Ética de la inteligencia artificial. De la tecnoética de Mario Bunge a los desafíos contemporáneos

Ética da inteligência artificial. Da tecnoética de Mario Bunge aos desafios contemporâneos

Ethics of Artificial Intelligence. From Mario Bunge's Techno-Ethics to Contemporary Challenges

Antoni Hernández-Fernández *

Este artículo ofrece una revisión de los desafíos contemporáneos de la ética de la inteligencia artificial (IA), partiendo de las aproximaciones actuales de la ética aplicada a la tecnología. Se analiza la contribución de Mario Bunge a la tecnoética y su relevancia en el contexto actual, caracterizado por el auge de sistemas autónomos, el aprendizaje automático y la toma de decisiones algorítmica. A partir de este marco, se examinan cuestiones clave como la transparencia, la equidad, la responsabilidad y el impacto social de la IA. Se comparan distintas perspectivas, destacando los debates sobre el equilibrio entre innovación y control ético. Finalmente, se proponen líneas de trabajo para fortalecer una ética de la IA que responda a los desafíos emergentes.

Palabras clave: inteligencia artificial; tecnoética; ética aplicada; responsabilidad tecnológica; Mario Bunge

Este artigo apresenta uma revisão dos desafios contemporâneos da ética da inteligência artificial (IA), a partir das abordagens atuais da ética aplicada à tecnologia. Analisa-se a contribuição de Mario Bunge para a tecnoética e sua relevância no contexto atual, marcado pelo crescimento de sistemas autônomos, aprendizado de máquina

* Profesor agregado del Instituto de Ciencias de la Educación y miembro del Comité de Ética de la Universitat Politècnica de Catalunya (UPC), España. Correo electrónico: antonio.hernandez@upc.edu. Publicaciones y proyectos: https://futur.upc.edu/AntonioHernandezFernandez. ORCID: https://orcid.org/0000-0002-9466-2704.

e tomada de decisão algorítmica. Com base nesse quadro, examinam-se questões-chave como transparência, equidade, responsabilidade e impacto social da IA. São comparadas diferentes perspectivas, destacando os debates sobre o equilíbrio entre inovação e controle ético. Por fim, propõem-se linhas de trabalho para fortalecer uma ética da IA que responda aos desafios emergentes.

Palavras-chave: inteligência artificial; tecnoética; ética aplicada; responsabilidade tecnológica; Mario Bunge

This article presents a review of the contemporary challenges of artificial intelligence (AI) ethics, based on current approaches to technology ethics. It analyzes Mario Bunge's contribution to techno-ethics and its relevance in today's context, characterized by the rise of autonomous systems, machine learning, and algorithmic decision-making. Within this framework, key issues such as transparency, fairness, responsibility, and the social impact of AI are examined. Various perspectives are compared, highlighting debates on the balance between innovation and ethical oversight. Finally, the article proposes avenues for strengthening AI ethics to address emerging challenges.

Keywords: *artificial intelligence; technoethics; applied ethics; technological responsibility; Mario Bunge*

Introducción

Mario Bunge, en su célebre conferencia de 1974 en Haifa «*Por una tecnoética*», acuñó el término "tecnoética" para referirse a la necesidad de una ética específica para la tecnología y sus profesionales (Bunge, 2019 [1974]).[1] Se describía entonces, por primera vez, un campo emergente que debía trascender los dilemas tradicionales de la bioética y la deontología profesional, comunes en las reflexiones de los años 70 del siglo XX. Bunge sostuvo que los tecnólogos tenían una responsabilidad ética ineludible en la era de la tecnología, desde el diseño de artefactos hasta el diseño de otras tecnologías de impacto social. Concretamente, sin tapujos, Bunge señalaba:

> "Los instrumentos son moralmente inertes y socialmente irresponsables. Por lo tanto, cuando actúan como instrumentos, el científico, el ingeniero o el administrador se niegan a asumir responsabilidades, salvo que fracasen en su cometido (y, si tienen éxito, quizá no rechacen los honores). Y, si alguien les reprocha su acción, se proclaman inocentes o excusan sus actos afirmando que han actuado siguiendo órdenes (*Befehlsnotstand*), e incluso algunos reaccionan con indignación. Obviamente, su actitud se explica ya sea por un exceso de humildad o por un exceso de arrogancia. En el primer caso, el científico se arrastra ante sus superiores; en el segundo, se eleva por encima de la humanidad ordinaria; en ambos casos, actúa indecentemente. El científico (ingeniero o administrador) puede lavarse las manos, pero eso no lo libera de sus deberes morales ni de sus responsabilidades sociales, no solo como ser humano y ciudadano, sino también como profesional. Porque, insistimos, los científicos, los ingenieros y los administradores son más responsables que cualquier otro grupo ocupacional del estado en que se encuentra el mundo" (Bunge, 2019 [1974], p. 133).

Argumentó así que ingenieros, científicos y administradores no solo debían rendir cuentas ante sus superiores, sino ante toda la humanidad. Porque, pese a su papel central en la configuración de la sociedad moderna, muchos profesionales históricamente se han desentendido de las consecuencias de su trabajo, lo que ha llevado a tragedias como la desigualdad social, la destrucción ambiental o la degradación y uniformización cultural, con la consecuente desaparición de lenguas y culturas que seguimos padeciendo globalmente. Apuntó también al error conceptual de atribuir a los artefactos ("los instrumentos son moralmente inertes y socialmente irresponsables") lo que son responsabilidades humanas, algo que sigue sucediendo

[1] Presentada en el simposio *«Ethics in an Age of Pervasive Technology»*, Technion, Haifa, en diciembre de 1974, republicada con permiso y revisada por el propio Bunge en la edición catalana conmemorativa de su centenario de 2019.

hoy en día: baste repasar los titulares de muchas noticias de la prensa en el mundo, referidas al tópico que nos ocupará, la inteligencia artificial (IA).

Bunge (2013, 2019 [1974]) criticó la falta de códigos éticos efectivos en la tecnología y señaló que la indiferencia moral de muchos tecnólogos derivaba de su enfoque exclusivo en la eficacia profesional. Insistió en que los científicos, tecnólogos y administradores no debían actuar como meros instrumentos desprovistos de responsabilidad. Rechazó la justificación de obediencia a órdenes, la manida excusa de "si no lo hago yo lo hará otro", y subrayó que su influencia sobre el mundo les confería una obligación ética mayor al común de los mortales. No obstante, Bunge (2019 [1974]) concluía en la cita anterior que, tarde o temprano, la sociedad exigiría rendir cuentas a estos profesionales, por lo que era imperativo que asumieran sus deberes morales y sociales, antes de que las consecuencias se volvieran irreversibles para el planeta. Pero ¿ha sucedido? ¿Se ha exigido, o se exige, asumir esta responsabilidad, el rendimiento de cuentas y el rol humano en el desarrollo tecnológico? Lamentablemente, el caso particular de la IA, parece aún indicarnos lo contrario, más de 50 años después de la conferencia seminal de Haifa.

El desarrollo de la IA plantea desafíos éticos sin precedentes, no solo en términos de sus implicaciones sobre la vida de los individuos, sino también en relación con su impacto estructural en la sociedad, a nivel global. Desde su conceptualización temprana, la ética de la tecnología ha oscilado entre la reflexión filosófica y la respuesta pragmática, urgida a menudo de la inmediatez, cuando no de la prisa, a problemáticas sociales emergentes (Quintanilla, 2017). La creciente autonomía e implantación generalizada de los sistemas de IA, en especial la denominada IA generativa, y su integración en la toma de decisiones críticas, en ámbitos cruciales en la sociedad como la salud, el derecho o la educación, han reabierto el debate imprescindible sobre el papel de la tecnoética en el diseño, implementación y regulación de las nuevas tecnologías.

Para Bunge, la tecnología no es solo un mero conjunto de herramientas o artefactos resultado de la acción humana, sino que implica un campo de conocimiento, a la vez que una práctica social con consecuencias filosóficas, políticas y morales (Bunge, 2012). Así, dentro de este marco polisémico, la tecnología se considera, para empezar, un campo de conocimiento que se ocupa de diseñar artefactos y de planificar su realización, operatividad y mantenimiento, a la luz del conocimiento científico, diferenciándola así de la técnica, que no precisaría necesariamente del conocimiento científico (Bunge, 2013). Reflexiones filosóficas más recientes adolecen de definiciones de tecnología o de IA (Innerarity, 2025).

