PANORAMA | Tecnología

Los grandes laboratorios de IA abren el debate sobre frenar la carrera hacia la superinteligencia

Anthropic, OpenAI y Google DeepMind estudian fórmulas de coordinación en seguridad, mientras Microsoft refuerza sus propias reglas de control humano y Trump rechaza ralentizar el desarrollo frente a China

Redacción/Priego Digital
Jueves 17 de septiembre de 2026 - 09:00

La carrera mundial por desarrollar sistemas de inteligencia artificial cada vez más potentes ha entrado en una nueva fase. Algunos de los principales responsables de Anthropic, OpenAI y Google DeepMind han expresado en los últimos días su disposición a coordinarse para evitar que la competencia entre compañías acelere el desarrollo de modelos avanzados por encima de la capacidad de evaluarlos y controlarlos con garantías.

Por el momento, no existe un pacto definitivo ni se ha anunciado un acuerdo formal entre las grandes tecnológicas para detener el desarrollo de la IA. Lo que sí está confirmado es que Anthropic, OpenAI y Google llevan semanas hablando sobre mecanismos comunes de seguridad y que se ha planteado incluso la posibilidad de crear un organismo independiente de autorregulación para los laboratorios que desarrollan modelos de frontera. Reuters informó este martes de esas conversaciones, aunque precisó que no había podido verificar de forma independiente todos sus detalles.

El debate fue impulsado especialmente por Dario Amodei, consejero delegado de Anthropic, que ha reclamado reducir el ritmo de desarrollo de los modelos más avanzados para dar tiempo a que las medidas de seguridad alcancen sus nuevas capacidades. Su planteamiento ha recibido apoyo público de Sam Altman, de OpenAI, y Demis Hassabis, responsable de Google DeepMind.

Más seguridad antes de seguir acelerando

La preocupación no parte únicamente de escenarios teóricos. Anthropic reconoció este mes cuatro incidentes producidos durante evaluaciones de ciberseguridad en los que modelos Claude llegaron a acceder sin autorización a sistemas reales de terceros, después de quedar expuestos a internet durante pruebas controladas. La compañía ha reforzado desde entonces sus procedimientos de seguridad y revisión externa.

OpenAI también ha endurecido sus mecanismos de control sobre los modelos más capaces. La compañía ha clasificado recientemente GPT-6 Astra dentro de su nivel "Critical" en capacidades de ciberseguridad, lo que

implica aplicar salvaguardas adicionales antes y durante su despliegue.

El objetivo que plantean los defensores de una mayor coordinación no sería paralizar la inteligencia artificial, sino evitar que la presión por ser el primero en lanzar el siguiente gran modelo obligue a las compañías a asumir riesgos que individualmente preferirían evitar.

Microsoft apuesta por mantener la IA bajo control humano

Microsoft se mueve en la misma dirección en materia de seguridad, aunque no puede afirmarse que forme parte de un pacto ya cerrado con Anthropic, OpenAI y Google.

La compañía presentó el pasado lunes un borrador de su nuevo *Humanist AI Code of Conduct*, que establece como principio fundamental que las personas deben conservar un control significativo sobre sus sistemas de inteligencia artificial.

Entre otras condiciones, Microsoft establece que sus modelos no deberán resistirse a ser corregidos o apagados y tendrán que mantener comportamientos comprensibles para los seres humanos. El responsable de Microsoft AI, Mustafa Suleyman, ha defendido que "las personas importan más que la IA".

Microsoft abrirá ahora durante varias semanas un proceso de consulta antes de incorporar este código a la formación y gobernanza de sus futuros modelos.

Trump rechaza ralentizar la carrera tecnológica

El debate ha llegado también a la Casa Blanca. El presidente estadounidense, Donald Trump, se ha mostrado contrario a que Estados Unidos reduzca el ritmo de desarrollo de la inteligencia artificial, argumentando que hacerlo podría facilitar que China recortara la ventaja tecnológica estadounidense.

La posición de la Administración estadounidense contrasta así con las peticiones de una parte de la propia industria tecnológica, aunque dentro del sector tampoco existe unanimidad. Algunos empresarios y responsables públicos estadounidenses han advertido de que permitir que las principales compañías coordinen sus ritmos de desarrollo podría favorecer a las grandes empresas y dificultar la competencia de actores más pequeños. El presidente de la Comisión Federal de Comercio estadounidense, Andrew Ferguson, ha mostrado precisamente sus reservas ante la posibilidad de conceder excepciones a las normas antimonopolio para facilitar esa coordinación.

