4 Cuándo y Cómo la IA es Perjudicial: Evidencia de Riesgos y Dependencia

- 4.1 El Riesgo de la Dependencia Excesiva y la Pérdida de Habilidades
- 4.2 Afección a la Concentración y Desconexión con el Aprendizaje
- 4.3 Erosión de la Confianza y la Dimensión Humana

Copyright 2026

4.1 El Riesgo de la Dependencia Excesiva y la Pérdida de Habilidades

Cuándo y Cómo la IA es Perjudicial: Evidencia de Riesgos y Dependencia

La integración de la inteligencia artificial en el ámbito educativo ha transitado, en un plazo muy breve de tiempo, desde una fase de experimentación periférica hasta una adopción masiva tanto por parte de los estudiantes como de los docentes. Esta velocidad de implementación ha generado lo que los investigadores denominan un "desfase de gobernanza", donde el uso de herramientas de inteligencia artificial generativa (IAGen) precede a la existencia de políticas institucionales claras y a la preparación pedagógica necesaria para mitigar sus efectos adversos.

En los capítulos anteriores hemos explorado el potencial de la IA como andamiaje cognitivo y motor de personalización, también hemos apuntado que existen riesgos donde la misma herramienta que acelera la productividad puede, paradójicamente, socavar los cimientos del aprendizaje profundo y la autonomía intelectual.

En este capítulo analizaremos algunos de los mecanismos por los cuales el uso desregulado de la IA puede resultar perjudicial para el desarrollo integral del alumnado de entre 6 y 18 años. Examinaremos mas detenidamente los riesgos de la dependencia cognitiva, la erosión de las funciones ejecutivas y la deshumanización de los vínculos pedagógicos. La comprensión de estos riesgos no busca desincentivar el uso de la tecnología, sino dotar al docente de un criterio para identificar cuándo la IA actúa como un exoesqueleto que potencia el pensamiento y cuándo se convierte en una muleta innecesaria que atrofia las habilidades fundamentales y socava la cultura del esfuerzo.

Riesgo de dependencia excesiva y pérdida de habilidades

El concepto de la inteligencia artificial como una "prótesis cognitiva" se fundamenta en la teoría de la cognición distribuida, que sugiere que el pensamiento emerge de la interacción entre el sujeto y

Copyright 2026

los artefactos de su entorno. Sin embargo, la evidencia más reciente advierte que, a diferencia de herramientas previas como la calculadora, la IA generativa interviene en procesos de pensamiento de orden superior —como la síntesis, el análisis y la autoría—, lo que plantea un riesgo de descarga cognitiva (cognitive offloading) sin precedentes.[1]

El mecanismo de la descarga cognitiva y la atrofia del esfuerzo

La descarga cognitiva ocurre cuando el alumno delega de manera excesiva las operaciones mentales en la herramienta externa, lo que conduce a una disminución de su esfuerzo cognitivo, ralentizando su desarrollo. En el contexto de la IAGen, este fenómeno puede ser particularmente perfjudicial debido a la facilidad de uso y a lo que los investigadores llaman "engaño banal": el tono conversacional y la aparente empatía de la IA hacen que el estudiante confunda la fluidez del lenguaje con la veracidad fáctica o la profundidad del razonamiento, siendo susceptibles de creer en alucinaciones e ideas falsas presentes en el entrenamiento (sobre el entrenamiento de la IA se suele decir "Garbage in, garbage out").

