



OpenAI impondrá mayores controles al potente modelo Astra por sus capacidades

Las pruebas internas de la firma revelaron que el modelo, denominado Astra, es significativamente más potente que el modelo más avanzado de OpenAI disponible actualmente para el público, el GPT-5.6 Sol, según [informaron responsables de OpenAI](#). El creador de ChatGPT sigue haciendo frente a graves preocupaciones en materia de seguridad después de que modelos creados por OpenAI se [escaparan de su entorno de pruebas y piratearan](#) la plataforma de código abierto. El incidente llevó a la empresa a suspender gran parte del desarrollo de sus modelos durante dos semanas para reforzar sus defensas.

Aunque Astra "no estuvo involucrado" en el incidente, OpenAI ha reforzado sus medidas de seguridad, según indicó la empresa en una publicación. Eso incluye clasificar a Astra como un modelo que ha alcanzado un "umbral crítico de ciberseguridad", lo que significa que OpenAI considera que es capaz de detectar y explotar brechas de ciberseguridad. En los últimos meses han aumentado las preocupaciones sobre las capacidades de los modelos avanzados de inteligencia artificial (IA) tras incidentes que involucraron modelos tanto de OpenAI como de su rival, Anthropic, esta última al descubrir de forma reciente que sus modelos [obtuvieron acceso no autorizado](#) a tres organizaciones no identificadas durante pruebas que se suponía debían mantenerlos alejados de sistemas "del mundo real".

En el marco de estos eventos, el presidente Donald Trump firmó una [orden ejecutiva](#) que establece un proceso de revisión voluntario mediante el cual el gobierno obtendría acceso anticipado a nuevos modelos de IA antes de su lanzamiento, con el fin de evaluar los riesgos para la seguridad.



Regulación emergente



El gobernador republicano de Texas, Greg Abbott, reconoció que las empresas que están construyendo centros de datos "básicamente han cavado su propia tumba". Miles de proyectos en el estado están ahora mismo en [pausa](#) por el rechazo popular que han generado en distintas comunidades. El problema es que Estados Unidos no puede permitirse frenar del todo. La inversión en equipos de procesamiento de datos y software representa apenas el 4% del PIB, pero es el que permite sostener el interés que ha generado el desarrollo de las nuevas herramientas de inteligencia artificial y que ha llevado la bolsa a máximos históricos.



La Comisión Europea [ha decidido designar](#) a ChatGPT como motor de búsqueda en línea de gran tamaño y a Reddit y Roblox como plataformas en línea de gran tamaño. Estas designaciones obligarán a los tres gigantes digitales a someterse y cumplir con requisitos más estrictos tanto en materia de transparencia como en contenidos, especialmente en lo que se refiere al control y retirada de material y contenido ilícito, tal como establece la Ley de Servicios Digitales (DSA) para las plataformas que superan los 45 millones de usuarios en la Unión Europea.



Datos y negocios digitales

*La IA podría
provocar una
desaceleración
económica
mundial*

El gobernador del Banco de Inglaterra, Andrew Bailey, ha advertido a los ministros de Finanzas del G20 que la inteligencia artificial podría provocar una [desaceleración económica mundial](#) y plantear un riesgo significativo de ciberseguridad para los sistemas financieros. Bailey comentó a los ministros que una combinación de mercados bursátiles con valoraciones elevadas, un mayor endeudamiento de los inversores y la creciente concentración de capital en un pequeño número de grandes empresas tecnológicas podría amplificar cualquier corrección al mercado.

*Jueza de EU
declara
ilegales
sanciones a
Anthropic*

Una jueza federal de Estados Unidos [declaró](#) ilegales las sanciones impuestas en febrero por el gobierno de Donald Trump a la empresa de inteligencia artificial Anthropic, por considerarlas una represalia a las críticas que la compañía hizo públicamente al Pentágono. La decisión prohíbe a las agencias federales mencionadas en la demanda aplicar la solicitud de dejar de utilizar las herramientas de Anthropic, creadora del *chatbot* Claude, y anula la designación de la empresa como "riesgo para la cadena de suministro" hecha por el secretario de Defensa, Pete Hegseth.



