EL MERCURIO

Fecha: 14/06/2026 Audiencia: 320.543 Sección: EDUCACION
Vpe: $7.531.034 Tirada: 126.654 Frecuencia: 0
Vpe pág: $20.570.976 Difusión: 126.654
Vpe portada: $20.570.976 Ocupación: 36,61%

Análisis a más de 700 ensayos universitarios calificados tanto por profesores como por modelos generativos:

Al evaluar trabajos universitarios, la IA tiende a premiar el estilo por sobre el contenido

■ Una nueva investigación concluye que estas herramientas suelen valorar características lingüísticas —como la longitud de los textos o la riqueza del vocabulario— y prestan menor atención a la calidad del razonamiento.

PEXELS / CREATIVE COMMONS

Los autores creen necesario investigar más sobre este tema, dado que "las universidades están sometidas a una enorme presión para reducir la carga laboral de sus académicos y mejorar su eficiencia, mientras intentan satisfacer expectativas estudiantiles cada vez más altas. Por ello, algunas podrían comenzar a recurrir a la IA para evaluar trabajos y tareas".

M. CORDANO

Por sí sola, la inteligencia artificial sigue siendo "demasiado superficial e inconsistente para evaluar adecuadamente los trabajos universitarios". Por ello, la decisión final sobre cómo se califican las tareas "debería siempre recaer en un evaluador humano".

Esa es la conclusión a la que llegó un equipo de académicos liderado por la U. de Cambridge, que usó algunos de los modelos de IA generativa más avanzados —entre ellos, Claude y ChatGPT, en sus versiones actualizadas en abril de 2026— para corregir 761 ensayos de estudiantes de Psicología de tres universidades del Reino Unido. Aunque no hubo un *prompt* (instrucción para interactuar con estos sistemas) específico, a cada modelo se le dio a entender que debía tomar el papel de un evaluador experimentado de cierta casa de estudios y que su labor era corregir textos de estudiantes de pregrado.

Aunque la precisión de la IA para calificar ensayos "no fue uniformemente alta", los investigadores observaron que logró coincidir con la categoría de nota asignada por evaluadores humanos, como sobresaliente o aprobado, en un 35% a 65% de los casos, dependiendo de la universidad analizada.

Sin embargo, los grandes problemas aparecieron frente a los extremos: la IA tendía a asignar notas muy bajas a trabajos que los evaluadores humanos consideraban excelentes. En cambio, sobrevaloraba ensayos que para las personas que corrigieron estaban entre los de peor desempeño.

Y es que, a diferencia de los humanos, los sistemas de IA mostraron una "hipersensibilidad a las características lingüísticas". Según explica el estudio, estos sistemas pusieron notas más altas según la extensión de los textos, así como la complejidad del vocabulario usado y oraciones formadas. Todo esto, independientemente de la calidad académica de los diversos ensayos.

"Nos sorprendió que ninguna de las estrategias que probamos logró superar este sesgo", comenta a "El Mercurio" Deborah Talmi, psicóloga y académica de la U. de Cambridge. "Los profesores adquieren su experiencia a través de vivencias prácticas y directas, mientras que los modelos de lenguaje de gran tamaño (LLM) solo pueden basarse en el lenguaje. La especulación es que esta es la razón por la que son más sensibles que los seres humanos a las características lingüísticas de un texto", dice.

Adopción en Chile

Consultado por el uso de herramientas de IA para apoyar procesos de corrección en Chile, Brayan Díaz, del Cenia, responde que va en alza y que su percepción es que el fenómeno continuará "aumentando mucho".

María Fernanda Rodríguez, de la U. Finis Terrae, agrega que, a nivel país, "diversas instituciones de educación superior han publicado marcos conceptuales, lineamientos y buenas prácticas en torno al uso de la IA, lo que evidencia un creciente interés de las instituciones por explorar estas herramientas y por desarrollar orientaciones para su uso responsable", pero que, "a nivel empírico, aún no existe suficiente evidencia nacional que establezca el grado de adopción de la IA en procesos de corrección y retroalimentación".

Originalidad y juicio crítico

"La IA todavía tiene dificultades para distinguir entre una escritura sofisticada y una argumentación profunda", plantea sobre los resultados María Fernanda Rodríguez, doctora en Ciencias de la Ingeniería y académica de la Facultad de Educación y Ciencias Sociales de la U. Finis Terrae. Así, los hallazgos del estudio "nos recuerdan que los sistemas de IA funcionan a partir de patrones estadísticos del lenguaje. Pueden reconocer señales asociadas a textos académicos de calidad, pero eso no es equivalente a evaluar con profundidad la originalidad de una idea, la profundidad de una interpretación o la calidad de un juicio crítico. Por esto, todavía debemos ser cautos cuando se trata de delegarles decisiones evaluativas de alto impacto".

> ***Una dependencia excesiva** de los mejores modelos de IA actualmente disponibles conduciría a una evaluación más homogénea de los estudiantes, que subestima la excelencia y favorece el estilo lingüístico por sobre la solidez del juicio académico".*
>
> **DEBORAH TALMI**
> ACADÉMICA U. DE CAMBRIDGE

Rodríguez agrega que los profesores evalúan "elementos como la pertinencia de los argumentos, las conexiones que establece el estudiante, la capacidad de cuestionar supuestos y de construir una posición propia. En complemento, la evaluación académica incorpora elementos de la experiencia profesional y conocimiento disciplinar que no siempre están explícitos en una rúbrica". Además, quienes enseñan tienden a evaluar la trayectoria de sus estudiantes.

"Suelen interpretar un trabajo considerando procesos previos, avances, dificultades y contextos específicos. Estos componentes disciplinares, profesionales y relacionales siguen siendo una fortaleza eminentemente humana", explica la especialista.

Brayan Díaz, investigador del Centro Nacional de Inteligencia Artificial (Cenia) y del Centro de Investigación para la Mejora de los Aprendizajes (Cima) de la Facultad de Educación de la U. del Desarrollo, indica que "no solo hay que pensar en la enseñanza como pararse frente a la clase y dictar. El proceso es mucho más complejo; uno como profesor evalúa y aprende de sus estudiantes, logra identificar fortalezas y debilidades. Si automatizamos todo el proceso, el docente tiene menos información para generar prácticas más efectivas y personalizadas".

De cualquier forma, Díaz advierte que esto no supone que no se pueda buscar un equilibrio, aplicando la IA no como reemplazo, sino como complemento.

"Aunque no tengo dudas de que la IA pasará a formar parte de todos los aspectos de nuestra vida, incluida la evaluación académica, no puedo predecir cómo se desarrollará exactamente ese proceso ni cómo funcionará en la práctica. A corto plazo, una aplicación sencilla sería utilizarla para permitir que los alumnos interactúen con la retroalimentación entregada por sus docentes; por ejemplo, para solicitar explicaciones más detalladas o aclaraciones adicionales. Esto podría aumentar su involucramiento con la retroalimentación", señala a este diario Yael Benn, académica de la U. Metropolitana de Manchester y otra de las autoras del estudio.

Sobre los usos de la IA en este contexto, Rodríguez agrega que "el desafío es asegurar que estas herramientas fortalezcan y no reemplacen los procesos de reflexión, deliberación y juicio profesional, que siguen siendo centrales en educación".