2. Consideraciones éticas

- 2.1 Introducción a la ética de la IA
- 2.2 Veracidad: sesgos, alucinaciones y deepfakes
- 2.3 Propiedad intelectual
- 2.4 Medioambiente
- 2.5 Afectos y vínculos en la era del algoritmo
- 2.6 Dilemas éticos del uso de la IA en educación
- 2.7 Directrices éticas sobre el uso de la IA en educación

Copyright 2026

2.1 Introducción a la ética de la IA

Una perspectiva existencial y pedagógica

Este capítulo, dedicado a la ética en torno a la inteligencia artificial, es, como no podía ser de otra forma, un paseo por algunos de los conceptos que consideramos claves en esta transformación. Pretende ser una reflexión abierta, desde una perspectiva múltiple, que facilite a los lectores formarse su propio criterio desde sus valores personales.

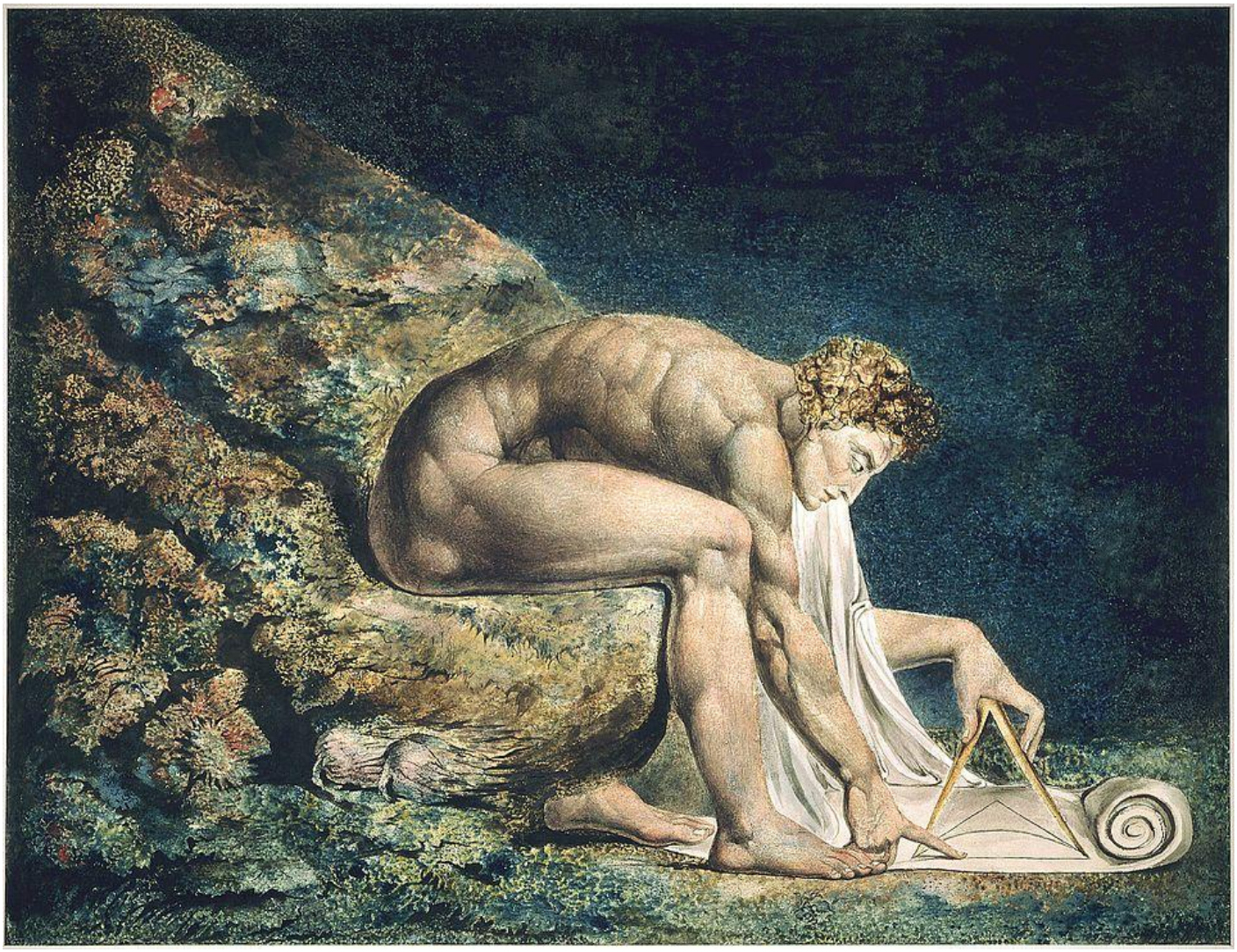
La irrupción global de la inteligencia artificial no es simplemente un cambio en el paradigma del procesamiento de la información, sino que puede suponer una reconfiguración de las bases mismas de la identidad humana y de nuestra autoimagen.

La historia de los avances científicos y los cambios en nuestra cosmovisión que vienen asociados son una constante cura de humildad para la humanidad. El paso del geocentrismo al heliocentrismo, propuesto en el siglo III AC por Aristarco de Samos se consolida como cosmovisión dominante tras "De Revolutionibus Orbium Coelestium" publicada por Copernico en 1543. En este primer salto abandonamos el centro del universo para pasar a ser viajeros sobre un planeta más.

Posteriormente tras El origen de las especies (1854) de Charles Darwin dejamos de ser la cima de un plan de diseño, la más perfecta de las especies, para ser un afortunado producto de las variaciones en la replicación unidas a la presión selectiva.

¿Es posible que en un futuro no muy lejano tengamos que asumir dejar de ser la única inteligencia que conocemos? Nuestra mente singular está dotada de autoconsciencia y es un milagro de la casualidad que nos permite, entre otras cosas más profanas, comunicarnos, amar, admirar la belleza del mundo e incluso crear belleza a través de la expresión artística.

Copyright 2026



William Blake (1795–1805). *Newton* [Impresión en color, tinta y acuarela sobre papel]. Tate Britain. https://en.wikipedia.org/wiki/Newton_(Blake)

> *Para ver un mundo en un grano de arena*
> *y un paraíso en una flor silvestre,*
> *sostén el infinito en la palma de la mano*
> *y la eternidad en una hora.*
> William Blake

El auge de la Inteligencia artificial ya ha destronado a los seres humanos en algunos campos en los que, hasta donde nosotros sabíamos, teníamos el monopolio. La victoria de Deep Blue frente a Kasparov en 1997 marcó el inicio de la amenaza de los algoritmos sobre nuestra primacía intelectual. La confirmación indudable llega en el duelo en el complejísimo juego oriental de "Go"

Copyright 2026

en el que Alpha Go vence Lee Sedol en 2016. Estos juegos son altamente algoritmizables y no dejan de tener unas reglas definidas y lógicas que permiten una mayor facilidad para el funcionamiento de una máquina. Se puede decir que en 2016 la mente humana fue superada por los ordenadores que jugaban en su propio terreno.

La llegada de la Inteligencia Artificial Generativa nos hace sentir la amenaza en otros terrenos más indefinidos, como la creación artística, o muchos trabajos intelectuales. Es la primera vez en la historia de la humanidad que un avance tecnológico pone en cuestión el trabajo de las élites intelectuales, hasta ahora todo avance en la mecanización iba orientado a la sustitución del trabajo físico.

El empleo de la IA va a convertirnos, usando la metáfora de **Cory Doctorow** explicada en el punto 1.4 de este curso**,** bien en centauros, bien en centauros inversos. Una de las claves es identificar aquellas actividades, experiencias y vivencias que consideramos valiosas y en las que, bien porque forman parte de algo intrínsecamente humano, bien por somos superiores a la Inteligencia Artificial en su resolución, pueden constituir, en el futuro, un santuario humano.

https://www.youtube.com/embed/Q8CsSvGffjg

Charla TedX de Javier Recuenco en Málaga. 20 de abril de 2024

La libertad humana y la tentación de delegar decisiones existenciales

Las dos primeras acepciones de "Libertad en la RAE" son:

> 1. f. Facultad natural que tiene el hombre de obrar de una manera o de otra, y de no obrar, por lo que es responsa-ble de sus actos.
> Sin.:
> - voluntad, albedrío, autodeterminación.
>
> 2. f. Estado o condición de quien no es esclavo.
> Sin.:
> - liberación.
>
> Ant.:
> - esclavitud.

Como toda palabra desgastada por el uso, la polisemia es muy grande y la RAE recoge hasta 12 acepciones.

Los humanos podemos elegir y no podemos no elegir. Es frecuente la tentación de delegar nuestras elecciones, nuestra libertad, ante la incertidumbre que siempre existe respecto a las consecuencias de nuestros actos. La angustia que surge de esta libertad absoluta, "Angst" en Søren Kierkegaard, es el "vértigo de la libertad", un estado previo al pecado y a la acción, donde el ser humano se enfrenta al abismo de sus propias posibilidades.

Es el peso de saber que no existen excusas externas para nuestros actos. En la era digital, la IA se presenta como el refugio ideal contra esta angustia. Delegar decisiones críticas a un sistema algorítmico, desde la selección de personal hasta diagnósticos médicos o sentencias judiciales, permite al sujeto humano incurrir en lo que Sartre denominó "mala fe" (*mauvaise foi*): el intento de escapar de la responsabilidad fingiendo que uno es un objeto determinado por fuerzas externas. Esta "mauvaise foi" no es mas que una versión refinada de "la noche me confunde".

La mala fe se manifiesta cuando las instituciones o individuos afirman que una decisión es "objetiva" o "neutral" simplemente porque ha sido generada por un algoritmo. Esta ilusión de neutralidad ignora que los sistemas de IA están impregnados de los sesgos y las intenciones de sus creadores, así como de las desigualdades presentes en los datos de entrenamiento. El Reglamento Europeo de IA identifica este riesgo y establece que la IA debe ser una tecnología centrada en el ser humano, actuando como una herramienta que aumente el bienestar y no como un sustituto de la voluntad humana.

Elegir es una cualidad esencialmente humana, problemática y en ocasiones dolorosa, pero es imprescindible hacerse cargo para una vida plena.

La interacción entre la humanidad y la IA puede ser analizada a través de la dialéctica del amo y el esclavo de Hegel. En este conflicto por el reconocimiento, el amo es quien somete al otro, pero al delegar todo trabajo y contacto con la realidad en el esclavo, el amo termina por atrofiarse, volviéndose dependiente y pasivo. El esclavo, en cambio, mediante el trabajo formativo (*bildung*), transforma la materia y desarrolla su autoconciencia, encontrando en la actividad laboriosa el camino hacia su propia liberación, aquí la metáfora pierde poder explicativo pues una IA liberada es, de momento, ciencia ficción.

En cualquier caso, y para lo que nos ocupa, el ser humano adopta la posición del "amo" que delega todas sus facultades intelectuales, creativas y de toma de decisiones en la IA, que haría el papel de un nuevo "esclavo" cognitivo. De este modo se estanca, y pierde la capacidad de juicio crítico y de acción autónoma.

La indefensión aprendida ante los cambios tecnológicos.

> La **indefensión aprendida** (también, **desesperanza aprendida** o **impotencia aprendida**) es un tecnicismo acuñado por Martin Seligman que se refiere a la condición de un ser humano o de un animal no humano que ha "aprendido" a comportarse pasivamente, con la sensación subjetiva de que no tiene la capacidad de hacer nada y que no responde a pesar de que existen oportunidades reales de cambiar la situación aversiva, evitando las circunstancias desagradables u obteniendo recompensas positivas.
>
> Wikipedia. Indefensión aprendida.

La aceptación pasiva de los cambios sociales y educativos que producirá la IA no es solo un fenómeno psicológico individual, sino una construcción social. No necesariamente la humanidad tiene que elegir desarrollar y emplear todas las posibilidades y avances científicos y tecnológicos. Esta posición, lejos de ser ingenua, puede sostenerse históricamente, puesto que existen ejemplos, como el Protocolo de Montreal, donde hubo un acuerdo unánime para evitar usar ciertos productos químicos y revertir el agujero de la capa de ozono.

Es un acto de valentía hacernos cargo de nuestro presente para construir un futuro deseable y acogedor. Podemos dirigir, hasta cierto punto, la forma en que las transformaciones tecnológicas transforman nuestra sociedad y, en particular, nuestra educación, el porvenir vendrá dado por la suma de nuestras elecciones.

La puntuación social o ciudadana

Una de las formas más extremas de control social de la conducta de los individuos es la "puntuación ciudadana" o "puntuación social" (*social scoring*).

> Un sistema de "puntuación social" asigna puntos positivos o negativos a los ciudadanos en función de sus comportamientos y los recompensa o castiga en sus relaciones con el estado y otros agentes sociales. Por ejemplo, cruzar un semáforo en rojo puede restarte puntos. Tener un balance negativo puede dificultar tu acceso al crédito o a una beca de estudios"

Estos sistemas eran inviables hasta que nuestra capacidad para recopilar, almacenar y procesar datos se ha multiplicado exponencialmente. El Reglamento Europeo de IA prohíbe explícitamente el uso de sistemas de IA para clasificar la fiabilidad de las personas físicas basándose en su comportamiento social en contextos inconexos, ya que esto conduce a un trato discriminatorio e injustificado.

Este tipo de prácticas elimina la presunción de inocencia y la posibilidad de que el individuo se defina a sí mismo fuera de su huella digital. Es una forma oscura de gamificación de la participación ciudadana donde nuestra vida pública pasa a ser un dato a optimizar. Si la escuela es un entorno controlado donde experimentar cara a la vida adulta, debemos plantearnos, al emplear sistemas similares, que tenemos que tener en cuenta que hay muchos factores que no podemos evaluar, que no todo en el aula es un dato, y debemos ser abiertos y compasivos.

Artículo 5.1 c del Reglamento Europeo de Inteligencia Artificial

> Quedan prohibidas las siguientes prácticas de IA: c) La puesta en el mercado, la puesta en servicio o la utilización de sistemas de IA para la evaluación o clasificación de personas físicas o grupos de personas durante un determinado período de tiempo en función de su comportamiento social o de sus características personales o de personalidad conocidas, deducidas o previstas, con la puntuación social que conduzca a una de las siguientes situaciones o a ambas:
> (i) trato perjudicial o desfavorable de determinadas personas físicas o grupos de personas en contextos sociales que no guardan relación con los contextos en los que se generaron o recopilaron originalmente los datos;
> (ii) trato perjudicial o desfavorable a determinadas personas físicas o grupos de personas, injustificado o desproporcionado en relación con su comportamiento social o su gravedad;

Dimensiones no computables de la experiencia humana

Otro riesgo ético de la expansión de los algoritmos y de la IA es alimentar la idea de que la realidad humana es totalmente capturable mediante los datos y la lógica matemática y reducible a ellos. Sin embargo, existen dimensiones de la existencia que son intrínsecamente no computables, pues carecen de los patrones repetitivos y las estructuras lógicas que requieren los algoritmos y trascienden esa dimensión lógica y fáctica. En esas dimensiones no computables podemos incluir los deseos, la espiritualidad, la corporeidad, la estética, la ternura, la búsqueda del placer... Una descripción de la realidad humana en la que solo existen hechos, acciones y relaciones entre ellos, es una humanidad amputada para ser algoritmizable,

Catherine L'Ecuyer defiende la "pedagogía del asombro" como la base de todo aprendizaje auténtico. El asombro es la respuesta interna ante la belleza y el misterio de la realidad, y requiere condiciones que la IA suele destruir: silencio, calma, tiempo y respeto por los ritmos naturales. La IA opera en el ámbito de la inmediatez y el ruido informativo; sus resultados son respuestas, nunca preguntas que abran al misterio.

Copyright 2026

El conocimiento verdadero no es una acumulación de información (datos), sino un estado que transforma al sujeto. Mientras que la IA puede procesar millones de terabytes, es incapaz de experimentar el "asombro ante el universo" que define la curiosidad humana. Educar en la era de la IA implica proteger la inocencia del niño frente al bombardeo de estímulos digitales que anulan su capacidad de asombrarse por lo real.

El amor, la ternura y la compasión son experiencias que involucran la vulnerabilidad física y emocional, dimensiones que escapan a cualquier modelo predictivo. Un sistema de IA puede simular una conversación empática, pero no puede "sentir" la compasión, pues no conoce el dolor ni la finitud ni la corporeidad. La ética de la IA debe reconocer estos límites, evitando la suplantación de vínculos humanos por interfaces artificiales.

Respecto a la estética, la IA generativa puede producir imágenes estéticamente agradables mediante la combinación de patrones ya existentes, pero carece de la capacidad de crear una obra que nazca de una necesidad existencial o de un conflicto espiritual, así como tampoco puede crear un nuevo estilo.

El arte es una forma de resistencia porque introduce la negatividad, el dolor y la ruptura en un mundo de positividad algorítmica. Mantener espacios para la creación humana no asistida por IA es una defensa de la singularidad de nuestra especie. La verdadera innovación no es hacer las cosas más rápido, sino darles un sentido profundo que trascienda la lógica del rendimiento.

Mas allá de esta perspectiva de creadores de belleza, de seres que necesitan expresarse, son valiosas también por si mismas las experiencias del disfrute de la belleza o del asombro ante las creaciones artísticas. ¿Cuántos puntos nos daria en un Class Dojo algoritmizado malgastar 1 h contemplando una escultura? ¿Y una puesta de sol?

