

No hay educación sin aprendizaje: por qué la IA educativa debe aprender a preguntar

Senén Barro Ameneiro

Director del CiTIUS-Centro Singular de Investigación en Tecnologías Inteligentes de la Universidad de Santiago de Compostela

La irrupción de los modelos grandes de lenguaje en la educación ha abierto posibilidades inéditas, pero su uso también presenta riesgos que conviene no ignorar. El principal es la delegación cognitiva, que puede hacer que no se desarrollen adecuadamente las capacidades intelectuales, especialmente de aquellos más vulnerables por estar madurando cognitivamente, lo que, paradójicamente, puede ir en contra del objetivo de la educación, que es aprender. Este artículo defiende que los sistemas de IA destinados a la educación deben ser diseñados y entrenados de forma explícita para promover el aprendizaje activo -a través del diálogo socrático, la reflexión guiada y la dificultad necesaria y productiva durante el aprendizaje- y no para ser usados como generadores compulsivos de respuestas que acaban haciendo que el estudiante simplemente externalice su esfuerzo cognitivo. La evidencia científica disponible respalda esta tesis con solidez creciente. Sobre esa base, el artículo propone cinco principios de diseño para una IA verdaderamente educativa y reclama estándares pedagógicos exigibles a cualquier sistema de IA que aspire a entrar en las aulas.

Este artículo forma parte de Apuntes Universidad es un espacio que reúne textos con mayor desarrollo centrados en el análisis de cuestiones clave de la Educación Superior. Su objetivo es aportar contexto, ordenar ideas y ofrecer claves de interpretación que contribuyan a una comprensión más amplia y fundamentada de los debates que afectan a la universidad.

EL PROBLEMA: QUÉ ES LA EDUCACIÓN Y CUÁLES LOS RIESGOS DE INTRODUCIR EN ELLA LAS MÁQUINAS

En el Fedro de Platón, Sócrates advierte de que la escritura producirá olvido porque se confiará en signos externos en lugar de ejercitar la memoria. Cuando Gutenberg presentó la imprenta en el siglo XV, los teólogos temieron que la Biblia quedara al alcance de manos sin preparación. Tras la aparición de la calculadora de bolsillo, muchos docentes de matemáticas decían que los estudiantes dejarían de pensar. Cuando llegó Google, muchos predijeron el fin de la memorización y, con ella, del conocimiento profundo. Hoy, con los modelos de lenguaje generativo - ChatGPT, Gemini, Claude y sus sucesores- estamos ante un nuevo episodio de esa tensión histórica. Pero hay algo que distingue la IA de todas las tecnologías anteriores: por primera vez la tecnología articula la información y el conocimiento, los argumenta y los presenta como si pensara (Barro, 2026a). Eso lo cambia todo y aumenta radicalmente las oportunidades, pero también los riesgos.

No se trata simplemente del riesgo de la trampa académica, del plagio o de la falta de honradez intelectual. Es algo más profundo y difuso: el riesgo de que millones de estudiantes acaben delegando en una máquina el trabajo mental que les haría mejorar y madurar intelectualmente. El riesgo, en suma, de que confundamos la respuesta con el aprendizaje.

La educación no es ir acumulando palabras y sus definiciones en un diccionario mental, sino comprender el mundo para entenderlo y actuar sobre él.

Esta distinción -respuesta versus aprendizaje- es el corazón del debate que este artículo quiere plantear. Y es un debate que no puede reducirse a si los estudiantes hacen trampas con la IA, sino a algo más fundamental: ¿para qué han sido diseñadas las IA que usamos en la educación?

LA CIENCIA DEL APRENDIZAJE: LO QUE SABEMOS SOBRE CÓMO APRENDE EL CEREBRO

Antes de hablar de IA conviene recordar lo que décadas de ciencia cognitiva nos han enseñado sobre el aprendizaje humano. Esta base es imprescindible para juzgar si la IA actual ayuda o perjudica.