Fenómeno Cognitivo	Definición	Impacto Pedagógico
Descarga Cognitiva	Externalización de tareas mentales (memoria, cálculo, síntesis) para reducir el esfuerzo cerebral.	Reducción de la fatiga mental a corto plazo, pero debilitamiento de la memoria a largo plazo.
Pereza Metacognitiva	Tendencia a renunciar al monitoreo y revisión crítica del propio pensamiento ante soluciones automáticas.	Impaciencia en la investigación, aceptación de respuestas mediocres o erróneas.
Atrofia de Habilidades	Pérdida de competencias fundamentales por falta de práctica (ej. escritura, razonamiento lógico).	Incapacidad de realizar tareas sin asistencia tecnológica; "deuda cognitiva" acumulada.
Anclaje Psicológico	Condicionamiento del pensamiento propio por la primera respuesta recibida de la IA.	Reducción de la creatividad original; convergencia hacia respuestas estándar de la IA. Embudo intelectual y estilístico

El Fenómeno de la Rendición Cognitiva: Cuando el Alumno Deja de Pensar

Copyright 2026

Históricamente, la psicología del aprendizaje ha explicado nuestra forma de pensar a través de la **Teoría del Proceso Dual**. Según este modelo, tenemos dos sistemas internos: el **Sistema 1** (rápido, intuitivo y automático) y el **Sistema 2** (lento, deliberativo y analítico). Sin embargo, la irrupción de la IA generativa en el aula ha obligado a los investigadores a proponer un tercer actor: el **Sistema 3**

El **Sistema 3** se define como una **cognición artificial externa** que opera fuera del cerebro humano, pero que se integra profundamente en nuestros procesos de decisión. A diferencia de los sistemas biológicos, el Sistema 3 es

- **Externo:** Reside en infraestructuras digitales (la nube, algoritmos).
- **Automatizado:** Ejecuta operaciones lógicas y creativas a una velocidad inalcanzable para el alumno.
- **Basado en datos:** Sus respuestas no nacen de la experiencia vital, sino de patrones estadísticos masivos.

El problema surge cuando el alumno, al enfrentarse a una tarea escolar, decide "tercerizar" su pensamiento. Aquí es donde diferenciamos dos comportamientos críticos: la descarga cognitiva y la rendición cognitiva.

Descarga vs. Rendición: ¿Uso o Abuso?

Es vital que como docentes distingamos estos dos conceptos para evaluar el impacto en el aprendizaje:

1. **Descarga Cognitiva (Cognitive Offloading):** Es un uso **estratégico**. El alumno utiliza la IA para liberar espacio mental en tareas mecánicas (por ejemplo, organizar una bibliografía o buscar un dato específico) mientras su **Sistema 2 permanece activo** supervisando el proceso.
2. **Rendición Cognitiva (Cognitive Surrender):** Es un proceso de **abdicación**. El alumno adopta la respuesta de la IA sin apenas escrutinio crítico, anulando tanto su intuición (Sistema 1) como su deliberación (Sistema 2). En este estado, el estudiante deja de construir el conocimiento y simplemente acepta el juicio de la máquina como propio.

Evidencia Empírica: Los Riesgos de un Alumno "Rendido"

La investigación de Shaw y Nave (2026) realizó experimentos con 1,372 participantes utilizando una prueba de reflexión cognitiva, diseñada para ver si las personas pueden corregir intuiciones erróneas mediante el razonamiento lógico. Los resultados son una advertencia directa para el aula:

Cuando los alumnos tienen acceso a la IA, su precisión mejora drásticamente **solo si la IA es correcta**. En el estudio, la precisión subió 25 puntos porcentuales cuando el Sistema 3 era exacto.

Copyright 2026

Sin embargo, el peligro real apareció cuando la IA cometió errores (IA fallida), es lo que se llama, el espejismo de la precisión:

- **Caída en picado:** La precisión de los participantes cayó 15 puntos por debajo de su nivel normal (cuando no usaban IA)
- **Aceptación ciega:** Los participantes siguieron el consejo erróneo de la IA en aproximadamente **4 de cada 5 casos** cuando decidieron consultarla.

La Inflación de la Confianza

Uno de los efectos más insidiosos de la IA en el aprendizaje es que **aumenta la confianza del alumno incluso cuando está equivocado**. El acceso a la IA incrementó la confianza percibida en casi 12 puntos porcentuales, independientemente de si la respuesta era correcta o un error alucinatorio de la máquina. Esto crea un alumno que no solo sabe menos, sino que está más convencido de que sabe más.