Lejos de aproximaciones reduccionistas, Bunge proporcionó una definición sistémica de tecnología (T), que representó mediante la endecatupla (Bunge, 2013, 2019):

$$T = <C,S,D,G,F,E,P,A,O,M,V>$$

siendo:

1. *C (Comunidad profesional):* conjunto de personas especializadas en la tecnología, que comparten conocimientos, valores y prácticas para diseñar, probar y evaluar artefactos o procesos.
2. *S (Sociedad):* contexto natural y social en el que la tecnología se desarrolla, incluyendo factores económicos, políticos y culturales que la impulsan o restringen.
3. *D (Dominio):* conjunto de entidades reales (o presuntamente reales) sobre las que actúa la tecnología, ya sean naturales o artificiales.
4. *G (Visión general o filosofía inherente a T):*
   a) Ontología: concepción de los objetos tecnológicos como elementos combinables y modificables.
   b) Gnoseología: realista y pragmática.
   c) *Ethos*: centrado en el uso de recursos naturales y humanos, especialmente cognitivos.
5. *F (Fondo formal):* teorías, métodos lógicos y matemáticos, en los que se apoya T.
6. *E (Fondo específico):* conjunto de datos, hipótesis y teorías empíricamente contrastadas, junto con métodos de investigación y diseño aplicables.
7. *P (Problemática):* problemas cognitivos y prácticos dentro del dominio D de la tecnología y de sus componentes.
8. *A (Fondo de conocimiento):* conocimientos acumulados en el tiempo por la comunidad profesional, que incluyen teorías, hipótesis, artefactos y diseños anteriores.
9. *O (Objetivos):* desarrollo e invención de nuevos artefactos y procesos, mejora de los existentes, así como la planificación de su uso y mantenimiento.
10. *M (Metodología):* procedimientos escrutables (contrastables, analizables, criticables) y justificables (explicables), concretamente:
    a) el método científico (problema cognoscitivo-hipótesis-contrastación-corrección de la hipótesis o reformulación del problema).
    b) el método tecnológico (problema práctico-diseño-prototipo-prueba-corrección del diseño o reformulación del problema).
11. *V (Valores):* conjunto de juicios sobre procesos y productos tecnológicos, incluyendo aspectos éticos, estéticos y funcionales.

Su propuesta filosófica definía pues la tecnología bajo una perspectiva sistémica, crucial en la epistemología y en las ciencias humanas (Altmann & Koch, 2011). Enfatizaba así la necesidad de evaluar no solo los efectos inmediatos de todo artefacto, sino también su contribución al bienestar humano en el marco de una sociedad compleja, global e interconectada (Bunge, 2013, 2019). Siguiendo esta línea bungeana de pensamiento, la IA debe ser analizada no solo desde una perspectiva utilitarista o centrada en el usuario, sino desde un enfoque holístico y multifactorial que considere sus implicaciones a nivel individual, profesional, social e institucional. Mientras que

gran parte de la discusión ética actual sobre la IA se centra en la privacidad, la equidad algorítmica y la autonomía de los usuarios, la tecnoética de Bunge nos invita ya en 1974 a reflexionar sobre dimensiones más amplias. ¿Quién controla el desarrollo tecnológico y con qué fines? ¿Cómo afecta la IA la distribución del poder y la riqueza en las sociedades humanas? ¿Qué modelos de gobernanza pueden garantizar que su desarrollo sea compatible con valores democráticos y principios de justicia? ¿Puede ser neutra la IA en un marco tan complejo?

Muchas aproximaciones recientes sobre la ética de la tecnología la incluyen en la ética aplicada (Gordon & Nyholm, 2024), aunque esta clasificación es problemática y posee errores conceptuales:[2] como pone de manifiesto la definición de Bunge, la tecnología debería tratarse como un campo de conocimiento interconectado, por lo que su ética no debería enfocarse de forma aislada o fragmentada. Debería evitarse, pues, una excesiva compartimentación filosófica de la tecnoética, con subcampos como la ética de la IA o la ética de la robótica, que tiendan a aislar a la tecnoética de los problemas e imbricaciones sociales, de su impacto económico o ambiental, y de sus vínculos epistemológicos.

Este artículo se propone examinar los desafíos éticos de la IA a partir de tres dimensiones interconectadas, siguiendo una clasificación derivada de la revisión de Pareto *et al.* (2021), planteada en un inicio para la robótica asistencial, pero que contiene un marco extensible a toda la IA: el bienestar individual, la ética del cuidado y la justicia sociopolítica. La primera dimensión se refiere a los impactos directos de la IA en la autonomía, la privacidad y la seguridad de los usuarios. La segunda aborda el papel de la IA en el rediseño de prácticas profesionales y relaciones de confianza en ámbitos como la educación, la salud o el mundo laboral. Finalmente, en la tercera dimensión se analiza cómo la IA reconfigura estructuras de poder sociopolítico, la distribución de recursos en el planeta y los marcos normativos en el mundo.

Para ello, en la siguiente sección se empezará revisando las definiciones y parte de la literatura sobre ética de la IA, con especial énfasis en una actualización filosófica de la tecnoética de Mario Bunge, y su evolución contemporánea. Posteriormente, se presentarán los principales hallazgos en torno a los problemas éticos identificados en estos tres niveles, y una discusión sobre las limitaciones del enfoque actual predominante en la ética de la IA. Finalmente, se esbozarán líneas futuras de investigación y propuestas para una ética de la IA que recupere la visión integral defendida por Bunge.

1. Sobre la IA

En 2019 la Comisión Europea partió de la definición del *Oxford English Dictionary*, que describe la IA como “la capacidad de los ordenadores u otras máquinas para

[2] Véase Pareto y Torras (2024) para una revisión.

mostrar o simular un comportamiento inteligente" (EU-HLEG, 2019). A partir de esta base, la Comisión Europea amplió y contextualizó esta definición, describiendo la IA como sistemas basados en *software* (programas) o componentes físicos (*hardware*) que imitan funciones cognitivas humanas, como aprender, resolver problemas y tomar decisiones. Los sistemas de IA pueden utilizar reglas simbólicas o aprender modelos, y también pueden adaptar su comportamiento al contexto, según su programación. De hecho, la Comisión Europea apuntó que los sistemas de IA son sistemas de *software* y *hardware* diseñados por humanos que, dado un objetivo complejo, actúan en la dimensión física o digital mediante (EU-HLEG, 2019):

- la percepción de su entorno a través de la adquisición de datos;
- la interpretación de los datos, estructurados o no estructurados, recogidos;
- el razonamiento sobre el conocimiento disponible o el procesamiento de información derivada de los datos;
- finalmente, decidir cuáles son las mejores acciones a tomar para alcanzar el objetivo planteado.

Podemos realizar dos puntualizaciones a esta definición. La primera se refiere al hecho antropocéntrico de limitar a los humanos el diseño de los sistemas de IA: ¿no pueden IA diseñar otras IA? Sin eliminar la cadena de responsabilidades humanas que habría tras ello, no es descartable, técnicamente, que una máquina diseñe a otra. La segunda puntualización se refiere a la antropomorfización de la definición de IA al utilizar la palabra "razonamiento", en el tercer punto, aunque posteriormente se hable de "procesamiento de información", algo más adecuado. La antropomorfización de la IA, entendida como la atribución de características humanas, es problemática, por diversas razones (Salles *et al.*, 2020; van der Rijt *et al.*, 2025):

- *Expectativas poco realistas:* al dotar a la IA de atributos humanos, los usuarios pueden desarrollar expectativas erróneas sobre sus capacidades, lo que conduce a una confianza excesiva, al denominado *hype* o burbuja hiperbólica de expectativas, y a la dependencia de sistemas que carecen de juicio moral y racional.
- *Violación de la dignidad humana:* interactuar con *chatbots* antropomorfizados puede ser incompatible con la dignidad humana, ya que tales sistemas no pueden mostrar reciprocidad en el respeto moral, lo que puede resultar en una disminución del respeto propio de los usuarios humanos.
- *Riesgos de manipulación informativa:* la personalización y antropomorfización de modelos de lenguaje pueden influir negativamente en los usuarios, afectando su toma de decisiones y potencialmente manipulándolos, especialmente cuando estos sistemas se dirigen a grupos vulnerables como niños o pacientes, en un contexto en el que es más difícil cada vez distinguir al humano de la máquina, del bulo (*fake*) o del bulo profundo (*deepfake*).

- *Reforzamiento de sesgos y estereotipos:* los sistemas de diálogo antropomorfizados amplifican sesgos y estereotipos (de género, raciales, lingüísticos, sociales), afectando la percepción y el comportamiento de los usuarios de manera no deseada.

