

Una ilusión

● Últimamente han emergido con fuerza los modelos “razonadores” de IA, como el chino Deepseek R1 o las versiones “o” de ChatGPT, con grandes avances. Estos LLM utilizan técnicas que simulan razonamientos paso a paso similares a los humanos; incluso podemos leerlos mientras lo hacen. Sin embargo, investigaciones recientes cuestionan seriamente esa capacidad.

Un estudio de Apple (The Illusion of Thinking, 2025) evaluó su desempeño en rompecabezas clásicos como la Torre de Hanói. El hallazgo central fue que, cuanto más complejo es el problema, menos se esfuerzan los modelos y más fallan. Incluso con instrucciones precisas, fracasan sistemáticamente a partir de cierto nivel de dificultad. Peor aún, en desafíos simples suelen encontrar la solución al comienzo, pero la sobreescriben con razonamientos erróneos. En los problemas complejos, colapsan con niveles de precisión cercanos a cero. Su desempeño parece depender más de la familiaridad con los datos de entrenamiento que de una comprensión real.

Los humanos también se equivocan, claro, pero una tecnología útil debe ser confiable. Si la IA no resuelve problemas de forma consistente, no es verdaderamente inteligente. Confundir lenguaje con pensamiento puede llevarnos a depositar confianza

donde no hay comprensión. Tal vez aún estemos lejos de una inteligencia artificial general (AGI), pero todavía estamos a tiempo de decidir qué clase de inteligencia queremos construir.

Fernando Roa