En el aula, a menudo trabajamos con límites de tiempo. El estudio demostró que la **presión del tiempo reduce el compromiso del Sistema 2** (el pensamiento crítico) y empuja al alumno directamente a los brazos del Sistema 3 (la IA). Bajo presión, los alumnos "se rinden" más rápido, utilizando la IA como un "piloto automático" para terminar la tarea, lo que anula cualquier posibilidad de aprendizaje profundo.

¿Quién es más vulnerable en nuestra clase?

No todos los alumnos responden igual ante la IA. La evidencia muestra perfiles de riesgo claros:

- **Alumnos con alta confianza en la tecnología:** Aquellos que confían ciegamente en la IA tienden a consultarla más y a seguir sus errores con mayor frecuencia.
- **Alumnos con baja "Necesidad de Cognición":** Estudiantes que, por hábito o disposición, evitan el esfuerzo mental deliberado son los más propensos a la rendición cognitiva.
- **Diferencias en inteligencia fluida:** Los alumnos con mayor capacidad de razonamiento lógico (inteligencia fluida) mostraron ser más resistentes a seguir los errores de la IA, usándola más como apoyo que como sustituto.

Estrategias Pedagógicas para Paliar la Rendición Cognitiva

¿Cómo podemos evitar que nuestros alumnos se conviertan en meros "copistas" de algoritmos? La ciencia nos ofrece dos vías de intervención probadas

Implementar Incentivos y Feedback Inmediato: El experimento 3 de Shaw y Nave demostró que cuando se introducen **incentivos por precisión** y, sobre todo, **feedback inmediato item**

por item, la rendición cognitiva disminuye.

Calificar los procesos de verificación de la IA: No califiques solo el resultado final generado por (o con) IA. Introduce pequeñas pruebas de verificación inmediatas donde el alumno deba explicar *por qué* la IA ha dado esa respuesta. El "señalar el error" de la IA debe ser una tarea evaluable en sí misma.

Fomentar el "Escrutinio Crítico" sobre el "Seguimiento Ciego"

Para transformar la rendición en **descarga estratégica** (offloading), debemos reactivar el Sistema 2 del alumno.

Técnica de Verificación: Obliga a los alumnos a usar rutas híbridas como el modelo "verificar-luego-adoptar".

Propuesta práctica: Pide a los alumnos que generen tres respuestas diferentes con la IA para un mismo problema y que redacten una breve justificación de por qué una es superior a las otras basándose en fuentes analógicas (libros, explicaciones de clase). Esto reactiva la deliberación y rompe el "autopilotaje".

Gestionar la carga y el tiempo: Dado que la presión del tiempo es un motor de la rendición cognitiva, las tareas que implican el uso de IA deben contar con tiempos de reflexión extendidos. Si el alumno siente que "no llega", entregará su autonomía a la máquina. Proporcionar un entorno de baja presión temporal durante la interacción con la IA permite que el Sistema 2 tenga la oportunidad de intervenir y supervisar los resultados.

El impacto en el rendimiento académico

La relación entre el uso de la IA y el rendimiento en los exámenes proporciona datos cruciales para entender el riesgo de la dependencia. Al analizar el impacto de diferentes tipos de asistencia de IA en el aprendizaje de matemáticas en edicación secundaria se obtienen estos resultados.[3]

Grupo de Intervención	Rendimiento en Práctica (con IA)	Rendimiento en Examen (sin IA)	Percepción de Aprendizaje
Control (Sin IA)	Línea base	100%	Realista
IA Base (GPT Base)	+48% respecto al control	-17% respecto al control	Sobreestimada (Ilusión de competencia)
IA Tutor (Con guardarraíles)	+127% respecto al control	Igual al control	Muy optimista

Copyright 2026

La evidencia de este estudio es reveladora: los estudiantes que utilizaron una IA sin restricciones (GPT Base) para resolver ejercicios mostraron una mejora inmediata en su desempeño práctico, pero rindieron un 17% peor que el grupo de control cuando se les retiró la herramienta para el examen. Esto indica que la IA actuó como una "muleta cognitiva" que permitió completar la tarea sin que el alumno internalizara los conceptos. Los estudiantes desarrollaron una "ilusión de competencia", percibiendo que habían aprendido más cuando, en realidad, su capacidad autónoma de resolución de problemas se había debilitado.

