

Pequeños modelos de lenguaje: 10 Tecnologías Emergentes 2025

Por: Will Douglas Heaven. 07/05/2025

Los grandes modelos lingüísticos desataron el poder de la IA. Ahora es el momento de hacer más con menos

QUIÉNES: Instituto Allen de Inteligencia Artificial, Anthropic, Google, Meta, Microsoft, OpenAI.

CUÁNDO: Ahora.

No nos confundamos: en el mundo de la IA, **el tamaño importa**. Cuando OpenAI lanzó GPT-3 allá por 2020, era el **modelo lingüístico** más grande jamás construido. La empresa demostró que bastaba con aumentar el tamaño de este tipo de modelo para que el rendimiento se disparara. Esto dio lugar a un *boom* tecnológico que se ha mantenido gracias a modelos cada vez más grandes desde entonces. Como dijo Noam Brown, investigador científico de OpenAI, ante el público de TEDAI San Francisco en octubre, «el increíble progreso de la IA en los últimos cinco años puede resumirse en una sola palabra: escala».

Pero a medida que las ganancias marginales que logran estos modelos más y más grandes son cada vez menores, los investigadores están descubriendo cómo **hacer más con menos**. Para determinadas tareas, los modelos **modelos de lenguaje reducidos** (SML, por las siglas en inglés de *small language models*) que se entrenan con conjuntos de datos más específicos pueden rendir tan bien como los grandes o incluso mejor. Esto supone una gran ventaja para las empresas deseosas de desplegar la IA para un puñado de aspectos específicos. **No hace falta tener todo Internet en el modelo para responder al mismo tipo de petición una y otra vez.**

OpenAI ofrece GPT-4o y **GPT-4o mini**; Google DeepMind tiene Gemini Ultra y **Gemini Nano**

; y Claude 3, de Anthropic, está disponible en tres tamaños: Opus, Sonnet y **Haiku**. Microsoft es pionera en una gama de pequeños modelos de lenguaje llamados **Phi**.

Cada vez son más las empresas de menor tamaño que también ofrecen modelos pequeños. La empresa de inteligencia artificial Writer afirma que su último modelo lingüístico iguala el rendimiento de los modelos más grandes de primer nivel en muchos parámetros clave, a pesar de que en algunos **casos sólo tiene la vigésima parte de parámetros** (los valores que se calculan durante el entrenamiento y determinan cómo se comporta un modelo).

Los modelos más pequeños son más eficientes, por lo que son **más rápidos de entrenar y ejecutar**. Es una buena noticia también para quienes están preocupados por el clima: como los modelos más pequeños funcionan con una fracción de la potencia que requieren sus primos gigantes, **consumen menos energía**.

Estos pequeños modelos también viajan bien: pueden funcionar directamente en nuestros bolsillos, sin necesidad de enviar peticiones a la nube. **Lo pequeño es lo próximo**.

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ](#)

Fotografía: Technology review

Fecha de creación

2025/05/07