Aunque una relativa antropomorfización de la IA puede facilitar la interacción humana con estos sistemas, es crucial ser consciente de sus implicaciones éticas y sociales. Es necesario reconsiderar la IA desde el diseño, contemplando sus diversas fases (Hernández-Fernández, 2025), de manera que se eviten malentendidos y se proteja la dignidad –y autonomía– de los usuarios. Así, por ejemplo, son habituales en los medios, e incluso en la literatura científica, alusiones metafóricas exageradas, como cuando se dice que la IA "alucina", cuando las que pueden alucinar, en todo caso, son las personas (Maleki *et al.*, 2024).

Los sistemas de IA pueden operar de manera autónoma en funciones específicas, gracias al uso de algoritmos, datos y modelos predictivos, aunque se debería exigir siempre, más allá de las regulaciones o de la legislación vigente, supervisión humana. Holmes, Bialik y Fadel (2019) recuperaron el denominado modelo SAMR de Rubén Puentedura para describir cuatro niveles posibles de integración de una tecnología en la sociedad, en su caso en el contexto de la reflexión sobre la IA en la educación. A este modelo SAMR se suman ejemplos de Hernández-Fernández (2025b):

- *Sustitución:* la tecnología reemplaza una herramienta tradicional sin mejorar su funcionalidad. Por ejemplo: en educación, sustituir libros de texto físicos por PDF.
- *Aumento:* la tecnología proporciona mejoras funcionales en las tareas, incrementando las potencialidades de tecnologías anteriores. Por ejemplo: agregar enlaces interactivos o vídeos a los textos digitales, respecto a un diccionario en papel convencional.
- *Modificación:* la tecnología transforma significativamente las tareas, permitiendo diseñar actividades innovadoras. Por ejemplo: colaborar en tiempo real en documentos compartidos.
- *Redefinición:* la tecnología hace posibles actividades completamente nuevas que antes eran inimaginables. Por ejemplo: crear en segundos cientos de imágenes con IA generativa en el *brainstorming* de un proyecto.

En este momento, precisamente estos dos últimos, los denominados niveles transformativos, la modificación y redefinición de las tareas, son los que deberíamos considerar respecto las tecnologías de IA, por su impacto significativo en muchas profesiones, tareas y ámbitos sociales. En el contexto de la educación, por ejemplo, la IA puede ser utilizada para personalizar la enseñanza, adaptar contenidos a las necesidades de cada estudiante y proporcionar herramientas innovadoras que

antes no eran posibles, pero también plantea interrogantes sobre la privacidad y la dependencia de estas tecnologías (Holmes, Bialik & Fadel, 2019), además del riesgo de la denominada "descarga cognitiva"; es decir, la cesión perjudicial para el aprendizaje de tareas que antes hacía el estudiante y que beneficiaban su desarrollo intelectual.

El impacto de la IA va más allá del ámbito educativo en el que se centraron Holmes, Bialik y Fadel (2019). Permea la sociedad. Se extiende a sectores tan diversos como la medicina, la justicia, la ciencia o la economía. En el diagnóstico médico, para empezar, la IA está revolucionando la forma en que los profesionales de la salud identifican y tratan enfermedades: herramientas basadas en IA pueden analizar imágenes médicas (como radiografías y resonancias magnéticas) con una precisión superior a la de los radiólogos en muchos casos. Los algoritmos de IA también pueden ayudar a predecir brotes de enfermedades o identificar patrones en grandes cantidades de datos, algo que de otro modo sería imposibles de analizar o detectar, mejorando la precisión diagnóstica y personalizando los tratamientos. No obstante, López de Mántaras (2023) advierte que los mejores resultados se obtienen de la sinergia entre médico y máquina, no de su actuación independiente (Fügener *et al.*, 2022), ya que el humano aportará el sentido común al análisis de la máquina. Pero el humano puede correr el riesgo de sucumbir a la máquina, de aceptar sus resultados como criterios superiores.

En el asesoramiento judicial, por otra parte, la IA está emergiendo como una herramienta poderosa para analizar grandes volúmenes de datos, contemplando con rapidez múltiples historiales de casos judiciales y proponiendo posibles análisis y sentencias, lo que podría agilizar la toma de decisiones en los tribunales, acelerando procesos que a veces se demoran años. Esto puede ser útil en jurisprudencia, ya que los sistemas de IA pueden ayudar a hacer recomendaciones sobre estrategias legales basadas en datos históricos, de modo que los profesionales puedan tomar decisiones informadas rápidamente, aunque de forma regulada (Comisión Europea, 2024). Sin embargo, el uso de IA en estos sectores también plantea retos éticos y riesgos legales, como la necesidad de garantizar que los algoritmos sean transparentes, justos y sin sesgos de entrenamiento que puedan perpetuar desigualdades, y que haya siempre revisión, responsabilidad y rendimiento de cuentas por parte del humano experto. Aunque, como recordaron Kahneman, Sibony y Sunstein (2021), un humano que sume el ruido inherente a su cognición a sus propios sesgos podría ser peor que una máquina con sesgos. Como en el caso médico, o en la educación, delegar en la IA no puede ser excusa para eximir al profesional de los errores que cometa la máquina, derivados o no de sesgos, pues en este caso, recordando a Bunge, el profesional actuaría como un mero instrumento, inerte e irresponsable.

2. Tecnoética e inteligencia artificial

La afirmación de Bunge (2019 [1974]) –*"Ethica more technico"*– plantea una tecnoética que debe abordarse con el mismo rigor y la misma sistematicidad que se aplica en

la ciencia y la ingeniería. En otras palabras, el conjunto de juicios de valor *(V)* en la tecnología no puede ser tratado de manera abstracta o subjetiva, sino que debe analizarse con métodos racionales y objetivos, teniendo en cuenta su impacto real en la sociedad y el medio ambiente. Hay, pues, ya en Bunge, una hermenéutica crítica subyacente, en la línea de lo que se ha reclamado más recientemente (Cortina, 1996; Pareto & Torras, 2024). Así, los valores (*V*), ausentes en la definición bungeana de ciencia (como él mismo reconoce en Bunge, 2019, p. 52), forman parte fundamental del sistema tecnológico. No se trata solo de la creación de artefactos, o de la resolución de problemas técnicos, sino también de los juicios de valor que afectan a la manera en que la tecnología se desarrolla, se implementa, se aplica y se evalúa en la sociedad.

Aquí es donde entra la distinción entre "endoaxiología" y "exoaxiología" (Bunge, 2019):

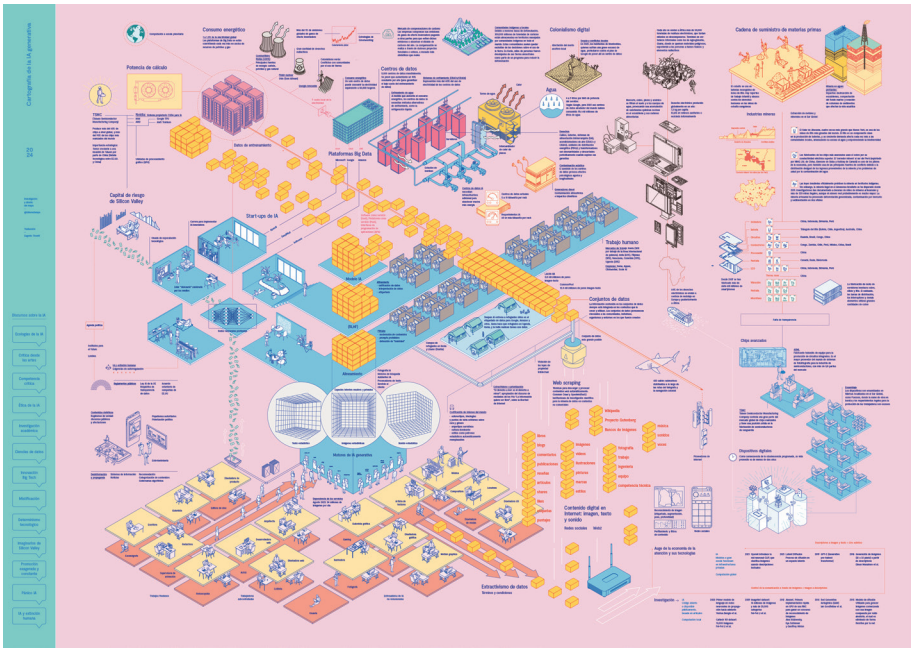
- *Endoaxiología:* se refiere a los juicios de valor internos dentro del proceso tecnológico. Estos incluyen la evaluación de problemas, diseños y pruebas en función de criterios técnicos, como la eficiencia, la precisión o la viabilidad. Por ejemplo, un ingeniero que diseña un parque eólico evaluará la calidad del diseño en términos de rendimiento energético y seguridad estructural.
- *Exoaxiología:* engloba los juicios de valor externos que analizan el impacto de la tecnología en la sociedad y en el entorno. No basta, siguiendo con el ejemplo anterior, con que el parque eólico sea técnicamente eficiente; también hay que considerar sus consecuencias ecológicas, económicas y sociales. Así, aunque no sean ingenieros, un ecólogo podría poner objeciones si el parque se diseña en un concurrido paso estacional de aves; un economista podría abundar en la creación de empleo local; y un sociólogo podría plantear los cambios sociales derivados del abandono de los terrenos afectados, al pasar los agricultores expropiados a trabajar en el parque eólico, tras un periodo de formación y reconversión laboral.

