EL MERCURIO

Fecha: 18/04/2026 Audiencia: 320.543 Sección: OPINION
Vpe: $2.417.090 Tirada: 126.654 Frecuencia: 0
Vpe pág: $20.570.976 Difusión: 126.654
Vpe portada: $20.570.976 Ocupación: 11,75%

"Mythos" y los peligros emergentes

El 7 de abril, la empresa de inteligencia artificial (IA) Anthropic publicó un documento en que describía las potencialidades de la nueva versión de su sistema Claude, Mythos Preview. El anuncio causó impacto, pues Anthropic declaró que, por ahora, no distribuirá al público este nuevo sistema, debido a los riesgos que supondría para los sistemas computacionales de empresas, gobiernos y personas. Esto creó expectación, además de preocupación, particularmente en la banca, que depende de que sus sistemas informáticos no se vean comprometidos por actores adversarios.

Anthropic fue formada por exempleados de Open AI, creador de Chat GPT, que tenían diferencias éticas y filosóficas con el liderazgo de aquella firma. El número de compañías que pagan por su producto se acerca al de Chat GPT, pero con una tendencia al alza, a diferencia de la tendencia más estática de su competidor.

La empresa ha estado en la prensa en los últimos meses. En febrero impactó el mercado accionario norteamericano, al ofrecer una versión de Claude orientada a la industria: se temió que pudiera reemplazar los servicios ofrecidos por empresas especializadas de *software* y por proveedores de datos financieros. Luego, tuvo un conflicto con el Pentágono, al imponer limitaciones al uso de su IA en tareas ilegales, lo que aumentó su popularidad.

Su más reciente versión de Claude tendría una aptitud especial para tareas relacionadas con seguridad informática. En sus pruebas descubrió miles de vulnerabilidades en *softwares* de uso común en las empresas. Estas vulnerabilidades —que llevaban años y hasta décadas sin detectarse— tienen el potencial de bloquear sistemas informáticos o permitir que sean controlados por agentes externos. Pero además, Mythos podría elaborar por sí mismo, en forma autónoma, el *software* requerido para explotar esas vulnerabilidades, hasta obtener el control total de un sistema. Ante ese riesgo, Anthropic decidió no ofrecer el producto, al menos, hasta que las principales debilidades en los sistemas informáticos más vitales sean subsanadas.

El problema es que una vez que existe un programa de IA que puede realizar estas labores, pronto podrán hacerlo programas de otros actores, por lo que la decisión de la firma solo sirve en el corto plazo. Hay muchos países que participan en una guerra sorda de sabotaje informático, como Rusia, Ucrania, China, Israel, Irán, Corea del Norte, Estados Unidos, India, países europeos y otros más de Asia. Saber que existe la posibilidad de una IA con alta capacidad para explotar vulnerabilidades es un aliciente para investigar en esa dimensión.

> *¿Qué hace Chile en esta materia? No parece una preocupación especial de los gobiernos.*

Esto releva la pregunta: ¿Qué hace Chile en esta materia? No parece ser una preocupación especial de los gobiernos. Aunque se han establecido políticas de ciberseguridad, nuestros desarrollos tecnológicos son muy básicos. Es posible que no haya capacidad técnica o recursos, pero ello arriesga dejarnos en una situación de dependencia, no solo ante otros Estados, sino también frente a bandas criminales que utilicen las nuevas herramientas.