Medio	Fecha
gestion.pe	17/11/2023 11:05:00
	Valorización
	US $ 2,353.00

PROHIBIDA SU REPRODUCCIÓN O DISTRIBUCIÓN

revia

Conoce más sobre revia y encuentra un asesor de confianza en tu ciudad

Quiero más info

Herejías Económicas | Germán Alarco

Capacidades y riesgos de la Inteligencia Artificial (IA) de frontera o vanguardia

GERMÁN ALARCO

Gran Bretaña acaba de, en la primera semana de noviembre, a una Cumbre Internacional de gobiernos, organizaciones internacionales, científicos y la sociedad civil, el objetivo era discutir las capacidades y riesgos de la IA, especialmente de frontera o de vanguardia.

El Departamento de Ciencia, tecnología e Innovación británico (DSIT) creó un documento de discusión, entre otros documentos, una relatoría 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 estuvieron presentes, en cambio, en el Perú seguimos mirándonos al ombligo, ignorando las grandes tendencias internacionales.

Antecedentes

El documento anota que estamos en medio de una revolución tecnológica que afectará fundamentalmente la forma en que vivimos, trabajamos y nos relacionamos con otros. La IA promete transformar casi todos los aspectos de nuestra existencia, en el desarrollo económico, hacer que el transporte sea más seguro y fluido, ayudar a los servicios públicos, acelerar y mejorar el diagnóstico y tratamiento de enfermedades como el cáncer y otras dolencias.

La IA de vanguardia está transformando la productividad y los servicios de software. Los asistentes digitales pueden escribir textos con fluidez y extensión, escribir bien, obtener buenas calificaciones en exámenes, generar artículos de noticias convenientes, traducir muchas lenguas, acceder documentos externos, entre otras capacidades.

Riesgos

Sin embargo, anota el documento, estas enormes oportunidades también ofrecen más que suficiente amenaza la estabilidad global y socavar nuestros valores. Para aprovechar las oportunidades, debemos proveer 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 realicen conjuntamente estos riesgos, que son complejos y difíciles de predecir, para mitigar los peligros inevitables y garantizar que la IA beneficie a la sociedad.

Contenido

El gobierno británico cree que se necesita más investigación y trabajo en la IA de frontera para: Describir el estado actual y las tendencias clave relacionadas con las capacidades de la IA de frontera; predecir y evaluar las capacidades de IA de frontera posibles evolucionar en el futuro y revelar algunos riesgos clave. Existe un incertidumbre en torno a las capacidades y los riesgos de la IA, incluso algunos expertos.

El informe cubre muchos riesgos, pero enfatiza que el riesgo general es que perdamos el control de la IA en esta tecnología que nos negaría permanentemente a nosotros y a las generaciones futuras los beneficios positivos. Al discutir los otros riesgos, lo hacen con el fin de implantar acciones para mitigarlos.

Definición básica

Según el documento, definir la IA es un desafío, ya que sigue siendo un campo en evolución. En el informe, definen las IA de frontera o de vanguardia como modelos de IA de propósito general (modelos de lenguaje) capaces de realizar una amplia variedad de tareas y 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, Bard y Claude. Es importante recalcar que, aunque hoy, esto es el LLM, la definición de IA de frontera también no está respaldada por el LLM y puede estar cambiando con otras tecnologías.

Cómo funcionan IA

En el informe se anota que las empresas de vanguardia en inteligencia artificial (OpenAI, DeepMind y Anthropic) desarrollan grandes modelos de lenguaje (LLM) como GPT-4 y Claude 2 que pueden procesar y producir texto, o multimodales, que pueden procesar imágenes, o multimodales, de entrenamiento de texto. A medida que les, palabra por palabra a continuación.

Al inicio del pre entrenamiento predice aleatoriamente, pero a medida que ve más datos, aprende de sus errores y mejora su predicción. Después de un ser finalizado el entrenamiento previo, el modelo es significativamente inteligente y puede predecir de forma precisa la siguiente palabra de un documento de texto elegido al azar.