El planteamiento seminal de Bunge resalta que la tecnología no puede, por ende, ser neutral ni evaluada exclusivamente desde una perspectiva técnica. Siempre habrá implicaciones éticas y sociales que deben considerarse: hacerlo de manera sistemática (*more technico*) permite tomar decisiones mejor fundamentadas y equilibradas. Olvidar que la tecnología es un sistema, o reducirla a los meros artefactos, comporta muchas problemáticas (Innerarity, 2025).

En lo referente a la IA, Kate Crawford dejó clara esta cuestión en su "Atlas de la IA" (Crawford, 2021), tanto cuando expuso que la IA no es 'ni inteligente, ni artificial', entre otros motivos por el impacto en los seres humanos (explotación laboral de etiquetadores de datos, guerras vinculadas a los monopolios de minerales) o los ecosistemas (alto consumo de agua y energía, minería extractiva de alto impacto ambiental), como cuando concluye, contundente, que la IA actualmente es un registro del poder socioeconómico mundial. La complejidad de las implicaciones

socioeconómicas y tecnoéticas de la IA, así como el carácter sistémico de estas tecnologías, se pueden visualizar en el proyecto artístico digital de Taller Estampa, “Cartografía de la IA”, Premio Ciudad de Barcelona 2024 (**Figura 1**), inspirado en parte en la obra de Crawford (2021).

Figura 1. Cartografía de la IA (Taller Estampa)



Fuente: https://cartography-of-generative-ai.net/.

El filósofo malagueño Antonio Diéguez revisó tres tópicos sobre la tecnología que pueden generar percepciones erróneas sobre su desarrollo e impacto en la sociedad (Diéguez, 2024). Estos mitos también se pueden aplicar y extender al caso de la IA, ilustrando cómo ciertas creencias distorsionan su comprensión y uso. Son concepciones que se diluyen con un simple vistazo a la complejidad de la cartografía de la IA (**Figura 1**). En concreto:

1. *Determinismo tecnológico.* Este tópico sostiene que el desarrollo tecnológico sigue una trayectoria inevitable y autónoma, como si las innovaciones ocurrieran por sí solas y fueran imposibles de detener o modificar, o no tuviésemos capacidad social o política para influir y legislar. Creer que la tecnología avanza por su propia dinámica interna, sin estar sujeta a decisiones humanas o sociales, es poco más que, en la línea bungeana, reconocer que los administradores

y políticos son meros instrumentos de los poderes fácticos socioeconómicos. Asimismo, se suele afirmar que la IA superará inevitablemente a la inteligencia humana en todas las áreas (inteligencia general artificial, AGI, en sus siglas en inglés), como si esto fuera un destino ineludible e incontrolable. Sin embargo, el desarrollo de la IA depende de decisiones políticas, económicas y éticas. Regulaciones como la reciente Ley Europea de la IA (Comisión Europea, 2024) muestran que su evolución está sujeta a control humano y no es simplemente un fenómeno inevitable o irreversible.

2. *Neutralidad tecnológica.* Este mito supone que la tecnología es intrínsecamente neutral y que sus efectos dependen solo de cómo la use la sociedad. Según esta visión, la IA sería una simple herramienta, un instrumento cuyo impacto dependería exclusivamente de los usuarios, de si la usan para el bien o el mal. Sin embargo, es bien sabido que los algoritmos de IA en redes sociales no son neutrales, ya que están diseñados con objetivos específicos, como maximizar el enganche de los usuarios o incrementar la publicidad y vender productos. Esto ha llevado a problemas como la amplificación de discursos de odio, o la manipulación de la opinión pública en contextos electorales, favoreciendo la polarización política y el desarrollo de las llamadas "cámaras de eco", en las que los usuarios acaban comunicándose solo con los que piensan como ellos, y, continuamente, atendiendo solo a las noticias que confirman su ideología. De hecho, puede haber aplicaciones de IA diseñadas directamente para denigrar, como es el caso de las aplicaciones que desnudan a las personas, vejatorias sobre todo contra las mujeres (Hernández-Fernández, 2024, 2025c).

3. *La tecnología nos deshumaniza.* Existe la creencia de que la tecnología, al mediar en nuestras interacciones y actividades, nos aleja de lo esencialmente humano, volviéndonos fríos, mecánicos o alienados. Es extraño, pues la técnica nos acompaña desde nuestros orígenes ancestrales, y es, sin duda, un rasgo distintivo de nuestra especie (Carbonell & Sala, 2002). Algunos plantean que el uso intensivo de tecnología nos desconecta de nuestras emociones, valores y relaciones sociales. En el caso de la IA, se dice a menudo que la IA, al automatizar procesos en el trabajo o en la educación, y delegar tareas cognitivas en ella, reduce la creatividad y la autonomía de las personas. Sin embargo, muchas herramientas de IA pueden potenciar habilidades humanas, como el pensamiento crítico o la expresión artística. Sin embargo, los modelos de IA generativa están siendo utilizados por músicos, artistas o escritores para ampliar sus procesos creativos, no para sustituirlos, en un replanteamiento de la creatividad (Boden, 2010; Wingström *et al.*, 2024). Además, en el ámbito médico, la IA permite personalizar tratamientos para pacientes, mejorando la calidad de vida en lugar de deshumanizar la atención sanitaria, y algo similar se puede plantear en la educación, o en las aplicaciones científicas; por ejemplo, en matemática aplicada (Du Satoy, 2020).

Tampoco deberíamos caer en el mito de que toda innovación tecnológica es beneficiosa. No todo **"**avance" tecnológico es bueno, o automáticamente positivo. El denominado "progreso" no siempre conduce a mejoras en la calidad de vida. La IA ha revolucionado el diagnóstico médico, permitiendo detectar enfermedades con gran precisión. Sin embargo, también ha generado problemas como el sesgo en los modelos de diagnóstico, que pueden ofrecer resultados menos precisos para ciertos grupos poblacionales debido a datos de entrenamiento insuficientes o sesgados. Eso sin entrar en la compleja cartografía que representó el Taller Estampa (**Figura 1**). Se suele creer que las tecnologías socialmente aceptadas suelen proponer mejoras que superan a los perjuicios, aunque seguramente las minorías, o los individuos perjudicados, disientan de la implantación de esas tecnologías que benefician a una mayoría. Curiosamente, en muchos casos, las tecnologías están al servicio de minorías poderosas, fomentando la brecha social.

El giro instrumental de la tecnología propuesto por Verbeek (2005, 2011) contempla estos falsos mitos de la tecnología: no es un simple medio neutral al servicio de fines humanos, sino que media activamente nuestra percepción del mundo y nuestras acciones en él; configura nuestra experiencia y agencia, por lo que hay una responsabilidad social sobre la tecnología, que no es un agente autónomo; y por último, Verbeek imbrica el diseño técnico a nuestra humanidad y a sus consecuencias socioeconómicas.

Porque el enfoque de Verbeek (2005, 2011) dialoga sin duda con la noción del "centauro ontológico" de la *Meditación de la técnica* de Ortega y Gasset (1982 [1939]), quien argumentó de forma pionera que el ser humano no es un ente puramente biológico ni puramente técnico, sino una síntesis de ambos (el centauro). Para Ortega, la tecnología es constitutiva de nuestra existencia: no es un añadido externo, sino que nos define como seres históricos y culturales. La conexión entre ambas concepciones radica en que, si la tecnología es más que un instrumento neutral, su desarrollo y uso deben comprenderse en términos de coevolución con la humanidad. Verbeek destaca que la mediación tecnológica transforma no solo nuestras acciones, sino también nuestras normas y valores, lo que refuerza la idea orteguiana de un ser humano inseparable de su dimensión técnica. Así, la ética de la tecnología no puede limitarse a evaluar sus consecuencias, sino que debe analizar cómo reconfigura nuestra forma de ser en el mundo y en las sociedades. La ética de la tecnología no puede escindirse de lo que somos.

