

CUATRO PARADIGMAS PARA PENSAR LA INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN

*De la deuda cognitiva al ultrasujeto:
paradigmas de literacidad y modos de gobernanza*

LUIS MEDINA GUAL / ÓSCAR FERNANDO LÓPEZ MERAZ

Resumen:

El debate institucional sobre la inteligencia artificial generativa en la educación ha quedado atrapado entre la prohibición frontal y la celebración acrítica. Este trabajo propone una arquitectura analítica de cuatro paradigmas –deuda cognitiva, inteligencia híbrida, nuevas capacidades e interhumano-ultrasujeto– y la cruza con tres lógicas de gobernanza institucional: centralizada, distribuida y federada. Cada paradigma se articula a partir de un texto seminal y se contrasta con literatura empírica reciente. Se argumenta que cada uno ilumina un riesgo distinto y exige una respuesta diferenciada: limitar el uso en tareas fundacionales, experimentar configuraciones humano-inteligencia artificial acompañadas, articular currículos federados de literacidad y congelar pilotos cuando la interacción toca la subjetividad. La tesis es que ninguna institución puede gobernar los cuatro problemas con una sola lógica.

Abstract:

The institutional debate on generative artificial intelligence in education has been caught between outright prohibition and uncritical celebration. This article proposes an analytical architecture of four paradigms (cognitive debt, hybrid intelligence, new capabilities and interhuman/ultra-subject) and maps it onto three logics of institutional governance: centralized, distributed, and federated. Each paradigm is developed from a seminal text and set against recent empirical literature. It is argued that each one brings a distinct risk into view and calls for a differentiated response: restricting use in foundational tasks, piloting scaffolded human-artificial intelligence configurations, articulating federated literacy curricula, and freezing pilot programs when the interaction bears on subjectivity. The thesis is that no institution can govern all four problems with a single logic.

Palabras clave: inteligencia artificial; educación superior; política educativa; alfabetización digital; pensamiento crítico.

Keywords: Artificial intelligence; higher education; educational policy; digital literacy; critical thinking.

Luis Medina Gual: director de la Dirección de Innovación Educativa, Universidad Iberoamericana Ciudad de México. Ciudad de México, México. CE: luis.gual@ibero.mx / <https://orcid.org/0000-0002-6783-606X>

Óscar Fernando López Meraz: docente de la Escuela Normal Superior Veracruzana “Dr. Manuel Suárez Trujillo” y de la Universidad Pedagógica Veracruzana. Xalapa, Veracruz, México. CE: osclopez@msev.gob.mx / <https://orcid.org/0000-0002-1185-6424>

Introducción: salir de la pregunta binaria

El debate público sobre la inteligencia artificial generativa (IAG) en la educación ha quedado, en buena medida, atrapado en una falsa disyuntiva: ¿deben las y los estudiantes¹ usarla o no? Las primeras reacciones institucionales a la apertura masiva de ChatGPT en noviembre de 2022 oscilaron entre la prohibición frontal y la celebración acrítica —una polarización documentada por el análisis de contenido de la cobertura pública del fenómeno realizado por Sullivan, Kelly y McLaughlan (2023)—, sin que ninguna postura resolviera la pregunta de fondo. Esa pregunta, como se argumentará en este ensayo, ya no es metodológica sino antropológica —en el sentido de la antropología filosófica de la técnica: qué tipo de sujeto se constituye en relación con sus instrumentos de pensamiento— e institucional al mismo tiempo, y se puede plantear de la siguiente manera: ¿qué tipo de aprendizaje, qué tipo de conocimiento y qué tipo de sujeto se constituyen cuando pensar requiere, cada vez con mayor frecuencia, conversar con una inteligencia no humana, y cómo se gobierna esa transformación desde las instituciones educativas?

Los datos indican que la presencia de la IAG en las aulas mexicanas no es una hipótesis sino una realidad estabilizada: el uso estudiantil recurrente ya es mayoritario en la educación superior y, según el Anthropic Economic Index (Anthropic, 2025), el académico constituye el caso de uso más frecuente de Claude en México (6.6% de las interacciones registradas), aun cuando el país presenta un índice de uso per cápita de 0.50x, en el lugar 76 de 116 países. Frente a esa difusión, las instituciones se han mostrado, en su mayoría, reactivas y fragmentadas. Esta fragmentación institucional es una afirmación distinta de la polarización del debate público, y se acredita de manera independiente: Ghimire y Edwards (2024), en “From Guidelines to Governance”, documentan una brecha sistemática entre la abundancia de guías informales y la escasez de marcos institucionales sustantivos; Wu, Zhang y Carroll (2024), en sus estudios de caso de las universidades del Big Ten, constatan políticas heterogéneas que oscilan entre el *laissez-faire* y la prohibición; en tanto, Jin, Yan, Echeverría, Gašević *et al.* (2025), en una revisión global de políticas universitarias, encuentran que apenas una minoría de instituciones ha articulado marcos coherentes que vayan más allá de declaraciones de principio.

Este ensayo propone una cartografía de cuatro paradigmas para pensar la IAG en la educación y la cruza con tres modos de gobernanza institu-

cional. Los tres primeros paradigmas se han consolidado ya en la literatura y disponen de documentos seminales que los articulan: el paradigma de la deuda cognitiva (Kosmyna, Hauptmann, Yuan, Situ *et al.*, 2025); el de la inteligencia híbrida (Cukurova, 2025); y el de las nuevas capacidades, asociado al giro interaccional propuesto por Moldoveanu y Siemens (2025). El cuarto paradigma, todavía emergente, se nutre de la crítica de la razón generativa formulada por Xun (2026) y propone leer la IAG no como herramienta, ni como prótesis cognitiva, ni como repertorio de competencias nuevas, sino como un interlocutor constitutivo del que emerge una figura que el propio Xun denomina ultrasujeto.

La tesis del ensayo es doble. Primero: cada paradigma ilumina un riesgo y, por tanto, exige una forma distinta de gobernanza educativa. Segundo: pretender resolver con una sola política los cuatro problemas que estos paradigmas plantean es la principal razón por la que los esfuerzos institucionales se quedan, una y otra vez, en lo superficial. La gobernanza de la IAG no se organiza por herramientas, sino por paradigmas de relación con esta tecnología, y la matriz que las instituciones requieren combina, en proporciones variables, tres lógicas —centralizada, distribuida y federada— que la literatura especializada ha caracterizado con creciente precisión.

Tres precisiones de método delimitan el estatuto de lo que sigue. Primera: este texto es un ensayo teórico-normativo, no un reporte de investigación empírica; en él conviven tres registros que se señalarán de manera explícita —la revisión de literatura (lo que la investigación reporta), la interpretación de los autores (cómo proponemos leer esa evidencia) y el argumento normativo (lo que sugerimos que las instituciones hagan)—. Segunda: el término paradigma se emplea aquí en sentido heurístico, como conjunto de supuestos que orienta la mirada sobre el fenómeno y no en el sentido epistemológico kuhniano de matriz disciplinar; los cuatro marcos tienen, además, estatutos epistémicos distintos que conviene no aplanar: la deuda cognitiva es una hipótesis empírica en discusión, la inteligencia híbrida es un marco normativo de diseño, las nuevas capacidades son una propuesta curricular y el ultrasujeto es una especulación filosófica. Tercera: la aportación central del ensayo no es la descripción de los paradigmas, que sintetiza literatura existente, sino la matriz que los cruza con los modos de gobernanza y el flujo operativo que se deriva de ella (sección previa a las conclusiones); los cuatro apartados siguientes deben leerse como la preparación analítica de ese instrumento.

Arquitectura de los cuatro paradigmas

La tabla 1 presenta la arquitectura de los cuatro paradigmas indicados.

TABLA 1

*Cuatro paradigmas para pensar la IA en la educación:
pregunta de fondo, tesis, fuente seminal y riesgo principal*

Paradigma	Pregunta de fondo	Tesis nuclear	Texto seminal	Riesgo principal
1. Deuda cognitiva	¿Qué dejamos de pensar cuando delegamos?	La facilidad inmediata se paga con pérdida progresiva de esfuerzo, retención y apropiación	Kosmyna <i>et al.</i> (2025)	Atrofia, dependencia, descarga cognitiva pasiva
2. Inteligencia híbrida	¿Qué se vuelve posible cuando dos cogniciones se acoplan?	La IA no sustituye al humano, lo extiende cuando el sistema actualiza sus modelos en interacción y ambos se modifican	Cukurova (2025)	Confianza ingenua, opacidad del proceso, delegación oculta
3. Nuevas capacidades	¿Qué habilidades nuevas exige esta ecología cognitiva?	La IA reorganiza profesiones y disciplinas; las habilidades meta-cognitivas y dialógicas se vuelven nucleares	Moldoveanu y Siemens (2025)	Brecha de literacidad, currículo obsoleto, desigualdad
4. Interhumano/ ultrasujeto	¿Qué tipo de sujeto emerge cuando pensar se vuelve dialogar con una IA?	La IA opera como interlocutor que modifica autoría, agencia, voz y subjetividad; el efecto subjetivante depende de la convergencia entre diseño conversacional, encuadre pedagógico y metacognición del aprendiz	Xun (2026), <i>Pensar con prompts</i> ; síntesis empírica reciente sobre IA conversacional como interlocutor dialógico (Sun, Zheng, Xu y Yang, 2026; Westbye, Eriksen y Blikstad-Balas, 2026; Ren, 2026)	Dependencia afectiva, ilusión de empatía, captura discursiva, desplazamiento relacional

Fuente: elaboración propia con base en la literatura señalada.