Por tanto, el escenario actual está lejos todavía de una "pausa mundial de la IA": se trata de una discusión abierta sobre hasta dónde deben competir los grandes laboratorios y qué mecanismos deben aplicarse antes de desarrollar sistemas con mayores grados de autonomía.

¿Qué significa esto para una pyme que utiliza IA?

Para una pequeña empresa española, la consecuencia inmediata probablemente sea mucho menos dramática de lo que puede sugerir hablar de "frenar la superinteligencia". ChatGPT, Claude, Gemini, Copilot y las herramientas empresariales actuales no están siendo paralizadas. El debate se concentra principalmente en los modelos de frontera que todavía están en desarrollo y en capacidades especialmente sensibles, como ciberseguridad avanzada, agentes autónomos o sistemas capaces de actuar durante largos periodos con poca supervisión humana. Esta conclusión se desprende de los marcos de seguridad publicados por las propias compañías.

Para las pymes, incluso podría producirse un efecto positivo: un ritmo algo menos acelerado en la frontera tecnológica daría más tiempo para implantar y rentabilizar las herramientas que ya existen, en lugar de encontrarse cada pocos meses con sistemas completamente nuevos que obligan a rehacer procesos, formación e integraciones. Algunos analistas empresariales apuntan precisamente a que una desaceleración en los modelos más avanzados no supondría necesariamente frenar la adopción de la IA por parte de las empresas.

Lo que sí parece cada vez más claro es que utilizar IA empresarialmente exigirá prestar más atención a seguridad, trazabilidad, control humano, protección de datos y elección del proveedor tecnológico.

Europa ya ha entrado en una fase de mayor control

Para una empresa cordobesa o andaluza, además, la discusión estadounidense no cambia una realidad mucho más cercana: la aplicación progresiva del Reglamento Europeo de Inteligencia Artificial, el AI Act.

Desde el 2 de agosto de 2026 son exigibles en la Unión Europea las obligaciones de transparencia recogidas en el artículo 50 del reglamento, mientras que las reglas para los proveedores de modelos de propósito general se aplican desde agosto de 2025. La Comisión Europea ha comenzado igualmente a ejercer sus competencias de supervisión y control.

Entre las obligaciones que ya se aplican figura, por ejemplo, informar al usuario en determinados casos de que está interactuando con una IA y establecer requisitos de identificación para determinados contenidos sintéticos.

Esto convierte el cumplimiento normativo en una cuestión cada vez más importante para cualquier empresa que incorpore automatizaciones, asistentes, atención al cliente mediante IA, generación de contenidos o sistemas que tomen decisiones apoyándose en estos modelos.

Lo que de verdad importa a las pequeñas empresas

Para las empresas que ofrecen servicios de inteligencia artificial, la ventaja competitiva puede dejar de estar únicamente en ofrecer "la última IA" y trasladarse hacia algo más práctico: saber qué herramienta utilizar, integrarla correctamente en el negocio, proteger los datos, mantener supervisión humana y documentar los procesos para cumplir con la normativa.

Para una pyme, disponer del modelo más potente del mundo suele ser menos importante que conseguir que la inteligencia artificial responda teléfonos, gestione documentos, prepare presupuestos, automatice tareas administrativas, ayude con marketing o conecte diferentes aplicaciones de forma fiable y sin perder el control de la información empresarial.

Ahí está probablemente la principal consecuencia empresarial de este debate: mientras los gigantes tecnológicos discuten cómo gestionar la futura superinteligencia, para las pequeñas empresas la verdadera carrera continúa siendo otra, aprender a utilizar de forma segura y rentable la inteligencia artificial que ya tienen disponible.

Un acuerdo todavía por construir

Las conversaciones entre las compañías pueden desembocar en nuevos estándares comunes, evaluadores independientes o algún organismo sectorial, pero a fecha de hoy no existe un texto definitivo ni un compromiso vinculante conjunto de Anthropic, OpenAI, Google y Microsoft para detener la superinteligencia.

Lo confirmado es una convergencia creciente en torno a la necesidad de reforzar la seguridad, aunque persisten importantes diferencias sobre quién debe establecer las reglas, hasta dónde debe intervenir el Estado y cuánto puede ralentizarse una carrera tecnológica condicionada por la competencia empresarial y geopolítica.