



GERMÁN  
ALARCO

Profesor de la  
Universidad del pacífico

El Departamento de Ciencia, tecnología e Innovación británico preparó como base para las discusiones, entre otros documentos, uno relativo a la IA que se comenta en esta nota. En la reunión participaron algunas universidades de la región, pero a nivel gubernamental de América Latina, solo Brasil y Chile. Desafortunadamente, como siempre, en el Perú seguimos mirándonos al ombligo, ignorando las grandes transformaciones internacionales.

#### ANTECEDENTES

El documento anota que estamos en medio de una revolución tecnológica que alterará fundamentalmente la forma en que vivimos, trabajamos y nos relacionamos unos con otros. La IA promete transformar casi todos los aspectos de nuestra economía y sociedad: avanzar en el descubrimiento de fármacos, hacer que el transporte sea más seguro y limpio, mejorar los servicios públicos, acelerar y mejorar el diagnóstico y tratamiento de enfermedades como el cáncer y mucho más.

La IA de vanguardia está transformando la productividad y los servicios de software. Los sistemas más avanzados pueden escribir textos con fluidez y extensión, escribir bien, obtener buenas calificaciones en exámenes escolares, generar artículos de noticias convincentes, traducir muchos idiomas, resumir documentos extensos, entre otras capacidades.

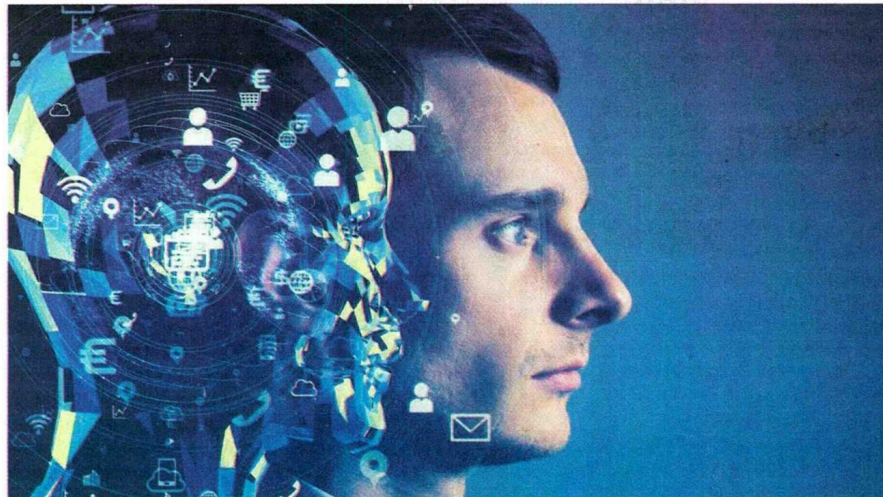
#### RIESGOS

Sin embargo, anota el documento, estas enormes oportunidades conllevan riesgos que podrían amenazar la estabilidad global y socavar nuestros valores. Para aprovechar las oportunidades, debemos comprender y abordar los riesgos. La IA plantea riesgos que no respetan las fronteras nacionales.

Es importante que los gobiernos, el mundo académico, las empresas y la sociedad civil trabajen juntos para sortear estos riesgos, que son complejos y difíciles de predecir, para mitigar los peligros potenciales y garantizar que la IA beneficie a la sociedad.

# Capacidades y riesgos Artificial (IA) de frente

**GRAN BRETAÑA CONVOCÓ**, la primera semana de noviembre, a una Cumbre Internacional de gobiernos, organizaciones internacionales, Universidades y la industria. El objetivo era discutir las capacidades y riesgos de la IA, especialmente de frontera o de vanguardia.



#### CONTENIDO

El gobierno británico cree que se necesita más investigación sobre el riesgo de la IA. El informe explica por qué. Describe el estado actual y las tendencias clave relacionadas con las capacidades de IA de vanguardia, y luego explora cómo las capacidades de IA de frontera podrían evolucionar en el futuro y revisa algunos riesgos clave. Existe una gran incertidumbre en torno a las capacidades y los riesgos de la IA, incluidos algunos expertos.

El informe cubre muchos riesgos, pero enfatiza que el riesgo general es una pérdida de confianza y confiabilidad en esta tecnología que nos negaría permanentemente a nosotros y a las generaciones futuras sus beneficios

transformadores positivos. Al discutir los otros riesgos, lo hacen con el fin de impulsar acciones para mitigarlos.

#### DEFINICIÓN BÁSICA

Según el documento, definir la IA es un desafío, ya que sigue siendo una tecnología que evoluciona rápidamente. Para los propósitos de la Cumbre, se define IA de frontera o de vanguardia como modelos de IA de propósito general altamente capaces que pueden realizar una amplia variedad de tareas e igualar o superar las capacidades presentes en los modelos más avanzados de la actualidad.

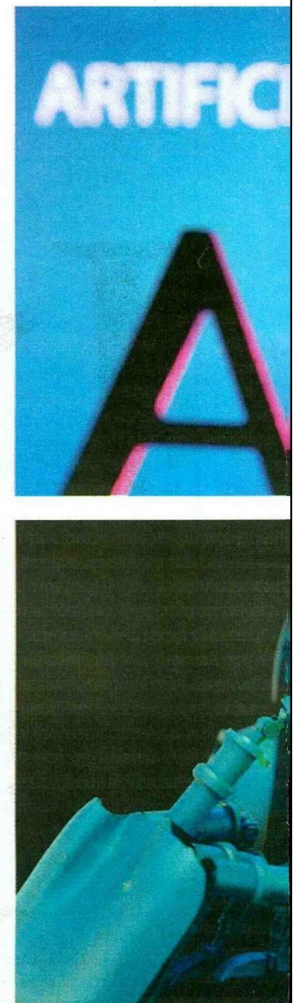
Hoy en día, esto incluye principalmente modelos de lenguajes grandes (LLM) como los que subyacen a ChatGPT, Claude y Bard. Sin embargo, es

importante señalar que, tanto hoy como en el futuro, los sistemas de IA de frontera podrían no estar respaldados por LLM, y podría estar respaldado por otras tecnologías.

#### CÓMO FUNCIONAN IA

En el informe se anota que las empresas de vanguardia en inteligencia artificial como OpenAI, DeepMind y Anthropic desarrollan grandes modelos de lenguaje (LLM) como GPT-4 en dos fases: preentrenamiento y ajuste. Durante la capacitación previa, un LLM lee millones o miles de millones de documentos de texto. A medida que lee, palabra por palabra, predice qué palabra vendrá a continuación.

Al inicio del pre-entrenamiento predice aleatoriamente, pero a medida que



ve más datos, aprende de sus errores y mejora su rendimiento predictivo. Una vez finalizado el entrenamiento previo, el modelo es significativamente mejor que los humanos a la hora de predecir la siguiente palabra de un documento de texto elegido al azar. Durante el ajuste fino, la IA pre