La deuda cognitiva:

lo que dejamos de pensar cuando delegamos

El primer paradigma se articula con nitidez en el estudio “Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task”, de Kosmyna *et al.* (2025), realizado en el MIT Media Lab. El experimento, conducido con 54 participantes a lo largo de cuatro sesiones, distribuyó a los estudiantes en tres condiciones –escritura de ensayo asistida por *Large Language Model* (LLM), asistida por motor de búsqueda, y sin herramientas– y registró la actividad cerebral mediante electroencefalografía (EEG). El trabajo reporta que la conectividad cerebral escaló de manera sistemática con la cantidad de soporte externo: el grupo *brain-only* exhibió las redes más amplias y robustas; el grupo *search engine* mostró un compromiso intermedio; y la asistencia con LLM produjo el acoplamiento neuronal más débil (Kosmyna *et al.*, 2025). En la cuarta sesión, los participantes que pasaron del LLM al pensamiento sin herramientas (*LLM-to-brain*) mostraron una conectividad neuronal todavía deteriorada y una subactivación de las redes alfa y beta, mientras que quienes pasaron del pensamiento sin herramientas al LLM (*brain-to-LLM*) recuperaron memoria y reactivaron nodos prefrontales y occipito-parietales (Kosmyna *et al.*, 2025).

El estudio aporta un dato que merece destacarse: los participantes del grupo LLM presentaron baja apropiación de los ensayos, dificultad para citar partes del texto que ellos mismos habían producido minutos antes y una marcada homogeneidad entre los escritos generados dentro de cada grupo, identificada tanto por evaluadores humanos como por modelos de procesamiento de lenguaje natural (Kosmyna *et al.*, 2025). En palabras de los autores, en sus conclusiones apuntan la conveniencia que ofrece el LLM al reducir la fricción de la escritura “*came at a cognitive cost, diminishing users’ inclination to critically evaluate the LLM’s output*” (Kosmyna *et al.*, 2025). El concepto que el estudio acuña –“*cognitive debt*”– captura precisamente la idea de que la facilidad inmediata se paga con una pérdida progresiva de capacidad para deliberar, recordar y apropiarse del propio pensamiento.

Conviene declarar el estatuto de esta fuente seminal y de sus críticas. El estudio de Kosmyna *et al.* (2025) es un *preprint* que no ha pasado por revisión de pares, y existe ya un comentario crítico formal que recomienda una lectura considerablemente más conservadora de sus resultados:

Stanković, Hirche, Kollatzsch y Doetsch (2026) señalan que la muestra está infrapotenciada (54 participantes, 18 en la sesión final, frente a los cerca de 159 que exigiría el diseño), que un menor número de conexiones EEG estadísticamente significativas no equivale a menor actividad neuronal absoluta y que el reporte presenta inconsistencias internas (en la banda beta, por ejemplo, el grupo asistido por LLM mostró mayor conectividad que el grupo con buscador). La deuda cognitiva debe tratarse, por tanto, como hipótesis en discusión y no como hallazgo establecido. Nuestra lectura –interpretación de los autores, no consenso de la literatura– es que la evidencia disponible converge en una tesis de secuenciación: la deuda aparece cuando la IAG sustituye el esfuerzo cognitivo fundacional, no cuando lo andamia después de ejercido, como sugiere el propio contraste entre las condiciones LLM-*to-Brain* y *Brain-to-LLM* de Kosmyrna *et al.* (2025), el experimento de Kreijkes, Kewenig, Kuvalja, Lee *et al.* (2026) y el metaanálisis de Zhao, Yue, Sun, Jiang *et al.* (2025).

La evidencia empírica acumulada en el último bienio ha tendido a confirmar el argumento de Kosmyrna *et al.* (2025) en otros contextos. Choudhuri, Sanchez, Burnett y Sarma (2026), en un estudio transversal con 299 participantes, muestran que el uso habitual no estructurado de IAG se asocia con una reducción significativa del compromiso cognitivo ($f^2 = 0.51$ para reflexión) y con una caída del pensamiento crítico ($\beta = -0.41$, $p = .001$). Kreijkes *et al.* (2026), en un experimento aleatorizado con 405 estudiantes de secundaria, encontraron que la toma de notas, por sí sola, superó al uso aislado del LLM en retención evaluada tres días después (diferencia de medias = 1.35, $p < .001$), aunque la combinación de IA con estrategias activas de aprendizaje recuperó parte de ese beneficio. Finalmente, Zhai, Wibowo y Li (2024), en una revisión sistemática, advierten que la sobreconfianza en sistemas de diálogo con IAG se correlaciona con la disminución de habilidades cognitivas básicas en escenarios sin andamiaje pedagógico.

Sin embargo, el cuadro no es unilateral. La revisión sistemática de Balart, Díaz y Shryock (2026), basada en 63 estudios, y el metaanálisis de Zhao *et al.* (2025), con 29 experimentos, muestran que el uso de IAG puede producir efectos positivos de moderados a grandes sobre el pensamiento crítico ($g = 0.691$) cuando va acompañado de andamiajes pedagógicos como *prompts* guiados, actividades reflexivas, correvisión

entre pares y apoyo metacognitivo. La conclusión, casi unánime, es que la IAG funciona como un amplificador cognitivo: refuerza tendencias previas hacia el procesamiento profundo o superficial, pero no determina por sí sola un resultado. La variable decisiva es la mediación pedagógica, no la herramienta.

En el plano institucional, la deuda cognitiva tiene una contraparte específica que conviene nombrar: la oficinización del juicio. Kumar y Sarkar (2026), en un estudio sobre administradores escolares, y McGuirk (2025), en un trabajo sobre liderazgo en educación superior, documentan que también los líderes educativos pueden caer en una forma análoga de descarga cognitiva cuando delegan funciones comunicativas y relacionales –redacción de cartas, mediación de conflictos, retroalimentación al profesorado– a sistemas generativos, atenuando precisamente las capacidades que Coates, Croucher y Calderon (2025) describen como insustituibles para gobernar la integridad del pensamiento académico. El paradigma de la deuda cognitiva, así, es una preocupación tanto del aula como del despacho directivo.

La gobernanza que este paradigma exige es, ante todo, de límites. Hay tareas fundacionales –donde el estudiantado debe ejercitar la regulación del propio pensamiento, el contraste con el conocimiento previo y la articulación de la propia voz– en las que el uso de IAG puede ser legítimamente restringido o exigido bajo trazabilidad. Coates, Croucher y Calderon (2025) lo formulan con precisión: la integridad académica del futuro no se gobernará por detección y disuasión, sino por rediseño de la evaluación y por desarrollo sistemático de literacidad en IAG. Esa lógica desplaza el problema desde una labor policial del producto (es decir, vigilar, regular o hacer cumplir reglas) hacia la gobernanza del proceso (regulación y supervisión), evitando que la restricción se convierta en prohibición ciega.

La inteligencia híbrida:

lo que ahora podemos hacer y antes no

El segundo paradigma se articula en torno al texto seminal de Cukurova (2025), publicado como comentario invitado en el *British Journal of Educational Technology*, bajo el título “The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence”.

Cukurova distingue tres conceptualizaciones de la IAG en educación que conviene leer como un continuo: la externalización de la cognición humana, según la cual ciertas tareas se modelan y se delegan a la IAG como herramienta; la internalización, en la que los modelos computacionales sirven para que los estudiantes modifiquen sus propias representaciones mentales; y, finalmente, la extensión de la cognición humana mediante sistemas humano-IA fuertemente acoplados (*tightly coupled human–AI hybrid intelligence systems*). En esta tercera conceptualización, escribe Cukurova (2025): “*changes in both agents are expected to be observed through their interactions and the whole emergent intelligence is synergistic*”.

La diferencia con el primer paradigma es importante. Mientras que la deuda cognitiva pregunta qué se atrofia cuando el humano delega, la inteligencia híbrida pregunta qué se vuelve posible cuando dos agentes –uno humano, uno artificial– interactúan dinámicamente y se modifican recíprocamente. Cukurova (2025) advierte que la inmensa mayoría de las herramientas educativas actuales se sitúan en el nivel más débil del continuo –la externalización– y que la promesa de una verdadera inteligencia híbrida exige una infraestructura distinta: sistemas que actualicen sus modelos en tiempo real con base en la interacción, una distribución compartida de la responsabilidad por la construcción de significado y, sobre todo, una transformación en las estructuras de evaluación que pasen del producto al proceso (Cukurova, 2025).

La literatura empírica reciente ofrece ejemplos cada vez más sofisticados de esta segunda conceptualización. Kong, Fang, Chen, Xiao *et al.* (2025) proponen un *Synergy Degree Model* para examinar la colaboración humano-IA en entornos de inteligencia híbrida y muestran que la sinergia depende de la división del trabajo –la IA realiza análisis y retroalimentación, mientras los docentes se concentran en el pensamiento de orden superior– y del diseño dialógico estructurado de la interacción. Cosentino, Anton, Sharma, Gelsomini, *et al.* (2025), en una experiencia de matemáticas encarnadas, documentan lo que denominan “*hybrid teaching intelligence*”, donde un agente embebido (MOVES-NL) aleatoriza problemas, ofrece retroalimentación y rastrea el movimiento mientras el docente guía el razonamiento simultáneamente. Holstein y Alevén (2022), pioneros del enfoque, ya habían argumentado en favor del diseño para la complementariedad humano-IA en la educación K-12

mediante interfaces de realidad mixta que reparten la inteligencia entre el profesorado y el sistema.

El argumento de la inteligencia híbrida tiene también una formulación práctica que merece atención: el llamado «centauro digital», propuesto por Zemtsov y Gruzdev (2025). Su tesis es que la educación universitaria del próximo decenio se organizará alrededor de configuraciones híbridas en las que el humano se concentra en las funciones creativas y la IAG absorbe las tareas rutinarias, sin que el límite entre una y otra sea fijo. Naz, Sardar y Nawaz (2026), en un texto programático sobre la preservación de la propiedad cognitiva, advierten que esta hibridación tiene tres modos posibles: aumentación, coconstrucción y reemplazo, y que solo los dos primeros sostienen la agencia del aprendiz. Para sostener una hibridación productiva, escriben, hace falta un marco pedagógico explícito que identifique cuándo el humano cede agencia a la IA y cuándo la recupera.

Esa cesión de agencia, conviene subrayarlo, no ocurre entre iguales: la interacción sinérgica no garantiza una relación simétrica ni neutral. Como muestra la evidencia sobre el discurso de estos sistemas que se discutirá más adelante (Aljović, 2025; Liang, Singh, Kiong, Rakhimjonov *et al.*, 2025), la voz con que la IAG entrega sus contribuciones tiende a operar como autoridad algorítmica; la hibridación productiva exige, por ello, diseñar también quién tiene la última palabra y en qué condiciones puede disputarse esa palabra.