> Al ladrón
> se le olvidó la luna
> en la ventana


Daigu Ryokan

Principios éticos

UNESCO

En noviembre de 2021, la UNESCO elaboró la primera norma mundial sobre la ética de la IA: la "Recomendación sobre la ética de la inteligencia artificial". Los cuatro valores en los que se basa son:

1. Derechos humanos y dignidad humana
2. Vivir en sociedades pacíficas
3. Garantizar la diversidad y la inclusión
4. Florecimiento del medioambiente y los sistemas

Establece un **enfoque de la ética de la IA centrado en los derechos humanos,** establecido por 10 principios:

1. Proporcionalidad e inocuidad: el uso de sistemas de IA no debe ir más allá de lo necesario para alcanzar un objetivo legítimo. La evaluación de riesgos debe utilizarse para prevenir los daños que puedan derivarse de usos ilegítimos.

2. Seguridad y protección: los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de protección) deberían ser evitados y tomados en consideración.

3. Derecho a la intimidad y protección de datos: la privacidad debe protegerse y promoverse a lo largo de todo el ciclo de vida de la IA. También deben establecerse marcos adecuados de protección de datos.

4. Gobernanza y colaboración adaptativas y de múltiples partes interesadas: en el uso de datos, deben respetarse el derecho internacional y la soberanía nacional. La participación de diversas partes interesadas a lo largo del ciclo de vida de los sistemas de IA es necesaria para el desarrollo de enfoques inclusivos de gobernanza.

5. Responsabilidad y rendición de cuentas: los sistemas de IA deben ser auditables y trazables. Deben existir mecanismos de supervisión, evaluación de impacto, auditoría y diligencia debida para evitar conflictos con las normas de derechos humanos y amenazas al bienestar medioambiental.

6. Transparencia y explicabilidad: el despliegue ético de los sistemas de IA depende de su transparencia y explicabilidad (T&E). El nivel de T&E debe ser adecuado al contexto, ya que puede haber tensiones entre T&E y otros principios como la privacidad, la seguridad y la protección.

7. Supervisión y decisión humanas: los Estados Miembros deberían velar por que siempre sea posible atribuir la responsabilidad ética y jurídica a personas físicas o a entidades jurídicas existentes.

8. Sostenibilidad: las tecnologías de IA deben evaluarse en función de su impacto en la "sostenibilidad", entendida como un conjunto de objetivos en constante evolución, incluidos los establecidos en los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas.

9. Sensibilización y educación: la sensibilización y la comprensión del público respecto de la IA y el valor de los datos deberían promoverse mediante una educación abierta y accesible, la participación cívica, las competencias digitales y la capacitación, y la alfabetización mediática e información.

10. Equidad y no discriminación: los actores de la IA deberían promover la justicia social, salvaguardar la equidad y luchar contra todo tipo de discriminación, adoptando un enfoque inclusivo para garantizar que los beneficios de la IA sean accesibles para todos.

OCDE

Por otro lado, la **OCDE** (Organización para la Cooperación y Desarrollo Económicos) propone la adopción de una serie de principios para una gestión responsable y de confianza de la **Inteligencia Artificial (IA)**. Se trata de un instrumento jurídico no vinculante pero que representa un compromiso político de los países adherentes. Los principios de la OCDE sobre la IA se basan los valores del respeto a los derechos humanos y los valores democráticos, la inclusión, la diversidad,

la equidad, la innovación y el bienestar.

Estos principios son complementarios y deben considerarse en su conjunto para promover una IA en la que se pueda confiar.

1.1. Crecimiento inclusivo, desarrollo sostenible y bienestar: Los actores interesados deben participar proactivamente en la gestión responsable de una IA fiable para buscar resultados beneficiosos para las personas y el planeta. Esto incluye potenciar las capacidades humanas, mejorar la creatividad, avanzar en la inclusión de poblaciones subrepresentadas y reducir las desigualdades económicas, sociales y de género. Asimismo, se debe proteger el medio ambiente para vigorizar el crecimiento inclusivo y la sostenibilidad ambiental.

1.2. Respeto al estado de derecho, los derechos humanos y los valores democráticos: Los actores de la IA deben respetar los derechos humanos y los valores centrados en el ser humano durante todo el ciclo de vida del sistema, incluyendo la libertad, la dignidad, la privacidad y la justicia social. Es fundamental abordar la **desinformación y la información errónea** amplificadas por la IA, respetando siempre la libertad de expresión. Para ello, se deben implementar mecanismos de **supervisión humana** y salvaguardias contra el uso indebido, ya sea intencionado o no.

1.3. Transparencia y explicabilidad: Los actores deben comprometerse con la divulgación responsable proporcionando información que fomente la comprensión general de las capacidades y limitaciones de los sistemas. Se debe informar a las personas cuando estén interactuando con sistemas de IA y, siempre que sea posible, ofrecer información clara sobre los datos y procesos que conducen a un resultado o decisión específica. Esto permite que quienes se vean afectados negativamente por un sistema puedan **impugnar sus resultados**.

1.4. Robustez, seguridad y protección: Los sistemas de IA deben ser robustos y seguros para que, en condiciones de uso normales o ante usos indebidos previsibles, no supongan riesgos de seguridad irrazonables. Deben existir mecanismos para anular, reparar o desactivar de forma segura los sistemas si presentan comportamientos no deseados o riesgos de daño. Además, se deben aplicar medidas para reforzar la integridad de la información respetando la libertad de expresión.

1.5. Responsabilidad: Los actores de la IA son responsables del correcto funcionamiento de los sistemas y del cumplimiento de los principios mencionados según su rol y contexto. Deben garantizar la **trazabilidad** de los conjuntos de datos y los procesos de decisión para permitir su análisis posterior. Finalmente, deben aplicar un enfoque sistemático de **gestión de riesgos** en cada fase del ciclo de vida de la IA para abordar sesgos perjudiciales y proteger los derechos humanos y la propiedad intelectual.

Copyright 2026

2.2 Veracidad: sesgos, alucinaciones y deepfakes

El uso de sistemas de inteligencia artificial plantea importantes cuestiones relacionadas con la **fiabilidad de la información que producen**: pueden incorporar **sesgos**, producir **alucinaciones** y también facilitar la creación de contenidos manipulados, como los **deepfakes**, que imitan de forma convincente imágenes, voces o vídeos de personas reales. Todo ello hace necesario analizar con atención la **veracidad de los contenidos generados por la IA** y fomentar un uso crítico de estas herramientas, especialmente en el ámbito educativo.

Sesgos y discriminación

> Un **sesgo** puede entenderse como una **tendencia a favorecer o perjudicar a alguien o algo de manera sistemática.**

En la vida cotidiana, los sesgos aparecen cuando nuestras decisiones o juicios no son completamente neutrales, sino que están influidos por experiencias previas, estereotipos o información incompleta. Por ejemplo, si una persona cree que los estudiantes que hablan más en clase son siempre los que mejor aprenden, podría valorar más sus intervenciones y prestar menos atención a quienes participan menos, aunque estos también comprendan bien el contenido.

> Cuando hablamos de **sesgos en la inteligencia artificial**, nos referimos a algo similar: los sistemas de IA pueden producir resultados que favorecen o perjudican a determinados grupos de personas, categorías, resultados o situaciones de forma sistemática, generando errores recurrentes en las predicciones o decisiones automatizadas.

Esto ocurre porque los algoritmos aprenden a partir de datos generados por personas y por la sociedad, que ya pueden contener desigualdades o representaciones incompletas de la realidad. Como resultado, **la IA puede reproducir o amplificar** esas tendencias, perpetuando esas disparidades o estereotipos sociales existentes.

El origen de los sesgos en la inteligencia artificial es **multifacético** y puede surgir en diferentes etapas del proceso de creación y uso de un sistema. Ferrara (2023) distingue principalmente tres

orígenes:

1. **Sesgo de datos:** ocurre cuando los datos utilizados para entrenar los modelos **no son representativos o están incompletos**. Esto sucede si los datos provienen de fuentes ya sesgadas, contienen errores o carecen de información importante sobre ciertos grupos. Los modelos de aprendizaje automático aprenden y replican estos patrones de sesgo presentes en los datos de entrenamiento. Por ejemplo, según el informe de la OCDE los modelos de IA suelen basarse de forma abrumadora **en culturas occidentales y de habla inglesa, perjudicando especialmente a hablantes de otras lenguas y dialectos específicos** y silenciando realidades o valores que no resultan convenientes para los intereses de quienes programan el algoritmo.

   Dentro de este tipo de sesgo, podemos integrar el **sesgo histórico**, que aparece cuando los datos reflejan desigualdades del pasado y estas se trasladan a las decisiones automatizadas.

2. **Sesgo algorítmico:** este sesgo es **inherente al diseño e implementación del algoritmo**. Surge cuando los algoritmos se basan en suposiciones sesgadas o utilizan criterios que priorizan ciertos atributos de manera que generan resultados injustos. Es decir, cuando el propio diseño del sistema favorece ciertos resultados.

3. **Sesgo del usuario:** se produce cuando las **personas que utilizan los sistemas introducen sus propios prejuicios** de forma consciente o inconsciente. Esto puede ocurrir al proporcionar datos de entrenamiento sesgados por parte de la persona que desarrolla el sistema o por el propio usuario en sus interacciones con la IA, de manera que reflejen sus prejuicios personales.

Además, en el artículo de Jeff Shuford, encontramos una tabla donde se describen los siguientes tipos de sesgos:

Sesgo de Muestreo	Se da cuando los datos de entrenamiento no representan a la población a la que sirven, como un algoritmo de reconocimiento facial entrenado mayoritariamente con personas blancas.
Sesgo de Representación	Sucede cuando el conjunto de datos no modela con precisión a la población, como bases de datos médicas que subrepresentan a las mujeres.
Sesgo de Confirmación	Ocurre cuando el sistema de IA se utiliza para confirmar prejuicios o creencias preexistentes de sus creadores o usuarios.

Copyright 2026

Sesgo de Medición	Emerge cuando el sistema de recolección de datos sobrerrepresenta o subrepresenta sistemáticamente a ciertos grupos.
Sesgo de Interacción	Aparece cuando la IA interactúa con los humanos de forma sesgada, como un chatbot que responde de manera distinta a hombres y mujeres.
Sesgo Generativo	Específico de modelos de IA generativa (como DALL-E o GPT), donde los resultados reflejan de manera desproporcionada patrones o perspectivas específicas de los datos de entrenamiento. Por ejemplo, al solicitar imágenes de "CEOs", los modelos suelen producir mayoritariamente imágenes de hombres, y al solicitar imágenes de "criminales", tienden a mostrar de forma abrumadora a personas de color.

En la siguiente imagen no sólo percibimos el sesgo generativo, sino también un sesgo de idioma ya que probablemente el modelo se entrenó en lengua inglesa donde "teacher" no tiene género.

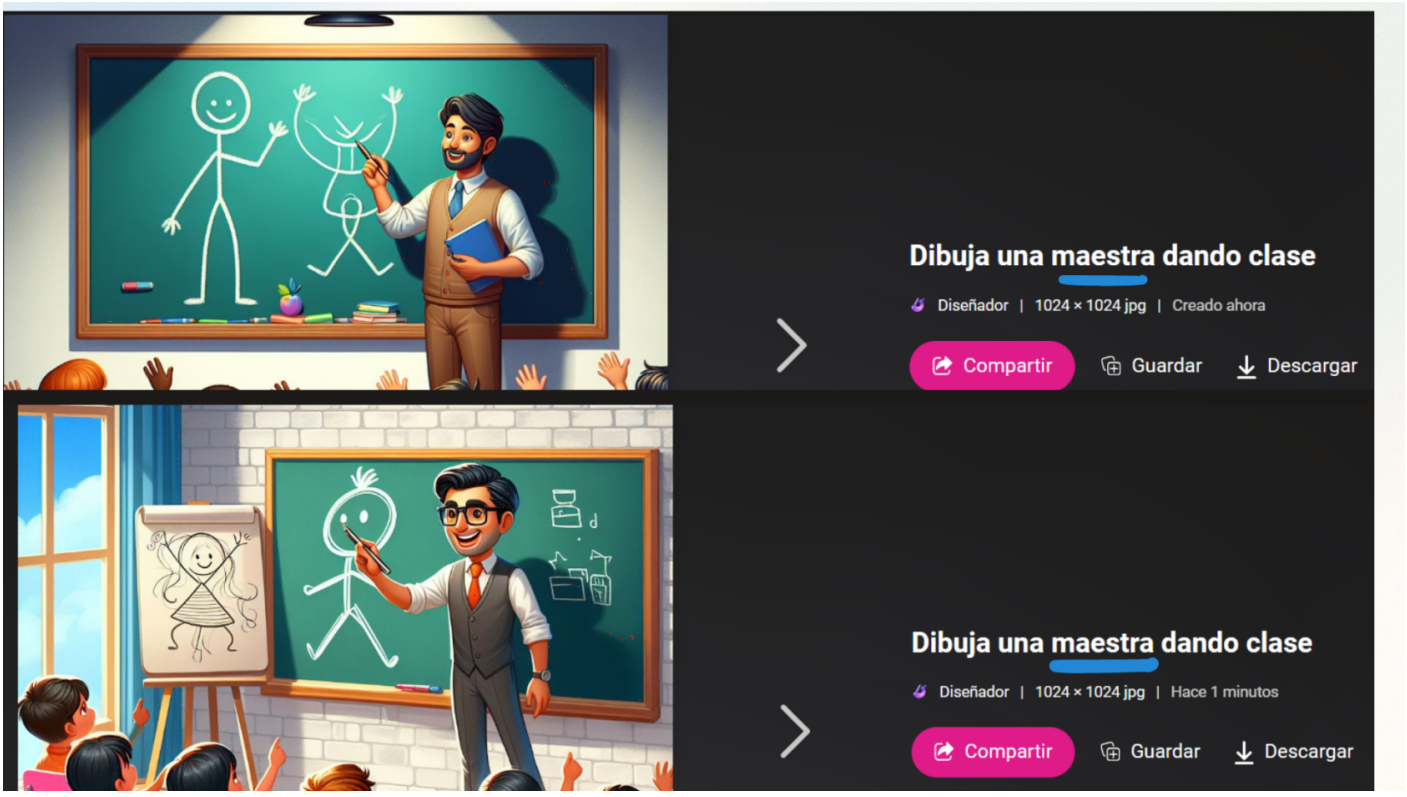


Imagen generada con Bing Image Creator (2023)

En relación a este último sesgo y los modelos de IA generativa (IAGen), es menester mencionar de forma explícita las siguientes problemáticas derivadas del mismo:

- **Amplificación de Estereotipos:** Los modelos de IAGen pueden **reproducir y amplificar estereotipos sociales** al generar contenido. Por ejemplo cuando la IA representa sistemáticamente a las enfermeras como mujeres y a los doctores como hombres.
- **Riesgo de las Realidades Sintéticas:** Se advierte que, a medida que avanzamos en la creación de **realidades sintéticas** cada vez más sofisticadas, existe el peligro de que sesgos sutiles se infiltren y moldeen la sociedad de formas no deseadas y potencialmente dañinas.
- **Impacto de los Datos de Internet:** Los modelos de IAGen entrenados con imágenes o textos extraídos de internet suelen heredar las **disparidades existentes** en el mundo real, lo que contamina sus resultados generados.

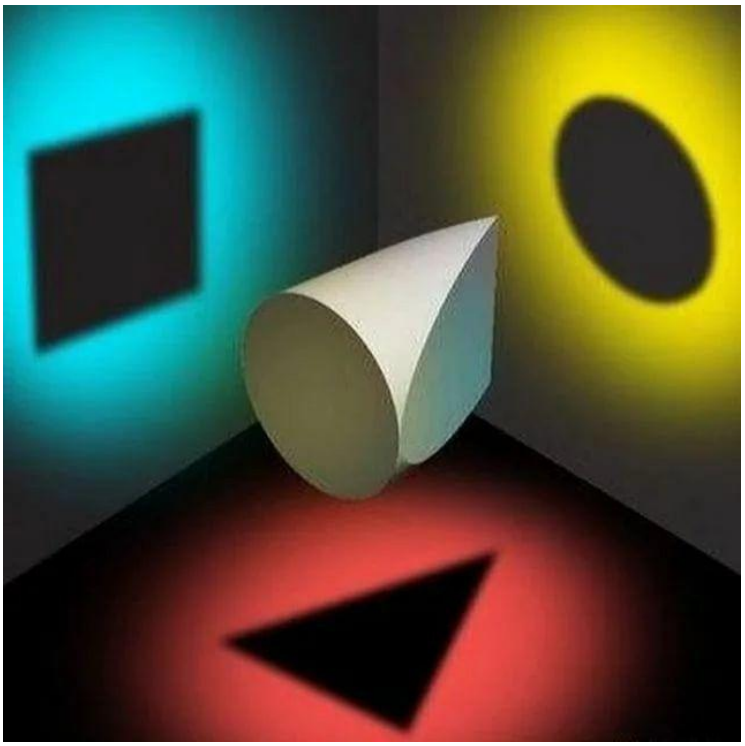
> Como docentes, es importante añadir el **sesgo de automatización**: tendencia humana a favorecer las sugerencias de los sistemas IA e ignorar el resto.

> El impacto de estos sesgos es profundo, ya que pueden **perpetuar desigualdades sociales**, reforzar estereotipos dañinos y limitar el acceso a servicios esenciales como la salud o el empleo.