Durante el ajuste fino, la IA puede se entrena aún más en conjuntos de datos altamente seleccionados, que se centran en tareas más especializadas o están estructurados para dirigir el comportamiento del modelo de manera más alineada con los valores del desarrollador y las expectativas del usuario.

Tareas posibles

Conversar con fluidez y detalladamente; Resumir documentos extensos; escribir largos secuencias de código que funcione con relativamente a partir de instrucciones en lenguaje natural; responder preguntas sobre lenguaje que requieran razonamiento de sentido común; y traducir documentos.

Asimismo, acelerar la investigación académica, por ejemplo, en ciencias de la vida, ciencia gráficos y calculando cantidades clave; dirigir las actividades de los robots mediante el razonamiento, la planificación y el control de movimientos, entre otras.

Por otra parte, los modelos son más útiles cuando se acompañan con otras herramientas y software. Los investigadores han creado programas de software llamados "andamios" que ayudan a los modelos a ejecutar tareas impulsoras agentes de IA autónomos. Se anota como ejemplos, encontrar información específica, apoyar la resolución de problemas complejos o jugarse de supervivencia de mundo abierto como Minecraft y Crafter.

El Futuro

Hay dudas en las respuestas que generan actualmente los LLM de IA, pero según el informe el progreso reciente de la IA ha sido rápido y probablemente continúa. Esto se debe a mejoras predecibles en el rendimiento de los modelos de IA de vanguardia cuando se desarrollan con más computación, más datos y mejores algoritmos.

También pueden surgir nuevas capacidades inesperadas. En un futuro no muy lejano podrían desarrollarse agentes avanzados de IA de uso general, aunque esto es un tema de debate, especialmente en lo que respecta al momento.

Sin embargo, existe una gran incertidumbre sobre el investigadores, si no la mayoría, no esperan que los sistemas de IA alcancen general el desempeño humano hasta dentro de veinte años y no están de acuerdo en que esto sea una sucesión gradual. Además, históricamente, y con frecuencia, hay bastante predicciones de avances inesperados en la IA que se materializan.

Riesgos

Se anota que debemos comprender los riesgos asociados con la IA de vanguardia para acceder y aprovechar de forma segura las oportunidades y beneficios que brinda la tecnología. Estos pueden ser factores de riesgo transversales, condiciones sociales que podrían agravar una serie de riesgos particulares. Por otra parte, hay riesgos específicos de IA de frontera, algunos actuales, otros futuros y pérdida de control.

Los riesgos transversales se refieren a los numerosos desafíos técnicos de larga data para construir sistemas de IA seguros, confiables y alineados y cómo tomar decisiones. Estos, según el informe, presentan fallas, vulnerabilidades y sesgos en su uso.

En general, los sistemas de IA de frontera no son sólidos, es decir, con frecuencia la falla en enfrentarse a entornos diferentes a sus datos de entrenamiento. Aunque la robustez de la IA es un área de investigación activa, en la práctica, la falta de robustez sigue siendo un problema sin resolver que afecta a todo tipo de modelos de aprendizaje automático.

Seguridad nacional

Algunos investigadores han argumentado que la industria de la IA tiene vínculos cercanos con industrias altamente centradas en la seguridad, como la vigilancia y la inteligencia militar. Sin embargo, los estándares de seguridad de la IA aún se encuentran en un nivel inicial. El trabajo de organizaciones de desarrollo de normas todavía está en otras muchas áreas.

Por otra parte, actualmente hay poca capacidad gubernamental y se está trabajando para construir un ecosistema maduro. Un desafío, según el informe, es que las normas se necesitan en un país y luego se implementan en otro, lo que aumenta la necesidad de un enfoque global. Hay también incentivos insuficientes para que los desarrolladores de IA inviertan en medidas de mitigación de riesgo.

