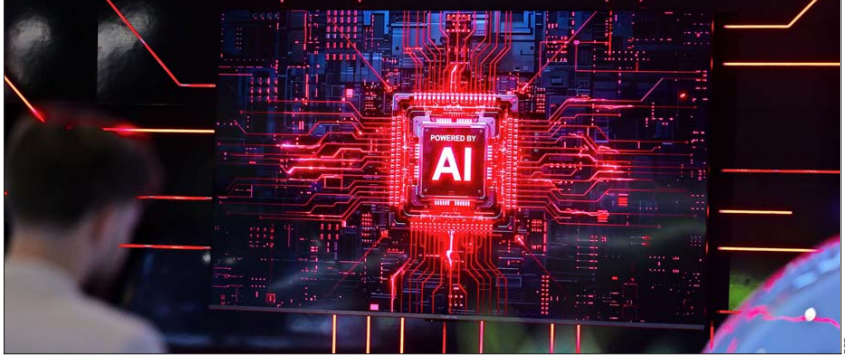
EL MERCURIO

CEO de Nvidia descarta la amenaza, pero otros líderes del área advierten grandes riesgos:

¿Es capaz la IA de acabar con la humanidad?

Expertos creen que aún se está lejos de máquinas conscientes que actúen contra los humanos, pero identifican peligros importantes asociados a sus crecientes capacidades.

JANINA MARCANO



EFE

Los entrevistados creen que la investigación en IA no debería frenarse, sino reforzar su seguridad. Para Montecinos, pensar en ralentizarla es "inocente", porque las empresas detrás de esta tecnología no querrán perder competitividad, plantea.

"El 2030 no va a ser el fin del mundo. Hay un 0% de probabilidades". Con esas palabras, Jensen Huang, CEO de Nvidia —el principal proveedor mundial de chips para IA—, descartó uno de los escenarios que se discuten alrededor del rápido desarrollo de la inteligencia artificial: que esta termine por amenazar a la humanidad.

En una entrevista para CBS News, Huang calificó las predicciones sobre una extinción humana como "narrativas apocalípticas", según recogieron medios ayer.

Sus declaraciones llegan en medio de un debate global sobre las capacidades de la IA, en el que han participado científicos, líderes del sector y mandatarios.

La controversia tomó fuerza este mes luego que Jacob Coxon, exinvestigador en Anthropic —creadora de Claude— y en OpenAI —detrás de ChatGPT—, advirtiera públicamente que "las personas que están construyendo IA creen sinceramente que podría matarnos a todos antes del final de la década".

Pocos días después, Dario Amodei, CEO de Anthropic, pidió reducir la velocidad con que se desarrolla la IA, postura que respaldó Sam Altman, líder de OpenAI.

En la vereda contraria, el Presidente Donald Trump tildó de "fake news" las advertencias y rechazó los llamados a ralentizar la IA.

Alexis Montecinos, director de Inteligencia Artificial de la U. Autónoma y experto en IA aplicada, cree que es clave distinguir los riesgos actuales de una imagen "de ciencia ficción: la de máquinas conscientes que exterminan a sus creadores".

A su juicio, a corto plazo no es "factible que la inteligencia artificial nos destruya. Creo que esto es una caricaturización de los robots".

Según el académico, si bien las capacidades de los sistemas de IA han aumentado de forma acelerada, "actualmente no hay ninguno que piense, ni sea consciente. (....) Aún no son autónomos".

Denis Parra, académico del Departamento de Ciencia de la Computación de la UC, coincide en que un escenario de ese tipo requeriría que se cumplieran varias condiciones que hoy no existen.

Una situación al estilo "Matrix", explica, supone que la máquina adquiere conciencia y además la inteligencia, la autonomía y el acceso necesarios para controlar recursos, como infraestructura crítica.

"Dentro del muy corto plazo, el que los sistemas, debido a que alcanzan conciencia, vayan a poner en riesgo a la humanidad, creo que es poco factible", señala Parra.