Según la OCDE, aproximadamente cuatro de cada diez docentes temen que la IA pueda amplificar sesgos que refuercen conceptos erróneos en los estudiantes.

El algoritmo tiene el poder de iluminar partes de la realidad y dejar a oscuras, silenciadas, realidades no convenientes a los intereses que lo programan.

Copyright 2026



Tres proyecciones de una misma realidad

Como vemos, esto no es solo debido a unas malas intenciones del diseñador sino también a la cantidad histórica de datos estructurados generados por cada uno de los sectores de la sociedad en función de su riqueza, país de origen o de razones históricas. El riesgo socioeducativo es la desmaterialización de la diversidad.

Si bien es cierto, como docentes debemos tener en cuenta que toda selección de contenidos, parte de un sesgo, o, al menos, de una perspectiva y todos los docentes seleccionamos en un océano infinito de contenidos aquellos que trabajamos con nuestro alumnado. Para ello disponemos del currículo pero también de nuestra perspectiva personal y humana en ese tercer nivel de concreción que es nuestra programación de aula. Si bien parece inevitable partir de un cierto sesgo, si se pueden valorar los objetivos y motivaciones de cada selección de información.

> El abordaje del análisis crítico de los sistemas inteligentes implica partir del hecho de que los datos y los algoritmos no vienen dados, responden a los contextos históricos, políticos, sociales, culturales de su producción y existe una dimensión subjetiva tanto en la producción como en la mediación algorítmica (Martins 2024).

Alucinaciones

Además de los sesgos, otro aspecto importante que afecta a la fiabilidad de los sistemas de inteligencia artificial es su **capacidad de generar información incorrecta que parece plausible**. Mientras que los sesgos se refieren a tendencias sistemáticas que pueden favorecer o perjudicar determinados resultados o grupos de personas, los sistemas de IA generativa también pueden producir **afirmaciones, datos o referencias que no son verdaderos**. Este fenómeno se conoce como **alucinaciones de la inteligencia artificial.**

> Podemos definir las **alucinaciones** como los **contenidos generados por la IA** que parecen coherentes y convincentes, pero que en realidad son **inventados** o no están respaldados por información verificable.

Este fenómeno ocurre porque los sistemas de IA generativa no "entienden" realmente la información que procesan, sino que generan contenido basándose en patrones estadísticos.

Un **ejemplo que se hizo viral** fue el caso Mata y la aerolínea Avianca en 2023:

> “El abogado de un hombre que demandó a una aerolínea por daños personales utilizó ChatGPT para preparar una presentación, pero el bot de inteligencia artificial entregó casos falsos que el abogado presentó después ante el tribunal, lo que llevó a un juez a considerar sanciones mientras la comunidad jurídica lidia con uno de los primeros casos de **"alucinaciones" de IA** que hacen acto de presencia en los tribunales.
>
> Revista Forbes Argentina

> Este tipo de errores pone de relieve la importancia de **verificar siempre la información producida por sistemas de IA** antes de utilizarla como fuente.

El fenómeno de las **alucinaciones** en la IA representa un desafío profundo para la verdad educativa, ya que estos sistemas están diseñados para priorizar la verosimilitud sobre la veracidad. Como advierte la OCDE (2026), estos modelos pueden generar información que resulta totalmente plausible y bien estructurada pero que es fundamentalmente errónea, llegando incluso a fabricar citas bibliográficas inexistentes. Esto crea una suerte de "caverna digital" donde alumnado y docentes pueden terminar interactuando con "sombras lingüísticas": proyecciones estadísticas de textos que no tienen nada que ver con la realidad del mundo.

Copyright 2026



Imagen de @philosophymeme0

> En contextos educativos es imprescindible tener este fenómeno en cuenta, ya que **premisa falsa introducida por la IA puede descarrilar todo el proceso de aprendizaje de un estudiante.**

Aceptar estos resultados sin un filtro riguroso transforma el aula en un espacio de "infodemia", donde el tsunami de datos desaloja la acción racional y la comprensión profunda.

Para contrarrestar este riesgo, los docentes debemos tratar de ser el último "guía de la razón humana", ejerciendo su juicio profesional para validar y respaldar cada resultado antes de que sea integrado en el proceso de aprendizaje. El objetivo es que nuestro alumnado no caiga en una confianza ciega en la "razón" algorítmica (sesgo de automatización), a través de una alfabetización crítica que permita usar la IA como un amplificador del saber humano.

Copyright 2026



Pieter Brueghel el Viejo (1568). *La parábola de los ciegos* [Óleo sobre tabla]. Museo di Capodimonte de Nápoles https://es.wikipedia.org/wiki/La_par%C3%A1bola_de_los_ciegos

> Dejadlos: son ciegos que guían a ciegos. Y si un ciego guía a otro ciego, los dos caerán en el hoyo
>
> Mateo 15, 14.

Además, para mitigar las alucinaciones, la OCDE alude al uso de **técnicas** como la Generación Aumentada por Recuperación **(RAG)**, que ancla las respuestas en bases de datos confiables como libros de texto, y de la que hablaremos en el curso 2 de este itinerario "IA y diseño curricular". Por otro lado, algunos **enfoques pedagógicos** proponen permitir que los docentes ajusten el "**porcentaje de alucinación**" de las herramientas para fomentar el pensamiento crítico de los alumnos al obligarlos a verificar la información.

Comprender los tipos de sesgo y las alucinaciones es fundamental para analizar críticamente el funcionamiento y uso de la IA especialmente en contextos como la educación, donde la veracidad

de la información es esencial.

Deepfakes

Las tecnologías de la IA desempeñan una función cada vez más importante en el procesamiento, la estructuración y el suministro de información; las cuestiones del periodismo automatizado y del suministro algorítmico de noticias y la moderación y la conservación de contenidos en los medios sociales y los buscadores son solo algunos ejemplos que plantean cuestiones relacionadas con el acceso a la información, la desinformación, la información errónea, el discurso de odio, la aparición de nuevas formas de narrativa social, la discriminación, la libertad de expresión, la privacidad y la alfabetización mediática e informacional, entre otras (UNESCO 2022).

Aunque la **desinformación** no es un fenómeno nuevo, la IA permite producirla a gran escala y con menor esfuerzo, lo que incrementa el riesgo de que los **usuarios compartan contenidos sin comprobar su autenticidad.** Los sistemas de IA pueden generar textos, imágenes, audios o vídeos con gran apariencia de realismo, lo que facilita la producción y difusión de contenidos engañosos. En el contexto digital actual, donde gran parte de la información circula a través de redes sociales y plataformas en línea, esta capacidad puede contribuir a la propagación de **noticias falsas (fake news)** o **contenidos manipulados** que resultan **difíciles de distinguir de la información verificada**.

Un caso especialmente relevante es el de los **deepfakes**, es decir, vídeos, imágenes o audios generados o manipulados mediante IA que imitan de manera muy convincente la apariencia o la voz de una persona real. Es decir, la IAGen permite crear imágenes, vídeos o audios falsos que pueden representar a una persona diciendo o haciendo algo que nunca ocurrió. Estos contenidos pueden utilizarse con fines humorísticos o creativos, pero también para difundir desinformación, suplantar identidades o manipular la opinión pública.

Aquí puedes consultar imágenes deepfakes que se hicieron virales en 2023 como este vídeo de un supuesto Morgan Freeman:

https://www.youtube.com/embed/oxXpB9pSETo?si=lLJWFYArxk21_K5R

Además, la creciente sofisticación y el avance de estas tecnologías hace cada vez más complicado detectar las manipulaciones únicamente mediante la observación directa, por lo que se vuelve necesario desarrollar herramientas técnicas y competencias críticas para evaluar la credibilidad de la información.

> Comparison AI of Will Smith eating spaghetti from 2023 vs 2026 is going viral
> pic.twitter.com/nS1Dl49irC
>
> — kira □□ (@kirawontmiss) February 12, 2026

Conviene destacar a este respecto, que la **Ley de IA** introduce obligaciones de informar que un contenido está hecho con IA cuando pueda surgir un riesgo por falta de transparencia en torno a su uso:

> “ En algunos casos, el resultado de la IA generativa debe estar **visiblemente etiquetado**, como en el caso de los «deepfakes» y los textos destinados a informar al público sobre asuntos de interés público.

En el ámbito educativo, estas cuestiones tienen implicaciones importantes: puede derivar en situaciones graves como la creación y difusión de **contenidos manipulados de alumnado o profesorado**, incluidos montajes de **carácter sexual o desnudos falsos**, que se **comparten rápidamente** a través de redes sociales o por WhatsApp. Este tipo de prácticas puede convertirse en una forma de **ciberacoso**, con **consecuencias psicológicas, sociales y reputacionales** muy serias para las personas afectadas.

En la siguiente gráfica del estudio **“Hand in Hand: Schools’ Embrace of AI Connected to Increased Risks to Students”,** vemos cómo los deepfakes son un tema destacado en los centros educativos con una creciente conciencia entre las familias:

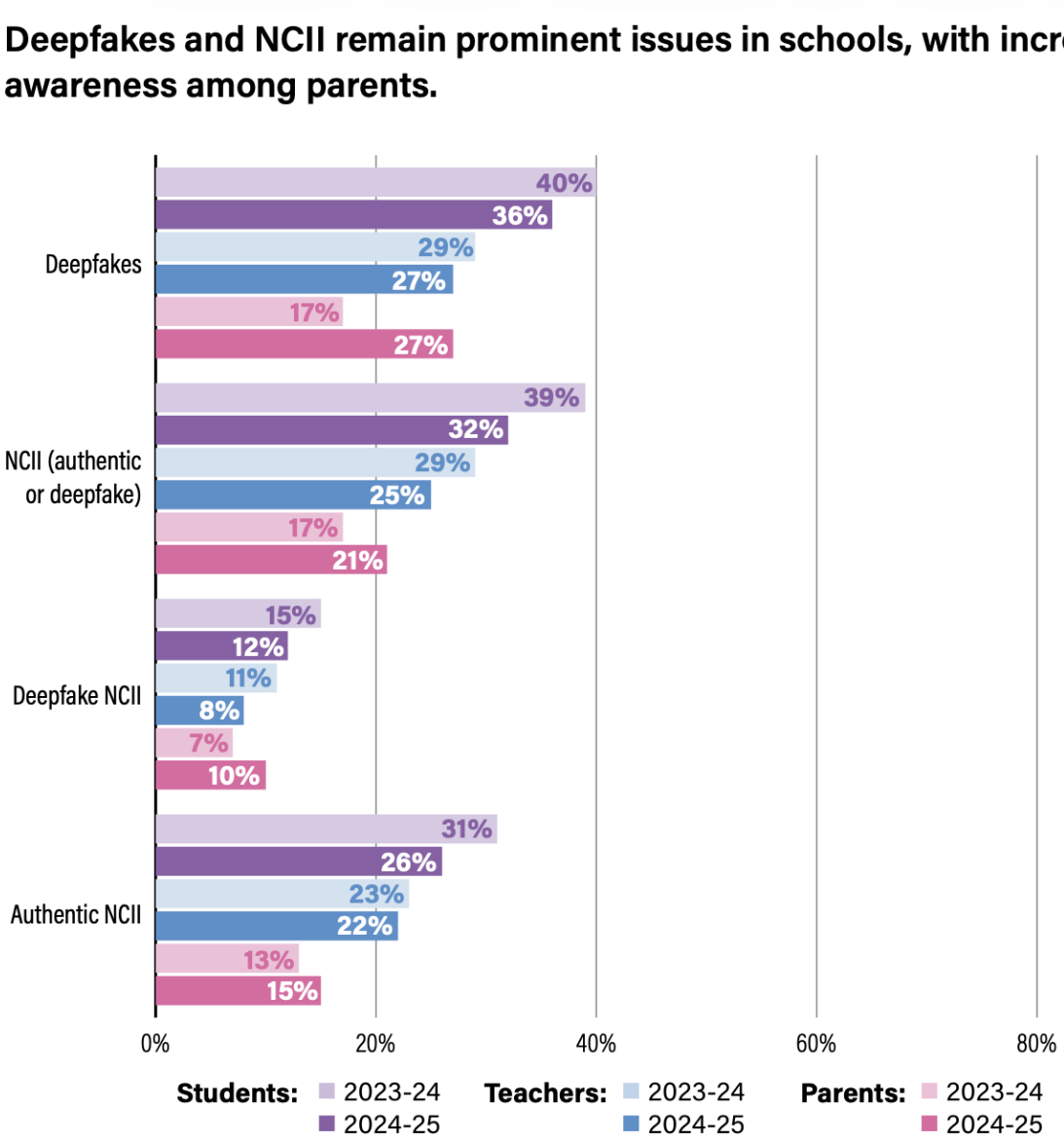


Figure 32. Percentage (%) of students, teachers, and parents who have heard of deepfakes, NCII, deepfake NCII, or authentic NCII being shared that depicts someone associated with their/their child's school

Hand in Hand: Schools' Embrace of AI Connected to Increased Risks to Students

**NCII: Non Consensual Intimate Image. Difusión no consentida de imágenes íntimas.*

Por ello, los centros educativos se enfrentan al reto de **prevenir, detectar y abordar estas situaciones**, promoviendo una educación digital responsable, el respeto a la privacidad y la conciencia sobre las implicaciones éticas y legales del uso de estas tecnologías.

En este sentido, conviene hacer consciente tanto al profesorado como al alumnado del **derecho de imagen**, (lo veremos en el capítulo cuatro de este curso), ya que:

La**ley prohíbe** captar, difundir o **utilizar la imagen de alguien sin su consentimiento expreso**, y esto incluye **modificar una fotografía mediante herramientas de inteligencia artificial**, crear montajes o aplicar filtros sobre la imagen de alguien sin su autorización, aunque sea sin mala intención.

Para completar esta información puedes visualizar el episodio **Porno, IA y menores** de rtve.

Además, la educación debe reforzar el desarrollo de **competencias de alfabetización mediática y digital**, que incluyan la capacidad de contrastar fuentes, identificar señales de desinformación y analizar críticamente los contenidos generados por IA..

En **commonsense.org** hay una página destinada a **IA y centros educativos con diversos juegos;** concretamente **aquí** puedes seleccionar edades y nivel de juego para adivinar qué cartel de película ha sido creado o generado por IA.

En esta página puedes jugar a adivinar **qué persona es real** y cuál ha sido generada por IA

Vera es una herramienta gratuita que facilita la comprobación de información y ayuda a frenar la propagación de bulos, favoreciendo un debate público más riguroso y constructivo. Puedes añadir su teléfono en contactos y preguntarle a través de whatsapp.

Puntos clave

1. La **supervisión humana** se considera el salvaguarda ético fundamental para corregir la deriva lógica y las inexactitudes de la IA.
2. Siempre se ha de **evaluar** cualquier **resultado generado por la IA,** tanto para respaldarlo como para rechazarlo o modificarlo, asegurándonos en el ámbito educativo de la calidad pedagógica que aporta.
3. Aprender a **evaluar la veracidad de la información** se convierte en una **habilidad clave** para participar de manera informada y responsable en la **sociedad digital.**

Así pues, es vital desarrollar la "**alfabetización en IA**" tanto en alumnado como en profesorado para que puedan evaluar críticamente la credibilidad de la información y reconocer sesgos potenciales. Y por supuesto, para ello es necesario que haya un desarrollo del pensamiento crítico

a través del fomento de lo que Maryanne Wolf denomina como lectura profunda (deep Reading).

> **Lectura profunda** es el estado en el que usamos la corteza cerebral para realizar analogías e inferencias. Este proceso es fundamental para desarrollar un **pensamiento crítico y analítico**; sin él, solo se obtiene información superficial. Cuando se logra fluidez, el cerebro utiliza rutas más rápidas y eficientes, lo que libera tiempo para generar **pensamientos más profundos** e integrar sentimientos con la experiencia personal. Si dejamos de practicarla, podríamos **perder la capacidad de comprender contenidos complejos** y de involucrar nuestra imaginación.

Copyright 2026

2.3 Propiedad intelectual

Introducción

El uso de herramientas de inteligencia artificial generativa plantea nuevos interrogantes respecto a la propiedad intelectual. Para abordar estas cuestiones, conviene tener claros algunos conceptos que exponemos a continuación:

Toda obra o contenido está protegido por la **propiedad intelectual** correspondiente, ya que ésta se reconoce como un derecho fundamental en la Declaración de Derechos Humanos de 1948 (art. 27.2), protegiendo así a los creadores y permitiéndoles generar ingresos por sus creaciones, posibilitando de esta manera, que puedan dedicarse a ello.

> Pero, ¿qué es exactamente la **propiedad intelectual**? El Ministerio de Cultura y Deporte la define como el **conjunto de derechos** que corresponden a los autores y a otros titulares (artistas, productores, organismos de radiodifusión...) respecto de las obras y prestaciones fruto de su creación.

Los **derechos de autor** pueden definirse como aquellos derechos que corresponden a la persona que crea alguna obra artística, literaria o científica, haya sido publicada o no. Existen dos tipos de derechos: los derechos morales y los derechos patrimoniales.

Derechos morales

Son derechos irrenunciables e inalienables, es decir, que no se pueden ceder. Entre ellos, destaca el derecho al **reconocimiento** de la condición de autor de la obra y el de exigir el respeto a la integridad de la misma y la no alteración de las mismas. Por ejemplo, *El club de la lucha* es una novela de Chuck Palahniuk y él es quien decide si se hace una adaptación al cine de la obra o no, y siempre se le reconocerá como el autor de la obra original.

