LITORALPRESS

Fecha: 13-12-2024
Medio: La Tercera
Supl. : La Tercera - Pulso
Tipo: Noticia general
Título: **Prepárese para el "pensamiento largo", el próximo salto adelante de la IA**

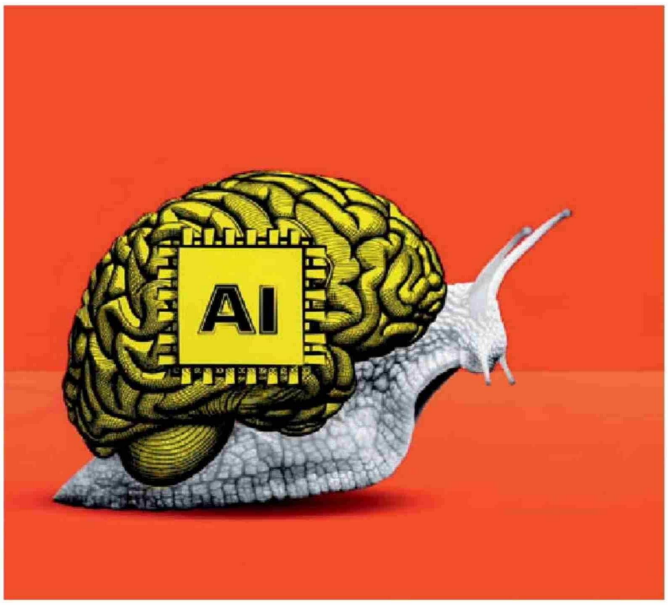
Pág. : 11
Cm2: 738,1
VPE: $ 7.342.944

Tiraje: 78.224
Lectoría: 253.149
Favorabilidad: No Definida

Prepárese para el "pensamiento largo", el próximo salto adelante de la IA

Una nueva generación de modelos de IA se tomará su tiempo para razonar, proporcionando respuestas más fiables a preguntas cada vez más complejas.

Steven Rosenbush /
THE WALL STREET JOURNAL



El director ejecutivo de Nvidia hizo una referencia al final de la última llamada de resultados de la compañía que merece mucha más atención de la que recibió, incluso si fue comprensiblemente eclipsada por los US$35.100 millones en ingresos trimestrales, impulsados en un 94% por la voraz demanda de *chips* de IA por parte de los clientes.

"Estamos en los inicios de esta revolución generativa de la IA, como todos sabemos", comentó Jensen Huang en la conferencia. "Y estamos al principio de una nueva generación de modelos básicos capaces de razonar y de pensar a largo plazo", indicó.

El "pensamiento largo" no llegó al *zeitgeist* (conjunto general de ideas, de creencias de un período) cuando el ChatGPT de OpenAI sorprendió al mundo por primera vez hace dos años con respuestas rápidas a preguntas sobre casi cualquier cosa. Sin embargo, tiene el potencial de reducir o eliminar los errores que a menudo salpicaban esas respuestas.

La idea es exactamente lo que parece, al menos al más alto nivel: los modelos de inteligencia artificial de larga duración están diseñados para tomarse más tiempo para "pensar" los resultados que generan para nosotros. Serán lo bastante inteligentes como para informarnos de sus progresos y pedirnos nuestra opinión sobre la marcha.

Eso puede significar dedicar unos segundos más a un problema, o mucho, mucho más tiempo, como Huang indicó en otro comentario revelador en junio pasado.

"En muchos casos, como saben, estamos trabajando en aplicaciones de inteligencia artificial que funcionan durante 100 días", comentó en la feria Computex de Taipei.

A medida que se desarrolle la capacidad de razonamiento de los modelos, se espera que la IA evolucione mucho más allá de la tecnología actual que trabaja en nuestro nombre en atención al cliente o automatización, o de los agentes aún más sofisticados que están empezando a aparecer.

Las capacidades de razonamiento a largo plazo de OpenAI avanzaron en septiembre con el lanzamiento de sus modelos de la serie o, de los que dijo que están diseñados para pasar más tiempo pensando antes de responder, razonando sobre tareas complejas y resolviendo "problemas más difíciles que los modelos anteriores en ciencia, codificación y matemáticas".