El esfuerzo como ingrediente esencial

El psicólogo Robert Bjork acuñó en 1994 el concepto de dificultades deseables (*desirable difficulties*) para describir una paradoja fundamental del aprendizaje: las condiciones que hacen que el aprendizaje parezca más fácil y fluido en el momento suelen producir una retención y transferencia inferiores a largo plazo, mientras que las condiciones que generan dificultad y esfuerzo -aunque resulten menos cómodas- producen un conocimiento más duradero y transferible (Bjork, 1994; Bjork & Bjork, 2011). Seguramente todos lo hemos vivido en carne propia.

La práctica de recuperación espaciada, la intercalación de temas y la generación activa de respuestas en lugar de la relectura pasiva, comparten un denominador común: son cognitivamente exigentes. Requieren que el cerebro trabaje. Y es precisamente ese trabajo el que construye las estructuras mentales que llamamos conocimiento.

En palabras del propio Bjork: *"las condiciones que mejoran el rendimiento en el momento a menudo no sirven para sostener el aprendizaje a largo plazo"*. Esto es algo que los profesores intuyen desde siempre, pero que la neurociencia ha confirmado con rigor creciente. Y que la IA, diseñada para dar respuestas inmediatas y sin fricción, ignora por completo en su configuración por defecto. Inspirándome en una conocida frase que muchos peregrinos extranjeros repiten a lo largo del Camino de Santiago: *No pain, no brain - no pain, no glory*, dicen los caminantes-.

El peligro de la delegación cognitiva

La delegación cognitiva (Barro, 2026b) -el uso de herramientas externas para reducir la carga mental- puede ser beneficiosa cuando libera recursos para dedicarlos a tareas intelectualmente más exigentes y que aportan un valor añadido al resultado. El problema surge cuando esa delegación es sustitución y la tecnología se usa simplemente para no pensar.

Un estudio de método mixto publicado en 2025 en la revista *Societies* (Gerlich, 2025) analizó mediante encuestas y entrevistas a 666 participantes de distintos grupos de edad y niveles educativos. Sus resultados revelaron una correlación negativa significativa entre el uso frecuente de herramientas de IA y las habilidades de pensamiento crítico, con la delegación cognitiva -cognitive offloading, en el estudio- como factor mediador. Los participantes más jóvenes mostraron mayor dependencia de la IA y puntuaciones más bajas en pensamiento crítico. Además, el nivel educativo más alto se asoció con mejores habilidades de pensamiento

crítico independientemente del uso de IA, lo que sugiere que la formación profunda ofrece cierta protección. Conviene subrayar, en todo caso, que se trata de un estudio correlacional basado en autoinformes, por lo que no permite establecer relaciones de causa y efecto.

El fenómeno recuerda al llamado "efecto Google": la tendencia a no memorizar información que sabemos que podemos recuperar fácilmente en cualquier momento (Sparrow et al., 2011), aunque conviene notar que intentos de réplica posteriores de este experimento han arrojado resultados no concluyentes. Con la IA generativa, este efecto se amplifica cualitativamente: ya no delegamos solo el almacenamiento de información, sino también su síntesis, su argumentación y su articulación. Las funciones cognitivas más elevadas -que la conocida Taxonomía de Bloom situaría en los niveles de análisis, evaluación y creación- quedan en manos de la máquina.

Solo se puede delegar sin coste, o con un coste reducido, una capacidad que ya está consolidada. La calculadora no daña a quien ya domina la aritmética, pero la delegación de la síntesis y la argumentación sí puede perjudicar a quien aún no las ha desarrollado suficientemente. Esto conecta de nuevo con las "dificultades deseables" de Bjork: el esfuerzo que la delegación elimina es justamente el que consolida la capacidad. Es más, la corteza prefrontal -sede de las funciones ejecutivas y la metacognición, precisamente lo que los LLM permiten delegar- no madura hasta pasados los veinte años (Gogtay et al., 2004), lo que, a expensas de lo que la ciencia vaya evidenciando

con el necesario rigor, hace plausible la vulnerabilidad diferencial, más perjudicial para los jóvenes, sobre todo en edades más tempranas.

LA EVIDENCIA EMPÍRICA: LO QUE MUESTRAN LOS ESTUDIOS RECIENTES

En los últimos dos años se ha acumulado una evidencia científica preocupante sobre el uso acrítico de LLM en educación, pero también -y esto es crucial- una evidencia esperanzadora sobre lo que ocurre cuando se diseñan correctamente.