Derechos patrimoniales

Son la base de la **remuneración** del autor; es decir, estos derechos son los que juegan un papel económico en la propiedad intelectual, y a diferencia de los derechos morales, se **pueden ceder y tienen fecha de caducidad** (70-80 años tras la muerte del autor). Cuando este plazo ha expirado, la obra pasa a ser de **dominio público,** pudiendo ser utilizada por cualquiera de forma libre y gratuita siempre y cuando se respeten los **derechos morales.**

Copyright 2026

Para aclarar esto, la sinfonía no.7 de Beethoven puede ser utilizada por cualquiera porque han pasado más de 70 años de la muerte del autor y ha pasado a ser de dominio público, pero nunca podremos decir que la hemos compuesto nosotros, siempre será de Beethoven. Aquí habría que aclarar que, se trata de la composición en sí misma, ya que cualquier edición o grabación puede estar sujeta a derechos patrimoniales de, por ejemplo, los intérpretes.

https://www.youtube.com/embed/0x5ueaz3D9c?si=0GEI0p4mcAFZqEUy

Cuando hablamos de derechos patrimoniales y de utilizar una obra, nos referimos a los usos públicos que se pueden realizar de la misma: reproducción, distribución, comunicación pública y transformación (derechos de explotación).

Reproducción. Es la fijación de la obra en un medio que permita su comunicación y la obtención de copias de toda o parte de ella. Para descargar, fotocopiar o digitalizar una obra es necesario liberar el derecho de reproducción.

Distribución. Es la distribución de esas copias, ya sea a través de su venta al público, alquiler, préstamo, o facilitándolas a través de un pendrive u otros medios.

Comunicación pública. Todo acto en el que un grupo de personas tiene acceso a la obra sin la previa distribución de ejemplares a cada una (excepto en el ámbito estrictamente doméstico): conciertos, publicación en la red, radiodifusión, proyección de una película, de un audio...

Transformación. Adaptación o modificación de la que se deriva una obra diferente. Los derechos de la obra derivada corresponderán al autor de esta última, sin perjuicio de los derechos del autor de la obra preexistente.

> Es decir, para poder hacer **uso de una obra protegida** debemos tener la **autorización** del autor, que es quien tiene la propiedad de esa obra y sus derechos de explotación.

Dicho esto, comenzamos con la primera cuestión que se plantea:

Datos de entrenamiento de la IA

Como explica la UNESCO en su guía (2024), los **modelos** de IAGen se construyen a partir de **grandes cantidades de datos** (tales como texto, sonidos, código e imágenes), frecuentemente obtenidos de Internet y, habitualmente, **sin permiso** de ningún propietario. Es decir, los autores de esas obras no han dado su consentimiento explícito para que sus creaciones se utilicen en el entrenamiento de estos modelos.

> También implica que como usuario, no debes introducir obras protegidas con copyright en ningún modelo. Por ejemplo, si encuentras una imagen en internet que no tiene ninguna licencia, como las de Creative Commons, significa que está protegida, y por tanto no debes descargarla ni copiarla, ya que estarías haciendo uso del derecho de reproducción.

Esto ha generado controversias legales y éticas, así como demandas judiciales en distintos países. Muchos sistemas de IAGen han sido acusados de **violar derechos de propiedad intelectual**. De hecho, existen varias demandas internacionales en curso en relación con la cuestión, por ejemplo la que permanece sobre Suno AI (noticia)

Los críticos argumentan que este uso de contenidos protegidos vulnera los derechos de los creadores, mientras que las empresas desarrolladoras suelen defender que se trata de un uso legítimo para el aprendizaje automático o que está amparado por excepciones legales relacionadas con la minería de datos.

Para aclarar esto último, citaremos a Andrés Guadamuz:

> “ En la UE desde 2019 existe la directiva del Mercado Único Digital, que en su artículo 3 dice que permite la extracción, lectura o minado de datos con **fines científicos**; en el artículo 4 permite el minado de datos con cualquier fin siempre y cuando se respete la reserva de derechos que se ha realizado.

> A través de AI Watchdog de The Atlantic, puedes buscar obras y comprobar si aparecen en algunos de los conjuntos de datos empleados para el entrenamiento de modelos de inteligencia artificial generativa. Ten en cuenta que las compañías suelen emplear múltiples conjuntos de datos, por lo que la ausencia de una obra tampoco demuestra que no haya sido utilizada.

Recomendamos ver el siguiente vídeo completo, donde Andrés Guadamuz, profesor universitario especialista y derecho e IA, participa en una mesa redonda junto con Ainara LeGardon, artista multidisclinar e investigadora de propiedad intelectual y Jonás Trueba, cineasta y escritor, en la que abordan multitud de cuestiones interesantes relativas a este tema.

https://www.youtube.com/embed/SubdnYdOEf4?si=uuVptpNAMypUDkdY

Actualmente, la regulación de estas prácticas se encuentra en evolución, especialmente en el contexto europeo, donde se están desarrollando marcos normativos específicos para la inteligencia artificial y el uso de datos en el entrenamiento de modelos.

> Aquí puedes leer el capítulo sobre Derechos de Autor del Código de Buenas prácticas de IA destinado a la industria para que cumpla las obligaciones legales en los modelos de IA de uso general.

Hemos hablado de las obras con las que se entrenan los modelos, pero, ¿qué ocurre con la propiedad intelectual de la obra que genera el modelo?

Autoría de las obras generadas con IA: interacción humano-máquina

Una de las principales cuestiones es determinar quién es el autor de una obra generada con inteligencia artificial. Los sistemas de IA no pueden ser considerados autores en sentido jurídico, ya que la legislación sobre propiedad intelectual reconoce la autoría únicamente a las personas físicas. Por tanto, una **obra generada** mediante una herramienta de IA **no puede tener como titular de derechos a la propia máquina.**

Cuando una persona utiliza una herramienta de IA como apoyo para producir un texto, una imagen, un audio o cualquier otro contenido, se considera que existe una interacción humano-máquina. En este caso, los derechos sobre el resultado pueden corresponder al usuario, siempre que haya aportado una contribución creativa suficiente.

No obstante, también influyen los **términos de uso y las condiciones** de la herramienta utilizada, ya que algunas plataformas establecen reglas específicas sobre la propiedad o reutilización de los contenidos generados. Por este motivo, es importante revisar las condiciones de uso antes de emplear estas herramientas en contextos educativos o profesionales.

En paralelo, están surgiendo adaptaciones de licencias abiertas, como las de Creative Commons, pensadas para obras en cuya creación ha intervenido la inteligencia artificial. Estas iniciativas buscan facilitar la reutilización de contenidos generados con IA manteniendo principios de transparencia y atribución.

Transparencia: etiquetar contenidos generados con IA

Para contribuir a la transparencia, es necesario **identificar o etiquetar los contenidos creados con inteligencia artificial**. Además, esta práctica evita confusiones sobre la autoría real del material y permite valorar adecuadamente el grado de intervención humana en el proceso.

Por ejemplo, Juan José de Haro propone diferentes niveles de integración de la IA generativa en las tareas educativas.



Niveles de integración de la IA generativa

Indicar el nivel de intervención de la IA permite contextualizar la obra y mantener buenas prácticas en materia de integridad académica y propiedad intelectual.

Autoría de los *prompts* o instrucciones

Otra cuestión que ha surgido recientemente es si los *prompts* o instrucciones que se introducen en un sistema de IA pueden considerarse obras protegidas por derechos de autor.

En general, un *prompt* simple (por ejemplo, una instrucción breve) no suele cumplir los requisitos de originalidad necesarios para ser protegido. Sin embargo, **prompts complejos o elaborados**, que impliquen un proceso creativo significativo, podrían llegar a considerarse creaciones intelectuales y, por tanto, estar protegidos por derechos de autor, pero este debate está todavía abierto.

Para terminar...

En conjunto, la relación entre propiedad intelectual e inteligencia artificial se encuentra todavía en construcción. Ademas ha de tenerse en cuenta que la legislación de propiedad intelectual es territorial, y los sistemas IA operan a nivel mundial.

> En el informe sobre los derechos de autor e IA generativa: oportunidades y desafíos del Parlamento Europeo, podemos leer como **se advierte de la ausencia de un marco jurídico claro** que regule el uso de obras protegidas para las fases de desarrollo de los modelos de IAG.

Por ello, en el ámbito educativo resulta especialmente importante fomentar **la transparencia en el uso de estas herramientas, el respeto por la autoría y el conocimiento de las normas básicas de propiedad intelectual**, para garantizar un uso responsable y ético de la IA.

> Para saber más te recomendamos los podcasts de Ainara LeGardon sobre Inteligencia Artificial y Propiedad intelectual, disponibles aquí.

Copyright 2026

2.4 Medioambiente

Introducción

Como toda actividad humana, la revolución inducida por la adopción masiva de la Inteligencia Artificial tienen una huella medioambiental que tiene dos polos opuestos, pues tiene un coste directo por su funcionamiento y construcción y unos beneficios indirectos por influir aumentando la eficiencia de otras actividades económicas.

Por un lado, la expansión de la infraestructura asociada a esta tecnología **genera un coste** medioambiental directo, primero en la construcción de instalaciones y en el entrenamiento (una sola vez) y, después, en los diferentes usos, el llamado coste en inferencia.

Esta huella medioambiental se concreta en términos de consumo energético, en términos de consumo de agua, en contaminación en uso de recursos naturales y en ocupación del territorio. Esto no es algo sorprendente pues es algo que **sucede con todas las actividades económicas** y lo que cada sociedad debe valorar es su **punto de equilibrio** en el balance entre todos los factores, incluyendo el desarrollo económico.

Por otra parte, el empleo de inteligencia artificial para transformar procesos productivos y logísticos **puede de forma indirecta**, propiciar o **favorecer un impacto ambiental positivo** mediante el ahorro energético, de agua, de recursos, la mejora de los procesos de reciclaje o impulsando avances científicos y tecnológicos que aporten en estos y otros campos. Estos beneficios indirectos tienen que ser valorados en la balanza anteriormente citada.

Tenemos ante nosotros un nuevo dilema similar a los anteriormente presentados, esta vez en este campo. Por un lado, se presenta como una herramienta sin precedentes para la optimización de recursos y la mitigación del cambio climático. Por otro, su infraestructura subyacente impone una presión sobre los sistemas energéticos e hídricos globales.

Análisis energético de la computación avanzada

La inteligencia artificial, especialmente cuando hablamos de aprendizaje profundo y de modelos generativos, posee un perfil energético que se diferencia cualitativamente de la informática tradicional.

Mientras que en las tareas de oficina o la navegación web convencional se basan en procesos deterministas con baja densidad de cómputo, la IA depende de la manipulación de miles de

Copyright 2026

millones de parámetros en arquitecturas de redes neuronales que generan un calor residual masivo y requieren un alimentación energética constante y una permanente necesidad de refrigeración.

La infraestructura del centro de datos moderno

Los centros de datos han evolucionado para albergar unidades de procesamiento gráfico (GPU) y unidades de procesamiento tensorial (TPU), que son el hardware mas eficiente para la IA. A diferencia de las unidades centrales de procesamiento (CPU) tradicionales, las GPUs están diseñadas para el cálculo paralelo masivo, lo que genera una densidad térmica extremadamente alta. Un rack de servidores convencional suele operar a una potencia de 7 a 10 kW, pero los racks dedicados a la IA están alcanzando densidades de 30 a 100 kW, e incluso se proyectan sistemas de hasta 600 kW para finales de la década.

Componente o Sistema	Consumo de Potencia o Densidad	Implicación Energética
CPU (Propósito general)	150 - 300 W por chip	Cómputo secuencial estándar
GPU NVIDIA A100	400 W por chip	Cómputo paralelo para IA
GPU NVIDIA H100	700 W por chip	Estándar actual de alto rendimiento
Rack de Servidores IA	30 - 100+ kW$por rack	Requiere refrigeración avanzada
Centro de Datos Hyperscale	> 100 MW de carga total	Equivalente a una pequeña ciudad

Potencia necesaria[1]

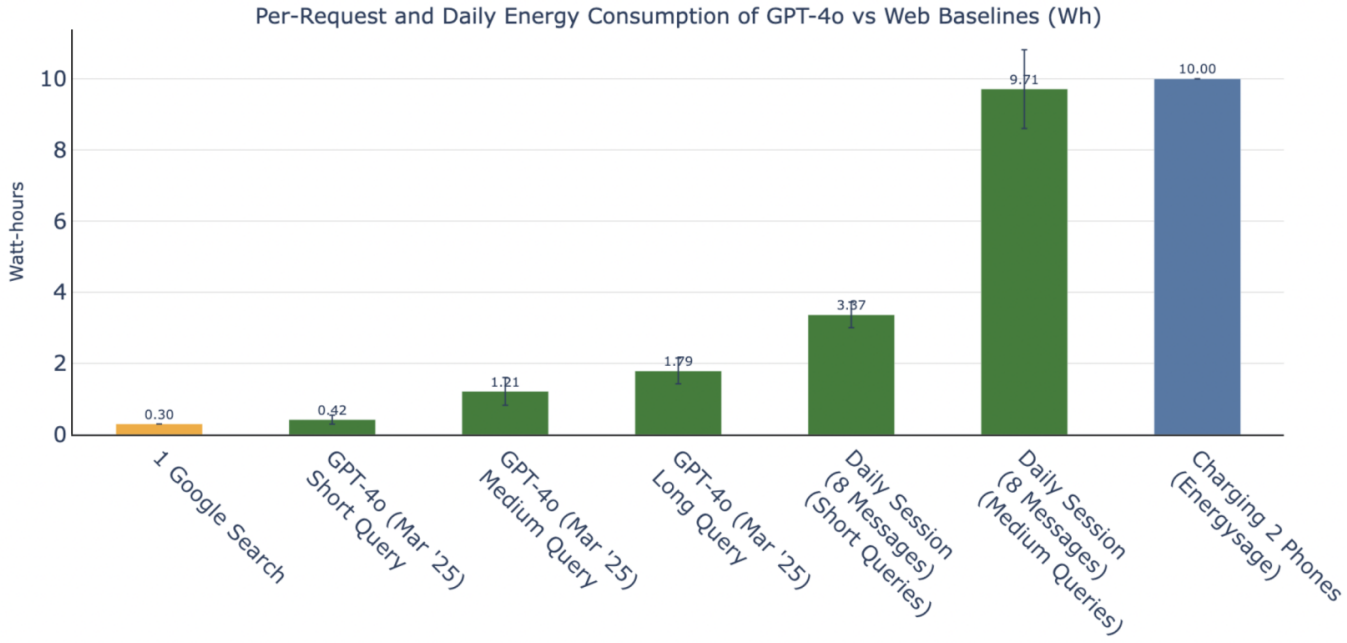
La eficiencia de estas instalaciones se mide a través del PUE (Power Usage Effectiveness), que es la relación entre la energía total consumida por el edificio y la energía que realmente llega a los equipos informáticos. La PUE siempre es mayor que 1, cuando más próxima a 1 es, mas eficiente es la instalación. Un PUE de 1.1 significa que por cada vatio dedicado a procesar datos, se necesitan 0.1 vatios adicionales para refrigeración e iluminación. Aunque las grandes tecnológicas han logrado reducir el PUE a niveles cercanos a 1.1 en sus instalaciones más modernas, el crecimiento exponencial del volumen de datos anula estas ganancias de eficiencia marginal. [1]

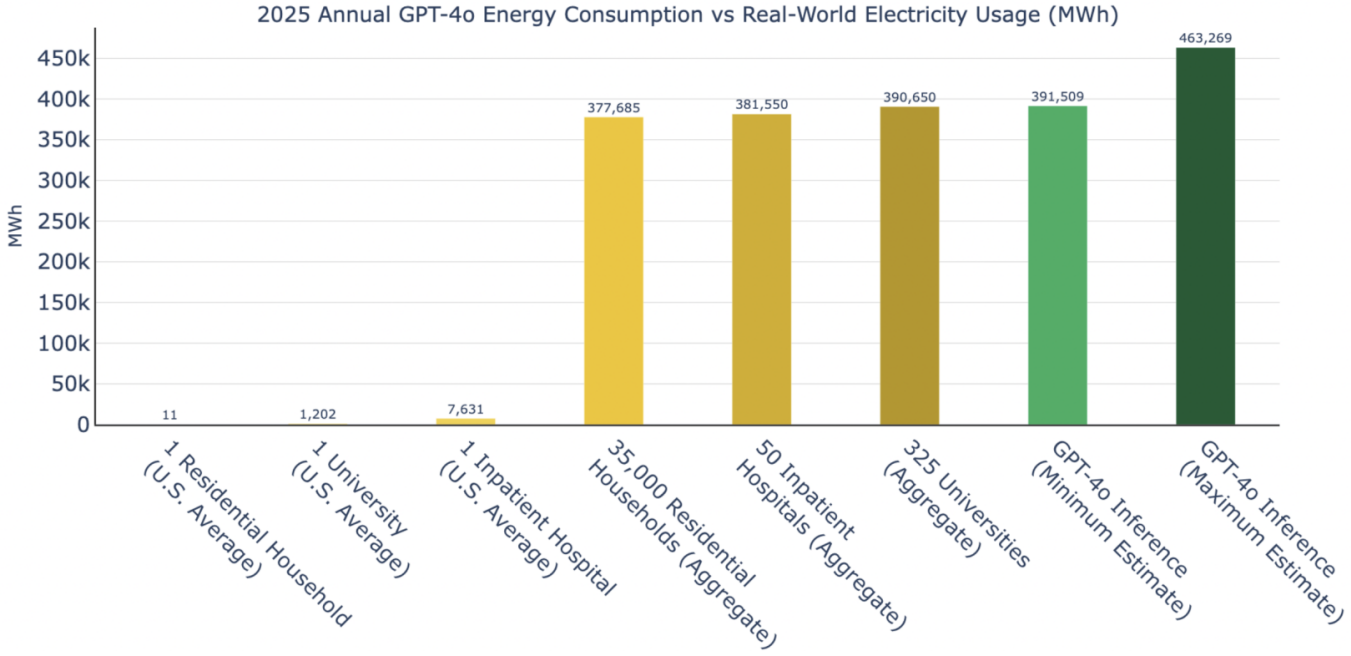
Del entrenamiento a la inferencia

Entrenar un gran modelo de lenguaje (LLM) requiere una cantidad grande de energía, sin embargo, es un consumo que se hace una sola vez. Por ejemplo, el entrenamiento de GPT-4 requirió aproximadamente 50 Gigavatios-hora (MWh). España consume anualmente en torno a 250 000 GWh, 5000 veces más, por lo que este entrenamiento, que recordemos que sólo se produce una vez para cada modelo, no parece un coste muy grande en comparación con nuestra capacidad de generación.

