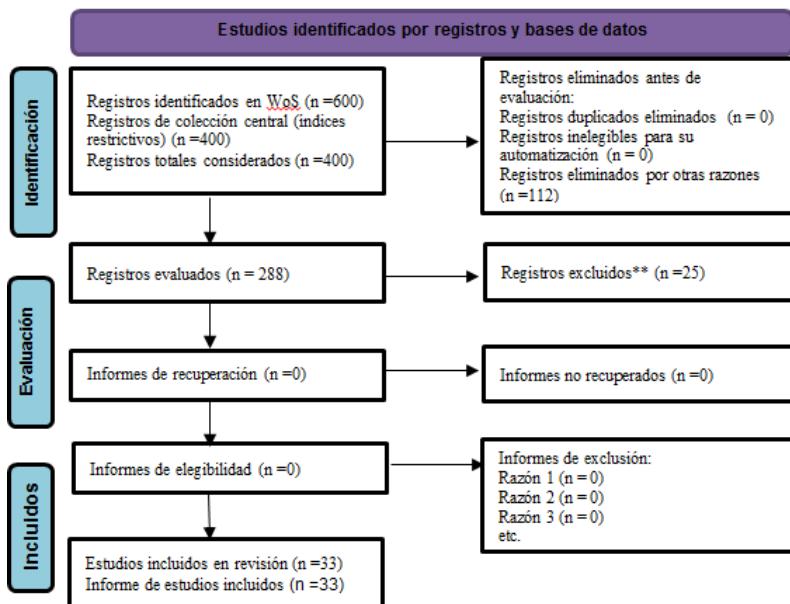


1 – no mencionados, 2 – mencionados no atendidos, 3 – enfoque especial, 4 – tema principal.

Fuente: elaboración propia

Figure 3:

Método PRISMA aplicado a la ética de AIA, sus sesgos y riesgos.



Fuente: elaboración propia (basado en Page et al, 2020).

CONCLUSIONES

Actualmente, el modelo de IA más relevante es el del enfoque empresarial estadounidense, pero antes del auge de la IA en 2022, el principal modelo de referencia occidental de IA era el modelo universitario de la UE (con fondos públicos para investigación, Foffano et al, 2023), debido a la preocupación por la ética y la transparencia de los algoritmos. Hubo una iniciativa para preservar este modelo en Europa (European Parliament, 2023), pero los resultados no han acompañado (siendo inferior el desarrollo de IA en la UE), puesto que las grandes tecnológicas se encuentran en EE.UU. (Bollerman, 2025). Este meta-análisis confirma que la transparencia de los algoritmos ha sido un debate relevante dentro de la literatura científica que aborda la ética de la IA y su aceptación social, pero desde el 2022, se viene focalizando la atención en el rendimiento y eficiencia de IA (favoreciéndose el modelo de EE.UU.). Por lo tanto, y a modo de prevención (para no perder el conocimiento generado antes de 2022), se puede formular una importante recomendación práctica: en tareas relacionadas con la vida y la salud, debería ser preferible una IA que priorice la transparencia al rendimiento, procurando imitar más el tipo de decisiones que toman los humanos al respecto (o sea, favorecer IA que simulen los procesos de toma de decisión humana y cuyo resultado sea fácilmente interpretable por humanos), evitándose así el problema de la caja negra insondable.

El riesgo de caja negra persiste hoy (Marcus y Teuwen, 2024) porque la corriente principal de la IA se centra en el enfoque empresarial estadounidense y sus redes neuronales profundas (DNN), como las transformadoras del campo, focalizadas en la eficiencia, la latencia, el rendimiento, etc. Sin embargo –como se viene señalando–, el enfoque universitario de la UE sigue siendo relevante para

sectores como la bioética, porque la transparencia de los algoritmos y la simulación de decisiones humanas son aspectos fundamentales (al preferirse la transparencia a la eficiencia, por afectada a la dignidad humana). En términos económicos, existen muchos costos de oportunidad con el modelo estadounidense, porque los límites éticos son exigidos por la dignidad según el Derecho Internacional de los Derechos Humanos (como *ius cogens* o derecho imperativo para todos), y es la forma más segura de mejorar y avanzar en el desarrollo de la IA (para controlar la IA en favor de los seres humanos, para restringir los monopolios tecnológicos, etc.). En este sentido, antes de 2022, existía una gran preocupación sobre la transparencia de los algoritmos y su funcionamiento bajo la presunción de soluciones con participación humana, por lo que no debería perderse en el ámbito biético.

Para futuras líneas de investigación, se pretende profundizar en los modelos, con atención a aspectos especiales, como el modelo universitario de IA en EE.UU. (Oh y Sanfilippo, 2025); la educación en IA y la integración de la diversidad y la discapacidad, así como analizar qué modelo tiene éxito en otras partes del mundo (Al-Rashaida et al, 2025; Buragohain y Chaudhary, 2025; Dumitru et al, 2025).