El coste cognitivo y en el rendimiento académico de la IA generativa

En un estudio reciente y muy citado del MIT Media Lab (Kosmyrna et al., 2025), aunque todavía en forma de preprint sin revisión por pares, los autores examinaron las consecuencias neurales y conductuales del uso de asistentes de IA en la escritura de ensayos. 54 estudiantes de cinco universidades de Boston fueron asignados aleatoriamente a tres grupos: uno que usaba ChatGPT, otro que usaba buscadores web convencionales y un tercero que escribía solo a partir de sus conocimientos. Los resultados son reveladores: los estudiantes del grupo LLM fueron incapaces en un 83% de los casos de citar correctamente una frase del ensayo que acababan de escribir minutos antes, frente a aproximadamente un 11% en cada uno de los otros dos grupos. Además, los ensayos de aquellos mostraron menor actividad cerebral medida mediante EEG.

Los autores lo denominaron acumulación de "deuda cognitiva" -término que yo cuestiono (Barro, 2026b)-: un coste

diferido que se paga cuando el estudiante tiene que enfrentarse al conocimiento sin el soporte artificial. Lo inquietante es que los estudiantes del grupo IA podían producir textos de alta calidad aparente en el momento, pero sin que ese proceso dejara huella cognitiva duradera en ellos.

Algunos estudios (Wecks et al., 2025), que analizan el impacto del uso de herramientas como ChatGPT en el rendimiento académico de estudiantes muestran efectos negativos especialmente acusados en los estudiantes con mayor potencial de aprendizaje, lo que sugiere que la IA perjudicaría más a quienes más podrían beneficiarse de un aprendizaje más intenso. En una línea complementaria, Lehmann, Cornelius y Sting (2025) mostraron experimentalmente que el efecto de los LLM depende de manera crítica del modo de uso: empleados como sustituto del propio razonamiento reducen el aprendizaje, mientras que usados como complemento pueden llegar a mejorarlo.

Estos hallazgos se alinean con los de investigadores que han observado cambios en el comportamiento de estudiantes de programación desde la llegada de los LLM: mayor longitud en las entregas, pero menor mejora entre versiones sucesivas y menor rendimiento global (Zhang et al., 2026). En palabras de los autores: "los hallazgos generan una alarma sobre la primera generación de graduados en la era de los LLM".

La confirmación a mayor escala de la distinción entre sustituir y complementar ha llegado de la mano de Strömberg y col. (2026), que analizaron los resultados de 26.811 estudiantes chinos de secundaria

a lo largo de 30 meses. El uso de la IA mejoró las notas de los deberes un 18% y redujo un 30% el tiempo de realización, pero recortó un 20% las notas de los exámenes mensuales a libro cerrado ya en un plazo de seis meses. Los datos combinan exámenes mensuales, exámenes de acceso a la enseñanza secundaria y a la universidad, así como las puntuaciones y el tiempo de realización de los deberes en nueve asignaturas. Los más afectados fueron los estudiantes de cursos inferiores, los de alto rendimiento y los chicos, sobre todo cuando el uso de la IA supuso una reducción del tiempo dedicado a los deberes. Quienes mantuvieron tiempos de dedicación similares a los de los no usuarios sufrieron pérdidas mínimas: el problema no es la IA en sí, sino delegarle el esfuerzo cognitivo sin más.

LA ALTERNATIVA: CUANDO LA IA APRENDE A PREGUNTAR

La evidencia no solo señala el problema: también apunta un posible camino, que, con mayor o menor intensidad, lleva dos milenios transitándose.

Sócrates en el siglo XXI

El método socrático -basado en la pregunta guiada, la refutación amable y el descubrimiento progresivo- es probablemente la tecnología pedagógica más antigua y mejor validada de la historia. Su premisa es radicalmente opuesta a la del oráculo que da respuestas: el docente no enseña declarando verdades, sino haciéndole preguntas al discípulo que le lleven a descubrirlas por sí mismo. Es lo que Sócrates llamaba mayéutica: el arte de ayudar a dar a luz las ideas que están latentes en la mente del estudiante.