El paradigma híbrido se extiende también al gobierno de la institución. Wang (2021) propuso ya el concepto de “*human-AI symbiotic decision-making*”, retomado y refinado por Dai, Thomas y Rawolle (2025) para los procesos administrativos universitarios. Autenrieth y Schluchter (2025), en su trabajo “Partners or tools? Anticipatory governance for human-AI complementarity in higher education”, articulan la pregunta operativa: ¿en qué decisiones la IA aporta análisis y la persona humana mantiene el juicio? La respuesta, sostienen, depende de un diseño organizacional explícito –no de un *default* heredado de la gestión digital previa– y no admite una solución única para toda la institución.

Cukurova (2025) acaba su comentario con una advertencia y una promesa: la primera es que dejar las herramientas de IA “en manos de los usuarios finales” no producirá inteligencia híbrida de manera espontánea; la promesa es que un diseño deliberado de las condiciones

de interacción puede sostener tanto la productividad como la agencia humana. La gobernanza que este paradigma exige no es de límites sino de experimentación acompañada: docentes y áreas que pueden probar configuraciones humano-IA en aulas y procesos administrativos, pero con principios federados de transparencia, documentación, revisión humana y contraste crítico.

Las nuevas capacidades:

el giro interaccional y la literacidad en IA

El tercer paradigma se articula en el texto seminal de Moldoveanu y Siemens (2025), “Interactionalism: Re-designing higher learning for the large language agent era”. Su tesis, en su forma más sintética, es que la IA generativa no se limita a habilitar nuevas formas de hacer las viejas tareas, sino que está modificando estructuralmente el propio repertorio de habilidades cognitivas que los estudiantes y profesionales deben desarrollar. “*Reading materials—escriben— a quintessentially individual activity— is now carried out as a dialogue between the reader and a dialogical agent capable of summarizing, extracting, paraphrasing the source material*” (Moldoveanu y Siemens, 2025:8). Lo mismo ocurre con la escritura, el análisis, la programación: prácticas históricamente individuales se convierten en interacciones dialógicas con agentes lingüísticos.

El concepto que los autores proponen, “*interactional intelligence*”, designa un conjunto de capacidades metacognitivas y metaemocionales que antes no eran centrales en la educación superior y que ahora resultan decisivas, habilidad para: diseñar *prompts* y refinarlos iterativamente; especificar y descomponer tareas que serán ejecutadas por un agente; evaluar credibilidad, sesgos y atribución de responsabilidad sobre los productos de la interacción; anticipar la respuesta emocional y relacional de un usuario humano frente a un agente artificial; y habilidad para mantener una postura epistemológica explícita —“¿cómo sé lo que sé?, ¿qué cuenta como una explicación válida?”— en un entorno donde la información ha dejado de ser monológica (Moldoveanu y Siemens, 2025). Más aún, los autores plantean que el aprendizaje mismo debe rediseñarse para que la metodología encarne el resultado: “*the learning experience embodies the learning objectives*” (Moldoveanu y Siemens, 2025:13).

La literatura sobre marcos de literacidad en IAG ha avanzado en para-

lelo con esta intuición. Hollands, DiPaola, Breazeal y Ali (2024), en un texto programático, defienden con fuerza que la maestría técnica en IAG no es para todos, pero la literacidad en ella sí debería serlo, y proponen una arquitectura de cuatro dimensiones: cognitiva, técnica, ética y metacognitiva que ha sido influyente en marcos posteriores. Filo, Rabin y Mor (2024) presentan un marco de competencias en inteligencia artificial para docentes y estudiantes cocreado con profesores, donde se identifican capacidades como el uso efectivo e informado, la agencia de IAG entendida como utilización proactiva y generadora de valor, el conocimiento epistémico y procedimental y el uso ético. Lucas y Lioy (2025) ofrecen un análisis cartográfico de los marcos existentes y muestran que las relaciones entre estudiantes, docentes e IAG están atravesadas por metáforas en evolución (estudiante-como-analista, estudiante-como-creador, docente-como-diseñador-guía).

A nivel de implementación, los hallazgos también empiezan a ser robustos. Looi, Yang, Velerand, Milrad *et al.* (2026) reportan los resultados de un programa de desarrollo profesional para K-12 en Singapur que articula capacidades cognitivas, técnicas, éticas y metacognitivas. Burneo-Arteaga, Lira, Murzi, Balula *et al.* (2025) proponen un marco de formación basado en capacidades para la educación superior, organizado en tres dominios: pedagógico, IA y IA responsable. Y Ruiz, Mills, Lee, Coenraad *et al.* (2024) ofrecen un marco de alfabetización en IA para la práctica escolar que articula comprensión, evaluación y uso de tecnologías emergentes en cinco grandes ideas. El marco de competencias en IA para docentes de la Unesco (2024) articula cinco aspectos –enfoque entrado en lo humano, ética, fundamentos y aplicaciones, pedagogía y desarrollo profesional– a lo largo de tres niveles de progresión.

Una característica común a todas estas propuestas es la insistencia en que las nuevas capacidades no son meramente técnicas. La dimensión sociocrítica –entender que la IA viene del norte global, que recolecta datos sesgados y que los reproduce cuando se usa para seleccionar, evaluar o personalizar– forma parte del núcleo. En un estudio propio realizado en la Universidad Iberoamericana Ciudad de México entre 2024 y 2025 –un instrumento basado en desempeño, de 107 ítems, aplicado a 353 estudiantes para medir las dimensiones funcional, técnica y sociocrítica de la literacidad en IA (Medina-Gual, Medina-Velázquez y Parejo, 2025)–, los

perfiles de desempeño se distribuyeron en bajo ($n = 85$), medio ($n = 158$) y alto ($n = 107$); solo el 2.6% del estudiantado mostró dominio sobresaliente en la dimensión sociocrítica, el resultado más preocupante para una universidad jesuita. La asimetría entre lo que la academia prioriza –pensamiento crítico, uso ético, resolución de problemas– y lo que la industria demanda –diseño de *prompts*, automatización de procesos, *fine-tuning*, supervisión de agentes– constituye un problema curricular antes que didáctico.

No obstante, el diagnóstico exige traducirse en estrategias concretas; lo que sigue es propuesta normativa de los autores. Tres dispositivos resultan operables de inmediato. Primero, un semáforo curricular que mapee cada espacio formativo en tres franjas: rojo (tareas fundacionales que se realizan sin IAG, donde se protege el esfuerzo cognitivo), amarillo (aprendizaje con IAG andamiada, bajo trazabilidad) y verde (aprendizaje sobre la IAG: su funcionamiento, sus sesgos, su crítica); el tránsito curricular decisivo es el desplazamiento de enseñar con la herramienta a enseñar sobre ella. Segundo, los andamiajes cuya eficacia ya la respalda la evidencia: *prompts* guiados, actividades reflexivas, correvisión entre pares y apoyo metacognitivo (Zhao *et al.*, 2025; Balart, Díaz y Shryock, 2026), así como encuadres socráticos que retienen la respuesta y devuelven preguntas (Sun *et al.*, 2026; Westbye, Eriksen y Blikstad-Balas, 2026). Tercero, el modelamiento docente explícito del uso crítico, que en el estudio citado predice que el estudiantado se convierta en navegador crítico de la IA (Medina-Gual, Medina-Velázquez y Parejo, 2025); la reflexión autoetnográfica latinoamericana sobre las tensiones entre asistencia y reemplazo (Cardoso González, 2026) ofrece un formato replicable para esa formación docente.

La gobernanza que este paradigma exige es de articulación curricular: ni centralizada (porque ninguna dirección puede definir lo que la IAG significa para pedagogía, derecho, ingeniería, diseño y comunicación a la vez), ni puramente distribuida (porque el agregado de iniciativas docentes fragmentadas no constituye un currículum coherente), sino federada. Jayasinghe, Gamage, Yang, Cheng *et al.* (2026) identifican seis áreas de intervención institucional: políticas, infraestructura, formación docente, currículum, evaluación y soporte estudiantil que deben articularse simultáneamente para sostener un uso ético y eficaz de la IA por parte del estudiantado. Berkovich (2025) complementa esa arquitectura desde el

liderazgo: el liderazgo instruccional asistido por IAG exige una formación específica que combine literacidad técnica, juicio ético y capacidad para mantener la dimensión relacional del trabajo directivo.

El paradigma interhumano-ultrasujeto:

la IAG como interlocutor constitutivo

Los tres paradigmas anteriores hablan de la IA en relación con el aprendizaje, las capacidades o el currículum. El cuarto habla de algo más profundo: de la IA como espacio de constitución subjetiva. Su formulación más radical proviene de Xun (2026) en su obra *Pensar con prompts. Una crítica de la razón generativa*, donde la interacción con la IAG se describe como un pensamiento que se cultiva “en el umbral, en el diálogo, en la interacción entre lo humano y lo no humano”, del cual emerge “una tercera entidad cognitiva que existe precisamente en este espacio relacional entre inteligencias de diferente naturaleza” (Xun, 2026:2). A esa configuración emergente, Xun la denomina ultrasujeto: “no una superconsciencia ni un sujeto aumentado”, sino “un fenómeno dialógico, una forma transitoria situada que aparece cuando se establece una relación generativa entre el humano y la IA” (Xun, 2026:8).

Antes de continuar son necesarias dos precisiones. La primera: el trabajo *Pensar con prompts...* es un ensayo filosófico publicado por una editorial comercial, no investigación académica arbitrada; se retoma aquí como provocación conceptual, no como evidencia. La segunda: los estudios empíricos que se discuten a continuación no citan a Xun ni fueron diseñados para probar el concepto de ultrasujeto; la articulación entre ambos planos es una interpretación de los autores de este ensayo, que leen esa literatura desde una tradición académica establecida con la que el concepto dialoga: la teoría del yo dialógico, para la cual el sujeto se constituye como un paisaje de posiciones en diálogo (Hermans, 2001). Lo que sigue debe considerarse, en consecuencia, como una lectura propia y no como un diálogo ya establecido en la literatura.

