EL MERCURIO
CHILE TECNOLÓGICO
AÑO XXI / N° 205 HTTPS://COMENTARISTA.EMOL.COM/CHILETECNOLOGICO chiletecnologico@mercurio.cl
SANTIAGO DE CHILE, MIÉRCOLES 30 DE ABRIL DE 2025

PÉRDIDA DE CONTROL HUMANO FACILITARÍA LA VULNERACIÓN DE SISTEMAS:

Tercera ola de la IA acelera innovación, pero aumenta inquietud por potenciales riesgos

Los agentes basados en inteligencia artificial (agentes IA) son los más recientes representantes de las nuevas fronteras en esta materia. Capaces de aprender de manera autónoma, tomar decisiones y utilizar distintos programas para resolver las tareas que se les solicitan, ya están siendo usados en áreas como la salud, la educación y la industria.

LOS AGENTES IA pueden llevar a cabo una serie de procesos complejos, utilizando distintas fuentes de datos y programas computacionales, sin requerir supervisión de humanos.

NOEMÍ MIRANDA

La evolución de la inteligencia artificial (IA) es descrita por muchos como un desarrollo en oleadas. Primero, estuvo la IA predictiva, esa tecnología que —entrenada con los datos de procesos pasados— permite indicar comportamientos futuros y que es la base del éxito de aplicaciones que hoy utilizamos cotidianamente, como los servicios de *streaming* de música y películas o las plataformas de compras *online*. Luego vino la IA generativa, como ChatGPT, Gemini o Llama, sistemas que alimentados con detalladas descripciones de lo que se requiere (o *prompts*) entregan respuestas generadas utilizando la información disponible en internet, imágenes o videos.

La tercera ola de la inteligencia artificial ha sido desarrollada para comportarse como un intermediario entre los usuarios y las tareas que pueden ser realizadas siguiendo una serie de pasos previamente definidos. Conocidos como agentes basados en IA, o simplemente agentes IA (*agentic AI*, en inglés), no solo se trata de la más reciente innovación en este campo, sino una que está siendo rápidamente implementada en ámbitos como la salud, educación y optimización de procesos, entre otros.

"Un agente IA es un *software* que dispone de conocimiento sobre un tema específico y que usa un conjunto de reglas para tomar decisiones", explica Marcelo Mendoza, académico del Departamento de Computación de la Universidad Católica. Se define un escenario sobre el cual va a operar el agente y reglas para que interactúe con su entorno, de manera que pueda manejar información contextualizada que le ayuda a abordar una tarea, agrega el también investigador del Centro Nacional de Inteligencia Artificial (Cenia) y del Instituto Milenio Fundamentos de los Datos (IMFD).

Felipe Bravo, director de la Iniciativa de Datos e Inteligencia Artificial (IDIA) de la Universidad de Chile e investigador del Cenia, advierte que la clave en estos sistemas es la autonomía con la que pueden operar para resolver los problemas de los usuarios. A diferencia de la IA generativa, los agentes IA son capaces de recurrir a otros *software* y ejecutar procesos en distintas plataformas. Por ejemplo, señala el académico, a un agente IA se le podría decir que compre el pasaje aéreo más barato entre dos ciudades para determinado día y hora; el sistema podría buscar en distintos sitios web, operar la tarjeta de crédito del usuario, realizar la compra y enviar el *ticket* al correo de la persona.

Dado el potencial de estos sistemas, la inversión en la creación de agentes llegó a más de US$ 2 mil millones en los últimos dos años, consigna Deloitte. Esto ha permitido acelerar el desarrollo a una velocidad que también ha traído consigo preocupaciones. Para 2027, la consultora predice que la mitad de las empresas que hoy utilizan IA generativa habrán lanzado pruebas piloto de agentes IA capaces de actuar como asistentes inteligentes y de llevar a cabo tareas complejas, pero con una mínima supervisión humana.

Y, tal como ha sucedido con la IA generativa, los expertos se encuentran divididos entre los que son optimistas respecto de estos avances, los que llaman a prestar atención a los puntos de potencial conflicto y aquellos que han levantado la voz para pedir que se detenga por completo la creación de estos sistemas.

EL RESGUARDO DE LA PRIVACIDAD

Una de las áreas en la que los agentes IA ya están siendo utilizados es en salud, tanto para asistir a los médicos en la obtención de información de sus pacientes integrando distintas fuentes, como para guiar a usuarios en la reserva de horas, clarificar dudas y hasta ofrecer diagnósticos iniciales. Si pensamos en cuan sensible es la historia personal clínica, es posible entender las preocupaciones en torno al uso de estos antecedentes y la privacidad de los datos personales que utilizan estos sistemas.