La buena noticia es que los LLM son técnicamente capaces de emular este método. Un trabajo presentado en la conferencia ECAI 2024 (Favero et al., 2024) - que organizamos en Santiago de Compostela- demostró que, configurando el modelo con indicaciones específicas orientadas al diálogo socrático, se obtenían diferencias estadísticamente significativas ($p < 0,001$) respecto a tutores IA no socráticos en métricas automáticas de calidad del diálogo y del pensamiento crítico. Ahora bien, la evaluación se realizó sobre un estudiante simulado por otro modelo de lenguaje, no sobre estudiantes reales. La primera evidencia con estudiantes universitarios reales ha llegado después, cuando Xi, Zhang y Wang (2026) encontraron que un agente conversacional socrático basado en LLM mejoraba tanto el rendimiento académico como el pensamiento reflexivo.

Otros investigadores (Macina et al., 2023; Chevalier et al., 2024) han explorado antes la integración del método socrático en LLM educativos, subrayando cómo el cuestionamiento iterativo y la guía adaptativa pueden fomentar el pensamiento crítico. El proyecto MathDial, por ejemplo, desarrolló un tutor de matemáticas basado en diálogos de múltiples turnos anclados en problemas de razonamiento matemático de varios pasos, que lograba orientar a los estudiantes hacia la solución sin dársela directamente.

El caso más conocido y a mayor escala de aplicación de estos principios es Khanmigo, el tutor de IA desarrollado por Khan Academy. A diferencia de los chatbots genéricos, Khanmigo está explícitamente diseñado para no dar respuestas

directas, sino para guiar al estudiante con preguntas, pistas y retroalimentación hacia el descubrimiento propio, aunque hoy en día no me consta que exista una evaluación independiente y revisada por pares que acredite ganancias de aprendizaje comparables a las de un tutor humano.

Sal Khan, fundador de Khan Academy, ha señalado que el advenimiento de la IA educativa es en cierto sentido un retorno a los orígenes de la educación, cuando la tutoría personalizada era la norma. Como él mismo recuerda: Alejandro Magno fue tutelado por Aristóteles, quien fue estudiante de Platón, quien fue discípulo de Sócrates, el inventor del método que hoy deberíamos replicar a escala digital.

MARCO TEÓRICO: POR QUÉ EL DISEÑO PEDAGÓGICO DE LA IA IMPORTA MÁS QUE LA TECNOLOGÍA

Un análisis reciente de Favero, Pérez-Ortiz, Käser y Oliver (2026) titulado "AI in Education Beyond Learning Outcomes" articula con claridad el argumento: la IA puede reforzar o erosionar el aprendizaje según cómo esté diseñada y sea usada. Los autores identifican cuatro dimensiones en las que la IA mal diseñada puede perjudicar a los estudiantes: cognición, autonomía, bienestar emocional y ética. La descarga cognitiva excesiva, la agencia reducida del estudiante, la dependencia emocional y las preocupaciones sobre privacidad y sesgo se retroalimentan mutuamente.

Su conclusión es inequívoca: "los sistemas de IA pedagógicamente alineados, éticamente fundamentados y centrados en el ser humano pueden sostener el

razonamiento esforzado, apoyar la autonomía del aprendiz y preservar la interacción social significativa". El problema es que ese diseño ha de ser deliberado y es costoso. No es el que resulta por defecto de entrenar un LLM para predecir el siguiente token de forma convincente.

Por otra parte, existe una tentación intelectualmente respetable de comparar los LLM con la calculadora o con el libro de texto: herramientas que también externalizaron una parte del esfuerzo cognitivo, pero que no eliminaron el aprendizaje. La analogía es parcialmente válida, pero oculta una diferencia crucial.

En general la calculadora externaliza la ejecución de cálculos que el estudiante ya ha comprendido conceptualmente y en la mayoría de los casos sabe resolver por su cuenta. El libro de texto proporciona información que el estudiante debe procesar, interpretar y sintetizar. El LLM, en cambio, puede externalizar directamente el proceso de síntesis y comprensión.