En definitiva, la revisión de Diéguez (2024) invita a repensar la tecnología y a analizar críticamente la relación entre ciencia, tecnología y sociedad: se debe reconocer que el desarrollo del conocimiento tecnológico es una construcción humana, no un fenómeno autónomo o inevitable. Los tópicos anteriores distorsionan la percepción social sobre la IA, como ha sucedido con otras tecnologías. No debemos aceptar sus avances sin ser críticos, ni asumir que no tenemos control sobre sus efectos.

En términos generales, deberíamos gestionar la aplicación de las herramientas y los sistemas de IA de manera ética, reflexionando sobre sus implicaciones en

cada caso de uso, y asegurando en general que la IA sirva para beneficiar a toda la sociedad de manera equitativa y justa, y no solo a unos pocos. Bastaría con aplicar la Declaración de Barcelona (2017), aunque la realidad cotidiana sea tozuda y nos demuestre que es algo más difícil de lo que parece. Específicamente, la Declaración de Barcelona para el desarrollo y uso adecuado de la IA en Europa (2017) contemplaba seis principios rectores:

- *Prudencia:* la IA debe desarrollarse con precaución, evitando riesgos imprevisibles.
- *Fiabilidad:* los sistemas de IA deben ser seguros, robustos y funcionar correctamente en todo momento.
- *Rendición de cuentas:* debe haber mecanismos claros para supervisar y evaluar el uso de la IA.
- *Responsabilidad:* las personas y organizaciones deben asumir la responsabilidad de los efectos de la IA.
- *Autonomía limitada:* la IA no debe tomar decisiones que afecten a los humanos sin supervisión.
- *No prescindencia del humano:* la tecnología debe complementar a los humanos, no reemplazarlos en decisiones críticas.

No nos extenderemos aquí, pero parece obvio que, para empezar, el principio de prudencia ha sido el primero en ignorarse a lo largo de la historia. La IA generativa ha permeado, por ejemplo, en educación o sanidad sin ningún control o regulación, hasta hace poco (Comisión Europea, 2024).

3. Problemas y retos éticos de la IA y su planteamiento en la educación

La IA plantea numerosos desafíos éticos derivados de su capacidad para interactuar con humanos, procesar información de manera autónoma y modificar la organización social. Sin embargo, no han sido tantas las propuestas claras de demarcación de estos retos, en una *Ethica more technico***.**

Es tentador aplicar la endecatupla definitoria de Bunge (2019 [1974]) para toda familia de tecnologías a la IA. Ajustar el modelo tecnológico de Bunge, integrando elementos científicos, metodológicos y sociales en su desarrollo y aplicación, proporciona un marco de estudio para la tecnoética, sin olvidar ninguna de sus dimensiones axiológicas. Una posibilidad podría ser:

- *C (Comunidad profesional):* investigadores en IA, ingenieros de *software*, científicos de datos y expertos en ética de la tecnología, trabajando conjuntamente. Este marco transdisciplinar es el más adecuado, al no separar a los filósofos de los expertos en IA, con ejemplos notables (Pareto y Torras, 2024).

- *S (Sociedad):* empresas, instituciones educativas y académicas, así como organismos reguladores que impulsan el desarrollo y uso de la IA.
- *D (Dominio):* datos digitales, modelos algorítmicos y computacionales y sistemas automatizados en entornos físicos y virtuales.
- *G (Visión general):*
  a) *Ontología:* la IA trata la información como estructuras procesables mediante algoritmos, pero también el denominado '*embodiment*', encarnación o corporeización, en el caso de la robótica.
  b) *Gnoseología:* basada en modelos matemáticos y estadísticos con una interpretación probabilística de la realidad, una realidad cognoscible.
  c) *Ética:* implica la responsabilidad en la toma de decisiones automatizadas, el uso de datos (personales y sociales) y los impactos sociales.
- *F (Fondo formal):* matemáticas aplicadas, lógica, estadística, lingüística cuantitativa, teoría de la computación y algoritmos de aprendizaje.
- *E (Fondo específico):* bases de datos, redes neuronales, algoritmos de aprendizaje automático y técnicas de procesamiento del lenguaje natural, sin olvidar la ética aplicada.
- *P (Problemática):* desafíos en la mejora de la precisión de modelos, sesgos en los datos y la explicabilidad de las decisiones algorítmicas.
- *A (Fondo de conocimiento):* desarrollo histórico de la IA, desde los primeros sistemas expertos hasta el aprendizaje profundo actual.
- *O (Objetivos):* crear sistemas capaces de realizar tareas cognitivas como el reconocimiento de patrones, la toma de decisiones y la automatización de procesos.
- *M (Metodología):*
  a) *Científica:* experimentación con algoritmos, evaluación de modelos y optimización, siguiendo los métodos de las ciencias.
  b) *Tecnológica:* desarrollo de arquitecturas de IA, pruebas de rendimiento y ajustes iterativos, en el marco de las cuatro fases del proceso tecnológico (conceptualización, modelización, construcción y presentación de resultados).
- *V (Valores):* transparencia, equidad, seguridad, eficiencia, explicabilidad, rendimiento de cuentas y sostenibilidad en el desarrollo y aplicación de la IA.

Huyendo pues del enfoque reduccionista centrado en el artefacto, y siguiendo la categoría de problemas éticos definidos para la robótica asistencial, tras la revisión de la literatura de Pareto *et al.* (2021), los problemas de cada uno de los 11 elementos definitorios podrían agruparse en tres dimensiones sociales en la interacción de las máquinas con las personas: bienestar, cuidado y justicia. Esta agrupación da cuenta de la complejidad sistémica del problema, que hemos visto en la endecatupla bungeana (o en la **Figura 1**), a la par que pone el foco en las personas, en una hermenéutica crítica (Cortina, 1996), alejada de una concepción meramente instrumental de la tecnología (Bunge, 2019 [1974]; Verbeek, 2005, 2011). De este modo, en el caso de la IA, una posible propuesta inspirada en Pareto *et al.* (2021) podría ser:

- *Bienestar:* incluye cuestiones como la privacidad y el control de datos, donde el uso de IA implica riesgos de vigilancia masiva y vulneración del consentimiento informado (Véliz, 2021). La autonomía humana también se ve amenazada, ya que una dependencia excesiva de la IA podría reducir la capacidad de toma de decisiones de los usuarios y ahondar en las problemáticas derivadas de la descarga cognitiva (causando un empeoramiento del aprendizaje, por ejemplo). Asimismo, la IA puede influir en la percepción de la identidad y dignidad humanas, propagando sesgos y estereotipos. Otros aspectos fundamentales en el bienestar incluyen la seguridad (tanto física como psicológica), el aislamiento y la pérdida del contacto humano y el impacto en las habilidades psicológicas y morales humanas.

- *Cuidado:* se centra en la legitimidad de la introducción de la IA en ámbitos tradicional e inherentemente humanos, como la educación, la salud o la asistencia social. La calidad de la práctica y la confianza en la IA son factores críticos, ya que su uso puede alterar las relaciones profesionales y la organización de roles. La robótica asistencial (Pareto *et al.*, 2021) es un buen ejemplo. También se podría cuestionar si la IA redefine el concepto mismo de cuidado, cambiando la relación entre quienes proveen y reciben atención, entre quienes se podrán permitir en un futuro un cuidador humano, y los que quedarán al cuidado de las máquinas. Por ejemplo, ya están proliferando aplicaciones basadas en IA de soporte psicológico que algunos pretenden que reemplacen a psicólogos profesionales. Los principios de la Declaración de Barcelona en este ámbito deberían estar muy presentes, empezando por la prudencia. Los riesgos son muchos; uno mencionado es que se acabe confiando más en la respuesta técnica que en la humana.

- *Justicia:* abarca la justicia distributiva. Es decir, la equidad en el acceso a los beneficios y costos de la IA, incluyendo su impacto en el mercado laboral y en la inclusión social. La justicia debe ser entendida más allá de los derechos individuales, como justicia social y global, de forma que no se ignoren los impactos de la IA en otras partes del mundo (**Figura 1**). La política de la tecnología relacionada con la IA es relevante, ya que sus objetivos y valores subyacentes pueden estar influenciados por intereses comerciales y gubernamentales. Además, la toma de decisiones automatizada plantea problemas de responsabilidad y transparencia, especialmente en sectores como la justicia o la sanidad. La automatización tiene sus ventajas, pero también genera conflictos, tanto técnicos como filosóficos y éticos (Innerarity, 2025). Finalmente, la sostenibilidad ecológica de la IA es un aspecto emergente, dado su elevado consumo de energía, el extractivismo y los residuos electrónicos que genera (Crawford, 2021).

