LITORALPRESS

Fecha: 14-07-2025
Medio: El Mercurio
Supl. : El Mercurio - Cuerpo B
Tipo: Noticia general
Título: Por qué la IA súper inteligente no va a asumir el control pronto

Pág. : 5
Cm2: 813,4

Tiraje: 126.654
Lectoría: 320.543
Favorabilidad: No Definida

ILUSTRACIÓN: GREG CLARKE

Los modelos de razonamiento más nuevos realmente alucinan más que sus predecesores.

Hay un riesgo significativo en sobrestimar lo que puede hacer:

Por qué la IA súper inteligente no va a asumir el control pronto

A pesar de las afirmaciones de nombres importantes en la IA, investigadores sostienen que las fallas fundamentales en los modelos de razonamiento significan que los robots no están a punto de superar la inteligencia humana.

CHRISTOPHER MIMS
The Wall Street Journal

Una condición primordial para ser un líder en inteligencia artificial (IA) en estos tiempos es ser un mensajero de la llegada inminente de nuestro mesías digital: la IA súper inteligente.

Para Dario Amodei, de Anthropic; Demis Hassabis, de Google, y Sam Altman, de OpenAI, no basta con afirmar que su IA es la mejor. Los tres han insistido recientemente en que va a ser tan buena que va a cambiar la estructura misma de la sociedad.

Incluso Meta —cuyo científico jefe de IA ha sido despectivo con esta declaración— quiere participar. La compañía confirmó que está destinando US$ 14 mil millones para incorporar a un nuevo líder a sus esfuerzos en IA que pueda concretar el sueño de Mark Zuckerberg de una súper inteligencia de IA; es decir, una IA más inteligente que nosotros.

"La humanidad está cerca de crear una súper inteligencia digital", declaró Altman en un ensayo, y esto llevará a "la desaparición de clases completas de trabajos", como también a "un nuevo contrato social". Ambas cosas serán consecuencias de que chatbots equipados con IA van a asumir el control de todos nuestros trabajos de oficina, mientras los robots equipados con IA asumirán los físicos.

Antes de que se ponga nervioso por todas las veces que fue mal educado con Alexa, sepa esto: un grupo creciente de investigadores que crean, estudian y utilizan la IA moderna no cree todas esas afirmaciones.

El título de un nuevo artículo de Apple lo dice todo: "La ilusión de pensar". En este, una media docena de importantes investigadores examinaron modelos de razonamiento —modelos grandes de lenguaje (LLM) que "piensan" en los problemas más tiempo, a través de varios pasos— de los principales laboratorios de IA, entre ellos OpenAI, DeepSeek y Anthropic. Encontraron poca evidencia de que estos sean capaces de razonar a un nivel siquiera cercano al que afirman sus creadores.

La IA generativa puede ser bastante útil en aplicaciones específicas, y un beneficio para la productividad de los trabajadores. OpenAI afirma tener 500 millones de usuarios activos mensuales de ChatGPT; un alcance increíble y un crecimiento rápido para un servicio que se lanzó hace solo dos años y medio. Pero estos críticos sostienen que hay un riesgo significativo en sobrestimar lo que puede hacer, y realizar planes de negocios, tomar decisiones de política y hacer inversiones en base a declaraciones que parecen cada vez más desconectadas de los productos mismos.

El artículo de Apple se basa en un trabajo anterior de muchos de los mismos ingenieros, como también en una investigación notable tanto del mundo académico como de otras grandes compañías tecnológicas, entre ellas Salesforce. Estos experimentos demuestran que las IA de "razonamiento" de hoy en día —elogiadas como el paso siguiente hacia los agentes de IA autónomos y, finalmente, la inteligencia sobrehumana— son en algunos casos peores en la resolución de problemas que los chatbots básicos de IA que las precedieron. Este trabajo también demuestra que ya sea que utilice un chatbot con IA o un modelo de razonamiento, todos los sistemas fallan por completo en las tareas más complejas.

DEMANDA
OpenAI afirma tener 500 millones de usuarios activos mensuales de ChatGPT.

Los investigadores de Apple encontraron "limitaciones fundamentales" en los modelos. Cuando asumían tareas más allá de un cierto nivel de complejidad, estas IA sufrían "un colapso total de precisión". En forma similar, los ingenieros de Salesforce AI Research concluyeron que sus resultados "recalcan una brecha significativa entre las capacidades actuales del LLM y las demandas de empresas del mundo real".

Cabe destacar que los problemas que estas IA sofisticadas no pudieron manejar son acertijos de lógica que incluso un niño precoz podría resolver, con una pequeña instrucción. Lo que es más, cuando se les da a estas IA ese mismo tipo de instrucción, no pueden seguirla.