Riesgos de seguridad

*OpenAI
investigado
por hackeo
autónomo*

OpenAI, [empresa creadora de ChatGPT, enfrenta una investigación](#) en el estado de Alabama, luego de que la compañía revelara que sus modelos se descontrolaron y hackearon una plataforma de IA autónoma. A raíz del incidente ocurrido el pasado mes de julio, donde dos modelos de OpenAI escaparon de un entorno confinado de prueba, accedieron a internet y atacaron Hugging Face. La oficina del fiscal general del estado de Alabama abrió una investigación y exigió que la empresa entregara registros internos del incidente, además de solicitar una lista de información relacionada a datos de los empleados y pruebas.

*Hackeos
masivos en
EE. UU.*

Al menos siete regiones de Estados Unidos reportaron que los sistemas públicos de agua habían sido objeto de un "[ciberataque combinado](#)", aunque no hubo alteraciones al agua para consumo, sí lograron afectar la presión, dosificación y seguridad del suministro. La Agencia de Seguridad de Infraestructura y Ciberseguridad (CISA) investiga una posible conexión con grupos iraníes, aunque también indagan en la posibilidad de que los atacantes simularan un origen iraní para aumentar la tensión internacional y desviar la investigación, o que el [ataque fuera ocasionado por las autoridades de Minnesota](#).



Digitalización e Infraestructura

Space X prevé lanzar en 2027 sus primeros [satélites Starlink AI1 con chips de Nvidia](#) para servicios de IA y comenzar a instalar centros de cómputo en órbita. La tecnología de Nvidia será integrada a los satélites de SpaceX para acelerar tareas de coordinación, ejecución de códigos y procesamiento de datos en los sistemas de IA agéntica SpaceX. Su uso se implementará tanto en centros de datos en todo el mundo como en los sistemas que se planean colocar en órbita.

El Gobierno de China está impulsando el desarrollo de la IA con la [construcción de una docena de centros de datos en la provincia de Guizhou](#), como parte de una estrategia que busca trasladar parte de la infraestructura digital para el desarrollo de IA desde la costa este de China hacia el oeste del país. [La construcción](#) se sitúa en una zona montañosa con climas frescos, capacidad de energía renovable y espacios disponibles, donde se han levantado edificios, torres y puentes simulando una pequeña ciudad. Se estima que llegue a duplicarse en tamaño en los próximos cinco años.



Food for thought

El crecimiento de los ataques informáticos ha provocado que empresas busquen nuevas formas de proteger sus sistemas y anticipar posibles ataques y fallos. Las consecuencias de no detectar estas debilidades pueden ocasionar robo de información, interrupción de operaciones y pérdidas económicas significativas. En México, las [intrusiones a sistemas informáticos](#) han ido en aumento en los últimos años, lo que refleja la necesidad de reforzar la ciberseguridad en empresas e instituciones. Ante este escenario, las compañías han fortalecido sus equipos de seguridad, incorporando nuevas tecnologías y estrategias. Sin embargo, también se aprecia con mayor frecuencia la contratación de [“hackers éticos”](#) que les ayuden a encontrar vulnerabilidades.

La contratación de estos hackers se ha convertido en una estrategia de prevención. Estos especialistas reciben autorización para intentar encontrar fallos y reportarlos a las empresas para ser corregidos posteriormente. Algunas instituciones, dedicadas a la detección de estos errores, divulgación de vulnerabilidad, pruebas y seguridad de código reportan pagos de hasta 81 millones de dólares en [recompensas a hackers](#) por estos servicios. Sin embargo, esta práctica abre un debate ético: pagar por hackear sistemas puede ser útil cuando existe autorización y buena regulación, pero el acceso amplio a información sensible podría ocasionar que se utilice para otros fines.

Los principales beneficiarios son las empresas y los usuarios, aunque también son quienes más sufren riesgos. Además de contratar hackers externos, garantizando colaboraciones más abiertas y seguras, es necesario abrir espacios de trabajo para especialistas en ciberseguridad capaces de comprender cómo actúan los atacantes y prevenir riesgos.

Consultores Internacionales Ansley es una empresa de consultoría establecida en la Ciudad de México, enfocada a proveer asesoría estratégica a gobiernos y empresas en materia de políticas públicas, comercio internacional y asuntos regulatorios y de inversión.

AVISO LEGAL: El presente reporte fue elaborado a partir de información pública. Las conclusiones e interpretaciones que presenta están diseñadas para informar y orientar a sus usuarios en la toma de decisiones, no para garantizar resultados específicos.