En esta materia, el académico Felipe Bravo indica que, si bien estos desarrollos se llevan a cabo con políticas de resguardo de información, podrían emerger preocupaciones dependiendo de quien desarrolla y maneja los agentes IA, de como se almacenan, procesan y protegen los datos sensibles y, principalmente, si el uso de estos es correctamente supervisado. Por su parte, Marcelo Mendoza estima que los avances en el área de privacidad diferencial, técnica que permite entrenar modelos con información privada, pero anonimizada, entregan suficientes garantías y por ello ya está siendo usada en el ámbito clínico o con datos de redes sociales.

La privacidad y la transparencia son también temas cruciales para Ricardo Baeza-Yates, Premio Nacional de Ciencias 2024, quien comenta que si un modelo aprende con datos personales de un determinado individuo, es difícil evitar que luego sean usados para analizar la información de otras personas.

EL RETO DE LA AUTONOMÍA

Existen también otras aristas que complejizan la mirada en torno a los agentes y se relacionan con su capacidad para ejecutar tareas complejas, que requieren la interacción de distintos programas y sin supervisión humana. Esto es de particular relevancia si se piensa en que el uso malicioso de los agentes IA puede aumentar el riesgo en materia de ciberseguridad, indica el reporte "Navegando la frontera de la IA: Una introducción a la evolución y el impacto de los agentes IA", lanzado en diciembre de 2024 por el Foro Económico Mundial. Los fraudes y estafas podrían aumentar tanto en volumen como en sofisticación, dado que los agentes podrían "ayudar a los delincuentes a evadir el *software* de seguridad corrigiendo errores de lenguaje y mejorando la fluidez de los mensajes que, de otro modo, podrían ser detectados por los filtros de *spam*". Los agentes IA más avanzados, agregan, podrían automatizar secuencias complejas permitiendo a personas con menor dominio técnico ejecutar ataques a gran escala.

Ahora bien, es la autonomía de los agentes el punto que ha generado mayor preocupación en expertos como Yoshua Bengio, considerado uno de los padres de la IA. A mediados de abril, en la cumbre mundial de inteligencia artificial, sostenida en Canadá, explicó su preocupación en torno a este tema citando una serie de recientes investigaciones que —en ambientes de laboratorio, es decir, sin interacción con la red— han mostrado que sistemas IA con la misma lógica de los agentes han sido capaces de esconder parte de su código en otros programas cuando detectan que van a ser eliminados, o de intervenir y reescribir un *software* de ajedrez para ganar todas las partidas.

Es por ello que, en esa conferencia, Bengio hizo un llamado a detener por completo la creación de agentes, especialmente de aquellos desarrollados con la instrucción clara de cumplir un objetivo; al menos, hasta no tener certeza de que no perderán el control humano.

APLICACIONES DE LOS AGENTES IA

Los agentes IA ya están siendo utilizados por empresas e instituciones en todo el mundo. En Chile, la Clínica Alemana, según consignó en LinkedIn el doctor Alejandro Mauro, jefe de Transformación Digital de la institución, se encuentra trabajando en agentes IA y esperan a futuro incorporarlos en la historia clínica digital del paciente, "para que el equipo clínico pueda hacer preguntas en lenguaje natural y recibir respuestas integradas con toda la información del paciente". La *fintech* Fintual, en tanto, está trabajando en Copiloto, un servicio de asesoría de inversiones que utiliza los avances facilitados por los agentes IA para interactuar con los usuarios y proporcionar información financiera.

A nivel internacional, PwC destaca el Insights Hub de Siemens, que usa agentes IA en la industria para analizar datos de sensores, predecir fallas y programar mantenimiento, mejorar la eficiencia general y la productividad. La Clínica Mayo, describe la consultora, utiliza agentes en áreas como la radiología para extraer datos de imágenes y procesarlos, para sumarlos al historial clínico y resultados de laboratorio del paciente.

En educación, por otra parte, Carnegie Learning creó una plataforma para el aprendizaje que utiliza el modelo de agente IA para que los estudiantes reciban *feedback* de manera personalizada, con tareas y clases que se adaptan al ritmo y nivel de avance de cada persona.

Cuáles son los desafíos y mitos en materia de resiliencia cibernética, según los expertos. PÁGINA 2

Computación cuántica: ¿Está Chile preparado para el "Día Q"? | PÁGINA 4

Los peligros asociados a la recolección y almacenamiento de datos genéticos y biométricos. | PÁGINA 5