Un artículo de Beale (2025) sobre pedagogía dialógica para los LLM propone marcos concretos de indicaciones que transforman la interacción del LLM con el estudiante: en lugar de que la pregunta del estudiante genere una respuesta, genera una pregunta de vuelta -una pista, una hipótesis que el estudiante debe evaluar-. La tecnología existe o puede crearse. Lo que puede faltar es la voluntad de hacerlo y usarla adecuadamente.

La zona de desarrollo próximo y el soporte inteligente

Lev Vygotsky introdujo en los años 30 del siglo pasado el concepto de zona de

desarrollo próximo: el espacio entre lo que un estudiante puede hacer solo y lo que puede hacer con el apoyo adecuado. El aprendizaje real ocurre precisamente en ese espacio, cuando el papel del docente (o de la máquina) es suficientemente desafiante para empujar al estudiante más allá de su zona de confort, pero no tan exigente como para producir frustración paralizante.

La IA, diseñada con criterio pedagógico, podría ser el apoyo más personalizado y paciente de la historia, detectando en qué punto del camino está cada estudiante y ajustando el nivel de sus preguntas en tiempo real. Pero la IA diseñada sin ese criterio -la que simplemente completa el texto de la pregunta con la respuesta más probable- no solo no ayuda, sino que construye un atajo que dificulta el aprendizaje.

PROPUESTA: QUÉ SIGNIFICA DISEÑAR IA QUE ENSEÑA A PENSAR

La tesis de este artículo no es que la IA sea mala para la educación. Es que la IA educativa debe ser diseñada con un propósito radicalmente diferente al de los sistemas de IA general. Y eso implica cambios en cada nivel del diseño, comenzando por la ideación misma de la IA.

Principios de diseño para una ia verdaderamente educativa¹

¹ Esta formulación en cinco principios es una síntesis propia, pero no reinventa la rueda, sino que integra tradiciones consolidadas de la ciencia del aprendizaje. El primero recoge el método socrático o mayéutica. El tercero combina explícitamente la zona de desarrollo próximo de Vygotsky de la que hemos hablado, con las "dificultades deseables" de Bjork. El segundo y el

- Primero, la IA educativa debe tener prohibido dar respuestas directas antes de que el estudiante haya intentado razonar. Esto no es una limitación técnica: es una decisión de diseño pedagógico.
- Segundo, debe detectar el nivel de comprensión real del estudiante, no el nivel de su pregunta. Un estudiante que pregunta correctamente puede ni siquiera entenderla si no es suya. El diálogo socrático permite detectar la diferencia.
- Tercero, debe ajustar la dificultad de sus preguntas a la zona de desarrollo próximo del estudiante. Si son demasiado fáciles no habrá aprendizaje y si son difíciles en exceso habrá frustración paralizante. El punto óptimo -la dificultad deseable de Bjork- requiere, una vez más, personalización.
- Cuarto, debe favorecer la metacognición: ayudar al estudiante a ser consciente de cómo está pensando y progresando, qué sabe y qué no y cómo puede mejorar su propio proceso de aprendizaje. Esto va más allá del contenido: es ese tan manido como importante aprender a aprender.

cuarto se apoyan en la tradición metacognitiva inaugurada por Flavell (1979) y desarrollada por Bjork. El quinto, finalmente, se inscribe en la línea del "fracaso productivo" (*productive failure*) de Kapur (2008) - coautor, por cierto, de MathDial- y en la concepción constructivista del error como parte del aprendizaje.

- Quinto, debe apreciar el error como información y parte del aprendizaje, no como fracaso. El error en ocasiones evidencia que el estudiante está trabajando en la zona de desarrollo próximo. Un sistema que evita los errores porque prefiere dar la respuesta correcta de inmediato no resulta educativo porque no ayuda a aprender.

Lo que el sistema educativo debe exigir

Los centros educativos, las administraciones y los responsables de política educativa tienen aquí una responsabilidad que no pueden delegar en el mercado. El mercado producirá la IA que mejor se venda, con la ayuda inestimable de la publicidad y el márketing. Pero no necesariamente la más demandada -sea de forma espontánea o inducida-, será la más educativa. Igual que muchos estudiantes prefieren matricularse en las materias más fáciles dentro de las que son optativas, podría triunfar la IA en educación que mejores respuestas dé a nuestras preguntas, sin someternos al esfuerzo de aprender.

