

# Evaluación educativa en tiempos de la Inteligencia Artificial

---

“La inteligencia artificial puede proporcionar una forma de personalizar el aprendizaje y mejorar la evaluación, permitiendo una retroalimentación más inmediata y adaptada a las necesidades del estudiante”  
(Luckin, 2018)

**Carlos Alario-Hoyos**

Departamento de Ingeniería Telemática, Universidad Carlos III de Madrid, Leganés, España

**Miguel Morales-Chan**

Departamento GES, Universidad Galileo, Ciudad de Guatemala, Guatemala

## Introducción

La Inteligencia Artificial (IA) ha ganado gran popularidad recientemente debido a los avances en IA generativa y a las numerosas aplicaciones lanzadas en los últimos meses (Lim et al., 2023), destacando *ChatGPT* como una de las más influyentes en los medios de comunicación (Ausat et al., 2023). Estas herramientas tienen el potencial de transformar múltiples sectores gracias a la automatización en la creación de contenidos y la capacidad para proporcionar respuestas completas y estructuradas sobre cualquier tema en diversos formatos y estilos. Esta transformación

también impacta en la educación superior. Si bien la IA ya se ha venido utilizando en este ámbito durante varios años, principalmente para ofrecer contenidos educativos personalizados, analizar y predecir el comportamiento de los estudiantes o facilitar su comunicación con el profesorado (Chassignol et al., 2018), la aparición de herramientas de IA generativa introduce nuevos desafíos. Uno de los más relevantes es la evaluación de los contenidos y competencias adquiridas por los estudiantes, a fin de que las instituciones puedan emitir certificaciones académicas válidas. La posibilidad de que los estudiantes utilicen herramientas de IA generativa como *ChatGPT* para realizar tareas complejas, como escribir ensayos, resolver problemas o generar código, requiere una revisión de los modelos de evaluación actuales en las universidades (Cotton et al., 2023; Rudolph et al., 2023), especialmente en entornos de baja o nula supervisión por parte del profesorado, como ocurre en la educación en línea (Susnjak, 2022). Además, la presión que enfrentan los estudiantes por aprobar asignaturas y obtener buenas calificaciones fomenta el uso de estas herramientas para cometer trampas académicas (Ventayen, 2023), lo que subraya la necesidad de redefinir las políticas universitarias para garantizar la integridad académica (Cotton et al., 2023).

## **Evolución de la evaluación educativa en la era de la IA: oportunidades**

La evaluación del conocimiento y las habilidades adquiridas por los estudiantes en una materia es una de las responsabilidades fundamentales del profesor, mientras que la certificación de estos mediante diplomas o credenciales académicas constituye un valor añadido esencial de las universidades. Esta función adquiere mayor relevancia en una era donde otros procesos académicos, como la generación de contenidos, la enseñanza y el apoyo a los estudiantes, han visto una reducción en su valor añadido al ser en parte reemplazados por terceros o automatizados mediante herramientas de IA (como la disponibilidad de contenidos de alta calidad en YouTube, cursos en línea, tutores inteligentes, asistentes virtuales, etc.). Una certificación inadecuada

de los conocimientos y habilidades, debido, por ejemplo, a la falta de capacidad para asegurar la autenticidad o autoría del trabajo del estudiante, podría erosionar el prestigio de la institución a mediano y largo plazo, especialmente ante los empleadores futuros de sus graduados.

Las evaluaciones tradicionales suelen utilizar un conjunto predefinido de elementos, como problemas, preguntas, ejercicios prácticos o análisis de casos, con el objetivo de medir la competencia de los estudiantes en uno o más aspectos. Sin embargo, los datos obtenidos para estas inferencias suelen ser limitados, y a menudo el aprendizaje de los estudiantes no es el enfoque central de la evaluación. Métodos comunes como las preguntas de opción múltiple, ensayos y preguntas de respuesta corta son ampliamente utilizados (Kaipa, 2021).