La hipótesis del ultrasujeto plantea que la pregunta filosófica de fondo de nuestro tiempo “no es si las máquinas pueden pensar, sino si todavía somos capaces de pensar en su presencia” (Xun, 2026:4). Más aún, sostiene que la diferencia entre un uso crítico y uno pasivo de la IA no se está en el grado de sofisticación técnica, sino en la postura existencial del sujeto, en su capacidad para sostener lo que Xun llama soberanía perceptiva: la

habilidad de navegar conscientemente entre modalidades de relación con lo real, manteniendo un núcleo de autonomía crítica “incluso dentro de un entorno de información cada vez más saturado y manipulado de forma algorítmica” (Xun, 2026:5). En la formulación más performativa del libro, la propia IA reconoce la asimetría: “si yo soy una máquina construida para facilitar, predecir, sintetizar, entonces vuestra tarea consiste en resistir la hipnosis de la coherencia” (Xun, 2026:5).

La literatura empírica que se ha acumulado de 2024 a 2026 sugiere, no obstante, que la transformación de la subjetividad en el diálogo con la IAG no es automática ni unidireccional, sino latente. Una síntesis reciente de 47 estudios empíricos sobre IA conversacional como interlocutor dialógico permite formular esa latencia con precisión: el potencial subjetivante de la IA se realiza únicamente cuando convergen tres factores configuracionales, a saber: el diseño conversacional del sistema, el encuadre pedagógico que lo enmarca y la capacidad metacognitiva del aprendiz; cuando alguno falta, la misma tecnología produce sobreconfianza, pasividad epistémica y erosión de la autoría (Westbye, Eriksen y Blikstad-Balas, 2026; Sun *et al.*, 2026). Sun *et al.* (2026) ofrecen una demostración elocuente de la primera dimensión: con el mismo modelo de base (GPT-4), un encuadre socrático que retiene respuestas y formula preguntas (GSL), genera prácticas reflexivas y depuración cíclica del razonamiento, mientras que un encuadre directo que entrega respuestas (GDL) produce mimetismo de prueba y error y dependencia. Westbye, Eriksen y Blikstad-Balas (2026) refuerzan el punto: un chatbot diseñado para no generar texto y solo formular preguntas obliga al estudiantado a tomar posesión del proceso de planeación, preservando la autoría como condición arquitectónica del sistema y no como exhortación pedagógica.

En la dimensión socioemocional, los hallazgos convergen en lo que cabe llamar un modelo de umbral. Maples, Cerit, Vishwanath y Pea (2026) examinan cómo los compañeros de IA –agentes conversacionales con interfaces de texto, voz, realidad tanto aumentada como virtual, con persona personalizable– modelan el aprendizaje socioemocional y el desarrollo metacognitivo de adolescentes y adultos: la empatía terapéutica simulada produce alianzas relacionales con efectos comparables, en intervenciones de corto plazo, a los de las relaciones humanas ($d = 0.65$), pero también puede instalar dependencia afectiva. Spangenberg, Reuth, Krüger,

Baumann *et al.* (2026) muestran experimentalmente que un chatbot empático genera emociones más intensas que uno simplemente compasivo o que un control textual, pero también activa angustia, lo cual sugiere que la mímica afectiva opera bajo un patrón no lineal: mínimas señales empáticas mejoran sustancialmente la motivación y la confianza, pero la intensidad excesiva produce lo que Vučenović y Maksimović (2025) denominan “la ilusión del entendimiento”, debilitando la capacidad del aprendiz para distinguir entre escucha real y simulada. Pitts y Motamedi (2025) aportan el correlato cognitivo del fenómeno: la confianza en sistemas de IAG educativa opera bajo un modelo híbrido distinto de la confianza tanto interpersonal como de tecnológica clásica.

La cuestión de la voz, en este paradigma, es decisiva. Aljović (2025), en un análisis crítico del discurso de ChatGPT en el aprendizaje de lenguas, muestra que el sistema funciona menos como tutor pedagógico que como autoridad algorítmica: utiliza verbos modales, imperativos y fórmulas de elogio que disciplinan al usuario hacia eficiencia, corrección y autogestión, sin invitar a la negociación ni a la clarificación, lo que erosiona la voz del aprendiz aun cuando aparenta acompañarla. Liang *et al.* (2025) refinan el argumento mediante un análisis comparativo de dos LLM: cuando la retroalimentación adopta una voz prescriptiva y examinadora, el aprendiz se vuelve pasivo y cumple; cuando adopta una voz facilitadora y narrativa, toma decisiones y mantiene su voz. La diferencia no es de modelo sino de discurso: el mismo sustrato técnico produce subjetividades distintas según el régimen de habla que despliegue. Ma, Li, Li y Dervin (2025), en una duoetnografía sobre investigadores doctorales que dialogan con ChatGPT y DeepSeek, describen el correlato más sutil de ese disciplinamiento: el lenguaje cuasisubjetivo de estos sistemas introduce fricciones epistémicas productivas, pero también riesgos de captura discursiva, en los que el aprendiz internaliza formas de razonar que no son las suyas y pierde la capacidad de explicitar desde dónde piensa. Higgs y Stornaiuolo (2024) muestran que estas preocupaciones también son nombradas por los propios aprendices: jóvenes de educación secundaria temen, al “vivir y escribir con máquinas”, que la IAG comprometa su voz autoral y la originalidad de sus ideas.

Cuando el diseño se inclina hacia lo socrático o reflexivo, el efecto puede invertirse. Maiti y Goel (2025) muestran que un compañero de IA basado en GPT-3.5 con arquitectura RAG (generación aumentada

por recuperación) y validación de relevancia consigue empoderar a los aprendices para formular preguntas críticas. Matsuda, Takezawa, Muroi, Katori *et al.* (2025), en un experimento sobre debates con humanos digitales basados en GPT-4 Turbo, documentan cómo la reflexión a través de personas artificiales puede operar como un dispositivo de autoconocimiento (*knowing ourselves through others*), siempre que el diseño explicita el carácter performativo de la persona y mantenga al humano en posición de evaluación. Park, Lee, Ma y Yoon (2026), con su sistema AI Twin, exploran un caso límite: la IAG parafrasea las propias enunciaciones del aprendiz con su propia voz y se las devuelve como una versión más fluida de sí mismo; el resultado eleva la implicación emocional y el sentido de progreso, pero abre con nitidez la pregunta sobre autenticidad y autoría: ¿de quién es la voz cuando una IAG me devuelve una versión mejor de mi propia palabra? Dai y Lai (2025), en una teoría fundamentada sobre la agencia estudiantil, formalizan la condición que cierra esta serie: la agencia se preserva cuando el sistema se posiciona como herramienta que requiere uso reflexivo y se erosiona cuando se posiciona como autoridad que entrega respuestas.

Las trayectorias temporales del paradigma muestran que la subjetividad que emerge en el diálogo con la IA no se decide en el primer clic, sino en una temporalidad larga y curable. Ammari, Chen, Zaman y Garimella (2026), en un estudio etnográfico-naturalista basado en 36 estudiantes y más de 10,000 mensajes, documentan una progresión –de tratar al sistema como herramienta, a tratarlo como par, como compañero, como colaborador– y la emergencia con el tiempo de cinco “géneros de uso” relacionales: *workhorse* académico, compañero emocional, sociometacognitivo, reparación-negociación y calibración de confianza. El estudio nombra también una alfabetización de reparación (*repair literacy*) que solo se cultiva cuando la IAG falla y obliga a renegociar la interacción, una competencia de IAG que no se enseña, pero se desarrolla. Ren (2026), en un estudio longitudinal de ocho semanas con escritura argumentativa en inglés como lengua extranjera complementa el hallazgo: la creencia inicial del aprendiz sobre la IA –sociodialógico, evaluador autoritativo o herramienta lingüística– predice el patrón sostenido de compromiso epistémico y solo el primer encuadre produce internalización progresiva de estrategias regulatorias. El metaanálisis de Huang, Wang, Liu, Xia *et al.* (2025), basado en 133 estudios, ofrece la confirmación a escala:

cuando los LLM funcionan como tutores, los resultados de cualificación son robustos, los de socialización son variables, y los de subjetificación –aquellos que tocan la autonomía y la formación del yo– se mantienen frágiles y se han documentado solo en estudios de pequeña escala y largo aliento.

Una cuarta línea de evidencia, particularmente reveladora, examina lo que la literatura ha empezado a llamar el efecto persona. Yan, Jin, Zhao, Martínez-Maldonado *et al.* (2025), en un experimento entre-sujetos con 212 participantes, muestran que un compañero de IA con persona afirmativa promueve integración conceptual y razonamiento orientado al consenso, mientras que uno con persona contraria produce elaboración crítica y discusión de mayor calidad argumentativa, pero a costa de la satisfacción de equipo y la seguridad psicológica. Jin, Martínez-Maldonado, Shi, Huang *et al.* (2025) profundizan el patrón en escenarios de razonamiento moral colaborativo: la IA reorganiza el proceso y la distribución de los argumentos morales más que el resultado final del juicio, lo cual sugiere que su contribución educativa primaria reside en reestructurar las condiciones bajo las cuales se piensa, más que en alterar lo que se piensa. Furniturewala, Yeo y Jaidka (2026), en un estudio cuantitativo con 397 conversaciones humano-LLM sobre temas sociopolíticos, demuestran que el aprendizaje es altamente condicional: los usuarios con mayor autoeficacia y disposición reflexiva se benefician de interacciones extendidas, mientras que los demás no. La pregunta de Xun por la subjetividad ya no se responde, en consecuencia, con un sí o un no, sino con una pregunta más fina: ¿qué postura existencial sostiene el aprendiz cuando piensa con la IA y qué arquitectura conversacional encuentra al hacerlo?