REFERENCIAS

Alonso MA, Sánchez-Bayón A & Castro-Oliva M (2023). Teoría austriaca del ciclo económico aplicada al caso español: del inicio del euro a la gran recesión y su recuperación. *Revista De Métodos Cuantitativos Para La Economía Y La Empresa*, 35, 280-310. <https://doi.org/10.46661/revmetodoscuanteconempresa.6837>

Alonso MA, Sánchez-Bayón A. y Gallego-Morales D (2024). Enhancing Visual Literacy and Data Analysis Skills in Macroeconomics Education. 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., Aishan, H., Abuwatfa, R. & Khan, A. (2025). AI Strategies for Inclusive Education Resources and Support, AI in Learning, Educational Leadership, and Special Education, 10.4018/979-8-3373-0573-8.ch012, p. 349-380.

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. y 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., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & 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. y 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>

Challoumis 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., Zhao, C., He, H., Steinhardt, J., Yu, Z., & 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. y 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., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & 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. DOI: 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:

doi/abs/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. *ББК 65.9 (4Укр) 261-18я431 Ф 59, 455.*

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. doi: 10.7861/clinmed.2019-0317. PMID: 32414724; PMCID: PMC7354034.

Rivas E, Núñez M, Rodríguez J y 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. DOI: 10.15366/bp2015.10.022

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

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

Sánchez-Bayón, A. (2021b). Urgencia de una filosofía económica para la transición digital: Auge y declive del pensamiento anglosajón dominante y una alternativa de bienestar personal, *Miscelánea Comillas. Rev. Ciencias Humanas y Sociales*, 79(155): 521-551. DOI: <https://doi.org/10.14422/mis.v79.i155.y2021.004>

Sánchez-Bayón, A. (2022a). De la Síntesis Neoclásica a la Síntesis Heterodoxa en la economía digital. *Procesos de Mercado*, 19(2): 277-306. <https://doi.org/10.52195/pm.v19i2.818>

Sánchez-Bayón, A. (2022b). ¿Crisis económica o economía en crisis? Relaciones ortodoxia-heterodoxia en la transición digital. *Semestre Económico*, 11(1): 54-73 doi: <http://dx.doi.org/10.26867/se.2022.1.128>

Sánchez-Bayón, A. (2022c). La Escuela Económica Española y su relación con los enfoques heterodoxos. *Semestre Económico*, 25(58): 1-28. <https://doi.org/10.22395/seec.v25n58a2>

Sánchez-Bayón, A. (2024). Ortodoxia versus heterodoxias sobre la colonización del Oeste estadounidense por empresas religiosas e ideológicas. *Carthaginensis*, 40(77): 117-156. DOI: <https://doi.org/10.15366/carthaginensis.40.77.117-156>

[https://doi.org/10.62217/carth.457.](https://doi.org/10.62217/carth.457)

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. DOI: 10.14422/pen.v81.i314.y2025.012

Sánchez-Bayón, A., Urbina, D., Alonso-Neira, M.A., & 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, M.A., 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.ijb2024.Spec.Ed.vol1.1.11094>

Sánchez-Bayón, A., Alonso, M.A., Miquel, A.B., Sastre, F.J. (2024b). Aprendizaje creativo e innovación docente sobre RSC 3.0, ODS y divisas alternativas. *Encuentros Multidisciplinares*, 78: 1-13

Sánchez-Bayón, A., Sastre, F.J. & Sánchez, L.I. (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., Pellejero, C., Luque, M. (2024d). Una revisión de la producción científico-académica sobre turismo en la Unión Europea (2013-23). *Iberian Journal of the History of Economic Thought*, 11(1): 55-64

Sánchez-Bayón, A., Miquel-Burgos, A.B., & Alonso-Neira, M.A. (2025). Experience of learning technovation 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>

Sastre Segovia, F. J., García Vaquero, M., Sánchez Bayón, A., & Mazier, A. (2024). ¿Recuperación económica española vía Pacto Verde Europeo? Evaluación de empleos verdes y sus capacidades. *Semestre Económico*, 13(1), 43-64. DOI: <https://doi.org/10.26867/se.2024.v13i1.162>

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 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.

Tajodin, A. y Sánchez-Bayón, A. (2024). ¿En qué falla la digitalización del sector de la salud?

Limitaciones y sesgos. *Revista Gestión De Las Personas Y Tecnología*, 17(51), 51-67. <https://doi.org/10.35588/w42e3v90>

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, p. 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. doi: 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.

Declaración de intereses y autoría

Se declara no tener ningún conflicto de intereses, que pueda haber influido en los resultados obtenidos o las interpretaciones propuestas. Todos los autores han participado en todos los apartados y fases del estudio.

Autor de correspondencia: Antonio Sánchez-Bayón (antonio.sbayon@urjc.es)

Financiación: investigación realizada con propios fondos de los investigadores (con apoyo de Jerusalem College of Technology beca №5969, más GESCE-URJC, GID TIC-TAC CEE-URJC y CIELO ESIC).

Agradecimientos: nuestro reconocimiento por su orientación a Dr. Avi Caspi (Jerusalem College of Technology - JCT), Prof. Dmitry Klokov (IRSN, Francia), y Dra. Ariella Richardson (JCT).