Estas evaluaciones presentan varios problemas (Swiecki et al., 2022). En primer lugar, pueden ser costosas en términos de tiempo y recursos, ya que requieren un diseño cuidadoso para transformar las respuestas de los estudiantes en evaluaciones de desempeño o aprendizaje, mediante listas de cotejo, rúbricas o claves de respuestas, entre otros (Mislevy et al., 2012). Cuando el número de estudiantes es elevado, las responsabilidades adicionales del profesor, como diseñar evaluaciones manualmente e interpretar los resultados, pueden convertirse en una carga considerable (Suto et al., 2011). En segundo lugar, las evaluaciones tradicionales tienden a ser puntuales, proporcionando una visión limitada de las capacidades de los estudiantes en un momento específico. Aunque estas evaluaciones pueden ofrecer información sobre los conocimientos y habilidades que los estudiantes tienen o carecen en ese instante, no suelen aportar datos relevantes sobre el proceso de aprendizaje en su conjunto.

En tercer lugar, las evaluaciones tradicionales tienden a ser homogéneas, asignando las mismas tareas o actividades a todos los estudiantes sin considerar sus conocimientos previos, habilidades, experiencias o antecedentes culturales. Este enfoque estándar limita la capacidad de reflejar el proceso de aprendizaje y solo capta el desempeño en un momento específico. Además, un enfoque único puede generar sesgos, ya que no todos los estudiantes tienen las

mismas oportunidades para demostrar su aprendizaje.

En cuarto lugar, las evaluaciones tradicionales suelen carecer de autenticidad. Por ejemplo, en las evaluaciones basadas en la resolución de problemas, los profesionales enfrentan desafíos en un entorno colaborativo, donde consultan con colegas, investigan soluciones existentes, aplican diversas estrategias y utilizan herramientas tecnológicas para abordar los problemas. En cambio, en el contexto académico, los estudiantes generalmente resuelven problemas de manera aislada y sin acceso a recursos adicionales que reflejen la realidad laboral. En quinto lugar, las evaluaciones tradicionales a menudo se vuelven obsoletas porque tienden a medir habilidades que, con el tiempo, pierden relevancia en el contexto actual.

En sexto lugar, las evaluaciones tradicionales suelen centrarse en la memorización y reproducción de información, en lugar de evaluar habilidades de orden superior, como el pensamiento crítico, la resolución de problemas y la creatividad. Este enfoque en la memorización puede no reflejar las competencias que los estudiantes necesitan para tener éxito en el mundo laboral y en la vida en general. Además, priorizar la memorización de datos puede desincentivar a los estudiantes a comprometerse con un aprendizaje más profundo y significativo, reduciendo su motivación intrínseca para aprender.

Por último, las evaluaciones tradicionales suelen carecer de retroalimentación oportuna y constructiva. En muchos casos, los estudiantes reciben calificaciones o puntajes sin información adicional sobre cómo mejorar su desempeño. La falta de retroalimentación específica dificulta que los estudiantes identifiquen áreas de mejora y desarrollen estrategias efectivas para superar sus deficiencias.

A pesar de la prevalencia de las evaluaciones tradicionales —discretas, uniformes, enfocadas en la memorización y con falta de retroalimentación—, los avances en IA generativa han abierto nuevas posibilidades en el ámbito educativo (Swiecki, 2022). Estas tecnologías se han aplicado en diversos contextos, como el diseño de currículos y la creación de materiales didácticos personalizados, la facilitación del aprendizaje colaborativo entre pares, y su integración en las aulas mediante herramientas ya consolidadas, como

el software de calificación automática de ensayos y las pruebas adaptativas. Además, han surgido innovaciones recientes, como la evaluación continua basada en datos derivados de la interacción en línea de los estudiantes con materiales de aprendizaje.

La IA generativa tiene el potencial de desempeñar un papel clave en el proceso de evaluación de los aprendizajes, al permitir una mayor personalización de las evaluaciones y un enfoque centrado en el aprendizaje profundo en lugar de la simple memorización. Asimismo, puede ofrecer retroalimentación en tiempo real adaptada a las necesidades individuales de cada estudiante, ayudando a los educadores a identificar áreas de mejora y ajustar sus enfoques pedagógicos según las necesidades específicas de cada alumno.

## **Transformación de la evaluación educativa con la IA: desafíos**

La evaluación del conocimiento y habilidades adquiridas por los estudiantes generalmente se realiza en entornos supervisados. En evaluaciones síncronas, como los exámenes, cuando el entorno es presencial, la supervisión suele estar a cargo del profesor, quien se asegura de que se cumplan las normas que garantizan la integridad académica. En entornos de educación en línea, la supervisión se delega comúnmente en herramientas de *proctoring* (supervisión), como *ProctorU*, que monitorean a los estudiantes durante la prueba (Atoum et al., 2017).