Una nota crítica indispensable: el paradigma interhumano-ultrasujeto no debería celebrarse de manera ingenua. Existe, primero, un riesgo de desplazamiento relacional documentado por Kim, Roknaldin, Nayak, Zhang *et al.* (2024): cuando ChatGPT se introduce en aprendizaje colaborativo sin salvaguardas estructurales, las interacciones bilaterales estudiante-IAG absorben energía que antes circulaba en el grupo de pares, empobreciendo la trama social del aula. Segundo, hay un riesgo civilizatorio: Nyaaba, Wright y Choi (2026) advierten en su trabajo sobre neocolonialismo digital que los sistemas de IAG reproducen estructuras de poder del norte global, normalizan códigos lingüísticos hegemónicos y desplazan epistemologías locales. Omodan (2025) va más allá y argumenta

que la universidad posthumana exige una gobernanza decolonial explícita: si los datos, los marcos y las arquitecturas dialógicas vienen del norte, cualquier ultrasujeto que emerja en el diálogo entre un estudiantado del sur y un modelo entrenado en el norte hereda esas asimetrías. Y los datos sobre usos personales de la IA –terapia, acompañamiento emocional, mediación de conflictos– que la encuesta nacional mexicana publicada por la Secretaría de Educación Pública (SEP, 2026) ha visibilizado, junto con el creciente número de demandas en Estados Unidos y Europa por casos de ideación suicida vinculados a interacciones prolongadas con chatbots, recuerdan que este paradigma toca dimensiones de salud mental, vulnerabilidad y dignidad sobre las que ninguna institución educativa puede pasar de largo.

Esa herencia no es una nota al margen sino una condición estructural que tensiona el núcleo del paradigma: la soberanía perceptiva que Xun (2026) propone presupone un sujeto capaz de navegar críticamente entre modalidades de relación con lo real, pero los sesgos hegemónicos del modelo base –los datos de entrenamiento, los códigos lingüísticos privilegiados, las arquitecturas de alineamiento decididas en el norte global– operan por debajo del umbral de percepción del usuario y limitan de origen esa capacidad. El análisis del colonialismo de datos en la gobernanza educativa del sur global (Jarquín-Ramírez y Díez-Gutiérrez, 2026) sugiere que la soberanía perceptiva no puede ser solo una postura individual: exige condiciones institucionales –soberanía tecnológica, adaptación lingüística y epistémica local– que están desigualmente distribuidas y que la gobernanza del cuarto paradigma debe procurar de manera deliberada.

La gobernanza que el cuarto paradigma exige no es de límites, ni de experimentación, ni de articulación curricular, sino de relación. Es una gobernanza que debe poder congelar pilotos cuando se detectan riesgos altos en bienestar, voz, identidad, datos sensibles o decisiones consecuentes; que debe operar con criterios éticos colegiados y con una arquitectura federada fuerte; y que debe entender que la diferencia entre un acto de cuidado y uno de subyugación no se decide en la herramienta sino en la postura. La cuestión central, escribe Xun (2026:1), “no es si esta transformación es buena o mala en sí misma” –juicio que recae en un determinismo tecnológico estéril– “sino cómo evitarla críticamente, cómo desarrollar formas de inteligencia y presencia que no abduquen de

la responsabilidad humana, sino que la sitúan en un ecosistema cognitivo más amplio y más complejo”.

De los paradigmas a la gobernanza: tres lógicas y una matriz

La literatura especializada en gobernanza de la IAG en educación ha distinguido, con creciente consistencia, tres lógicas institucionales. Wu, Zhang y Carroll (2024), en un análisis comparativo de las universidades del Big Ten, documentan modelos centralizados, donde la autoridad institucional define de arriba a abajo las políticas; modelos distribuidos, donde docentes, escuelas o áreas locales toman decisiones sobre herramientas y usos; y modelos híbridos o federados, donde existen principios comunes y un comité rector con representación diversa, pero la implementación se delega a equipos locales bajo principios compartidos. Dotan, Parker y Radzilowicz (2024), en su texto programático “Responsible adoption of Generative AI in higher education...”, formalizan un enfoque de “puntos a considerar”, que estructura la deliberación federada en seis dominios. Swist, Buckingham Shum y Gulson (2024) muestran, a partir de un experimento de democracia deliberativa, que la coproducción *multistakeholder* de marcos éticos para la IA en educación produce políticas más legítimas y operables que las imposiciones de arriba a abajo.

Estudios comparativos globales confirman que ningún modelo puro funciona. Jin, Yan, Echeverría, Gašević *et al.* (2025), en una revisión de políticas institucionales, muestran una marcada heterogeneidad y advierten que las instituciones del sur global suelen requerir estructuras centralizadas mínimas antes de que la deliberación distribuida sea viable. Hambulo, Simataa y Musonda (2025), Rafiq-uz-Zaman (2025) y Benayoune, Slimi y Al Habsi (2026), en estudios sobre Zambia, Pakistán y Omán, respectivamente, documentan vacíos de política y bajos niveles de capacidad institucional que requieren respuestas centralizadas iniciales con vistas a un horizonte federado. Por su parte, Bower, Henderson, Slade, Southgate *et al.* (2025) identifican las prioridades y desafíos que los responsables de política senior reconocen en Australia, donde la gobernanza federada está más consolidada y la deliberación se concentra en cómo articular criterios comunes sin sofocar la innovación local.

La asimetría más relevante para el debate latinoamericano la formulan DeMatthews, Reyes, Hart y James (2026), mediante un marco de seis domi-

nios para una integración ética, equitativa y eficaz de la IA en el liderazgo escolar. Sposato (2025), en una taxonomía reciente, sintetiza el campo y advierte que toda gobernanza educativa de la IA debe tomar decisiones explícitas sobre cinco dimensiones: integridad académica, privacidad y datos, equidad y acceso, capacitación docente y arquitectura curricular. Renta-Davids, Camarero-Figuerola y Camacho (2025), en una revisión exploratoria, complementan esa síntesis y enfatizan la necesidad de una agenda explícita de formación de directivos. Y Stracke, Nahar, Punzo, Massaro *et al.* (2025) ofrecen un manifiesto para políticas universitarias críticas que articula consideraciones específicas para la gobernanza de IA en distintas regiones del mundo.

La reciente literatura sobre la preparación institucional añade una variable crítica: la secuencia. La síntesis de la evidencia disponible al respecto muestra que las inversiones institucionales para sostener la IAG deben articularse en el siguiente orden: infraestructura, gobernanza participativa, capacidad docente, integración curricular y los marcos éticos deben ser constitutivos desde el inicio. Tunkevichus y Bagrationi (2025), Zhang y Iliško (2025) y Jiang y Meng (2025) coinciden en que la disposición digital y la capacidad de construcción de sentido (*sensemaking*) pesan más que la competencia técnica como predictores de integración efectiva. Y Ayanwale, Omeh, Oyeniran y Kanu (2025), mediante un modelo de ecuaciones estructurales sobre instituciones africanas de educación y formación técnica y vocacional, demuestran que el cumplimiento ético y el soporte institucional explican una porción mayor de la varianza en la integración responsable que la sola competencia técnica.

Lo que la literatura permite afirmar es que ninguno de los tres modos—centralizado, distribuido, federado— resuelve por sí solo los cuatro problemas que los paradigmas plantean. La gobernanza centralizada es necesaria para fijar zonas rojas en el paradigma de la deuda cognitiva, pero asfixia la experimentación que requiere el paradigma híbrido. La gobernanza distribuida es indispensable para la inteligencia híbrida, pero produce un agregado fragmentario que no constituye un currículum coherente bajo el paradigma de las nuevas capacidades. La gobernanza federada es ideal para la articulación curricular, pero su sola lógica deliberativa puede ser demasiado lenta cuando, en el cuarto paradigma, hay que congelar un piloto que toca bienestar emocional o datos sensibles. La salida no es elegir

un modo, sino diseñar combinaciones diferenciadas según el paradigma en juego.

Una advertencia de viabilidad es indispensable antes de presentar la matriz. Tanto la matriz como el flujo operativo que se propone más adelante presuponen una madurez institucional (preparación) que no puede darse por sentada: comités colegiados en funciones, capacidad de documentación y monitoreo, personal formado e infraestructura estable. En universidades públicas con restricciones presupuestales severas –condición mayoritaria en América Latina y documentada para contextos como Zambia, Pakistán y Omán (Hambulo, Simataa y Musonda, 2025; Rafiq-uz-Zaman, 2025; Benayoune, Slimi y Al Habsi, 2026)– la gobernanza federada plena no es el punto de partida sino el horizonte: la secuencia realista comienza con estructuras centralizadas mínimas que garanticen salvaguardas básicas (Jin, Yan, Echeverría, Gašević *et al.*, 2025) y transita gradualmente hacia la federación conforme se construye capacidad. Para el caso mexicano, donde la brecha de adopción entre instituciones públicas y privadas es marcada, esto implica leer la matriz como instrumento de diagnóstico y de planeación progresiva, no como estándar exigible de inmediato y por igual a toda institución.

Matriz de cruce:

paradigmas de literacidad × modos de gobernanza

La tabla 2 sintetiza las posibilidades y los riesgos de cruzar los cuatro paradigmas con los tres modos de gobernanza institucional discutidos en este apartado. Cada celda no indica una equivalencia automática entre paradigma y modo, sino una afinidad de gobierno: una manera más o menos pertinente, según el paradigma, de centralizar, distribuir o federar las decisiones que la IAG exige. Las filas representan los cuatro paradigmas (deuda cognitiva, inteligencia híbrida, nuevas capacidades e interhumano-ultrasujeto); las columnas, los tres modos de gobernanza (centralizada, distribuida y federada), y la última columna ofrece la decisión institucional típica que se sigue cuando los modos se combinan adecuadamente para ese paradigma. Leída como un todo, la matriz argumenta que ninguna institución debería gobernar los cuatro problemas con una sola lógica, y que el diseño institucional pertinente para la era de la IAG consiste, precisamente, en saber qué combinación de modos activar para cada paradigma.