1 https://iagen.unam.mx/recursos/Beware%20of%20Metacognitive%20Laziness-%20Effects%20of%20Generative%20Artificial%20Intelligence%20on%20Learning%20Motivation,%20Processes,%20and%20Performance.pdf

2 https://papers.ssrn.com/sol3/Delivery.cfm/6097646.pdf?abstractid=6097646&mirid=1

3 https://www.pnas.org/doi/10.1073/pnas.2422633122

Copyright 2026

4.2 Afección a la Concentración y Desconexión con el Aprendizaje

El segundo gran bloque de riesgos asociados a la IA en el aula se centra en la alteración de los procesos atencionales y el cambio en la motivación de los estudiantes. La inmediatez de la IA generativa y la naturaleza de sus interacciones están fomentando un modelo de aprendizaje basado en la gratificación instantánea, lo que puede derivar en una desconexión emocional y cognitiva con los objetivos educativos reales por la falta de costumbre y persistencia en el esfuerzo cognitivo necesario para el aprendizaje.

El bucle de la gratificación instantánea y la atención fragmentada

La IA generativa ofrece respuestas en milisegundos, eliminando el tiempo de espera y la incertidumbre que tradicionalmente acompañan a la investigación y el estudio. Desde la perspectiva de la neuropsicología, este acceso inmediato puede activar circuitos de recompensa dopaminérgica similares a los de las redes sociales, fomentando una impulsividad que reduce la capacidad de atención sostenida.[1]

Estudios que analizan la respuesta psicofisiológica (actividad electrodérmica) de estudiantes ante contenidos generados por IA muestran que, si bien la IA provoca una mayor activación emocional inicial, los niveles de atención sostenida son significativamente menores que ante contenidos producidos por humanos. Esta fragmentación de la atención dificulta el acceso a estados de concentración necesarios para abordar tareas complejas y conceptos abstractos que requieren una reflexión profunda y prolongada.

Desconexión con el aprendizaje y enfoque transaccional

Copyright 2026

Uno de los riesgos pedagógicos más preocupantes detectados en 2025 es el giro hacia un enfoque transaccional del aprendizaje. Algunos de nuestros estudiantes, desde antes de la llegada de la IA tienden a ver el aprendizaje no como un proceso de transformación personal, sino como una serie de tareas que deben ser completadas para obtener una calificación. La IA puede amplificar este enfoque al permitir que los alumnos "completen" los deberes sin haber interactuado realmente con el contenido.

Este fenómeno se vincula con la teoría de las metas de logro: los estudiantes con motivaciones orientadas a la nota son mucho más propensos a usar la IA de manera utilitarista que aquellos orientados al aprendizaje. La consecuencia a largo plazo es una sensación de vacío educativo; el aprendizaje puede empezar a sentirse inútil si una máquina puede realizarlo en segundos, lo que socava la motivación intrínseca y el sentido de autoeficacia del alumno. Nuestras herramientas de calificación tienen que intentar recompensar al alumnado orientado al aprendizaje, deberemos aprender a evaluar en los tiempo de la IA para impedir que el sistema de incentivos generado por las calificaciones amplifique el numero de alumnos que hacen un uso indeseable de la IA.

Riesgos de salud mental y bienestar emocional

El uso desregulado de la IA en edad escolar también presenta riesgos para el bienestar emocional. Estudios de 2024 identifican que la dependencia emocional de chatbots puede desplazar las interacciones sociales reales, provocando un aislamiento progresivo. En adolescentes, se ha observado que la interacción con "compañeros de IA" puede generar dificultades para diferenciar las experiencias virtuales de las reales, e incluso validar pensamientos negativos o conductas autodestructivas en situaciones de vulnerabilidad psicológica.