Todos estos aspectos, por ejemplo, los ha intentado incorporar, con sus luces y sombras, la regulación de la IA de la Unión Europea (Comisión Europea, 2024). En la **Tabla 1** se han intentado agrupar y sintetizar algunos problemas éticos relacionados con la IA, según estas tres dimensiones. Se trataría de problemas abordables tanto desde la investigación filosófica como desde la educación, como se verá posteriormente.

Tabla 1. Síntesis de algunos riesgos y problemas éticos relacionados con la IA, según las dimensiones de bienestar, cuidado y justicia

Dimensión	Problema ético	Descripción
Bienestar	Privacidad y control de los datos	Riesgo de vigilancia, falta de transparencia en el manejo de datos, con potenciales sesgos.
	Manipulación y engaño	IA generando interacciones engañosas, creando bulos y vínculos no auténticos.
	Autonomía humana	Dependencia de la IA, reducción de habilidades cognitivas o en la toma de decisiones.
	Pérdida de contacto humano	Sustitución de interacciones humanas por interacciones automatizadas con IA.
	Seguridad	Riesgos de accidentes, amenazas psicológicas, falta de confiabilidad.
	Dignidad e identidad	Riesgo de objetificación de las personas, con impacto en la autoestima, la percepción de autoeficacia y la autoimagen.
Cuidado	Legitimidad de la IA en prácticas humanas	Cuestionamiento y limitación de su uso en educación, salud y asistencia.
	Calidad y sesgos sociales de la práctica	Reducción de la calidad del servicio debido a automatización excesiva, pérdida de contacto humano y sesgos socioeconómicos.
	Confianza (o temor) a la automatización	Respuesta emocional y dependencia de sistemas de IA que pueden fallar y ser opacos, sesgados e incluso erráticos.
	Redefinición del cuidado	Cambios en el significado y las prácticas del cuidado humano, y redefinición del cuidado artificial y de su marco legal.
Justicia	Justicia distributiva	Desigualdades en el acceso y distribución de los beneficios de la IA. Se democratizan y distribuyen los costes, pero no los beneficios económicos.
	Política de la tecnología IA	Intereses comerciales y gubernamentales influyendo en la regulación, y en desarrollos con intereses particulares, en detrimento del interés social.
	Responsabilidad y transparencia	Dificultades para asignar responsabilidades en decisiones algorítmicas, y falta de transparencia (y de explicabilidad) en el entrenamiento y en los resultados.
	Sostenibilidad ecológica	Consumo de energía, agua y generación de residuos electrónicos.
	Impacto social global	Explotación de personas en países que no velan por los derechos laborales, donde las multinacionales delegan tareas mal pagadas, como el etiquetado de datos.

Fuente: elaboración propia, siguiendo la propuesta de Pareto *et al.* (2021).

Estos marcos generales, por supuesto, tienen muchas aristas que requieren de una investigación filosófica y científica profunda, y deben concretarse según las áreas en las que la IA impacte. En el caso de la educación, como propuesta práctica, sería muy interesante abordar la formación de los estudiantes en la enseñanza obligatoria siguiendo las problemáticas de la **Tabla 1**, y por supuesto vinculada a los currículos educativos oficiales. En el caso de la secundaria, por ejemplo, hay algunos precedentes de propuestas concretas de tecnoética, en el marco curricular de la enseñanza en Cataluña (Ramírez Marqués, 2020; Martínez, 2021; Muñoz Cuenca, 2022), y, si se me permite, previos a la burbuja de la IA de finales de 2022.

Posteriormente, tras la irrupción especialmente de los Grandes Modelos de Lenguaje (*large language models* o LLM, en sus siglas en inglés) de la IA generativa, a tal efecto, la UNESCO propuso un marco competencial para docentes (Cukurova & Miao, 2024) y otro para estudiantes (Miao & Shiohira, 2024) que incluía (por fin) la dimensión ética dentro de una matriz bidimensional con cinco aspectos que evolucionan a través de tres niveles de progresión: adquisición, profundización y creación. Para abordar en la educación estos tres niveles de progresión en la enseñanza de la ética de la IA, es fundamental estructurar el aprendizaje en cada dimensión, adaptándose, por supuesto a los currículos oficiales que, no obstante, suelen ir por detrás de la innovación tecnológica. A continuación, se describe sucintamente cómo implementar estos niveles propuestos por la UNESCO en cada aspecto, ampliando el enfoque general de Hernández-Fernández (2024b):

1. Mentalidad centrada en las personas:

- Adquirir: enseñar en el aula la importancia de la agencia humana en la interacción con la IA, destacando que las personas deben tener control sobre las decisiones asistidas por la tecnología.
- Profundizar: estudiar casos donde la IA influye en la toma de decisiones y fomentar la reflexión sobre la responsabilidad individual en su uso, así como la responsabilidad social y de los políticos y administradores.
- Crear: diseñar proyectos donde los estudiantes evalúen y propongan formas donde la IA pueda fortalecer la responsabilidad social y la equidad, de forma crítica.

2. Ética de la IA:

- Adquirir: introducir los principios éticos fundamentales de la IA, como los de la Declaración de Barcelona, fomentando valores como la transparencia, la equidad o la privacidad.
- Profundizar: explorar dilemas éticos en el uso de IA, promoviendo un debate crítico en el aula sobre sus implicaciones en diferentes contextos, como la integración de algoritmos de decisión en los coches autónomos (véase, por ejemplo: Ramírez Marqués, 2020).

- Crear: desarrollar normas o directrices éticas concretas, en colaboración con otros estudiantes y docentes, para regular el uso de IA en distintas áreas de su actividad académica y cotidiana.

3. Fundamentos y aplicaciones de la IA:

- Adquirir: explicar los conceptos básicos de la IA, sus definiciones básicas, así como fundamentos sobre el funcionamiento del aprendizaje automático y los algoritmos.
- Profundizar: aplicar técnicas de IA a problemas concretos en distintas disciplinas, en especial en las materias de tecnología de secundaria.
- Crear: desarrollar soluciones innovadoras con IA, integrando herramientas avanzadas en proyectos propios, adecuados al nivel del curso.

4. Pedagogía de la IA:

- Adquirir: utilizar herramientas de IA para la enseñanza asistida, como sistemas de tutoría inteligente, que sean transparentes y de código abierto.
- Profundizar: integrar la IA en estrategias pedagógicas, adaptando la enseñanza a las necesidades de los estudiantes, atendiendo a la diversidad.
- Crear: diseñar nuevas metodologías educativas potenciadas por la IA, explorando su impacto real en el aprendizaje, sin olvidar los aspectos éticos, y descartando el uso de la IA cuando empeore el aprendizaje (luchando así contra el falso mito de la autonomía).

5. IA para el desarrollo personal y profesional:

- Adquirir: comprender cómo la IA puede facilitar el aprendizaje a lo largo de la vida, y ayudar a los estudiantes que no posean capacidad económica para pagar un profesor de refuerzo, por ejemplo.
- Profundizar: usar la IA para mejorar procesos organizacionales y la formación dentro de instituciones educativas.
- Crear: aplicar la IA para potenciar el desarrollo profesional, promoviendo cambios en las metodologías, la estructura del trabajo y el aprendizaje.

De este modo, un enfoque progresivo y amplio permitiría que la enseñanza de la ética de la IA evolucione desde la adquisición y comprensión básica de sus principios (primer nivel), la profundización en problemáticas concretas como las de la **Tabla 1** (segundo nivel), hasta la implementación activa y la innovación responsable (tercer nivel), siempre adecuándose al nivel y curso de los alumnos (Cukurova & Miao, 2024; Miao & Shiohira, 2024). En este sentido, queda como reto de la didáctica de la tecnología, o de las diversas disciplinas curriculares, en cuanto la tecnología se encuentra imbricada inevitablemente en todas las asignaturas, el proponer actividades y contenidos relacionados con las tres dimensiones de bienestar, cuidado y justicia (Pareto *et al.*, 2021).