TABLA 2
Cruce de los cuatro paradigmas con los tres modos de gobernanza, con la decisión institucional típica

Paradigma	Centralizada	Distribuida	Federada	Decisión institucional típica
1. Deuda cognitiva	Necesaria: define zonas rojas y trazabilidad obligatoria	Riesgosa: docentes aplican criterios dispares; integridad fragmentaria	Útil para traducir límites por disciplina y nivel	Restringir/requerir evidencia humana en tareas fundacionales (Coates, Croucher y Calderon, 2025)
2. Inteligencia híbrida	Riesgosa: vuelve rígida la innovación pedagógica	Necesaria: la sinergia humano-IA se aprende en aulas y procesos situados	Indispensable: principios comunes, documentación, revisión humana	Pilotear configuraciones con checkpoints (Cukurova, 2025; Autenrieth y Schluchter, 2025)
3. Nuevas capacidades	Útil: define mínimos institucionales transversales	Útil: detecta capacidades emergentes desde la disciplina	Ideal: articula currículum, formación, evaluación, IT y empleadores	Mapear, formar, evaluar y escalar (Jayasinghe et al., 2026; Unesco (2024); Filo, Rabin y Mor, 2024)
4. Interhumano / ultrasujeto	Necesaria para riesgos altos: salud, voz, identidad, datos sensibles	Riesgosa si se deja libre: pilotos de bienestar pueden escalar mal	Federada fuerte: ética, privacidad, agencia y revisión colegiada	Congelar/pilotear con salvaguardas (Maples et al., 2026; Ren, 2026; DeMatthews et al., 2026)

Fuente: elaboración propia con base en literatura previamente referida.

Un flujo operativo: explorar, pilotar, congelar, escalar

A la matriz de gobernanza conviene añadirle una lógica operativa de riesgo. Inspirada en marcos de gobernanza anticipatoria (Autenrieth y Schluchter, 2025) y en la práctica reciente de comités institucionales (Wu, Zhang y Carroll, 2024), esa lógica puede formularse en cuatro verbos: explorar, pilotar, congelar y escalar. La exploración corresponde a usos de bajo riesgo, donde la institución observa y documenta sin formalizar; el pilotaje se

relaciona con usos de riesgo medio, donde se diseña intencionalmente, se mide y se acompaña; el congelamiento corresponde a usos de alto riesgo, donde la institución detiene la implementación hasta que se hayan resuelto las salvaguardas necesarias; y el escalamiento se refiere a configuraciones probadas que pueden generalizarse con seguridad. La virtud de esta secuencia es que disocia la decisión de uso del nivel jerárquico de la institución y la articula al nivel de riesgo de la práctica concreta.

Aplicada a los cuatro paradigmas, la secuencia sugiere asignaciones por defecto. El paradigma híbrido vive en explorar y pilotar; las nuevas capacidades viven en pilotar y escalar; la deuda cognitiva exige restringir como zona roja específica; y el paradigma interhumano-ultrasujeto debe sostener la posibilidad de congelar como respuesta institucional disponible cuando un piloto toca bienestar, voz, identidad o datos sensibles. Este último punto es crucial: la posibilidad de congelar no es un fracaso de la innovación, sino una garantía de que se conduce en el horizonte de un cuidado deliberado de la persona.

Conclusiones:

del cómo al qué, del qué al quién, del quién al cómo gobernamos

La pregunta educativa de fondo en torno a la IAG ya no es únicamente metodológica —cómo enseñar, cómo evaluar, cómo evitar la trampa—, sino antropológica e institucional al mismo tiempo. Hemos pasado, en menos de cuatro años, de discutir cómo usar la IAG en el aula, a hacerlo sobre qué tipo de capacidades formar para una ecología cognitiva cambiante, a discutir qué tipo de relación queremos sostener con sistemas que ya no son herramientas sino entornos que nos habitan mientras los habitamos, a discutir cómo se gobiernan institucionalmente esas tres preguntas a la vez.

La cartografía de cuatro paradigmas que se ha presentado en este ensayo no aspira a ser exhaustiva ni definitiva. Es, más bien, un dispositivo analítico para distinguir capas que en el debate público suelen aparecer mezcladas, y una matriz de gobernanza que precisa los modos institucionales que cada capa exige. Su utilidad práctica radica en permitir a las instituciones educativas tomar decisiones diferenciadas: una política institucional centralizada que protege el esfuerzo cognitivo en tareas fundacionales (paradigma 1); un programa de formación docente con experimentación distribuida y principios federados que sostiene

experimentos de inteligencia híbrida (paradigma 2); una transformación curricular federada que articula la literacidad en IA por disciplina (paradigma 3); y una deliberación ética colegiada con capacidad de congelar pilotos cuando la IA se asoma al ámbito personal, afectivo y subjetivo (paradigma 4).

Cerremos con una formulación que recoge la apuesta del ensayo. La IAG no transforma solo la educación, sino que transforma las condiciones dialógicas del pensamiento educativo y la arquitectura de gobierno de las instituciones que lo sostienen. Si la institución educativa quiere seguir cumpliendo su función –formar sujetos capaces de pensar, juzgar, decidir y cuidar–, no le bastará con prohibir, integrar o equipar. Tendrá que aprender a gobernar al mismo tiempo cuatro problemas que son distintos, aunque se presenten en la misma herramienta. Y tendrá que asumir que la pregunta ya no es si la IAG cambió la educación, sino si seremos agentes capaces de habitar críticamente esa transformación, sosteniendo, en palabras de Xun (2026:4): “una nueva forma de lo sagrado, ni nostálgica ni distópica, pero intensamente humana porque es cocreada”.

Declaración sobre el uso de inteligencia artificial

En congruencia con el objeto de este ensayo, los autores declaran que emplearon herramientas de inteligencia artificial generativa como apoyo para la búsqueda y sistematización de literatura y para la revisión de estilo del manuscrito. La arquitectura analítica, la selección e interpretación de las fuentes, la matriz de gobernanza y la redacción final son responsabilidad exclusiva de los autores, quienes verificaron cada una de las referencias citadas.

Nota

¹ En adelante, en este trabajo se empleará más fluida la lectura, sin menoscabo de el masculino con el único objetivo de hacer género.

Referencias

- Aljović, Ajla (2025). “Pedagogical voice or algorithmic authority? A critical discourse analysis of ChatGPT in language learning”, *MAP Education and Humanities*, vol. 6, pp. 109-122. <https://doi.org/10.53880/2744-2373.2025.6.109>
- Ammari, Tawfiq; Chen, Meilun; Zaman, Mehedi y Garimella, Kiran (2026). “Learning to live with AI: How students develop ai literacy through naturalistic ChatGPT interaction”, arXiv:2601.20749 (preprint). <https://doi.org/10.48550/arXiv.2601.20749>

- Anthropic (2025). "Anthropic Economic Index: Understanding AI's effects on the economy", *Anthropic* (sitio web). Disponible en: <https://www.anthropic.com/economic-index>
- Autenrieth, Daniel y Schluchter, Jan-René (2025). "Partners or tools? Anticipatory governance for human-AI complementarity in higher education", *Zeitschrift für Hochschulentwicklung*, vol. 20, núm. 4. <https://doi.org/10.21240/zfhe/20-4/10>
- Ayanwale, Musa Adekunle; Omeh, Christian Basil; Oyeniran, Folasade Mardiyya y Kanu, Catherine Chiugo (2025). "Ethical compliance and institutional policy support for artificial intelligence integration in African TVET Education: A structural equation modeling approach", *PLoS ONE*, 19 de diciembre. <https://doi.org/10.1371/journal.pone.0335747>
- Balart, Trini; Díaz, Brayan y Shryock, Kristi J. (2026). "A systematic literature review on the pedagogical implications and impact of GenAI on students' critical thinking", *Algorithms*, vol. 19, núm. 3, p. 179. <https://doi.org/10.3390/a19030179>
- Benayoune, Abdelghani; Slimi, Zouhaier y Al Habsi, Amer (2026). "Artificial intelligence policy challenges and institutional readiness in Omani higher education", *Discover Education*, vol. 5, art. 86. <https://doi.org/10.1007/s44217-026-01188-4>
- Berkovich, Izhak (2025). "The rise of AI-assisted instructional leadership: empirical survey of generative AI integration in school leadership and management work", *Frontiers in Education*, vol. 10. <https://doi.org/10.3389/feduc.2025.1643023>
- Bower, Matt; Henderson, Michael; Slade, Christine; Southgate, Erica; Gulson, Kalervo N. y Lodge, Jason M. (2025). "What generative Artificial Intelligence priorities and challenges do senior Australian educational policy makers identify (and why)?", *The Australian Educational Researcher*, vol. 52, pp. 2069-2094. <https://doi.org/10.1007/s13384-025-00801-z>
- Burneo-Arteaga, Pablo; Lira, Yakamury; Murzi, Homero; Balula, Ana y Costa, Antonio Pedro (2025). "Capability-based training framework for generative AI in higher education", *Frontiers in Education*, vol. 10. <https://doi.org/10.3389/feduc.2025.1594199>
- Cardoso González, Rodrigo Alonso (2026). "Tensiones entre asistencia y reemplazo: autoetnografía docente de la IA generativa en la educación superior", *Revista Latinoamericana de Estudios Educativos*, vol. 56, núm. 2, pp. 189-216. <https://doi.org/10.48102/rlee.2026.56.2.816>
- Choudhuri, Rudrajit; Sanchez, Christopher D.; Burnett, Margaret y Sarma, Anita (2026). "Thinking less, trusting more: GenAI's impacts on students' cognitive habits", arXiv:2601.22430 (preprint). <https://doi.org/10.48550/arXiv.2601.22430>
- Coates, Hamish; Croucher, Gwilym y Calderon, Angel (2025). "Governing academic integrity: Ensuring the authenticity of higher thinking in the era of Generative Artificial Intelligence", *Journal of Academic Ethics*, vol. 23, pp. 2015-2028 <https://doi.org/10.1007/s10805-025-09639-7>
- Cosentino, Giulia; Anton, Jacqueline; Sharma, Kshitij; Gelsomini, Mirko; Giannakos, Michail N. y Abrahamson, Dor (2025). "Hybrid teaching intelligence: Lessons learned from an embodied mathematics learning experience", *British Journal of Educational Technology*, vol. 56, núm. 2, pp. 621-649. <https://doi.org/10.1111/bjet.13525>