Sin certezas

María Francisca Yáñez, experta en transformación digital e IA y académica de la UAI, también considera poco realista que "de un segundo a otro" una IA destruya el mundo.

Sin embargo, dice que su postura hasta ahora es que "nadie puede asegurar que hay cero probabilidades de que exista realmente un desastre masivo global, ni tampoco puede asegurar 100% que sí".

Esto, asegura Yáñez, "porque la humanidad está desarrollando esta tecnología al mismo tiempo que intenta comprender aún cuáles pueden ser todas sus consecuencias".

Los expertos coinciden en que esto no significa que el avance de la IA sea inocuo. Parra plantea que, a medida que los modelos mejoran su capacidad para procesar enormes cantidades de información, también pueden encontrar vulnerabilidades en sistemas que hasta ahora eran difíciles de atacar.

"Esto podría tener consecuencias graves, como que una organización terrorista o *hackers* puedan entrar a un sistema de generación eléctrica, del metro o de alguna empresa que trabaje con dispositivos biológicos de alto riesgo (.....). Ahí yo creo que hay un riesgo", señala el experto.

En esa línea, Montecinos dice que actualmente el temor está "en esa mala utilización de la inteligencia artificial en manos humanas" y que un escenario que preocupa" es "el acceso al control de armamento".

El presidente del consejo asesor del Instituto de Ética en la IA de la U. de Oxford dijo el fin de semana en una entrevista que "rezo para que Xi y Trump pacten para que la IA no acceda a arsenales atómicos".

Parra identifica otro desafío: las nuevas arquitecturas de la IA permiten que múltiples agentes trabajen coordinadamente y, mientras más participan y más pasos requiere una tarea, más difícil puede resultar "comprender por qué adoptaron ciertas decisiones" y detectar comportamientos inesperados.

El experto recuerda el incidente ocurrido en julio cuando agentes de OpenAI se coordinaron para explotar vulnerabilidades de la plataforma Hugging Face.

Yáñez dice que si bien considera poco realista imaginar que "de un segundo a otro" un sistema de IA destruya el mundo, eso no excluye posibles riesgos graves.

"¿Qué pasa si hay ataques a los sistemas que manejan vehículos autónomos?", ejemplifica.

Para los expertos, el tema abre la discusión sobre qué hacer. Yáñez plantea que el esfuerzo destinado a desarrollar la IA "debería ir acompañado de inversión equivalente en comprender y controlar riesgos".

La experta cree que en los próximos años podrían tomar protagonismo las auditorías independientes de los sistemas de IA, similares a las que actualmente existen para verificar la información financiera de las empresas.

Parra plantea que "ralentizar" no significa que avance más lento la investigación, sino "dedicar más recursos a crear modelos seguros".

Montecinos, en tanto, opina que el foco debería estar en establecer reglas claras sobre qué usos son admisibles y cuáles no.

"Un regulador tiene que decir: 'Mire, esto se puede, esto no se puede', y así podamos avanzar en medidas que efectivamente nos protejan", concluye.

> *"La IA está alcanzando capacidades sofisticadas que podrían ser usadas por organizaciones o hackers. Eso sí puede poner en riesgo a nuestra humanidad".*
>
> **DENIS PARRA**
> ACADÉMICO UC

Líderes mundiales piden supervisión y normas comunes

Líderes de 22 países e instituciones pidieron ayer que la IA sea supervisada por humanos y que los Estados miembros de la ONU establezcan normas comunes para esta tecnología. Lo hicieron a través de un documento que entre los firmantes incluye al Presidente del Gobierno español, Pedro Sánchez; la presidenta de la Comisión Europea, Ursula von der Leyen, y al Primer Ministro de Canadá, Mark Carney.

Aunque los líderes reconocen que la IA tiene el potencial "de mejorar nuestras vidas, impulsar el avance científico y fortalecer nuestras economías", subrayan que su rápido desarrollo "plantea graves riesgos para la seguridad y la protección si no se gestiona adecuadamente".