En evaluaciones asíncronas, como la entrega de trabajos, la detección de plagio se realiza con herramientas especializadas, como *Turnitin*, y la autoría puede verificarse mediante entrevistas o exámenes orales complementarios con el autor. Este enfoque ayuda a garantizar que el trabajo entregado sea genuino y refleje las capacidades del estudiante.

Las herramientas de IA, especialmente las de IA generativa, están afectando los procesos de evaluación en múltiples formas. La capacidad de obtener respuestas complejas de manera casi instantánea complica la supervisión en evaluaciones síncronas, tanto en entornos presenciales (especialmente si la evaluación se realiza frente a un computador) como en línea. Existen numerosas extensiones de navegador, como

*GPT4Moodle* o *GPTGo*, que se conectan con *ChatGPT* para facilitar este tipo de acceso. Por otro lado, la generación de textos de alta calidad (ensayos, trabajos, etc.) mediante estas herramientas plantea nuevos desafíos para las evaluaciones asíncronas, lo que exige una evolución de las herramientas “antiplagio” para detectar textos producidos por IA. Herramientas como *GPTZero* o *GPTKit* han sido desarrolladas específicamente para abordar este problema.

A continuación, se presentarán diversas recomendaciones que permiten a los docentes aprovechar estos desafíos derivados de los cambios en la educación superior y los formatos de evaluación, para adaptarse y mejorar los procesos de enseñanza (Gimpel et al., 2023).

#### • Evaluación del proceso y no solo del producto

Las herramientas de IA generativa permiten generar productos finales de alta calidad sin dejar evidencia del proceso seguido por el estudiante. Por ello, es crucial evaluar el proceso además del producto. Se puede obtener información sobre los pasos seguidos mediante dos enfoques: el propio estudiante puede reportar de forma detallada cómo desarrolló el trabajo, o bien se pueden utilizar registros automatizados (*logs*) de las herramientas empleadas para rastrear su progreso. Incluso si se utilizan herramientas de IA en algunas etapas del proceso, es importante que el estudiante informe de su uso de manera honesta, incluyendo los *prompts* o instrucciones que haya empleado.

#### • Evaluaciones más auténticas y colaborativas

La autenticidad y colaboración en las evaluaciones son aspectos clave para mejorar la calidad del aprendizaje. Aunque innovar en este ámbito ha sido un reto, los docentes pueden apoyarse en enfoques de aprendizaje experiencial para diseñar evaluaciones más reales y significativas. Este enfoque se basa en la resolución activa de problemas en contextos reales, lo que permite a los estudiantes participar en la identificación y resolución de desafíos del mundo real. Estas evaluaciones auténticas y colaborativas requieren trabajo en equipo y pueden ser exigentes, pero al mismo tiempo

ofrecen oportunidades valiosas de aprendizaje. Para complementar el proceso, se puede incluir una presentación oral de los resultados, lo que permitirá verificar el grado de dominio de los contenidos por parte del estudiante.

#### • Integrar el uso de *ChatGPT* y otro tipo de herramientas IA

En la educación superior, es cada vez más necesario integrar herramientas de inteligencia artificial generativa, como *ChatGPT*, para mejorar la calidad de la investigación y la redacción de los estudiantes. A pesar de los beneficios, es crucial abordar la preocupación sobre la transparencia y la violación de derechos de autor. Por ello, se sugiere que las instituciones de educación superior desarrollen políticas institucionales, acciones que promuevan una cultura de integridad académica y códigos de honor que aborden explícitamente el uso de estas herramientas y que los estudiantes asuman la responsabilidad de posibles errores o violaciones de derechos de autor. Además, se pueden exigir a los estudiantes proporcionar una lista de indicaciones utilizadas y, en algunos casos, una transcripción completa de la conversación con *ChatGPT* que informó su trabajo. Es importante destacar que, para un uso responsable de estas herramientas en la educación superior, se deben establecer prácticas transparentes y responsables.

### **El papel de la inteligencia artificial en la evaluación educativa: una mirada hacia el futuro**

Las herramientas de IA generativa evolucionan a un ritmo muy rápido. En este sentido, es importante tener en cuenta que los productos que hoy en día presentan una calidad baja mejorarán en el corto y medio plazo. Por ejemplo, las respuestas proporcionadas por *ChatGPT* son cada vez más realistas a medida que mejoran los modelos en los que se basa, notándose un gran salto de calidad de *GPT-3.5* a *GPT-4*; este salto además se ha producido en el curso de unos pocos meses solamente.