Necesitamos estándares pedagógicos para la IA educativa, del mismo modo que existen estándares curriculares para los libros de texto. No basta con que una herramienta de IA sea confiable, sino que ha de poder demostrar que promueve y consigue el aprendizaje real: que aumenta la comprensión profunda, el pensamiento crítico y la autonomía intelectual del estudiante, y no solo su capacidad de producir textos de aspecto académico.

¿Y quién debe evaluar y certificar esos estándares, y cómo? No hace falta partir

de cero. Existen modelos consolidados de evaluación independiente mediante estudios controlados, como el What Works Clearinghouse estadounidense o la Education Endowment Foundation británica, y marcos que gradúan el nivel de exigencia probatoria, como los niveles de evidencia de la ley ESSA en Estados Unidos. En Europa, además, el Reglamento (UE) 2024/1689 -la llamada Ley de IA- ya clasifica los sistemas de IA utilizados en la educación como de alto riesgo y les exige requisitos verificables; sería natural apoyar en ese marco una certificación específicamente pedagógica, que acredite no solo que la herramienta es segura, sino que educa.

Personalizar la educación desde el docente

En julio de 2026 Anthropic lanzó un modelo de Claude para docentes verificados (Claude for teachers: Introducing Claude for Teachers - Anthropic). Está disponible en EE. UU. para lo que allí llaman K-12 (nuestra primaria y secundaria). Al menos durante un año estos profesores tendrán acceso gratuito y con la garantía de que los datos derivados de su uso no serán utilizados para el entrenamiento de modelos de IA. No es, en todo caso, un movimiento aislado: OpenAI (ChatGPT Edu) y Google (Gemini for Education) han lanzado ofertas análogas para el mundo educativo, señal de que el interés por este “mercado” es muy grande.

No se trata simplemente de un LLM que sabe mucho de educación, sino que maneja recursos adicionales para que realmente sea útil. Por ejemplo, mediante el grafo de conocimiento de Learning Commons, división de tecnología educativa de la Chan Zuckerberg Initiative, puede

acceder a los estándares académicos de los 50 estados y a las competencias y progresiones de aprendizaje que subyacen a cada estándar, apoyándose en currículos con base empírica (por ejemplo, los currículos educativos abiertos de la OpenSciEd y la Illustrative Mathematics, dos organizaciones estadounidenses sin ánimo de lucro). A partir de ahí permite al profesorado: la planificación de clases alineadas con el estándar de su estado; la diferenciación de materiales por niveles de preparación -por ejemplo, hacer adaptaciones para los estudiantes con planes individualizados o extensión para los más avanzados-; análisis de datos del aula -el docente entrega una carpeta con diagnósticos, asistencia, notas propias y Claude construye una imagen de dónde está cada estudiante-; y tareas programadas recurrentes -p. ej., proponer ajustes a la lección del día siguiente-.

Es una buena estrategia que avanza en la personalización, pero no a través del estudiante sino del profesorado, aportándole nuevos recursos y más tiempo para que realice tareas de mayor valor añadido para sí mismo y para los estudiantes.

Según la iniciativa SCALE de la Universidad de Stanford, la personalización de la educación mediante IA que se dirige al docente y no al estudiante cuenta por ahora con un respaldo causal más sólido -los primeros estudios sugieren que ahorra tiempo de preparación sin merma de la calidad instructiva- y esquivada en buena medida el problema de la delegación cognitiva. Ciertamente es que aparecen o se pueden acentuar otros riesgos, como son el sesgo de automatización -aceptar lo que la máquina aporta sin revisarlo-,

estandarización encubierta -los currículos embebidos definen qué se considera "bueno" como contenido docente- y posible erosión de la competencia de planificación si esta se delega sistemáticamente. También es un riesgo que los docentes queden cautivos de las empresas que comercializan estos sistemas, y a través de ellos lo acaben siendo los sistemas educativos en los que ejercen.

No nos engañemos, la prueba del algodón del valor de todas estas herramientas, sean para el docente o el discente, es verificar si hay una mejora objetivamente constatada del aprendizaje del estudiante. El resto es retórica. Por eso, hemos de ser cautos con esta y otras herramientas y someterlas a usos bien fundados, con usuarios bien formados y con procesos de evaluación rigurosos y objetivos.