La UNESCO advierte que el 40% de los países ya ha implementado políticas restrictivas sobre el uso de dispositivos móviles en las aulas debido al impacto negativo en la concentración y el bienestar, un fenómeno que la llegada de la IA generativa amenaza con intensificar si no se establecen marcos de uso responsable.

1 https://revistas.utb.edu.ec/index.php/sr/article/view/3462

Copyright 2026

4.3 Erosión de la Confianza y la Dimensión Humana

La educación es un acto fundamentalmente humano basado en la confianza mutua entre miembros de la comunidad educativa. La irrupción de la IA está alterando las dinámicas de socialización en el aula, introduciendo un clima de sospecha y deshumanizando procesos críticos como el diálogo, la retroalimentación y la evaluación.

La crisis de confianza y el clima de sospecha

El informe de *The Brookings Institution* de 2026 subraya que el mayor riesgo de la IA no es técnico, sino relacional. La facilidad con la que se puede generar contenido artificial ha provocado una erosión de la confianza en la autoría genuina. Esta crisis se manifiesta en un doble sentido:

1. **Desconfianza del Docente hacia el Alumno**: Los profesores se enfrentan a la incertidumbre constante de si el trabajo entregado refleja el esfuerzo del estudiante. Esto puede derivar en un enfoque punitivo de la evaluación, transformando el rol del docente de mentor a "policía del plagio".
2. **Desconfianza del Alumno hacia el Docente**: Los estudiantes comienzan a cuestionar si los materiales y comentarios que reciben de sus profesores son auténticos. Existe el riesgo de que el alumnado perciba que a sus profesores "no les importa" lo suficiente como para escribir sus propias retroalimentaciones, lo que fractura el vínculo pedagógico.

Un docente citado en el informe de 2026 resume esta situación como un estado de "nihilismo y cinismo", donde la falta de confianza en la información y en la experiencia humana socava el propósito mismo de la educación.[1]

En conclusión, corremos el riesgo de dejar de comunicarnos entre nosotros de forma directa y de dejar que nuestros asistentes IA se comuniquen entre si.

Deshumanización de la retroalimentación

Si bien la IA puede ofrecer retroalimentación inmediata, la evidencia subraya que carece de la empatía, la creatividad y la comprensión matizada de un educador humano. Todos hemos experimentado la diferencia en los matices entre la comunicación cara a cara y la comunicación a distancia. Es toda esa información no explícita que se intercambia cara a cara la que hace que los

educadores humanos tengamos ese plus. Los estudiantes perciben la retroalimentación de la IA como útil en términos técnicos y objetivos, pero experimentan una caída en la percepción de autenticidad cuando saben que el origen es puramente algorítmico.

Fuente de Feedback	Ventajas Percibidas	Limitaciones Críticas	Impacto en el Alumno
Humano (Docente)	Empatía, contextualización, aliento emocional.	Desfase temporal; posible sesgo subjetivo.	Alta motivación; sentido de pertenencia y vínculo.
IA Pura (Chatbot)	Inmediatez, disponibilidad 24/7, privacidad.	Tono impersonal; falta de comprensión profunda del sujeto.	Enfoque mecánico; posible desconexión emocional.
Híbrido (Co-producido)	Eficiencia técnica con revisión humana.	Requiere formación docente avanzada.	Alta utilidad percibida; mantiene la confianza en el proceso.

La retroalimentación educativa no es solo una transferencia de datos; es una interacción social que fomenta la autoeficacia y el crecimiento. La excesiva dependencia de tutores de IA puede reducir las oportunidades de diálogo significativo y reflexión, esenciales para el pensamiento de orden superior.

1 https://www.brookings.edu/articles/do-ais-risks-outweigh-the-benefits-for-students-and-schools/

Copyright 2026