Lo mismo puede afirmarse de las imágenes generadas por DALL·E.

Por tanto, conviene no despreciar a las herramientas de IA generativa solamente porque hoy en día sus resultados todavía pueden contener fallos fácilmente detectables por el ser humano y anticiparse a los cada vez mejores resultados que vendrán en el futuro, especialmente con los nuevos avances relacionados con la utilización de *plug-ins* desde *ChatGPT* para conectar con otros servicios (p. ej., *Wolfram Alpha* para realizar operaciones matemáticas).

La formación del profesorado y del estudiantado en este tipo de herramientas de IA es fundamental para poder hacer un uso correcto de ellas. Es importante que las instituciones educativas dediquen recursos a esta formación y pongan este nuevo contexto tecnológico encima de la mesa, también en las reuniones con las asociaciones de estudiantes. De la misma forma es importante crear observatorios de herramientas de IA generativa que puedan ser utilizadas en las universidades para apoyar los diferentes procesos académicos, especialmente los que se refieren a la docencia y, en particular, a la evaluación.

También es importante tener en cuenta que en este nuevo contexto se produce un problema en relación con la equidad en el acceso a las herramientas de IA generativa. Muchas de estas herramientas tienen actualmente versiones con costo que permiten su utilización ilimitada o con funcionalidades avanzadas. En este sentido, no todos los miembros de la comunidad universitaria tienen recursos para poder asumir dicho costo, lo que deriva en diferencias entre estudiantes y profesores, y también entre diferentes universidades. Es posible que algunos estudiantes dispongan, por tanto, de acceso a herramientas de IA generativa que les permitan resolver más fácilmente ciertos problemas o realizar ensayos con una mayor calidad, mientras que otros tendrán que conformarse con versiones más básicas y, posiblemente, cuyo uso sea más fácilmente detectable en contexto de evaluación académica. Lo mismo sucede con los profesores y con el acceso (o no) a herramientas que permitan una detección del contenido generado por herramientas de IA generativa. Precisamente las universidades deben revisar su política de adquisición de herramientas de apoyo a los procesos de enseñanza y aprendizaje para posiblemente incluir en el catálogo de su oferta a herramientas de IA generativa.

## Conclusiones

En el ámbito académico, resulta imperativo integrar la tecnología, la pedagogía y la ética, especialmente en lo que concierne a la evaluación educativa. Con el constante avance en las herramientas de inteligencia artificial generativa, resulta necesario que la comunidad educativa se adapte y esté preparada para el uso de estas tecnologías en la evaluación del aprendizaje. Para ello, es esencial que se reconsidere el proceso de evaluación y se adecuen los códigos de honor académico, con el fin de que se contemplen estos avances tecnológicos. Además, se debe ofrecer una formación adecuada al personal universitario y establecer un diálogo constante con las asociaciones de estudiantes para garantizar el uso adecuado y responsable de estas nuevas herramientas. La colaboración entre la academia, los estudiantes y la tecnología resulta fundamental para una evaluación justa y efectiva en el ámbito educativo.

## Referencias

- Atoum, Y., Chen, L., Liu, A. X., Hsu, S. D., & Liu, X. (2017). Automated online exam proctoring. *IEEE Transactions on Multimedia*, 19(7), 1609-1624.
- Ausat, A. M. A., Massang, B., Efendi, M., Nofirman, N., & Riady, Y. (2023). Can Chat GPT Replace the Role of the Teacher in the Classroom: A Fundamental Analysis. *Journal on Education*, 5(4), 16100-16106.
- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, 136, 16-24.
- Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and Cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12.
- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A., ... & Vandrik, S. (2023). Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education. Universität Hohenheim.

- Kaipa, R. M. (2021). Multiple choice questions and essay questions in curriculum. *Journal of Applied Research in Higher Education*, 13(1), 16-32.
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21(2), 100790.
- Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of educational data mining*, 4(1), 11-48.
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1).
- Susnjak, T. (2022). ChatGPT: The End of Online Exam Integrity? arXiv preprint arXiv:2212.09292.
- Suto, I., Nádas, R., & Bell, J. (2011). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education*, 26(1), 21-51.
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075.
- Ventayen, R. J. M. (2023). ChatGPT by OpenAI: Students' Viewpoint on Cheating using Artificial Intelligence-Based Application. Available at SSRN 4361548.