

Universitarios que usan ChatGPT creen entender más de lo que realmente pueden explicar

Un inédito estudio desarrollado por investigadores del Centro de Investigación Avanzada en Educación de la Universidad de Chile y el Centro Nacional de Inteligencia Artificial, realizado con más de 100 estudiantes, reveló que quienes interactuaron con ChatGPT sobreestimaron cuánto habían aprendido y luego elaboraron respuestas de menor calidad que quienes recibieron la misma información, pero solo la leyeron.

Una explicación clara, ordenada y convincente puede generar una sensación engañosa: creer que entendimos algo cuando todavía no somos capaces de explicarlo por nosotros mismos.

Ese es el fenómeno que aborda una investigación realizada por Ivan Aslanov, investigador del Centro Nacional de Inteligencia Artificial–CENIA en conjunto con Ernesto Guerra, académico del Instituto de Estudios Avanzados en Educación e investigador del CIAE de la U. de Chile, y Patricio Felmer, investigador CIAE y Premio Nacional de Ciencias Exactas 2011. El estudio exploró cómo interactuar con herramientas de inteligencia artificial como *ChatGPT* puede aumentar la “**ilusión de profundidad explicativa**”, es decir, la tendencia a sobreestimar cuánto entendemos sobre un fenómeno.

Para estudiarlo, los investigadores trabajaron con 102 estudiantes universitarios, divididos en tres grupos. Todos debieron abordar cuatro preguntas sobre fenómenos cotidianos: por qué existen los años bisiestos, cómo se fabrica el vidrio, cómo funcionan los cierres y por qué la Luna tiene fases.

La diferencia estuvo en qué información recibieron y a través

de qué medio. Un grupo interactuó con *ChatGPT*; otro leyó exactamente los mismos textos generados por el chatbot, pero sin saber conocer su procedencia; y un tercer grupo no recibió material adicional. Después, todos tuvieron que evaluar su capacidad de explicar cada fenómeno, y luego explicar los fenómenos por escrito. Finalmente, luego de leer su propia explicación, evaluaron de nuevo qué tan bien podían hacerlo.

La brecha entre “creo que sé” y “puedo explicarlo”

El resultado fue especialmente llamativo: quienes interactuaron con *ChatGPT* mostraron la mayor sobreestimación de su capacidad para explicar los fenómenos. La diferencia entre lo que creían poder explicar, antes y después de escribir sus respuestas fue significativamente mayor comparado en el grupo que simplemente había leído los mismos contenidos.

El efecto no se puede explicar simplemente por la calidad de la información. Los dos grupos recibieron los mismos contenidos. Lo que cambió fue la experiencia de interactuar con el chatbot.

Pero la investigación encontró algo más: las explicaciones posteriores también fueron de menor calidad en el grupo que utilizó GPT. Dos evaluadores independientes calificaron las respuestas en una escala de 1 a 4. El grupo GPT obtuvo un promedio de 2,47 puntos, frente a 2,75 del grupo que había leído los mismos textos sin interactuar con la IA. El grupo que no recibió material obtuvo 1,8 puntos.

El análisis lingüístico apuntó en la misma dirección. Las respuestas del grupo GPT tuvieron un promedio de 43,8 palabras, mientras que las del grupo que solo leyó los textos alcanzaron 59,6 palabras, mostraron una mayor diversidad de vocabulario. Además los textos escritos por este último grupo se parecían más al texto original, comparados con los textos escritos por el grupo GPT.

Para Ernesto Guerra, académico del IE e investigador CIAE de

la U. de Chile, esto plantea una distinción fundamental: “Tener acceso a una buena explicación no significa necesariamente haber incorporado ese conocimiento. Una persona puede reconocer y seguir una explicación cuando la tiene delante, pero no necesariamente reconstruirla después por sí misma”.

El hallazgo adquiere especial relevancia en un contexto en que la inteligencia artificial ya forma parte de las prácticas de aprendizaje de los estudiantes. Los resultados de PISA 2025 muestran que el uso de herramientas de IA para aprender está ampliamente extendido entre los estudiantes chilenos de 15 años, “lo que vuelve especialmente pertinente preguntarse no sólo cuánto utilizan estas herramientas, sino también qué ocurre con la comprensión que construyen a partir de ellas”, dice el experto.

¿El problema es ChatGPT?

Los autores son cuidadosos con las conclusiones. El estudio no demuestra que utilizar ChatGPT sea malo para aprender, ni que la IA produzca siempre peores resultados. Lo que muestra es un efecto específico: al interactuar con un chatbot, las personas pueden terminar con una confianza en su propia comprensión mayor que la que corresponde a su desempeño posterior.

Una posible explicación es que los usuarios atribuyan a su propio conocimiento parte de la información que en realidad está disponible en la herramienta. Otra posibilidad es que la forma de interacción de los chatbots –que presenta las respuestas como una conversación fluida– contribuya a generar una mayor sensación de comprensión o confianza. Los propios investigadores señalan que estas hipótesis requieren nuevos estudios.

La advertencia resulta especialmente relevante para la educación. En un escenario donde obtener una explicación es cada vez más fácil, el desafío ya no es solo conseguir la

respuesta, sino comprobar si realmente aprendimos.

En definitiva, el estudio deja una pregunta que va mucho más allá de ChatGPT: si una explicación está a un clic de distancia, ¿cómo sabemos que realmente la aprendimos? “En un mundo donde obtener información es cada vez más fácil, una de las habilidades más importantes puede ser saber distinguir entre tener una explicación disponible y haber aprendido realmente a explicar”, explica el académico del IE-CIAE.