CONCLUSIONES: LA IA QUE NECESITA LA EDUCACIÓN AÚN ESTÁ POR CONSTRUIR

La inteligencia artificial tiene el potencial de ser el tutor más paciente, más personalizado y disponible de la historia de la humanidad. En teoría, podría adaptar su nivel a cada estudiante, detectar sus lagunas conceptuales, proponer rutas de aprendizaje individualizadas y estar disponible en cualquier momento.

Pero ese potencial solo se realizará si quienes idean, diseñan, entrenan y despliegan estos sistemas entienden que su objetivo no es dar respuestas: es ayudar a formar personas que sepan sobre todo pensar y hacerlo de modo crítico, construyendo sus propias respuestas. Personas que sepan reconocer cuándo no saben, y

qué hacer con esa ignorancia; idealmente corregirla.

Todo lo anterior debe afirmarse, eso sí, con la honradez intelectual que la propia tesis de este artículo exige. La evidencia disponible es reciente y todavía imperfecta: algunos de los estudios más citados son preprints pendientes de revisión por pares o los resultados de una investigación que apenas está comenzando -los horizontes temporales de estudio rara vez superan los dos años y parte de los datos son correlacionales o autoinformados-. Quien defienda -como aquí se hace- que la IA educativa debe demostrar su valor mediante evaluaciones rigurosas, ha de aplicar la misma vara de medir a la evidencia que sostiene su diagnóstico. Lo notable, sin embargo, es la convergencia: métodos distintos, contextos distintos y muestras distintas -de los laboratorios de Boston a los institutos de secundaria chinos- apuntan en la misma dirección: cuando la IA sustituye el esfuerzo cognitivo del estudiante, el aprendizaje se tambalea.

Hay, además, un hallazgo contraintuitivo en esos mismos estudios que merece una reflexión propia. Tanto el trabajo de la Universidad de Bremen (Wecks et al., 2025) como el de la secundaria china (Strömberg et al., 2026) coinciden en algo inesperado: los más perjudicados por la delegación cognitiva no son los estudiantes rezagados, sino los de mayor rendimiento y potencial. El temor habitual -que la IA amplíe las brechas educativas favoreciendo a los ya aventajados- aquí parece que se invierte: la IA usada como

sustituto del esfuerzo no nivela hacia arriba, sino hacia abajo, y recorta precisamente donde más se podría ganar. La equidad educativa en la era de la IA no consistirá solo, por tanto, en garantizar el acceso a las herramientas; consistirá en garantizar que las herramientas no arrebatan a nadie -tampoco a los mejores- la parte del esfuerzo que les hace aprender más, mejor y más rápido.

La IA que hoy se usa masivamente en educación no está diseñada para ese propósito. Está diseñada para ser útil en el sentido más amplio del término: para dar información pertinente, bien organizada y escrita. Eso puede resolver dudas, pero aprender a pensar exige mucho más.

Sócrates nunca dio una respuesta directa cuando podía llegar a ella a través de nuevas preguntas. Dos mil cuatrocientos años después, ese sigue siendo el principio pedagógico más importante que conocemos. La IA educativa del futuro -la que de verdad merezca ese nombre- habrá aprendido también, como Sócrates, a hacer y hacerse preguntas, más que a dar respuestas, incluso cuando se ignora la verdad.

Esa IA aún está en gran medida por construir, pero si queremos sabremos cómo hacerla. Es necesario no olvidar, eso sí, que el aprendizaje tiene un coste, y si no lo pagas hoy lo harás mañana.