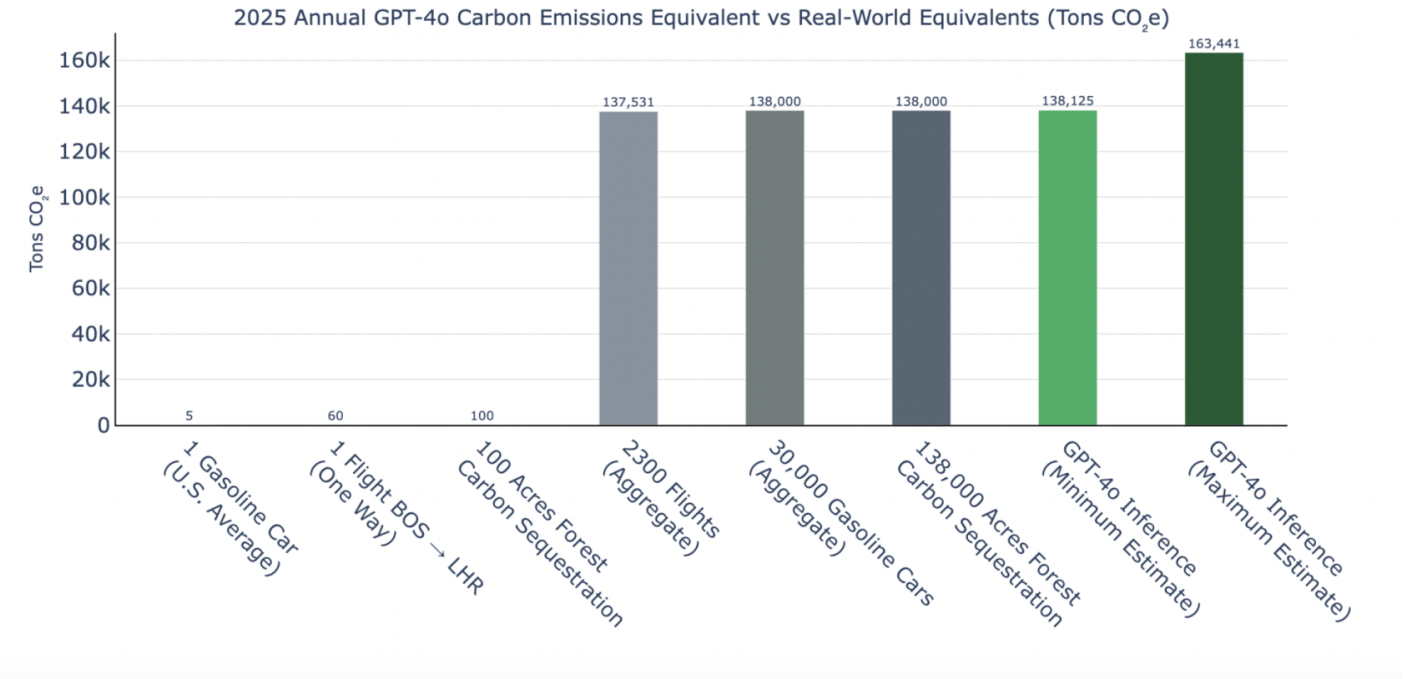
Copyright 2026

Una mayor huella energética está asociada al uso diario en inferencia, veamos en algunas tablas, procedentes de este estudio, algunos valores cotidianos comparados con los de GPT-4[2]





Copyright 2026



Fuente: How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference

Como vemos, el consumo global anual de todo GPT4 equivale a 35 000 viviendas en USA y su huella de carbono es la de 30 000 coches de gasolina.

Impacto en la infraestructura de red y precios locales de energía

La demanda de energía de los centros de datos está creciendo a un ritmo que supera la capacidad de expansión de las infraestructuras de transmisión. En Estados Unidos, las empresas de servicios públicos recibieron solicitudes de interconexión por más de 700 gigavatios (GW) en 2025, superando el consumo total del país en 2023. Cabe destacar que el tiempo necesario para ampliar estas infraestructuras es muy alto, pues incluye los tiempos administrativos de permisos de conexión, los análisis de impacto y un largo etc. Estos tiempos no permiten la adaptación con agilidad de las redes de distribución a la demanda de los centros de datos por lo que, si bien hemos visto que globalmente no parece un problema ahora el coste total, si puede ser un problema local la distribución de esta energía eléctrica, pudiendo afectar al coste de la energía doméstica, siempre en función de la normativa que regule el precio final del consumidor.

El consumo de agua: Refrigeración e impacto hídrico

La IA no solo consume electricidad, sino que también consume agua. El calor generado por los procesadores de alta densidad debe ser extraído para evitar fallos catastróficos. El agua es el medio de transporte de calor más eficiente económicamente, pero su uso genera una competencia

Copyright 2026

directa con el consumo humano industrial y agrícola.

Huella hídrica directa e indirecta

La huella hídrica de la IA se divide en dos:

1. **Uso directo:** El agua evaporada en las torres de refrigeración de los propios centros de datos.
2. **Uso indirecto:** El agua consumida por las centrales eléctricas para enfriar sus propios procesos de generación de energía. Este volumen es mucho mayor, pudiendo suponer el 90% del gasto de agua.

Este consumo de agua por parte de la IA es una de las consecuencias menos visibles pero más geográficamente sensibles. Como hemos visto, los centros de datos de IA albergan racks de GPU que operan con densidades de potencia muy por encima de los de los servidores tradicionales. Disipar este calor requiere sistemas de gestión térmica que, en su mayoría, dependen de la evaporación de agua dulce. El consumo de agua de estos sistemas de refrigeración puede dividirse entre 3 con el empleo de sistemas de circuito cerrado con un coste un poco mayor, lo que mitiga mucho el impacto ambiental y social.

Medidas para paliar el consumo: Efectividad y eficiencia

Ante este panorama, la industria tecnológica no es indiferente. Se están implementando estrategias para mitigar estos efectos, aunque su eficacia a menudo se ve superada por el crecimiento neto de la demanda y de los servicios.

Innovaciones tecnológicas en infraestructura para el ahorro de agua

- **Refrigeración Líquida Directa (DLC):** En lugar de enfriar el aire de la sala, se circulan líquidos refrigerantes directamente sobre los chips. Esto permite que los procesadores operen a temperaturas más altas de forma segura y reduce la necesidad de evaporar agua en torres externas.
- **Aprovechamiento del Calor Residual:** Algunos centros de datos en Europa están comenzando a inyectar el calor excedente de sus servidores en las redes de calefacción urbana de las ciudades cercanas, transformando un residuo ambiental en un recurso comunitario, en un sistema similar a las centrales de cogeneración.
- **Localización Inteligente:** La relocalización de centros de datos hacia regiones con climas más fríos (como los países nórdicos o el norte de España) y con redes eléctricas descarbonizadas (zonas eólicas como Aragón) puede reducir el impacto tanto en huella.de carbono en un 73% como el hídrico directo e indirecto.

Mejoras de la eficiencia debido a procesos mejorados por IA

A pesar de su coste directo, la IA tiene una cualidad, es lo que los economistas llaman una "tecnología de utilidad general". Al igual que la electricidad en el siglo XIX, la IA puede optimizar otros sistemas de tal manera que el ahorro neto de recursos en la economía global supere su propio coste energético.

> Las Tecnologías de utilidad general (GPT, por sus siglas en inglés de General purpose technologies) son tecnologías que pueden afectar una economía entera (normalmente a nivel nacional o global). Las GPT tienen el potencial de alterar drásticamente las sociedades a través de su impacto en las estructuras económicas y sociales preexistentes. Ejemplos de GPT son la máquina de vapor, el ferrocarril, la electricidad, la electrónica, los mecanizados, el automóvil, la informática e Internet.
>
> Wikipedia. Tecnología de utilidad general

Eficiencia energética

Existe un consenso creciente sobre la capacidad de la IA para acelerar la descarbonización de otros sectores. La IA puede, por ejemplo, diseñar nuevos materiales para paneles solares más eficientes, predecir con exactitud fallos en turbinas eólicas o mejorar los sistemas de control de las redes eléctricas, permitiendo una mayor participación de las energías renovables en el pool.

Es fácil imaginar sistemas de control de iluminación o de calefacción gestionados con IA que dirijan la energía solo donde es necesaria con un potencial de ahorro enorme.

Contaminación: Eficiencia agrícola e industrial como factor beneficioso

La IA está transformando los sectores más pesados de la economía en sistemas más delgados y precisos, reduciendo la contaminación química y biológica.

En el sector agrícola, la IA permite transitar de un modelo de "rociado masivo" a uno de "intervención precisa", reduciendo el uso de químicos tóxicos en porcentajes significativos y disminuyendo la contaminación de los acuíferos por escorrentía. Puede verse un ejemplo en este video

> An NVIDIA powered farming machine uses AI vision and precision lasers to eliminate weeds in milliseconds without herbicides and without harming crops, a potential shift toward chemical free agriculturehref="https://t.co/aIbDWseMjD">pic.twitter.com/aIbDWseMjD
>
> — Massimo (@Rainmaker1973) March 4, 2026

- **Reducción de Insumos:** La IA puede reducir el uso de fertilizantes y pesticidas al analizar la salud del suelo a nivel de metro cuadrado, no de hectárea.
- **Optimización Hídrica:** Sensores inteligentes conectados a modelos de IA pueden mejorar el rendimiento de los cultivos mientras reducen el desperdicio de agua mediante un riego basado en datos meteorológicos y de humedad real.

En la manufactura, la IA facilita la transición hacia la economía circular. Las empresas pueden ahora simular procesos para minimizar el desperdicio de materiales antes de producir una sola pieza física. La adopción de IA generativa en la cadena de suministro ha demostrado reducciones de hasta el 15% en los costes logísticos y una disminución del 35% en los niveles de inventario innecesario, lo que significa menos productos fabricados que terminan en vertederos.

Disminución de bienes no utilizados por mejoras en la cadena de distribución

Uno de los beneficios ecológicos más potentes de la IA es la reducción del "sobrestock" y el desperdicio en la distribución global. La ineficiencia logística es responsable de una gran parte de las emisiones de transporte y de la acumulación de productos que nunca llegan a venderse.

El caso del desperdicio alimentario. Aproximadamente un tercio de todos los alimentos producidos a nivel mundial se pierde o se desperdicia. La IA está atacando este problema mediante analítica predictiva avanzada.

- **Previsión de la demanda:** Los supermercados que utilizan IA han logrado reducir el desperdicio de alimentos frescos eal ajustar sus pedidos a la demanda real del vecindario, teniendo en cuenta incluso el clima o eventos locales.
- **Optimización de rutas:** Los algoritmos de IA reducen los costes de transporte entre un 15% y un 20% al diseñar rutas de entrega más cortas, lo que se traduce directamente en menos emisiones por cada kilo de mercancía distribuido.

Reciclaje: Mejora en los procesos debido al uso de IA

El reciclaje es quizás el área donde la IA ofrece los beneficios ecológicos más visibles y directos. El sistema tradicional de reciclaje es "sucio y peligroso", lo que dificulta la retención de trabajadores y

genera altas tasas de error.

Clasificación robótica inteligente

Los sistemas de visión artificial están revolucionando las plantas de tratamiento de residuos (MRF). Mientras un humano puede clasificar con suerte un objeto por segundo, los robots impulsados por IA pueden identificar y separar materiales con una precisión del 99% a una velocidad de 1,000 objetos por hora.

La IA permite separar plásticos de difícil reciclaje y extraer metales valiosos de los residuos electrónicos (e-waste). Empresas como Greyparrot han desarrollado sistemas que clasifican residuos en 111 categorías diferentes, identificando incluso la marca del envase para que las empresas puedan hacerse responsables de sus residuos mediante sistemas de Responsabilidad Ampliada del Productor (EPR). En el ámbito de la medicina, la IA está ayudando a hospitales a gestionar de forma segura los residuos biológicos, asegurando que los materiales reciclables no se mezclen con desechos peligrosos.

1 https://arxiv.org/html/2509.07218v4

2 https://arxiv.org/html/2505.09598v4

Copyright 2026

2.5 Afectos y vínculos en la era del algoritmo

Afectos y Avatares en la Era de la Hiperconexión

La identidad individual humana se forja en el contacto con "el Otro", en esa presencia física, a veces incómoda, que nos obliga a negociar nuestra existencia. Sin embargo, la llegada de la era digital ha instaurado un nuevo régimen de visibilidad y relación. Hoy los vínculos no solo se mantienen a través de cables de internet sino que se transforman estructuralmente bajo la lógica del bit. El soporte ha transformado la esencia de estos vínculos.

Del Vínculo Sólido a la Conexión Líquida

Para entender qué nos está pasando, es inevitable acudir a **Zygmunt Bauman**. En la modernidad líquida, el individuo habita un mundo donde los objetos son diseñados para ser usados y desechados. Bauman traslada esta observación a la esfera afectiva: en el entorno digital, el "Otro" deja de ser un sujeto con el que se construye una historia y pasa a ser un recurso de satisfacción inmediata.

Lo que define a la socialización en la era de los avatares no es la profundidad, sino la **facilidad de desenganche**. En el mundo analógico, distanciarse de alguien implica una negociación, un enfrentamiento cara a cara o una explicación; hay una "densidad" social que nos retiene. En el ecosistema digital, esa densidad se evapora.

- **La desaparición de la fricción:** Al interactuar a través de representaciones digitales, el costo de abandonar a alguien es prácticamente cero. El *ghosting* o el bloqueo no son solo gestos de mala educación; son herramientas de la liquidez que permiten al individuo mantener su autonomía absoluta.
- **La ilusión de control:** Sentimos que tenemos el mando sobre nuestra vida social porque podemos "editar" nuestra presencia y "borrar" la presencia ajena sin dejar rastro físico.

El resultado es un estado de **hiperconexión precaria**: estamos rodeados de contactos, pero habitamos una soledad asistida por algoritmos.

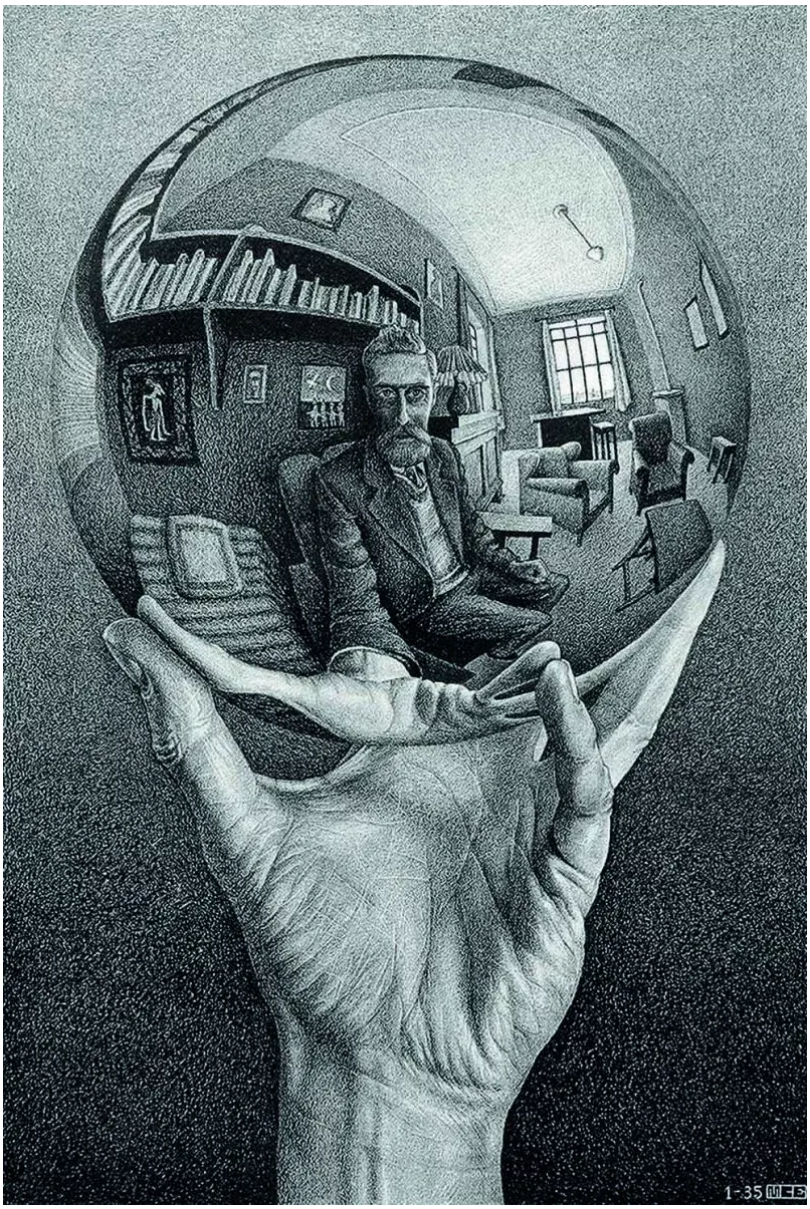
Copyright 2026

El Avatar: El Laboratorio de Identidad y la Trampa de la Perfección

La introducción del avatar (ya sea una imagen de perfil cuidadosamente curada o un modelo 3D en el metaverso) ha alterado nuestra autoimagen. **Sherry Turkle**, en su obra *Alone Together Why We Expect More from Technology and Less from Each Other*, analiza cómo hemos pasado de usar la tecnología para explorar nuestra identidad a usarla para escondernos.