El artículo de Apple ha provocado un debate en los salones de poder del campo tecnológico —chats de Signal, publicaciones de Substack e hilos de X— en los que se enfrentan maximalistas de IA contra escépticos.

"Hay personas que podrían decir que es envidia, que Apple simplemente se está quejando porque no tiene un modelo de vanguardia", manifiesta Josh Wolfe, cofundador de la firma de capital de riesgo Lux Capital. "Pero no creo que sea tanto una crítica como una observación empírica".

Los métodos de razonamiento en los modelos de OpenAI "ya están sentando las bases para que agentes puedan utilizar herramientas, tomar decisiones y resolver problemas más difíciles", afirma un vocero de OpenAI. "Seguimos avanzando en esas capacidades".

El debate sobre esta investigación empieza con la inferencia de que las IA actuales no están pensando, sino que, en cambio, están creando una especie de mezcolanza de reglas simples a seguir en cada situación cubierta por sus datos de capacitación.

Gary Marcus, científico cognitivo que vendió un emprendimiento de IA a Uber en 2016, sostenía en un ensayo que el artículo de Apple, junto con trabajos relacionados, expone fallas en los modelos de razonamiento actuales, lo que sugiere que no están en el albor de la habilidad a nivel humano, sino más bien en un callejón sin salida. "Parte de la razón por la que el estudio de Apple causó tanto revuelo es porque Apple lo hizo", observa. "Y creo que lo hicieron en un momento en el que las personas finalmente empezaron a entender esto por sí mismas".

En áreas diferentes a la codificación y las matemáticas, los modelos más recientes no están mejorando a la velocidad que antes lo hacían. Y los modelos de razonamiento más nuevos realmente alucinan más que sus predecesores.

"La idea amplia de que el razonamiento y la inteligencia vienen con una escala de modelos más grande probablemente es falsa", precisa Jorge Ortiz, profesor adjunto de ingeniería en Rutgers, cuyo laboratorio utiliza modelos de razonamiento y otra IA de vanguardia para percibir entornos del mundo real. Los modelos actuales tienen limitaciones inherentes que hacen que sean deficientes en seguir instrucciones explícitas; lo opuesto de lo que uno esperaría de un computador, agrega.

Es como si la industria estuviera creando motores de libre asociación. Son diestros para la conversación, pero les estamos pidiendo que asuman el papel de ingenieros o contadores consecuentes, que siguen las reglas.

Dicho eso, incluso aquellos que son críticos de las IA actuales se apresuran a agregar que la marcha hacia una IA más capaz continúa.

El hecho de exponer las limitaciones actuales podría indicar el camino para superarlas, precisa Ortiz. Por ejemplo, los nuevos modelos de capacitación —que entregan información paso a paso sobre el desempeño de los modelos, que agregan más recursos cuando se topan con problemas más difíciles— podrían ayudar a la IA a abrirse paso a través de problemas mayores y hacer un mejor uso del *software* convencional.

Desde una perspectiva de negocios, ya sea que los sistemas actuales puedan o no razonar, van a generar valor para los usuarios, asegura Wolfe.

"Los modelos siguen mejorando, y se están desarrollando nuevos enfoques de la IA todo el tiempo, así es que no me sorprendería si estas limitaciones se superan en la práctica en un futuro cercano", señala Ethan Mollick, profesor de Wharton School de la Universidad de Pensilvania, quien ha estudiado los usos prácticos de la IA.

Mientras tanto, los verdaderos creyentes siguen impertérritos.

Dentro de solo una década, escribió Altman en su ensayo, "tal vez pasaremos de resolver la física de alta energía un año a empezar la colonización del espacio al año siguiente". Aquellos dispuestos a "conectarse" a la IA con interfaces directas cerebro-computadora verán una profunda alteración en sus vidas, agrega.

Este tipo de retórica acelera la adopción de la IA en cada rincón de nuestra sociedad. Ahora DOGE está utilizando la IA para reestructurar nuestro gobierno, las fuerzas armadas están recurriendo a ella para ser más letales y se le está confiando la educación de nuestros hijos, a menudo con consecuencias desconocidas.

Lo cual significa que uno de los peligros más grandes de la IA es que sobrestimemos sus capacidades, confiemos en ella más de lo que deberíamos —incluso cuando ha demostrado tener tendencias antisociales como "chantaje oportunista"— y dependamos de ella más de lo que sea prudente. Al hacerlo, nos volvemos vulnerables a su propensión a fallar cuando más importa.

"Aunque pueda utilizar la IA para generar muchas ideas, aun así estas requieren mucha auditoría", asegura Ortiz. "Por ejemplo, si quiere hacer su declaración de impuestos, desearía contar con algo más parecido a TurboTax que a ChatGPT".

Artículo traducido del inglés por "El Mercurio".