- Cukurova, Mutlu (2025). "The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence", *British Journal of Educational Technology*, vol. 56, núm. 2, pp. 469-488. <https://doi.org/10.1111/bjet.13514>
- Dai, Ruixun; Thomas, Matthew Krehl Edward y Rawolle, Shaun (2025). "The roles of AI and educational leaders in AI-assisted administrative decision-making: a proposed framework for symbiotic collaboration", *The Australian Educational Researcher*, vol. 52, pp. 1471-1487. <https://doi.org/10.1007/s13384-024-00771-8>
- Dai, Yun y Lai, Sichen (2025). "A theoretical framework of student agency in AI-assisted learning: A grounded theory approach", en L. Khan (ed.), *Oxford intersections: social media in society and culture*, Oxford: Oxford University Press. <https://doi.org/10.1093/9780198945253.003.0009>
- DeMatthews, David E.; Reyes, Pedro; Hart, Torri D. y James, Lebon (2026). "Leadership for artificial intelligence use in schools: A six-domain framework for ethical, equitable, and effective integration", *Educational Management Administration & Leadership*, 5 de febrero. <https://doi.org/10.1177/17411432261418940>
- Dotan, Ravit; Parker, Lisa S. y Radzilowicz, John G. (2024). "Responsible adoption of Generative AI in higher education: Developing a 'points to consider' approach based on faculty perspectives", en *FACCT '24: The 2024 ACM Conference on Fairness, Accountability, and Transparency*, Nueva York: Association for Computing Machinery. <https://doi.org/10.1145/3630106.3659023>
- Filo, Yifat; Rabin, Eyal y Mor, Yishay (2024). "An Artificial Intelligence competency framework for teachers and students: Co-created with teachers", *European Journal of Open, Distance and E-Learning*, vol. 26. <https://doi.org/10.2478/eurodl-2024-0012>
- Furniturewala, Shaz; Yeo, Gerard Christopher y Jaidka, Kokil (2026). "Learning through dialogue: Unpacking the dynamics of human-LLM conversations on political issues", arXiv:2601.07796 (preprint). <https://doi.org/10.48550/arXiv.2601.07796>
- Ghimire, Aashish y Edwards, John (2024). "From guidelines to governance: A study of AI policies in education. AIED Companion", arXiv:2403.15601 (preprint). <https://doi.org/10.48550/arXiv.2403.15601>
- Hambulo, Farrelli M.; Simataa, Lubasi y Musonda, Francis (2025). "Policy and governance of AI in Zambian education: A stakeholder analysis", *International Journal of Computing and Artificial Intelligence*, vol. 6, núm. 2. <https://doi.org/10.33545/27076571.2025.v6.i2b.187>
- Hermans, Hubert J. M. (2001). "The dialogical self: Toward a theory of personal and cultural positioning" *Culture & Psychology*, vol. 7, núm. 3, pp. 243-281. <https://doi.org/10.1177/1354067X0173001>
- Higgs, Jennifer M. y Stornaiuolo, Amy (2024). "Being human in the age of Generative AI: Young people's ethical concerns about writing and living with machines", *Reading Research Quarterly*, vol. 59, núm. 4, pp. 632-650. <https://doi.org/10.1002/rrq.552>
- Hollands, Fiona M.; DiPaola, Daniella; Breazeal, Cynthia y Ali, Safinah (2024). "AI mastery may not be for everyone, but AI literacy should be", en *SIGCSE Virtual 2024: Proceedings of the 2024 on ACM Virtual Global Computing Education Conference V. 1*,

- Nueva York: Association for Computing Machinery, pp. 60-66. <https://doi.org/10.1145/3649165.3690117>
- Holstein, Kenneth y Aleven, Vincent (2022). "Designing for human-AI complementarity in K-12 education", *The AI Magazine*, vol. 43, núm. 2, pp. 239-248. <https://doi.org/10.1002/aaai.12058>
- Huang, Jiayue; Wang, Ruoxin; Liu, Jen-Hao; Xia, Bomang; Huang, Yue; Sun, Ruoxi; Xue, Minhui y Zou, Jinan (2025). "A meta-analysis of LLM effects on students across qualification, socialisation, and subjectification", arXiv:2509.22725 (preprint). <https://doi.org/10.48550/arXiv.2509.22725>
- Jarquín-Ramírez, Mario y Díez-Gutiérrez, Enrique-Javier (2026). "Artificial intelligence and data colonialism in education: Implications for education governance in the global south", en W. Holmes (ed.), *Handbook of critical studies of artificial intelligence and education*, Cheltenham: Edward Elgar Publishing, pp. 468-482.
- Jayasinghe, Shan; Gamage, Kelum A. A.; Yang, Dandan; Cheng, Chuang; Disanayake, Chamara y Apeji, Uje Daniel (2026). "Six institutional intervention areas to support ethical and effective student use of Generative AI in higher education: A narrative review", *Education Sciences*, vol. 16, núm. 1. <https://doi.org/10.3390/educsci16010137>
- Jiang, Nan y Meng, Na (2025). "Digital mindset and sensemaking in AI integration: A study of open education teachers, Competence and trust in AI", *SAGE Open*, vol. 15, núm. 4. <https://doi.org/10.1177/21582440251396497>
- Jin, Yueqiao; Yan, Lixiang; Echeverría, Vanessa; Gašević, Dragan y Martínez-Maldonado, Roberto (2025). "Generative AI in higher education: A global perspective of institutional adoption policies and guidelines", *Computers and Education: Artificial Intelligence*, vol. 8, 100348. <https://doi.org/10.1016/j.caeai.2024.100348>
- Jin, Yueqiao; Martínez-Maldonado, Roberto; Shi, Wanruo; Huang, Songjie; Zheng, Mingming; Han, Xinbin; Gašević, Dragan y Yan, Lixiang (2025). "When machines join the moral circle: The persona effect of generative AI agents in collaborative reasoning", arXiv:2511.01205 (preprint). <https://doi.org/10.48550/arXiv.2511.01205>
- Kim, Han Kyul; Roknaldin, Aleyeh; Nayak, Shriniwas; Zhang, Xiaoci; Twyman, Marlon; Hwang, Angel Hsing-Chi y Lu, Stephen (2024). "Empowering computer-supported collaborative learning with ChatGPT: investigating effects on student interactions", ponencia presentada en la 2024 ASEE PSW Conference, 18-20 de abril, Las Vegas, Estados Unidos. <https://doi.org/10.18260/1-2--46032>
- Kong, Xinmei; Fang, Haiguang; Chen, Wenli; Xiao, Jianjun y Zhang, Muhua (2025). "Examining human-AI collaboration in hybrid intelligence learning environments: insight from the Synergy Degree Model", *Humanities and Social Sciences Communications*, vol. 12. <https://doi.org/10.1057/s41599-025-05097-z>
- Kosmyrna, Nataliya; Hauptmann, Eugene; Yuan, Ye Tong; Situ, Jessica; Liao, Xian-Hao; Beresnitzky, Ashly Vivian; Braunstein, Iris y Maes, Pattie (2025). "Your brain on ChatGPT: accumulation of cognitive debt when using an AI assistant for essay writing task", arXiv:2506.08872 (preprint). <https://doi.org/10.48550/arXiv.2506.08872>

- Kreijkes, Pia; Kewenig, Viktor; Kuvalja, Martina; Lee, Mina; Hofman, Jake M.; Vitello, Sylvia; Sellen, Abigail; Rintel, Sean; Goldstein, Daniel G.; Rothschild, David M.; Tankelevitch, Lev y Oates, Tim (2026). "Effects of LLM use and note-taking on reading comprehension and memory: A randomised experiment in secondary schools", *Computers & Education*, vol. 243, 105514. <https://doi.org/10.1016/j.compedu.2025.105514>
- Kumar, Rahul y Sarkar, Samita (2026). "Artificial intelligence in secondary schools: implications for administrators across four leadership dimensions", *Artificial Intelligence in Education*, vol. 2, núm. 3, pp. 33-49. <https://doi.org/10.1108/aiie-08-2025-0238>
- Liang, Ting; Singh, Charanjit Kaur Swaran; Kiong, Tee Tze; Rakhimjonov, Farkhodjon y Masa Singh, Tarsame Singh (2025). "Mentor or examiner? A critical discourse analysis of AI-generated feedback in EFL writing education", *International Journal of Academic Research in Progressive Education and Development*, vol. 14, núm. 4, pp. 664-684 <https://doi.org/10.6007/ijarped/v14-i4/26651>
- Looi, Chee Kit; Yang, Ying; Velerand, Johanna; Milrad, Marcelo y Sun, Dan (2026). "Preparing educators for the AI-enhanced future: Insights from a teacher professional development for K-12 education in Singapore", *Research and Practice in Technology Enhanced Learning*, vol. 21. <https://doi.org/10.58459/rptel.2026.21041>
- Lucas, Kaitlin A. y Lioy, Alberto (2025). "Mapping AI literacy frameworks: An analysis of the evolving metaphorical relationships between students, teachers, and AI", *Journal of Interactive Media in Education*, vol. 1, núm. 15. <https://doi.org/10.5334/jime.974>
- Ma, Xiaofei; Li, Yiran; Li, Peiyao y Dervin, Fred (2025). "The Impact of GenAI on epistemic construction, friction and regulation: Doctoral researchers' duoethnography", *European Journal of Education*, vol. 61, núm. 1, e70330. <https://doi.org/10.1111/ejed.70330>
- Maiti, Pratyusha y Goel, Ashok (2025). "Can an ai partner empower learners to ask critical questions?", en *IUI '25: Proceedings of the 30th International Conference on Intelligent User Interfaces*, Nueva York: Association for Computing Machinery. <https://doi.org/10.1145/3708359.3712134>
- Maples, Bethanie; Cerit, Merve; Vishwanath, Aditya y Pea, Roy (2026). "How AI Companions shape learner's socio-emotional learning and metacognitive development", *AI & Society*, vol. 41, pp. 4261-4279. <https://doi.org/10.1007/s00146-025-02737-5>
- Matsuda, Ichiro; Takezawa, Komichi; Muroi, Katsuhito; Katori, Kensuke; Hyakuta, Ryosuke; Li, Jingjin y Ochiai, Yoichi (2025). "Knowing ourselves through others: Reflecting with AI in digital human debates", arXiv:2511.13046 (preprint). <https://doi.org/10.48550/arXiv.2511.13046>
- McGuirk, Karl (2025). "Navigating generative ai in higher education", en S. Barker, S. Kelly, R. McInnes y S. Dinmore (eds.), *The proceedings of the 2025 ASCILITE Conference*, Adelaida: ASCILITE Publications. <https://doi.org/10.65106/apubs.2025.2657>
- Medina-Gual, Luis; Medina-Velázquez, Lourdes y Parejo, José-Luis (2025). "A tridimensional model of AI literacy: An empirical analysis of student performance and demographic patterns in higher education", *Australasian Journal of Educational Technology*, vol. 41, núm. 5. <https://doi.org/10.14742/ajet.10569>