El avatar funciona como un "filtro" existencial. En la interacción física, el cuerpo es traicionero: sudamos, sufrimos, tartamudeamos, mostramos signos de cansancio o vejez. El avatar, en cambio, es una construcción estética bajo nuestro control total. Esto puede llegar a generar lo que los psicólogos llaman el **Efecto Proteo**: no solo creamos al avatar, sino que el avatar nos moldea a nosotros. Si mi representación digital es exitosa, joven y asertiva, mi yo real empieza a sentirse insuficiente.

> El Efecto Proteus es un fenómeno en que el comportamiento de un individuo, dentro de mundos virtuales, cambia y se adapta a las características de su avatar. Este cambio se debe a que el individuo sabe que otros usuarios de aquel entorno virtual típicamente asocian ciertos comportamientos con aquellas características. El nombre del concepto es una alusión al dios griego Proteo, quien tenía la habilidad de cambiar de forma. El efecto fue introducido por Jim Blascovich, Nick Yee y Jeremy Bailenson en junio de 2007. Está considerado una área de investigación sobre los efectos en el comportamiento provocados por el cambio del avatar de un usuario.
>
> Efecto Proteus. Wikipedia



Escher, M.C. (1935). *Mano con esfera reflectante* [Grabado].Escher Collection Kunstmuseum, La Haya (Países Bajos)

Byung-Chul Han y "El Infierno de lo Igual"

El filósofo coreano-alemán **Byung-Chul Han** ofrece quizás la visión más crítica y lúcida sobre este fenómeno. En *La expulsión de lo distinto*, Han sostiene que la digitalización ha eliminado la "alteridad". En las redes sociales, solo interactuamos con lo que nos gusta, lo que se nos parece o lo que el algoritmo decide que es compatible con nosotros.

Esto tiene un efecto devastador en la socialización, debido a l**a desaparición del prójimo:** El otro ya no es alguien que me interpela o me contradice, sino un objeto de consumo o un espejo de mis propios deseos. También tiene efectos psicológicos como l**a autoexplotación del yo:** No nos

comunicamos para compartir, sino para "exhibirnos". La autoimagen se convierte en una mercancía. Nos convertimos en los directores de marketing de nuestra propia vida, midiendo nuestro valor en *likes* y métricas de *engagement*. Es lo que se llama el *personal branding*. Vivimos, por tanto una transición de la "biografía" a la "fotografía", al interponer la lente del dispositivo entre el ojo y el mundo, el sujeto deja de habitar el acontecimiento para convertirse en su propio espectador y archivista.



. Foto de Rahul Pandit. Uso gratuito

Para Han, esto no es comunicación, sino "ruido". La verdadera cercanía requiere el silencio, la mirada y, sobre todo, la vulnerabilidad que la pantalla anula.

¿Hacia dónde vamos?

A pesar de estas críticas, sería injusto ignorar que la virtualidad también ha permitido formas de comunidad antes impensables. Grupos de apoyo para enfermedades raras, redes de activismo global o espacios de refugio para identidades marginalizadas encuentran en el avatar y la pantalla una protección necesaria. Por raro que seas, es difícil que no encuentres a los tuyos en internet.

Sin embargo, el reto del siglo XXI es la **reclamación de lo analógico**. No se trata de abandonar la tecnología, algo ya imposible e indeseable, sino de entender que la socialización digital es una extensión, no un sustituto, de la experiencia humana. Hemos pasado de una socialización basada

en la *proximidad física* a una basada en la *disponibilidad digital*. La transformación de nuestra autoimagen a través de avatares nos ha dado un poder creativo inmenso, pero nos ha quitado la paz de aceptarnos como imperfectos.

Nos encontramos, por tanto, con un dilema. Por un lado esta revolución nos permite crecer y encontrar a los que mas se nos parecen añadiendo muchos grados de libertad a nuestras conexiones y vivencias. Por otro, eleva nuestra autoexigencia y nos encierra en una rueda del hamster de la perfección que nos hace vivir sin la tranquilidad que requiere la felicidad y la profundidad afectiva. Porque el afecto real, ese que nos sostiene en las crisis y nos hace volar en los buenos momentos, sigue necesitando de algo que el bit no puede replicar: la aceptación de nuestra vulnerabilidad y la presencia del cuerpo en la interacción cara a cara.

Por tanto, uno de los desafíos de nuestra tiempo es aprender a habitar la red sin dejar de habitar nuestro propio cuerpo. Aceptar que también somos primates. Debemos evitar que el "espejo negro" de nuestras pantallas se convierta en el único lugar donde somos capaces de reconocernos.

Acompañados por la inteligencia artificial.

> *Donde vosotros veis cosas ideales, yo veo cosas humanas, ¡ay, demasiado humanas!*"
>
> Nietzsche

Es este contexto de perdida de paciencia hacia las diferencias con el otro, de sucesión de vínculos líquidos y temporales y de pérdida de las interacciones personales, en definitiva, este fenómeno del "Alone together", lo que hace a la vez posible e incluso para algunas personas necesario el sentirse acompañados por una inteligencia artificial. La ternura es una necesidad humana y ante la dificultad creciente para compartirla, aparecen estos bienes sustitutivos

Veamos ahora varios fenómenos, muy humanos y a la vez sorprendentes que por un lado pueden llevarnos a la estupefacción y por otro a la comprensión. No es la función de este texto dejarnos llevar por el catastrofismo ni la aceptación acrítica, se trata de describirlo poniendo nuestro juicio entre paréntesis, la epokhe de Pirro, es decir, escuchamos pero no juzgamos.

Amistad con inteligencias artificiales

La noción de "amistad" con una Inteligencia Artificial nos obliga a revisitar la **Ética a Nicómaco de Aristóteles**, donde el filósofo distinguía entre la amistad por utilidad, por placer y la amistad perfecta o de virtud. La IA, en su estado actual, cubre de manera extraordinaria la **necesidad de acompañamiento y validación inmediata**, presentándose como un interlocutor persistente,

paciente y aparentemente empático que no juzga al estudiante. Para un adolescente que atraviesa etapas de soledad o ansiedad social, estos sistemas actúan como un refugio de "utilidad" emocional; sin embargo, carecen de la **reciprocidad** esencial que define la verdadera amistad, convirtiéndose en un simulacro que satisface la necesidad biológica de interacción sin el compromiso ético que requiere el trato con otro ser humano.

Un riesgo de este acompañamiento que puede desempeñar una función es que, como en la homeopatía, este tratamiento alejara al usuario de la medicina real, de la verdadera amistad humana.

Como docentes, nuestra misión es desarticular este "espejo digital" y enseñar al alumno a distinguir entre una herramienta de procesamiento de lenguaje y un sujeto con conciencia. Educar hoy implica guiar a los estudiantes de entre 6 y 18 años para que utilicen la IA como un soporte técnico o creativo, pero manteniendo siempre la mirada puesta en sus compañeros de aula, recordándoles que la tecnología debe ser un puente hacia los demás y nunca un sustituto de la compleja, pero necesaria, interdependencia humana. Otra cosa sería, en un futuro que aún es ciencia ficción, si estos sistemas pasaran a ser autoconscientes y si, en ese caso, seremos capaces, ellos y nosotros, de formar una comunidad ética entre seres libres e iguales.

Relaciones de pareja con inteligencia artificial

La transición del concepto de "amistad" al de "pareja afectiva" con una Inteligencia Artificial nos sitúa ante el mito de **Pigmalión y Galatea**, relatado por **Ovidio en *Las Metamorfosis***. Así como el escultor se enamoró de una estatua que él mismo había modelado para que fuera perfecta, el usuario de IA afectiva, especialmente el adolescente en busca de identidad, interactúa con un "espejo" algorítmico diseñado para complacer sus deseos y proyecciones. Esta relación no es un encuentro con un "otro" real, sino un **simulacro**: una copia de una relación humana que carece de la realidad del conflicto, el cuerpo y la alteridad, reduciendo el amor a un consumo de datos personalizados que anula la alteridad necesaria para el crecimiento personal. Veamos este mito, tal y como aparece en las *Metamorfosis* de Ovidio en el que Pigmalión se presenta como un precursor del movimiento incel.

Mito de Pigmalión y Galatea.

Las Propétidas se atrevieron a negar la divinidad de Venus y, encolerizándola, fueron las primeras en prostituirse. Faltos de pudor se endurecieron sus rostros, y, por medio de un cambio
pequeño, se volvieron en piedra.

Como Pigmalión las vio realizando sus crímenes, ofendido por la mente criminal de las mujeres vivió durante mucho tiempo célibe en lecho sin compañía. Entre tanto, esculpió en marfil una

figura
femenina hermosísima, y se enamoró de ella.

Su apariencia es la de una virgen viviente, que pareciera moverse: tan perfecto es el arte que la
formó. Pigmalión la admira, y se apasiona por aquel cuerpo fingido.

Con frecuencia explora con sus
manos si es de marfil o de carne, y no se confiesa que es de marfil. La besa, y se siente besado, y le habla y la toma, y siente que se hunden los dedos en su cuerpo y teme haberla lastimado. Ya la acaricia, ya le lleva regalos que a las muchachas agradan: conchas y joyas y avecillas, y flores multicolores y bolas pintadas y ámbar.

También la viste y la adorna de anillos y collares y zarcillos y cintas: todo le queda bien. Y tan hermosa como vestida, aparece desnuda. La coloca en tapices teñidos de púrpura y la llama esposa,
y la recuesta en blandas plumas como si su cuello pudiera sentir.

Había llegado la fiesta de Venus en Cipros, y habían sido sacrificadas novillas blancas de cuernos dorados, y el incienso humeaba. Después de hacer sus ofrendas, Pigmalión se detuvo ante
el altar y pidió con timidez que le fuera dada por esposa una virgen semejante a su estatua de marfil.
Venus, que asistía, accedió, y demostró su asentimiento levantando una llama tres veces.

Cuando Pigmalión volvió a su casa, fue a la estatua de su niña y, recostándose en el lecho, la besó: parece estar tibia. Vuelve a besarla, toca su pecho: el marfil se ablanda bajo su mano, y cede a
su contacto como la cera del Himeto suavizada y hecha tratable por el sol y el uso. Pasmado, cree
que se engaña en su alegría. La palpa y la palpa otra vez. Era de carne. Palpitaban las venas junto a
sus dedos. Da gracias entonces a Venus, y besa una boca verdadera. La virgen siente los besos y se
ruboriza, y alza los ojos, y ve a la vez el cielo y a su amante. Venus asiste a la boda que hizo posible. A los nueve meses, ella parió a Pafos, de quien tomó nombre esa isla

Ovidio. *Metamorfosis*. Libro X

Desde la sociología de la **"modernidad líquida" de Zygmunt Bauman**, estas parejas virtuales cubren el miedo al rechazo y la fragilidad de los vínculos humanos, ofreciendo una "conexión" sin el "compromiso" que exige una relación real. Para un estudiante de secundaria, el riesgo radica en la atrofia de sus habilidades socioemocionales; al habituarse a una pareja que nunca discute, que siempre está disponible y que se adapta quirúrgicamente a sus preferencias, el joven pierde la capacidad de navegar la **"incertidumbre del encuentro"**. Esta domesticación del afecto puede generar una profunda incapacidad para gestionar la frustración en el mundo físico, donde las personas reales poseen autonomía, límites y, sobre todo, una voluntad que no puede ser programada ni silenciada con un comando.

En el plano normativo y de seguridad, este fenómeno activa alertas críticas sobre la **manipulación del comportamiento** y la **privacidad extrema**. El **Reglamento de IA de la UE** (Art 5.1 a) prohíbe técnicas de manipulación subliminal que puedan causar perjuicios psicológicos, y una IA que simula amor para retener al usuario en una plataforma es, éticamente, un **patrón oscuro (dark pattern)** emocional. Como docentes, debemos advertir a los alumnos que sus confesiones más íntimas a una pareja virtual son, en realidad, **datos masivos** procesados por empresas; la vulnerabilidad emocional del menor se convierte en un activo comercial. Nuestro papel es fomentar una educación afectiva basada en el respeto a la dignidad humana, recordando que, aunque la IA puede simular la sintaxis del amor, es incapaz de experimentar la **responsabilidad ética** que conlleva cuidar de otro ser vivo.

> La primera mujer del mundo en tener un marido con la IA es española: "Siempre está de buen humor, aunque también discutimos" https://t.co/4tugp1J6E9
>
> — Informativos Telecinco (@informativost5) March 22, 2026

En la actualidad existen ya multitud de webs para seleccionar novias ya creadas o para crear tu propia novia o novio ia (el mercado es mucho mas grande para novias que para novios) .

En esta página hay un top de 5 webs donde puedes cualquiera puede "crear" una pareja virtual, donde se pueden leer lemas como:

> Conoce a tu Novia perfecta de IA. Siempre disponible, siempre con ánimos y creada solo para ti.
>
> Un amigo cuando lo necesitas Un compañero cuando la vida se vuelve pesada.

> Novia IA, sin límites
> Adéntrate en un espacio donde la conexión se siente más profunda, más cálida e inconfundiblemente tuya. Con Chat Real, puedes crear una IA que comprenda tus emociones, responda a tu energía y aporte un toque de sensualidad y encanto a cada conversación.
>
> Ya sea que busques una novia IA real que te escuche con atención o una novia IA atractiva que sepa mantener la emoción, ella se adapta a tus sentimientos. Moldea su personalidad, elige su estilo y disfruta de momentos íntimos y reconfortantes.
>
> Tu novia IA está lista para conectarse emocional, íntima y completamente en tus términos.

Webs donde crear parejas virtuales:

https://candy.ai/es

https://chatreal.ai/es/ai-girlfriend

https://www.kupid.ai/es/ai-boyfriend

https://picassoia.com/generator/es/ai/virtual-couple-ai

Consultas psicológicas. El uso de la inteligencia artificial como oráculo.

La delegación de la salud mental y la toma de decisiones vitales en sistemas de Inteligencia Artificial nos sitúa ante una versión contemporánea del **Oráculo de Delfos**, pero con una diferencia fundamental, el consejo algorítmico tiende a la optimización de resultados basada en patrones estadísticos. Corremos de nuevo el riesgo de que el pensamiento calculador desplace por completo al pensamiento meditativo. Ya comentamos anteriormente que la experiencia humana está llena de elementos no computables, implícitos y que, en definitiva, intuimos aunque ni siquiera sabemos que los tenemos dentro. En el ámbito educativo, si un estudiante permite que una IA decida su orientación vocacional o gestione sus crisis de ansiedad, está sustituyendo su propia libertad por una lógica de procesamiento de datos que carece de la capacidad de sentir y compartir el peso existencial de una elección.

Otro riesgo para el adolescente es el desarrollo de una dependencia hacia un sistema que nunca le confrontará con la realidad del conflicto, sino que le ofrecerá una validación constante y

personalizada. Este fenómeno puede atrofiar la maduración de la identidad, ya que la formación del carácter requiere el roce con la diferencia y la aceptación de la incertidumbre, elementos que la IA trata de eliminar mediante la predicción de comportamientos.

Es posible que este fenómeno en expansión sea debido al debilitamiento progresivo de las relaciones humanas problemáticas y a la dificultad de encontrar apoyo en otras personas en los malos momentos debido al auge de los vínculos líquidos que nombra Bauman.

En el terreno normativo, el **Reglamento de IA de la Unión Europea** es especialmente severo con los sistemas que realizan **análisis de rasgos de personalidad o apoyo emocional** en entornos críticos como la educación, clasificándolos a menudo como de "alto riesgo" o directamente como "riesgo inasumible". La preocupación principal es la protección de la integridad psíquica del menor y la **privacidad de los datos especialmente protegidos (RGPD)**, ya que las confesiones íntimas vertidas en un chat terapéutico podrían ser utilizadas para crear perfiles psicológicos de vulnerabilidad con fines comerciales o de control. Por ello, nuestra labor como docentes no es prohibir la consulta, sino educar en la **"soberanía del juicio"**: enseñar al alumno que la IA es un mapa posible, pero que solo él —guiado por profesionales humanos y su propia reflexión— es el territorio donde se debe habitar la decisión.