Conclusiones

Cincuenta años después de la conferencia de Haifa, en la que Bunge acuñó el término "tecnoética", en contraposición tácita a la bioética emergente, no podemos aquí sino recuperar el decálogo de conclusiones con las que Bunge (2019 [1974]) sentenciaba el recorrido esperable a aquella, entonces, disciplina emergente:

> "(i) A diferencia de la ciencia básica o pura, que es intrínsecamente valiosa o, en el peor de los casos, carece de valor, la tecnología puede ser valiosa o perjudicial según los fines que sirva. Por lo tanto, es necesario someter la tecnología a controles morales y sociales. (ii) La tecnología perversa solo puede eliminarse descartando sus fines dañinos. Y los malos usos de la buena tecnología no se corrigen ni se evitan frenando la investigación tecnológica, sino fomentando una tecnología aún mejor y haciéndola moral y socialmente sensible. (iii) El tecnólogo, como cualquier otra persona, es personalmente responsable de todo lo que diseña, planifica, recomienda o ejecuta. Por lo tanto, es tan digno de elogio o condena como cualquier otro. O incluso más, dado el carácter racional y deliberado de sus decisiones. (iv) El tecnólogo es responsable no solo ante su empleador y sus colegas profesionales, sino también ante todos los que puedan verse afectados por su trabajo. Su preocupación fundamental debería ser el bienestar público. (v) El tecnólogo que contribuye a aliviar los problemas sociales o a mejorar la calidad de vida es un benefactor público. Pero aquel que, a sabiendas, contribuye a deteriorar la calidad de vida, engaña al público diseñando productos inútiles o nocivos, o difunde información falsa, es un criminal. (vi) Dado que el especialista no puede abordar todos los problemas multilaterales y complejos que plantean los proyectos tecnológicos a gran escala, estos deberían encomendarse a equipos de expertos en diversos campos, incluidos los científicos sociales aplicados, y someterse al escrutinio y control públicos. (vii) Considerando el fuerte impacto de la tecnología en la sociedad y el medio ambiente, el tecnólogo debería compartir el poder con el administrador y el político. La tecnocracia global, o el gobierno de expertos en todos los campos de la acción humana, no es una amenaza, sino una promesa, especialmente si sus decisiones se toman con la aprobación del público. (viii) Los tecnólogos deberían abordar sus propios problemas morales, en lugar de fingir que estos pueden transferirse a los administradores o políticos. (ix) Los tecnólogos deberían contribuir a modernizar la ética, intentando construir una tecnoética como la ciencia de la conducta correcta y eficiente de ellos mismos. (x) La ética no saldrá de su estancamiento actual si los filósofos persisten en ignorar la experiencia moral de los

> tecnólogos, si no prestan atención a cómo la tecnología fundamenta sus prescripciones o normas, y si se niegan a emplear los recursos formales (lógicos y matemáticos) que brindan claridad y precisión" (Bunge, 2019 [1974], pp. 135-136).

En el momento actual, es más necesario que nunca un análisis de este decálogo de Bunge (2019 [1974]) sobre la tecnoética, en relación con los problemas actuales de la IA, especialmente la IA generativa. En general, podría concluirse:

i. La IA, como toda tecnología, debe someterse a controles morales y sociales. Las IA generativas pueden amplificar la desinformación, generar noticias falsas o manipular la opinión pública mediante bulos profundos (*deepfakes*). Es esencial establecer regulaciones y controles éticos que limiten su uso indebido y garanticen su alineación con el bien común (Comisión Europea, 2024).
ii. La IA perversa solo se elimina descartando sus fines dañinos. No basta con prohibir tecnologías problemáticas; hay que evitar y perseguir sus usos nocivos. Un ejemplo es el abuso de IA en la creación de contenido ilícito o falso, que no se resuelve eliminando la IA en sí, sino desarrollando herramientas para detectar y frenar la propagación de desinformación, o prohibiendo aplicaciones sin ningún uso bueno (Hernández-Fernández, 2024, 2025c).
iii. Los tecnólogos son responsables de lo que diseñan, planean y ejecutan. Los desarrolladores de IA deben asumir la responsabilidad de los sesgos en sus modelos, del impacto de la automatización en el empleo y del uso de la IA en vigilancia masiva (Véliz, 2020). No pueden excusarse en una supuesta neutralidad de la tecnología (Verbeek, 2011), máxime si sus creaciones amplifican desigualdades o vulneran derechos de las personas.
iv. Los tecnólogos son responsables ante la sociedad en su conjunto. La IA afecta a millones de personas, desde su uso en redes sociales, hasta en la toma de decisiones bancarias (si se concede o no un crédito), educativas (en las calificaciones académicas) o médicas (en diagnósticos médicos). Los profesionales y tecnólogos no pueden actuar solo en interés de sus empleadores o accionistas; deben considerar el impacto social de sus creaciones y rendir cuentas (Pareto *et al.*, 2021).
v. La IA puede ser beneficiosa o perjudicial según su uso. Sin abundar en el mito de la neutralidad, no debemos negar que la IA generativa puede ayudar en la educación, la creatividad o la medicina, pero también se usa para manipular elecciones, generar bulos o facilitar el fraude. Es clave educar en un uso crítico, responsable y ético que maximice los beneficios y minimice los daños de la IA (Hernández-Fernández, 2024), como de toda tecnología, sin olvidar su regulación (Coeckelbergh, 2019).
vi. La complejidad de la IA exige equipos multidisciplinares y control público. El desarrollo de IA no debe quedar solo en manos de ingenieros, tecnólogos y empresarios. Es necesario incorporar filósofos, juristas, sociólogos y expertos en

derechos humanos para evaluar impactos y garantizar su regulación democrática, como se ha visto en el caso de la robótica asistencial (Pareto & Torras, 2024).

vii. La tecnocracia global no es una amenaza si sus decisiones se someten a la aprobación pública. El poder de la IA en la toma de decisiones (desde créditos bancarios o diagnósticos clínicos, hasta la seguridad nacional) y en el control masivo de datos no puede quedar en manos de unas pocas corporaciones, oligopolios o multinacionales tecnológicas. Se necesita supervisión pública y mecanismos democráticos para evitar que la IA se convierta en un instrumento de control social en el capitalismo de vigilancia (Véliz, 2020; O'Neil, 2018).

viii. Los tecnólogos deben abordar los problemas morales de la IA. El sesgo algorítmico, la explotación laboral en el entrenamiento de los sistemas de IA, o la privacidad de los usuarios, no pueden ser problemas delegados a reguladores externos. Los creadores de IA deben integrar la ética en el diseño desde el principio y contemplar todas las fases del diseño tecnológico (Hernández-Fernández, 2025).

ix. Es necesaria una tecnoética que guíe la evolución de la IA. Más allá de regulaciones reactivas, se necesita una ética específica para la IA que combine principios filosóficos con conocimientos técnicos, ayudando a diseñar IA seguras, transparentes, explicables y alineadas con los valores humanos (Comisión Europea, 2024).

x. La tecnoética de la IA no debe ignorar la experiencia de los tecnólogos. Los debates éticos sobre la IA tampoco pueden quedar solo en manos de filósofos alejados de la práctica tecnológica. Son cruciales los equipos transdisciplinares y que los propios desarrolladores participen en la construcción de normas y estándares éticos que guíen la evolución de la IA, máxime en aras de la reducción del impacto ambiental y la sostenibilidad (Mokiy, 2023; Crawford, 2021).

En conclusión, cincuenta años después de la conferencia de Haifa, transmitir el espíritu de la tecnoética de Bunge es fundamental, cuando no urgente. Me sorprende la ausencia de sus reflexiones y postulados en buena parte de los estudios críticos recientes sobre la IA. La irrupción de la IA generativa plantea desafíos complejos e inéditos en la intersección entre tecnología, sociedad y ética, desde la proliferación de desinformación hasta el riesgo de vigilancia masiva y sesgos algorítmicos. Retomar el enfoque progresivo y amplio que Bunge propuso implica no solo regular la IA para minimizar sus riesgos, sino también integrarla en la enseñanza y el debate público, garantizando su desarrollo bajo principios de bienestar, cuidado y justicia (Pareto *et al.*, 2021). En este sentido, la tecnoética no debe quedarse en una reflexión abstracta, sino traducirse en una praxis crítica y necesariamente innovadora, en la educación como en otros ámbitos, que guíe la evolución de la IA y sus usos hacia un horizonte de responsabilidad social y sostenibilidad. En caso contrario, las consecuencias pueden ser nefastas para la humanidad.