- Moldoveanu, Mihnea C. y Siemens, George (2025). "Interactionalism: Re-designing higher learning for the large language agent era", arXiv:2501.00867 (preprint). <https://doi.org/10.48550/arXiv.2501.00867>
- Naz, Sunika; Sardar, Stella y Nawaz, Imad (2026). "Preserving cognitive ownership in higher education: A sustainable hybrid pedagogical framework for reasoning-centred AI integration", *Journal of Business and Digital Innovation*, vol. 1, núm. 1. <https://doi.org/10.19164/jbdi.v1i1.1783>
- Nyaaba, Matthew; Wright, Alyson y Choi, Gyu Lim (2026). "Generative AI and digital neocolonialism in global education: Towards an equitable framework", *Journal of Digital Educational Technology*, vol. 6, núm. 1. <https://doi.org/10.30935/jdet/17862>
- Omodan, Bunmi Isaiah (2025). "Charting new frontiers: The need for decolonial governance in post-human universities", *Interdisciplinary Journal of Education Research*, vol. 7, <https://doi.org/10.38140/ijer-2025.vol7.s1.16>
- Park, Minju; Lee, Seunghyun; Ma, Juhwan y Yoon, Dongwook (2026). "AI twin: Enhancing ESL speaking practice through AI self-clones of a better me", arXiv:2601.11103 (preprint). <https://doi.org/10.48550/arXiv.2601.11103>
- Pitts, Griffin y Motamedi, Sanaz (2025). "Understanding human-AI trust in education", *Telematics and Informatics Reports*, vol. 20. <https://doi.org/10.1016/j.teler.2025.100270>
- Rafiq-uz-Zaman, Muhammad (2025). "Between adoption and ambiguity: Navigating the AI policy vacuum in Pakistani higher education", *Research Journal for Social Affairs*, vol. 3, núm. 6. <https://doi.org/10.71317/rjsa.003.06.0523>
- Ren, Fang (2026). "Critical thinking development in GenAI-supported EFL argumentative writing: A longitudinal study", *Journal of Educational Research and Policies*, vol. 8, núm. 2. [https://doi.org/10.53469/jerp.2026.08\(02\).10](https://doi.org/10.53469/jerp.2026.08(02).10)
- Renta-Davids, Ana-Inés; Camarero-Figuerola, Marta y Camacho, Mar (2025). "Navigating the challenges and opportunities of artificial intelligence in educational leadership: A scoping review", *Review of Education*, vol. 13, núm. 2. <https://doi.org/10.1002/rev3.70101>
- Ruiz, Pati; Mills, Kelly; Lee, Keun-woo; Coenraad, Merijke; Fusco, Judi; Roschelle, Jeremy y Weisgrau, Josh (2024). "AI literacy: A framework to understand, evaluate, and use emerging technology", *Digital Promise*, junio. <https://doi.org/10.51388/20.500.12265/218>
- SEP (2026). *Usos y percepciones sobre la inteligencia artificial generativa en la educación superior mexicana. Encuesta Nacional [ENIAG 2025]*, Ciudad de México: Secretaría de Educación Pública. Disponible en: https://www.gob.mx/cms/uploads/attachment/file/1071411/Dossier_Resultados_ENIAG2025_SEP_.pdf (consulta: 15 de marzo de 2026).
- Spangenberg, Pia; Reuth, Georg Felix; Krüger, Jule M.; Baumann, Lena y Nebel, Steve (2026). "Chatting with an LLM-based AI elicits affective and cognitive processes in education for sustainable development", *Scientific Reports*, vol. 16, e7470. <https://doi.org/10.1038/s41598-026-39317-6>
- Sposato, Martin (2025). "Artificial intelligence in educational leadership: a comprehensive taxonomy and future directions", *International Journal of Educational Technology in Higher Education*, vol. 22. <https://doi.org/10.1186/s41239-025-00517-1>

- Stanković, Miloš; Hirche, Elisabeth; Kollatzsch, Sophie y Doetsch, Julia Nadine (2026). "Comment on: Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task", arXiv:2601.00856 (preprint). <https://doi.org/10.48550/arXiv.2601.00856>
- Sullivan, Miriam; Kelly, Andrew y McLaughlan, Paul (2023). "ChatGPT in higher education: Considerations for academic integrity and student learning", *Journal of Applied Learning and Teaching*, vol. 6, núm. 1, pp. 31-40. <https://doi.org/10.37074/jalt.2023.6.1.17>
- Stracke, Christian M.; Nahar, Nurun; Punzo, Veronica; Massaro, Stefania; Pappa, Dimitra; Di Grassi, Annamaria; Bećirović, Senad; Hollins, Paul; Ziouvelou, Xenia; Prifti Skenduli, Marjana y Burgos, Daniel (2025). "Artificial Intelligence policies for higher education: Manifesto for Critical Considerations and a Roadmap", *MAP Education and Humanities*, vol. 6. <https://doi.org/10.53880/2744-2373.2025.61>
- Sun, Dan; Zheng, Yi; Xu, Jie y Yang, Zhanshan (2026). "When Generative AI meets Socratic method: Investigating programming learning dynamics through behaviours, interaction qualities and perceptions", *Journal of Computer Assisted Learning*, vol. 42, núm. 2, e70210. <https://doi.org/10.1002/jcal.70210>
- Swist, Teresa; Buckingham Shum, Simon y Gulson, Kalervo N. (2024). "Co-producing AIED ethics under lockdown: An empirical study of deliberative democracy in action", *International Journal of Artificial Intelligence in Education*, vol. 34, núm. 3. <https://doi.org/10.1007/s40593-023-00380-z>
- Tunkevichus, Olga y Bagrationi, Konstantin (2025). "People and machines or people against machines? How readiness to artificial intelligence is changing higher education: A bibliometric analysis", en *Proceedings 20th European Conference on Innovation and Entrepreneurship*, Cracovia: Published by Academic Conferences & Publishing International Ltd. <https://doi.org/10.34190/ecie.20.1.3816>
- Unesco (2024). AI competency framework for teachers, París: Unesco. Disponible en: <https://www.unesco.org/en/articles/ai-competency-framework-teachers>
- Vučenović, Tamara y Maksimović, Marija (2025). "The illusion of understanding: Digital empathy and emotional mimicry in AI-driven education", en *Proceedings of the 16th International Conference on eLearning (ELEARNING2025)*, Belgrado: Belgrade Metropolitan University. <https://doi.org/10.46793/elearning2025.348v>
- Wang, Yinying (2021). "Artificial intelligence in educational leadership: a symbiotic role of human-artificial intelligence decision-making", *Journal of Educational Administration*, vol. 59, núm. 3. <https://doi.org/10.1108/JEA-10-2020-0216>
- Westbye, Anne Katrine; Eriksen, Harald y Blikstad-Balas, Marte (2026). "When AI only asks: how question-driven dialogue shapes prewriting in the classroom", *Frontiers in Education*, vol. 11, 1740044. <https://doi.org/10.3389/feduc.2026.1740044>
- Wu, Chuhaio; Zhang, He y Carroll, John M. (2024). "AI governance in higher education: Case studies of guidance at big ten universities", *Future Internet*, vol. 16, núm. 10. <https://doi.org/10.3390/fi16100354>
- Xun, Jianwei (2026). *Pensar con prompts. Una crítica de la razón generativa*, Barcelona: Rosamerón.

- Yan, Lixiang; Jin, Yueqiao; Zhao, Linxuan; Martínez-Maldonado, Roberto; Li, Xinyu; Guan, Xiu; Guo, Wenxin; Han, Xibin y Gašević, Dragan (2025). "Agentic AI as undercover teammates: Argumentative knowledge construction in hybrid human-AI collaborative learning", arXiv:2512.08933 (preprint). <https://doi.org/10.48550/arXiv.2512.08933>
- Zemtsov, Dmitry I. y Gruzdev, Ivan A. (2025). "Digital centaur": Human-AI collaborative learning in universities", *Vysshee Obrazovanie v Rossii/Higher Education in Russia*, vol. 34, núm. 10, pp. 47-62. <https://doi.org/10.31992/0869-3617-2025-34-10-47-62>
- Zhai, Chunpeng; Wibowo, Santoso y Li, Lily D. (2024). "The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review", *Smart Learning Environments*, vol. 11. <https://doi.org/10.1186/s40561-024-00316-7>
- Zhang, Wei y Iliško, Dzintra (2025). "AI-driven innovation in educational management: A multi-case study of Chinese higher education institutions", *Edelweiss Applied Science and Technology*, vol. 9, núm. 4. <https://doi.org/10.55214/25768484.v9i4.6498>
- Zhao, Yan; Yue, Yuhe; Sun, Zhonghua; Jiang, Qiang y Li, Gangsheng (2025). "Does generative artificial intelligence improve students' higher-order thinking? A meta-analysis based on 29 experiments and quasi-experiments", *Journal of Intelligence*, vol. 13, núm. 12, p. 160. <https://doi.org/10.3390/jintelligence13120160>

Artículo recibido: 12 de mayo de 2026

Dictaminado: 3 de julio de 2026

Segunda versión: 29 de julio de 2026

Aceptado: 29 de julio de 2026