Concentración poder

Los investigadores han comenzado a explorar la probabilidad de una alta concentración de poder de mercado entre los desarrolladores de IA de frontera. Los altos costos fijables asociados con el entrenamiento de modelos de IA de frontera crean economías de escala y barreras significativas de entrada para los actores más pequeños.

Una concentración considerable del poder de mercado podría debilitar la competencia, reducir la innovación y las opciones de los consumidores. La pérdida de opciones del consumidor también significa que tienen menos voz en el uso de sus datos personales, una posible manipulación del comportamiento, vigilancia y una erosión de las normas democráticas.

Datos sesgados

El documento señala que existe una amplia gama de posibles daños sociales que ha llevado hasta ahora a un debate en torno a la ética de la IA, con una amplia investigación y preocupación. Aquí se centran sólo en unos pocos daños sociales, pero esto no es para restar importancia a los demás.

Ellos se enfocan en la degradación de la información (al representar falsamente personas y eventos) y en la discriminación laboral. Las comunidades ven la automatización y el desplazamiento en los mercados laborales como uno de los desafíos más probables a los que la IA puede el bienestar social, aunque eventualmente también puede generar mejoras en las condiciones laborales, reduciendo históricamente la demanda de mano de obra en trabajos más peligrosos.

Disrupción laboral

Los estudios sugieren que los sectores con mayor exposición a la disrupción del mercado laboral debido a la IA actual son: los financieros y legales, mientras que la industrial, la manufactura, la agricultura y la minería son las menos expuestas.

Los modelos de IA de frontera pueden contener y magnificar sesgos arraigados que no están claros, lo que refuerza desigualdades y estereotipos sociales e históricos. Estos sesgos, a menudo sutiles y profundamente arraigados, comprometen el uso equitativo y ético de los sistemas de IA, lo cual dificulta que la IA pueda mejorar la equidad en las decisiones.

Sin control

Un problema que se aborda en el informe, antes de sus conclusiones, es que la pérdida de control podría acelerarse si los sistemas de IA tienen medios para aumentar su propia inteligencia y reducir el control humano.

Al respecto, los sistemas actuales tienen algunas capacidades básicas que, al continuar el rápido progreso de la IA, podrían usarse para aumentar su propia inteligencia, aunque actualmente estas capacidades no son suficientes para plantear riesgos significativos.

Ir a más noticias

Cómo evitar que la inteligencia artificial nos haga perder el empleo

Principios para la inteligencia artificial: ¿cómo evitar que nos dominen las máquinas?

Una agenda para la inteligencia artificial: propuestas

Cómo evitar que la inteligencia artificial nos haga perder el empleo

Publicación de la Unión Europea sobre inteligencia artificial: ¿qué es y para qué sirve?

SOBRE EL AUTOR

GERMÁN ALARCO

Economista. Doctor en Economía y Ciencias Sociales (Universidad Paris 1 Panthéon-Sorbonne). Ex director del Instituto Nacional de Estadística e Informática (INEI). Ex director del Instituto Nacional de Innovación Agraria (INIA). Ex director del Instituto Nacional de Salud (INS). Ex director del Instituto Nacional de Desarrollo Social (INDESOL). Ex director del Instituto Nacional de Cultura (INC). Ex director del Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual (INDECOPI). Ex presidente del CONAS. Ex miembro de diversas comisiones de alto nivel del sector público. Más información sobre su experiencia.

ÚLTIMAS NOTICIAS

Telemundo En Vivo: cómo ver transmisión del Mundial 2023 por TV y online

Aeropuerto Chinchero tiene un avance de 8% en obras, estará listo en 24 meses

Jornada laboral de 40 horas, cuando entrará en vigencia en Ecuador

Aprende a identificar el billete de 5 soles y evita ser víctima de estafas

Controversia sobre obras de El Sauce: "Se ha ignorado a 25 familias"