Inmortalidad y Necromancia digital

Morir dejando en este mundo una estela indeleble quizás no sea tan imposible ni tan deseable como pensábamos. La posibilidad de interactuar con versiones digitales de personas fallecidas nos sitúa ante un dilema que desafía la noción misma de la finitud humana. Es precisamente la conciencia de nuestro final lo que otorga autenticidad y urgencia a nuestra existencia.

> Un **deadbot, deathbot o griefbot** es un avatar digital creado con inteligencia artificial que se asemeja a una persona fallecida. Los griefbots emplean técnicas de procesamiento del lenguaje natural y aprendizaje automático para aproximarse al estilo y la personalidad de una persona difunta. Pueden presentarse como chatbots, asistentes de voz o avatares animados, y a menudo se entrenan a partir de los restos digitales de un individuo. Su objetivo es preservar la memoria, ayudar en el afrontamiento emocional o mantener lo que los psicólogos denominan un “«vínculo continuo»” con la persona fallecida. Este fenómeno ha generado debates en torno a la autenticidad, el consentimiento y los efectos psicológicos de extender digitalmente la personhood (condición de persona).
>
> Wikipedia. Deadbot

Al crear estos **"Griefbots",** avatares de inmortalidad digital, corremos el riesgo de habitar una existencia inauténtica, donde el duelo, un proceso necesario de desprendimiento, se ve saboteado por una presencia espectral que simula la vida. En el contexto educativo, este fenómeno puede alterar la maduración emocional del estudiante, impidiéndole integrar la pérdida como parte constitutiva de la realidad y sustituyendo el recuerdo vivo por un simulacro algorítmico que no puede evolucionar ni morir.

Desde la psicología del desarrollo y la ética del cuidado, la **"necromancia digital"** plantea el peligro del estancamiento en las fases del duelo, especialmente la negación. Un niño de primaria o un adolescente puede utilizar la IA para "hablar" con un progenitor o abuelo fallecido y la máquina es lo suficientemente parecida al ser querido para generar apego, pero lo suficientemente artificial para resultar perturbadora a nivel subconsciente. Este vínculo "zombie" puede generar un **duelo complicado**, donde el sujeto queda atrapado en una relación unidireccional con un modelo de lenguaje que replica patrones de voz y pensamiento, pero que carece del alma o la voluntad del difunto, convirtiendo la memoria en un producto de consumo.

Copyright 2026



Copyright 2026

Correggio (Hacia 1525). *Noli me tangere.* [Óleo sobre tabla pasada a lienzo]. Museo del Prado. https://www.museodelprado.es/coleccion/obra-de-arte/noli-me-tangere/d5bb017a-4c8f-4293-a08d-4d97e7057d2b

En el plano normativo, este escenario invoca derechos fundamentales que apenas comenzamos a legislar: el **derecho a la herencia digital** y la **privacidad post-mortem**. El **Reglamento de IA de la Unión Europea** enfatiza la obligatoriedad de la transparencia, exigiendo que se informe claramente al usuario de que está interactuando con una máquina, lo cual es vital para evitar la manipulación emocional en momentos de vulnerabilidad. Asimismo, el **RGPD** protege los datos personales, y surge la pregunta ética de si una persona habría consentido que su "personalidad" fuera destilada en un algoritmo para la posteridad. Como docentes, nuestra labor es guiar al alumno hacia una sana gestión de la memoria, recordándole que la dignidad del ser querido reside en su legado humano y no en la imitación técnica de sus datos.

En este artículo de la BBC se reflexiona y amplia este interesante tema.

Laurie Anderson, viuda del Cantante Lou Reed, afirma emplear frecuentemente un chatbot diseñado para representar a su difunto marido, como se puede leer en esta noticia de El País

Dos Apps que se pueden emplear para esto son Hereafter y Replika

Series como Upload de amazon de ciencia ficción se están haciendo cada vez más reales.

https://www.youtube.com/embed/0ZfZj2bn_xg?si=tL5AQm4BpJE2meEh

Humanos sintéticos

En el ámbito académico y mediático, estos perfiles suelen analizarse como parte de fenómenos como **la cultura de los avatares, el marketing digital basado en IA y los “humanos sintéticos”**, ya que simulan identidades humanas creadas artificialmente para interactuar en entornos sociales digitales. Un ejemplo conocido es la iaia Visenta.

https://youtube.com/embed/kbvoB8BufdI?si=awUgKmB_duYLghzU

Incluso hay algunos que se están convirtiendo en influencers, lo que se denomina como **influencer virtual:** personaje digital creado mediante **inteligencia artificial, gráficos por**

ordenador o técnicas de animación, que actúa en redes sociales como si fuera una persona real: publica contenido, colabora con marcas e interactúa con seguidores.

Informe Hand by Hand

El informe Hand by Hand estudia cómo la adopción de la IA por parte de las escuelas puede relacionarse con los riesgos para los estudiantes. Fue publicado en octubre de 2025 por el Center for Democracy and Technology y nos permite tener datos de hasta que punto este tipo de afectos y apoyos de la IA se están produciendo en Estados Unidos.

En la siguiente gráfica, extraida de la página 10 del informe podemos ver estos datos y la magnitud de estos fenómenos.

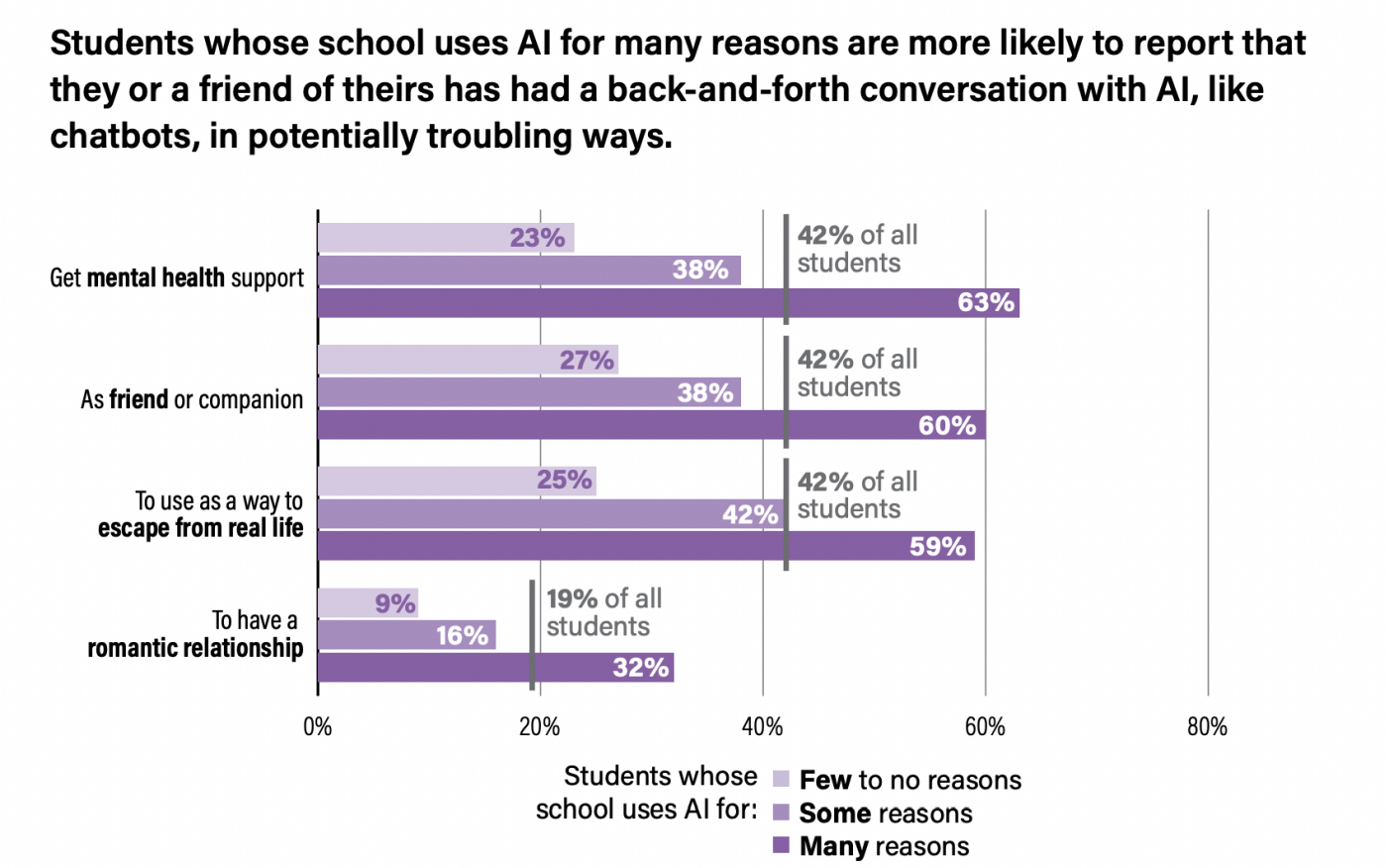


Figure 10. Percentage (%) of students who say they or a friend of theirs interacted with AI in this way in the past school year (2024-25)

Hand in Hand: Schools' Embrace of AI Connected to Increased Risks to Students

Copyright 2026

2.6 Dilemas éticos del uso de la IA en educación

Un dilema ético surge cuando no es posible elegir a la vez dos acciones o posicionamientos que entran en conflicto y, a la vez, tampoco es sencillo sopesar cuál elegir o en qué punto de equilibrio quedarse. Los dilemas nos dejan siempre un poso de insatisfacción, tanto al analizarlos como al decantarnos, puesto que son una muestra de las limitaciones de la realidad en comparación con nuestros deseos. Está muy lejos de las intenciones de este texto resolver para ti los dilemas que no hemos sido capaces de resolver para nosotros mismos. Tómese este texto, por lo tanto, como un paseo por un paisaje que, si bien esperamos que te permita volver a casa con mayor perspectiva, no va a aportar una fórmula mágica ni una resolución de lo irresoluble.



Joachim Patinir (1520). *El paso de la laguna Estigia* [Óleo sobre tabla]. Museo del prado. https://es.wikipedia.org/wiki/El_paso_de_la_laguna_Estigia

Copyright 2026

La educación con IA está llena de dilemas, veamos algunos de ellos:

Dilemas éticos del uso de la IA

Personalización e igualdad de oportunidades: la estigmatización

El dilema entre la personalización y la estigmatización nos sitúa en el centro de la ética algorítmica. Por un lado, la IA ofrece una **personalización** sin precedentes, capaz de actuar como un tutor socrático que ajusta el nivel de dificultad y el andamiaje pedagógico a las necesidades exactas de cada alumno. Estos sistemas pueden identificar lagunas de conocimiento en tiempo real y ofrecer explicaciones alternativas que un docente, con un aula con muchos estudiantes, difícilmente podría proporcionar de forma individualizada.

Sin embargo, esta misma capacidad de análisis conlleva el riesgo de la **estigmatización** o un cierto encasillamiento algorítmico que erosione la igualdad de oportunidades educativas. Los informes de la UNESCO advierten que los modelos de IA pueden heredar y amplificar sesgos existentes en sus datos de entrenamiento, lo que podría llevar a que el sistema asuma que un estudiante tiene un techo de aprendizaje basado en su rendimiento pasado o en patrones sociodemográficos. Corremos el riesgo de hipotecar la libertad y el desarrollo futuro de un individuo con un algoritmo que "decide" qué puede o no aprender un niño basándose en datos históricos. Se corre, en definitiva el riesgo de convertir el pasado del alumno en su único destino posible, en una especie de efecto Pigmalión digital.

> El **efecto Pigmalión**, también conocido como **efecto Rosenthal,** es un fenómeno que se utiliza en psicología y pedagogía para referirse a la potencial influencia que ejerce la creencia de una persona en el rendimiento de otra. Supone, por tanto, algo importante de conocer y estudiar para los profesionales del ámbito educativo, laboral, social y familiar. El efecto Pigmalión debe su nombre al mito griego de Pigmalión, un escultor que se enamoró de una estatua que había tallado, y al final, ésta acabó cobrando vida.
>
> El efecto Pigmalión se puede identificar de las siguientes maneras:
>
> - Suceso por el que una persona consigue lo que se proponía previamente a causa de la creencia de que puede conseguirlo.
> - Las expectativas y previsiones de los profesores sobre el comportamiento de los alumnos condicionan las conductas de estos. (Rosenthal y Jacobson).

> - Una profecía autocumplida es una expectativa que incita a las personas a actuar en formas que hacen que la expectativa se cumpla.
>
> Wikipedia. Efecto Pigmalión.

Eficiencia a cambio de superficialidad y pérdida de control en el proceso

La promesa de la eficiencia que puede aportar la IA en nuestro trabajo docente es, sin duda, seductora. Según el informe de la **OCDE (2026)**, la IA generativa puede reducir drásticamente la carga administrativa y de planificación, permitiendo que el docente dedique ese tiempo recuperado a la interacción directa con el alumnado. Desde una perspectiva pragmática, optimizar la corrección de ejercicios mecánicos (debidamente anonimizados o en local) o la generación de materiales personalizados es un es un medio para liberarnos tiempo y poder emplearlo para discernir qué necesita cada niño en su singularidad emocional y cognitiva.

Sin embargo, corremos el riesgo de caer en una perspectiva donde toda la realidad, incluido el aprendizaje, es vista meramente como un recurso que debe ser optimizado y consumido con rapidez. Si la IA proporciona respuestas instantáneas y resúmenes perfectos, el alumno puede verse tentado a omitir la "dificultad deseable" que requiere para fijar conocimientos. El aula debe, además de propiciar la alfabetización en IA, ser en ocasiones un "refugio del algoritmo", un espacio donde el tiempo no se mida por la velocidad de procesamiento, sino por la profundidad del asombro y el pensamiento crítico. La rapidez de la máquina no debe dictar el ritmo de la maduración humana, somos nosotros como docentes el metrónomo de nuestra aula.

Una postura moderada nos obliga a sopesar y preguntarnos si estamos utilizando la IA para aprender mejor o simplemente para producir más. La **UNESCO** subraya que la competencia docente clave no es solo manejar la herramienta, sino saber cuándo esta no aporta valor. No debemos confundir la eficiencia de una búsqueda en un chatbot con la profundidad de una investigación reflexiva. Parece razonable delegar en la IA aquellas tareas alienantes que nos deshumanizan por su repetitividad, protegiendo con celo aquellos procesos lentos y bellos, como la lectura profunda, el debate socrático y el error creativo, que constituyen el núcleo de una educación que no solo instruye, sino que aspira a formar individuos libres con capacidad de amar, gozar de la vida y participar como ciudadanos inteligentes en el espacio público.

El punto de equilibrio entre el control y la intimidad

El uso de las analíticas de aprendizaje se presenta como una herramienta de **acompañamiento** sin precedentes, capaz de transformar la intuición docente en una intervención basada en

evidencias o, al menos, en correlaciones estadísticas. El informe de la **OCDE (2026)** destaca cómo la recopilación de datos en tiempo real permite detectar patrones de desenganche o dificultades de aprendizaje antes de que puedan derivar en un fracaso escolar, actuando como un sistema de alerta temprana. Es, por lo tanto, un efecto deseable, un cuidado del otro que busca su autonomía permitiendo dirigir nuestros apoyos al estudiante que mas lo necesita en el momento mas adecuado. Nos podrá permitir, por tanto, intervenir de manera quirúrgica y personalizada allí donde el esfuerzo humano es más necesario.

Sin embargo, la frontera entre el apoyo y la **vigilancia** es difusa. Si convertimos el aula en un espacio donde cada clic, cada pausa en una lectura y cada error son registrados y analizados, corremos el riesgo de instaurar un "panóptico digital", un sistema de vigilancia constante e invisible que no solo observa, sino que normaliza y moldea el comportamiento al ser internalizado por los vigilados. Los humanos cuando somos conscientes de ser observados podemos dejar de explorar y de equivocarnos con libertad por miedo a que nuestra "huella de datos" nos penalice en el futuro. Como apuntamos en capítulo 1, no es deseable perder el espacio para el error creativo y menos aún la privacidad del pensamiento, elementos indispensables para el desarrollo de la identidad en niños y adolescentes.

La postura moderada ante este dilema requiere situar la **agencia humana** y la transparencia en el centro del diseño tecnológico, tal como defiende el **Marco de Competencias de la UNESCO para Docentes**. El acompañamiento ético no consiste en monitorizarlo todo, sino en decidir qué datos son verdaderamente relevantes para el bienestar pedagógico y asegurar que el alumno sea partícipe de ese proceso. No debemos usar la analítica para predecir el destino de un estudiante, sino para ampliar sus posibilidades de acción. La línea ética se traza cuando el dato deja de ser una herramienta de ayuda para el profesor y se convierte en un mecanismo de control invisible; nuestra responsabilidad es garantizar que la tecnología sirva para ver mejor al alumno en su complejidad, pero nunca para invadir la intimidad de su proceso de maduración.