BIBLIOGRAFÍA

- Barro, S. (2026a). Poden pensar as máquinas? Alvarellos Editora.
- Barro, S. (2026b). Delegación y coste cognitivos. Universidad, sí.
<https://www.universidadsi.es/delegacion-y-coste-cognitivos/>
- Beale, R. (2025). Dialogic Pedagogy for Large Language Models: Aligning Conversational AI with Proven Theories of Learning. arXiv:2506.19484.
- Beatty, S. (2024). En: Microsoft News (mayo 2024). Khan Academy and Microsoft partner to expand access to AI tools.
<https://news.microsoft.com/source/features/ai/khan-academy-and-microsoft-partner>
- Bjork, E. L. & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. En M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world* (pp. 56-64). Worth Publishers.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. En J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). MIT Press.
- Chevalier, A. et al. (2024). Language Models as Science Tutors. *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*, PMLR 235, pp. 8310-8335. arXiv:2402.11111. (Presenta Tutor-Chat, un corpus de 80.000 diálogos de tutoría multi-turno anclados en libros de texto STEM).
- DiCerbo, K. (2025). Entrevista publicada en Education Week y realizada por Rick Hess: "Can an AI-Powered Tutor Produce Meaningful Results?" Education Week, 29 de julio de 2025.
<https://www.edweek.org/technology/opinion-can-an-ai-powered-tutor-produce-meaningful-results/2025/07>
- Favero, L., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2024). Enhancing Critical Thinking in Education by means of a Socratic Chatbot. *ECAI 2024 Workshop on AI in Education and Educational Research (AIEER)*. arXiv:2409.05511.
- Favero, L., Pérez-Ortiz, J. A., Käser, T., & Oliver, N. (2026). AI in Education Beyond Learning Outcomes: Cognition, Agency, Emotion, and Ethics. arXiv:2602.04598.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906-911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Forero, M. G. & Herrera-Suárez, H. J. (2023). ChatGPT in the Classroom: Boon or Bane for Physics Students' Academic Performance? arXiv:2312.02422.
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), Article 6. <https://doi.org/10.3390/soc15010006>; corrección en *Societies*, 15(9), 252; <https://doi.org/10.3390/soc15090252>
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., III, Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings*

- of the National Academy of Sciences, 101(21), 8174-8179.
<https://doi.org/10.1073/pnas.0402680101>
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424.
<https://doi.org/10.1080/07370000802212669>
 - Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. arXiv:2506.08872.
 - Kulesa, A. C., Mission, M., Croft, M., & Wells, M. K. (2025, junio). Productive Struggle: How Artificial Intelligence Is Changing Learning, Effort, and Youth Development in Education. Bellwether. <https://bellwether.org/publications/productive-struggle/>
 - Lehmann, M., Cornelius, P. B., & Sting, F. J. (2025). AI Meets the Classroom: When Do Large Language Models Harm Learning? arXiv:2409.09047.
 - Macina, J. et al. (2023). MathDial: A Dialogue Tutoring Dataset with Rich Pedagogical Properties Grounded in Math Reasoning Problems. *Findings of EMNLP 2023*, pp. 5602-5621. arXiv:2305.14536.
 - Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea, 12.7.2024. El Anexo III clasifica como de alto riesgo determinados sistemas de IA en educación y formación profesional.
 - <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
 - Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, 333(6043), 776-778.
 - Strömberg, D., Lei, V., & Wu, Y. (2026). The Generative AI Learning Penalty: Evidence from Chinese Secondary Education. CEPR Discussion Paper No. 21577. Centre for Economic Policy Research.
<https://cepr.org/publications/dp21577>
 - Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
 - Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2025). Generative AI Usage and Exam Performance. *ECIS 2025 Proceedings*, Article 8. Paper ECIS2025-1421.
 - Xi, L., Zhang, Y., & Wang, Q. (2026). Investigating the effects of an LLM-based Socratic conversational agent on students' academic performance and reflective thinking in higher education. *Computers & Education*, 241, 105494.
<https://doi.org/10.1016/j.compedu.2025.105494>
 - Zhang, Y., Savelka, J., Goldstein, S. C., & Conway, M. (2026). Changes in Coding Behavior and Performance Since the Introduction of LLMs. En *Proceedings of LAK26: The 16th International Learning Analytics and Knowledge Conference* (pp. 844-851). ACM.
<https://doi.org/10.1145/3785022.3785075>

Univers**s**idad



Studia XXI

ESTRATEGIAS Y POLÍTICAS
STRATEGIES AND POLICIES

 www.universidadsi.es

 [Universidad, sí](https://www.facebook.com/Universidad_si)

 [@universidad_si](https://twitter.com/universidad_si)

 [Universidad, sí](https://www.linkedin.com/company/universidad-si)