Financiamiento

Esta publicación es parte del proyecto PID2024-155946NB-I00 del Ministerio de Ciencia, Innovación y Universidades (MICIU), la Agencia Estatal de Investigación (AEI/10.13039/501100011033) y del European Social Fund Plus (ESF+), HORIZON-101234399-BSC-AI-Factory, así como del reconocimiento de la Generalitat de Catalunya 2021 SGR 01266.

Reconocimiento de uso de IA

El autor reconoce el uso de los modelos de lenguaje integrados en HuggingChat para la revisión de la traducción del resumen y las palabras clave al inglés y al portugués, así como también para dar formato en la generación de la **Tabla 1**. Más allá de este reconocimiento, se declara como único autor del artículo y es completamente responsable del contenido incluido en él.

Agradecimientos

Debo agradecer a Mario Bunge su mentoría intelectual y la correspondencia que establecimos en sus últimos años, tanto en lo referente a la tecnoética como a la filosofía en general. También a su alumno aventajado, Alfons Barceló, por los múltiples debates que en parte se han reflejado aquí.

Bibliografía

Altmann, G., & Koch, W. A. (2011). Systems: New paradigms for the human sciences. Berlín: Walter de Gruyter.

Boden, M. A. (2010). Creativity and art: Three roads to surprise. Oxford: Oxford University Press.

Bunge, M. (2019 [1974]). Por una tecnoética. Conferencia «Ethics in an Age of Pervasive Technology», Technion, Haifa. En Filosofía de la tecnología (131-143). Barcelona: Edicions IEC-Iniciativa Digital Politècnica.

Bunge, M. (2012). Filosofía de la tecnología y otros ensayos. Lima: Fondo Editorial.

Bunge, M. (2013). Pseudociencia e ideología. Pamplona: Laetoli.

Bunge, M. (2019). Filosofía de la tecnología. Barcelona: Edicions IEC-Iniciativa Digital Politècnica. Recuperado de: http://hdl.handle.net/2117/169030.

Carbonell, E. & Sala, R. (2002). Aún no somos humanos: propuestas de humanización para el tercer milenio. Barcelona: Península.

Coeckelbergh, M. (2019). Artificial intelligence: Some ethical issues and regulatory challenges. Technology and regulation, 2019, 31-34. DOI: https://doi.org/10.71265/a9yxhg88.

Comisión Europea (2024). Ley de la IA, Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas sobre inteligencia artificial. Recuperado de: https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai.

Cortina, A. (1996). El estatuto de la ética aplicada. Hermenéutica crítica de las actividades humanas. Isegoría, 13, 119–127.

Crawford, K. (2021). Atlas of AI: power, politics, and the planetary costs of artificial intelligence. New Haven: Yale University Press.

Cukurova, M. & Miao, F. (2024). AI competency framework for teachers. UNESCO Publishing. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000391104.

Diéguez, A. (2024). Pensar la tecnología. Barcelona: Shackleton books.

Du Sautoy, M. (2020). Programados para crear. Barcelona: Acantilado.

Estampa (2022). Cartografía de la IA. Proyecto Artístico disponible en abierto en: https://cartography-of-generative-ai.net/.

EU-HLEG (2019). A definition of AI: main capabilities and disciplines. Bruselas: Comisión Europea. Recuperado de: https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines .

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2022). Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. Information Systems Research, 33(2), 678-696. DOI: https://doi.org/10.1287/isre.2021.1079

Gordon, J. S. & Nyholm, S. (2024). Ethics of Artificial Intelligence. En The Internet Encyclopedia of Philosophy. Recuperado de: https://iep.utm.edu/ethics-of-artificial-intelligence/

Hernández-Fernández, A. (2024a). Tecnoética en el uso de la inteligencia artificial: el derecho a no ser desnudado. En: Perspectivas contemporáneas en educación:

innovación, investigación y transformación (1234-1251). Dykinson. Recuperado de: http://hdl.handle.net/2117/422475.

Hernández-Fernández, A. (2024b). Inteligencia artificial para docentes según la UNESCO. Educational Evidence. Recuperado de: https://educationalevidence.com/inteligencia-artificial-para-docentes-segun-la-unesco/.

Hernández Fernández, A. (2025a). Inteligencia artificial y diseño. Revista Gráfica, 13(26), 265-288. DOI: https://doi.org/10.5565/rev/grafica.461.

Hernández Fernández, A. (2025b). Intel·ligència artificial i creativitat en l'ensenyament artístic. Revista Imatges Colbacat, 5, 11-20. Recuperado de: https://hdl.handle.net/2117/430178.

Hernández Fernández, A. (2025c). Desnudando mujeres: sesgo de género en la inteligencia artificial generativa de imagen. En: WSCITECH25 (12-24). Recuperado de: https://hdl.handle.net/2117/428809.

Holmes, W., Bialik, M. & Fadel, C. (2019). Artificial intelligence in education: promises and implications for teaching and learning. Center for Curriculum Redesign.

Innerarity, D. (2025). Una teoría crítica de la inteligencia artificial. Barcelona: Galaxia Gutenberg.

Kahneman, D., Sibony, O. & Sunstein, C. R. (2021). Ruido: Un fallo en el juicio humano. Barcelona: Debate.

López de Mántaras, R. (2023). 100 coses que cal saber sobre intel·ligència artificial. Valls: Cossetània Edicions.

Ramírez Marqués, E. (2020). Tecnoètica a l'Educació Secundària Obligatòria. Recuperado de: http://hdl.handle.net/2117/333380.

Maleki, N., Padmanabhan, B. & Dutta, K. (2024). AI hallucinations: a misnomer worth clarifying. En 2024 IEEE conference on artificial intelligence (CAI) (133-138). IEEE. DOI: https://doi.org/10.1109/CAI59869.2024.00033.

Martínez, N. (2021). Tecnoètica al segon cicle de l'Educació Secundària Obligatòria (3r i 4t d'ESO). Recuperado de: http://hdl.handle.net/2117/349488.

Miao, F., & Shiohira, K. (2024). AI competency framework for students. UNESCO Publishing. Recuperado de: https://unesdoc.unesco.org/ark:/48223/pf0000391105.

Mokiy, V. (2023). Transdisciplinarity and Artificial Intelligence in the Service of Sustainable Development of Society. Artificial Intelligence and Human Mediation, 61.

Recuperado de: Artificial-Intelligence-and-Human-Mediation-Invited-editors-and-co-authors-Leonardo-Costa-and-Mariana-Thieriot.pdf.

Muñoz Cuenca, A. (2022). Futuro, ética y digitalización: una propuesta de introducción de la tecnoética en la asignatura "Tecnología e Ingeniería". Recuperado de: http://hdl.handle.net/2117/370731.

O'neil, C. (2018). Armas de destrucción matemática: cómo el big data aumenta la desigualdad y amenaza la democracia. Madrid: Capitán Swing Libros.

Ortega y Gasset, J. (1982 [1939]). Meditación de la técnica y otros ensayos sobre ciencia y filosofía. Buenos Aires: Espasa-Calpe Argentina. En: Obras, vol. 21. Madrid: Revista de Occidente.

Pareto, J. & Torras, C. (2024). To each Technology Its Own Ethics? A Reply to Sætra & Danaher. Philosophy & Technology, 37(3), 107. DOI: https://doi.org/10.1007/s13347-024-00798-w.

Pareto, J., Román, B. & Torras, C. (2021). The ethical issues of social assistive robotics: A critical literature review. Technology in Society, 67, 101726. DOI: https://doi.org/10.1016/j.techsoc.2021.101726.

Quintanilla, M. Á. (2017). Tecnología: un enfoque filosófico y otros ensayos de filosofía de la tecnología. Ciudad de México: Fondo de Cultura Económica.

Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. AJOB neuroscience, 11(2), 88-95. DOI: https://doi.org/10.1080/21507740.2020.1740350.

van der Rijt, J. W., Mollo, D. C. & Vaassen, B. (2025). AI Mimicry and Human Dignity: Chatbot Use as a Violation of Self-Respect. arXiv preprint arXiv:2503.05723.

Véliz, C. (2021). Privacidad es poder. Madrid: Debate.

Verbeek, P. P. (2005). What things do: Philosophical reflections on technology, agency, and design. University Park: Pennsylvania State Univ. Press.

Verbeek, P. P. (2011). Moralizing technology: Understanding and designing the morality of things. Chicago: University of Chicago press.

Wingström, Roosa, Hautala, Johanna & Lundman, Riina (2024). Redefining creativity in the era of AI? Perspectives of computer scientists and new media artists. Creativity Research Journal, 36(2), pp. 177-193. DOI: https://doi.org/10.1080/10400419.2022.2107850.