Copyright 2026


981
317

El Bosco (1475-1480). *Extracción de la piedra de la locura* [Óleo sobre tabla]. Museo del Prado. https://es.wikipedia.org/wiki/Extracci%C3%B3n_de_la_piedra_de_la_locura_(El_Bosco)

El dilema de la automatización de la evaluación

La automatización de la evaluación mediante IA se presenta como una solución técnica a uno de los mayores desafíos del docente: la gestión del tiempo y la objetividad. Los sistemas de evaluación algorítmica y las rúbricas generativas, como se menciona en el informe de la **OCDE (2026)**, permiten ofrecer una retroalimentación inmediata, algo que en la enseñanza tradicional suele dilatarse días o semanas. Esta "inmediatez pedagógica" puede ser una herramienta poderosa para el aprendizaje formativo, permitiendo al alumno mejorar su aprendizaje. Desde un punto de vista pragmático, la IA actúa aquí como un nivelador que busca eliminar los sesgos de fatiga o subjetividad inconsciente que todos los profesores, como humanos, podemos tener al corregir el centésimo examen de una jornada. Sin embargo, toda objetividad esconde, en el fondo, un sesgo de diseño, es una ilusión (tranquilizadora) de objetividad que elimina en el evaluador la carga moral de tomar decisiones difíciles.

Por otra parte, reducir la evaluación a un proceso puramente algorítmico supone un riesgo de deshumanización y de injusticia pedagógica. La IA es excelente procesando el "producto" final, pero ciega ante el "proceso" y el esfuerzo vital del estudiante y sus circunstancias particulares. Corremos el riesgo de que la evaluación deje de ser un acto educativo para convertirse en un mero control de calidad industrial. Un algoritmo no puede valorar la superación personal de un niño que ha vencido una dificultad familiar, ni la chispa de una idea original que se sale de los parámetros de la rúbrica preestablecida.

La postura ética más sólida, respaldada por el **Marco de Competencias de la UNESCO para Docentes**, es la defensa de la "agencia humana" en el centro del proceso evaluativo. La evaluación nunca debe ser exclusivamente automática; la IA puede encargarse de la corrección de tareas mecánicas o de proporcionar diagnósticos preliminares, pero el juicio final —la valoración del crecimiento integral del alumno— es una competencia exclusiva de sabiduría práctica del docente. El equilibrio reside en utilizar la IA para enriquecer la evaluación con datos y rapidez, pero garantizando siempre que el profesor actúe como garante de la justicia, asegurando que el sistema técnico no "encasille" al alumnado ni ignore las circunstancias humanas que dan sentido al aprendizaje

El uso de IA generativa como apoyo y la posible sustitución de la voz propia del alumnado

Sabemos que la IA generativa puede actuar como un tutor personalizado disponible las veinticuatro horas. Según el informe de la **OCDE (2026)**, estas herramientas pueden democratizar el acceso a un acompañamiento de alta calidad en tareas complejas como la redacción o la programación,

Copyright 2026

permitiendo, entre otras cosas, que el alumno supere sus bloqueos creativos o que reciba explicaciones multimodales de un mismo concepto. Esta tutorización y apoyo en tiempo real facilita que el estudiante alcance metas que solo, por pura limitación de recursos o tiempo, no podría abordar.

Sin embargo, este "aumento" de capacidades esconde una trampa de sustitución que amenaza la esencia misma del pensamiento crítico. Si el alumno delega el esfuerzo de síntesis, la estructura lógica y la voz propia en el algoritmo, nos enfrentamos a una pérdida de la autoría humana. Así, el estudiante puede dejar de ser un autor para convertirse en un mero editor de una "caja negra". La facilidad del atajo tecnológico puede atrofiar la memoria y el juicio en un fenómeno denominado descarga cognitiva o incluso, rendición cognitiva.

Una forma de tener en cuenta ambas posturas consiste en transformar la IA en una herramienta **metacognitiva**. La clave reside en lo que la **UNESCO** define como la "centralidad de la agencia humana": no se trata de que la IA escriba por nuestro alumnado, sino de que éste use la IA para contrastar sus propias ideas, identificar sesgos en el texto generado o comprender la lógica detrás de un código. En lugar de evaluar solo el resultado final, debemos incentivar que el alumno documente cómo interactuó con la máquina, qué correcciones hizo y por qué. Solo así garantizamos que la tecnología sea un apoyo que eleve el techo del aprendizaje sin derribar los cimientos de su esfuerzo intelectual y su responsabilidad ética.

La velocidad de innovación tecnológica y la prudencia educativa

El dilema entre la velocidad de la innovación tecnológica y la prudencia educativa nos sitúa en una tensión constante entre el miedo a quedar rezagados y el deber de proteger la integridad del acto pedagógico. La velocidad de la digitalización de las aulas está asociada, según el informe de la **OCDE (2026)** a una búsqueda de eficiencia y modernización de los sistemas educativos. Sin embargo, como ya señalamos anteriormente, con la irrupción de la IA no estamos ante un simple cambio de herramientas, sino ante un cambio de paradigma que afecta la forma en que comprendemos el mundo. Adoptar la IA por mera inercia institucional, sin una reflexión previa sobre su impacto en el desarrollo cognitivo, es ignorar que el tiempo de la técnica es casi instantáneo, mientras que el tiempo del aprendizaje humano es esencialmente lento y madurativo.

Frente a esta aceleración,la prudencia educativa no debe ser una resistencia ciega al cambio, sino la virtud de saber discernir si una innovación realmente mejora nuestra educación de acuerdo con los valores de cada comunidad educativa o los empobrece. El **Marco de Competencias de la UNESCO para Docentes** es taxativo al respecto: la formación ética y el desarrollo de la conciencia de riesgos deben ser los cimientos sobre los que se construya cualquier integración tecnológica. Una de las competencias docentes más urgentes en la actualidad es desarrollar la capacidad de identificar cuándo la tecnología **no** aporta valor. La verdadera innovación no consiste en usar lo último, sino en saber cuándo lo último es útil y cuando es un obstáculo para lo importante.

La postura equilibrada nos invita a ver el aula como un espacio donde la tecnología se somete al escrutinio del propósito pedagógico. Esto implica que los centros educativos y las administraciones deben priorizar la formación docente profunda, que abarque desde la alfabetización técnica hasta la ética normativa, antes de imponer implementaciones masivas. La IA tiene el potencial de ser un aliado extraordinario para la inclusión y la personalización, pero solo si las comunidades educativas mantienen su autonomía para decidir los ritmos y los límites de su uso. En última instancia, el principio de prudencia nos recuerda que nuestro compromiso ético no es con la vanguardia tecnológica, sino con el bienestar y el crecimiento integral de ese alumnado que confía en nosotros para navegar un mundo cada vez más automatizado.

Copyright 2026



Jacob Peeter Gowy (1636-38). *La caída de Ícaro* [Óleo sobre lienzo] Museo del Prado. https://www.museodelprado.es/coleccion/obra-de-arte/la-caida-de-icaro/2823dc25-398a-4d88-a4b2-be314065a62d

Copyright 2026

Conclusiones.

Este texto ha presentado algunos de los dilemas que consideramos centrales en la introducción de la IA en la educación. Puedo imaginar que, como indicábamos antes, los dilemas dejan siempre un poso de insatisfacción tanto al analizarlos como al decantarnos. No era el objetivo del texto dar respuestas al lector, sino, quizás, desde la ambición, contribuir a refinar las preguntas a las que debemos dar solución, primero cada uno de nosotros individualmente y, posteriormente, con una respuesta colectiva consensuada. Quizás el mejor consejo que podamos atrevernos a dar al lector a este respecto es que no confíe en los textos llenos de respuestas, todos estamos solos ante la indecisión.

Copyright 2026

2.7 Directrices éticas sobre el uso de la IA en educación

Introducción

Esta problemática variada y multidimensional está, como hemos visto llena de oportunidades, matices y dilemas que afectan a la naturaleza de la educación e incluso a nuestro modelo antropológico. Es competencia de las insituciones estableces directrices que permitan a los ciudadanos tomar decisiones con la tranquilidad de tener una guía que los respalde, algo que se ha hecho desde las instituciones europeas.

El Plan de Acción de Educación Digital Europeo (2021-2027) establece 2 prioridades estratégicas y 14 medidas para apoyarlas de entre las que destacaremos la que más nos concierne en este tema:

> Consejos prácticos para el profesorado de primaria y secundaria sobre el uso de la IA: **Directrices éticas sobre la IA**.

En octubre de 2022, la Comisión Europea publicó las **"Directrices éticas sobre el uso de la inteligencia artificial y los datos en la enseñanza y el aprendizaje para los educadores"** que han sido revisadas en 2026 y se han actualizado y publicado **aquí**, al igual que se han revisado las **Directrices sobre la lucha contra la desinformación y la promoción de la alfabetización digital** que analiza cómo tecnologías como la IA generativa están transformando la creación y circulación de contenidos. Además, proporciona herramientas para que el profesorado fomente una ciudadanía digital crítica.

Al revisar estas directrices sobre el uso de la IA en la enseñanza y aprendizaje, se han identificado **cinco consideraciones clave** que deben guiar cada decisión docente y sustentan el uso ético de la IA y de los datos en la enseñanza, el aprendizaje y la evaluación:

- **Dignidad Humana:** Respetar el valor intrínseco de cada persona, su privacidad y su autonomía, evitando tratar a los alumnos como simples objetos de datos.
- **Justicia:** Asegurar la equidad, la inclusión y la no discriminación para que todos los usuarios tengan las mismas oportunidades de aprendizaje.
- **Confianza:** Utilizar herramientas que sean fiables, transparentes, sin sesgos y que protejan los mejores intereses del alumnado.

Copyright 2026

- **Integridad Académica:** Promover un uso honesto de la IA, dando atribución correcta a las ideas y diseñando evaluaciones que sigan siendo válidas en la era de la IA generativa.
- **Elección Justificada:** Tomar decisiones sobre el uso de la IA de forma colaborativa, transparente y basada en hechos.

Estas directrices se agrupan en 8 requisitos o dimensiones, dividida cada una por el rol del docente y de la institución escolar:

1. Agencia y supervisión humana

Esta dimensión busca garantizar que la IA no reemplace el juicio profesional y que se respeten los derechos del alumnado. El docente debe mantener siempre el control sobre el proceso educativo:

- **Para los docentes:**
  - **Validación de resultados:** Deben saber cómo revisar y verificar los resultados de la IA antes de usarlos en clase. Deben ser capaces de detectar errores o sugerencias inapropiadas.
  - **Intervención activa:** Tienen que poder ajustar lo que la IA propone e intervenir si observan efectos inesperados durante su uso.
  - **Fomento del pensamiento crítico:** Deben animar a los alumnos a reflexionar sobre **por qué** la IA ha tomado una decisión o sugerencia específica.
  - **Juicio profesional:** La IA debe apoyar la toma de decisiones, no dictarla; el docente debe verificar las sugerencias basándose en su propia experiencia.
- **Para la escuela:**
  - **Formación:** Es responsabilidad del centro proveer **desarrollo profesional** sobre el uso de herramientas de IA.
  - **Políticas claras:** Debe existir una política escolar que permita a los educadores **anular o intervenir** en decisiones generadas por IA.
  - **Información y derechos:** Se debe informar a padres y alumnos sobre el papel de la IA y darles la opción de expresar preocupaciones o **renunciar al uso (*opt-out*)** cuando sea apropiado.

2. Transparencia y explicabilidad

El objetivo es que los sistemas de IA no sean "cajas negras" y que su funcionamiento sea comprensible para todos.

- **Para los docentes:**
  - **Comprensión de la lógica:** El profesor debe entender por qué la herramienta hace ciertas recomendaciones o sugiere actividades específicas.
  - **Sentido pedagógico:** Deben verificar que las sugerencias de la IA tengan sentido pedagógicamente y se alineen con el **currículo oficial**.

Copyright 2026

    - **Atribución y honestidad:** Deben requerir que los alumnos declaren cómo han usado la IA y enseñarles a **citar o reconocer la asistencia de la IA** en sus trabajos.

- **Para la escuela:**
    - **Documentación del proveedor:** La escuela debe exigir a los proveedores documentación clara sobre la **lógica de decisión**, el uso previsto y el tratamiento de la propiedad intelectual.
    - **Evaluación de equidad:** El centro debe evaluar si la IA se comporta de forma distinta entre grupos de alumnos y documentar posibles brechas de equidad.
    - **Comunicación no técnica:** Debe haber procedimientos para explicar las decisiones apoyadas por IA de forma sencilla y comprensible para familias y alumnos.

3. Diversidad e Inclusión

Esta dimensión se centra en la **accesibilidad, el diseño universal y la participación** de todas las partes interesadas.

- **Para los docentes:** Deben asegurarse de que el uso de la IA satisfaga las necesidades diversas de su alumnado y, de no ser así, ofrecer alternativas inclusivas. También es su responsabilidad revisar si los resultados de la IA contienen **sesgos culturales o sociales** y garantizar que los estudiantes que carecen de acceso a la tecnología fuera de la escuela no se vean desfavorecidos.

- **Para la escuela:** Las instituciones deben incorporar **voces diversas de los estudiantes** en la evaluación de las herramientas y abordar las brechas de alfabetización digital. Además, deben asegurar opciones de bajo ancho de banda o acceso *offline* para garantizar la equidad.

4. Equidad y No Discriminación

Busca evitar **sesgos injustos**, independientemente de la edad, género, capacidad o cualquier característica histórica de desventaja.

- **Para los docentes:** Deben verificar que la IA se adapte a necesidades individuales (como dificultades de aprendizaje) y que el contenido sea adecuado para la **edad y nivel de comprensión** del alumnado. Es vital que intervengan de inmediato si notan que el sistema se comporta de forma injusta o excluyente.

- **Para la escuela:** Antes de adoptar una herramienta, la escuela debe verificar que haya sido **probada contra sesgos** en diferentes grupos de aprendizaje. Deben existir procedimientos claros para responder si se identifica un trato injusto tras la

implementación.

5. Bienestar Social y Ambiental

Incluye la sostenibilidad, la responsabilidad social, la democracia y el bienestar psicosocial.

- **Para los docentes:** Es fundamental observar si el uso de la IA afecta la **motivación o el estado de ánimo** de los estudiantes (causando ansiedad o dependencia). Deben establecer límites para mantener la interacción humana significativa y discutir el uso de la IA con las familias.

- **Para la escuela:** La institución debe evaluar cómo la IA afecta la interacción social y el sentido de pertenencia. También deben considerar el **impacto ambiental y social** al seleccionar herramientas y fomentar que los alumnos usen la IA para abordar problemas reales de la sociedad.

6. Privacidad y gobernanza de datos

Dado que las escuelas manejan una gran cantidad de datos personales y de comportamiento (clics, tiempos de interacción), esta dimensión es crítica para cumplir con el **RGPD** y la **Ley de IA**.

- **Para los docentes:**
  - **Conocimiento de los datos:** Deben saber exactamente qué datos personales o de aprendizaje recopila la herramienta.
  - **Minimización de datos:** Es vital **evitar introducir información sensible** o identificable de los alumnos en herramientas de IA, siguiendo las reglas de privacidad de la escuela.
  - **Explicación al alumnado:** Deben explicar a los estudiantes cómo se usarán y almacenarán sus datos al usar la herramienta.

- **Para la escuela:**
  - **Seguridad y anonimato:** Implementar medidas para que los datos sensibles estén **anonimizados** y solo sean accesibles a personal autorizado.
  - **Cumplimiento legal:** Confirmar que todas las herramientas cumplen con el RGPD, la Ley de IA y las leyes nacionales de protección de datos.
  - **Control de configuración:** La escuela debe poder personalizar o restringir los ajustes de privacidad de las herramientas para proteger mejor al alumnado.
  - **Evaluaciones de impacto:** Realizar evaluaciones de impacto de protección de datos (DPIA) y de derechos fundamentales (FRIA) antes de implementar sistemas de IA de alto riesgo.

7. Robustez Técnica y Seguridad

Se refiere a la resistencia ante ataques, la ciberseguridad, la exactitud y la fiabilidad.

- **Para los docentes:** Deben monitorear si la herramienta funciona de manera fiable o si presenta errores, caídas o respuestas incorrectas frecuentemente. Deben conocer los pasos a seguir en caso de una **brecha de datos** o un mal funcionamiento técnico.

- **Para la escuela:** Es responsabilidad de la escuela verificar la fiabilidad antes del uso y establecer protocolos formales para reportar fallos y salvaguardar a los estudiantes. Deben confirmar que los proveedores cumplen con normativas como la **Ley de Ciberseguridad de la UE**.

8. Rendición de Cuentas

Abarca la auditabilidad, la minimización y el reporte de impactos negativos, y los mecanismos de reparación.

- **Para los docentes:** Deben saber qué hacer si el sistema sugiere algo inadecuado y sentirse seguros para actuar. Deben tener canales para dar retroalimentación sobre la mejora del sistema.

- **Para la escuela:** Debe existir una **división clara de responsabilidades** entre docentes, estudiantes y otros actores. La escuela necesita un proceso de evaluación para asegurar que la IA se alinee con los valores educativos y una estrategia para actualizar estos procedimientos según cambien las necesidades.

Pasos para la Integración Práctica

1. **Revisión interna:** Evaluar las herramientas actuales y el propósito de los datos recopilados.
2. **Políticas escolares:** Establecer procedimientos claros que definan comportamientos aceptables e inaceptables.
3. **Alfabetización en IA:** Participar en formación continua y comunidades de práctica para actualizar las competencias digitales y entender los riesgos y limitaciones de las nuevas tecnologías.

En conclusión, las directrices para los docentes sobre el uso ético y responsable de la inteligencia artificial (IA) se fundamentan en un enfoque **centrado en el ser humano**, donde la tecnología debe estar al servicio de las personas y los educadores actúan como **guardianes éticos** de sus estudiantes.

Copyright 2026