Catherine Brownstein, profesora adjunta del Hospital Infantil de Boston y de la Facultad de Medicina de Harvard que investiga enfermedades extremadamente raras, sostuvo que las nuevas capacidades de razonamiento de OpenAI están acelerando su trabajo.

"Lo uso con frecuencia para reducir drásticamente las partes no tan divertidas de mi trabajo, como resumir otro estudio que podría o no ser relevante para la pregunta que estoy haciendo", comentó Brownstein. "También he podido hacer conexiones que probablemente no habría podido hacer, gracias a la capacidad de o1 de destilar conceptos genéticos complejos en explicaciones accesibles", agregó.

Pistas de lo que está por venir

La idea del pensamiento prolongado se basa en una dicotomía del pensamiento humano a la que el fallecido Daniel Kahneman se refirió como Sistema 1 y Sistema 2.

"El Sistema 1 funciona automáticamente y rápido, con poco o ningún esfuerzo y sin sensación de control voluntario", escribió el psicólogo ganador del Premio Nobel en su libro "Pensar, rápido y despacio". El Sistema 2 "asigna la atención a actividades mentales que requieren esfuerzo, incluidos los cálculos complejos", explicó. Puede adivinar qué domina la IA en estos momentos.

"La IA que estamos construyendo actualmente es básicamente como el Sistema 1 , sostuvo el científico cognitivo Gary Marcus en una conversación reciente. Las limitaciones inherentes a ese enfoque son parte de la razón por la que Marcus cree que la sociedad necesita barreras de protección de IA para evitar un "lío al estilo del Aprendiz de Brujo".

Así, el "pensamiento largo" es un esfuerzo por introducir la IA en el Sistema 2.

La capacidad de razonamiento de los nuevos modelos está aún en sus primeras fases, pero va camino de lograr avances significativos el año que viene, según Srinivas Narayanan, vicepresidente de ingeniería de OpenAI.

"Vamos a tener sistemas de IA que podrán hablar más fluidamente con nosotros, que también podrán visualizar el mundo real", afirmó Narayanan. "Y esta combinación de razonamiento y capacidades multimodales, creo, nos va a permitir construir aplicaciones *agénticas* más potentes el año que viene", realzó.

Salesforce, pionera del *software* como servicio, sigue aumentando la inversión en su motor de razonamiento Atlas, el cerebro de los agentes de IA que se pusieron a disposición del público en octubre, de acuerdo a Silvio Savarese, científico jefe y vicepresidente ejecutivo de investigación de IA de la empresa.

"Estamos impulsando a los agentes, y a nuestro propio Agentforce, hacia un razonamiento de tipo Sistema 2, lo que permite a la IA ofrecer perspectivas más profundas, impulsar acciones sofisticadas y crear un impacto significativo en todas las funciones empresariales", aseguró Savarese.

El auge de las aplicaciones basadas en modelos del Sistema 2 podría ayudar a rentabilizar la enorme inversión en IA. Como escribí en septiembre, David Cahn, socio de Sequoia Capital, asegura que la infraestructura de Nvidia necesita generar colectivamente US$600.000 millones en ingresos de por vida para justificar el gasto de las empresas en esos sistemas en el transcurso de un solo año, y no estaba ni mucho menos en camino de alcanzar pronto esa cifra.

Pero los modelos de razonamiento impulsarán al mismo tiempo la demanda de esa infraestructura de IA, incluidos los chips, el software y los centros de datos. Requieren un aumento de lo que se conoce como inferencia, o el tipo de cálculo que realizan los modelos de IA entrenados cuando responden a las peticiones de los usuarios. La inferencia también es un área en la que brillan las plataformas de Nvidia.

Y como dijo Nvidia el mes pasado en su llamada con los inversores: "El cómputo de inferencia escala exponencialmente con el pensamiento largo".

En otras palabras, "el pensamiento largo" forma parte del juego largo para la economía de la IA. WSJ

Traducido del idioma original por PULSO.