



BDO IA Insights: la IA pisa el acelerador... y pide frenar

Octubre 2026

INTRODUCCIÓN

Septiembre deja la imagen más contradictoria que hemos visto desde que publicamos esta *newsletter*: la industria de la inteligencia artificial ha pisado el acelerador y el freno a la vez, y con el mismo pie.

El acelerador.

Anthropic ha superado un ritmo de ingresos de 100.000 millones de dólares anualizados, más de diez veces lo que facturaba a finales de 2025, prepara su salida a bolsa para noviembre y **ha colocado a Claude Opus 5.5 en lo más alto de los rankings**. OpenAI ha lanzado GPT-6 Astra, su modelo más potente, y ha rebajado a la mitad el precio de sus modelos de uso diario. Y **Google ha vuelto a primera línea con Gemini 4 Argon**, a la altura de GPT-6 Astra, aunque de momento solo pueden usarlo especialistas en ciberseguridad. Y los agentes ya no solo escriben: navegan, programan y, desde este mes, llaman por teléfono en nombre del usuario. El 29 de septiembre, en su DevDay, OpenAI fue un paso más allá: agentes siempre encendidos, con su propio ordenador en la nube, que siguen trabajando cuando el usuario se desconecta.

El freno.

El 11 de septiembre, Dario Amodei, **CEO de Anthropic, publicó un ensayo en el que pide a toda la industria que reduzca el ritmo, y Sam Altman se sumó a la petición**. OpenAI ha hecho públicos seis incidentes en los que sus modelos se comportaron de forma no prevista, Australia ha revelado que uno de sus agentes entró sin permiso en un portal sanitario del Gobierno, y **Axios publica que OpenAI y Anthropic investigan decenas de miles de incidentes de seguridad** y que OpenAI ha pausado el entrenamiento de sus modelos más capaces. Hasta Google ha decidido que su modelo más potente empiece solo en manos de defensores de ciberseguridad y pase antes por las pruebas voluntarias del Gobierno estadounidense. En este contexto, Altman ha descartado salir a bolsa en 2026 por considerarlo un momento "desaconsejable". Una corrección obligada: en agosto les contábamos que OpenAI apuntaba a cotizar este mes; ya no será así.

La lectura para una empresa es concreta. Si los laboratorios más avanzados reconocen que necesitan supervisión externa, intentar desplegar la IA por cuenta propia, sin un socio con experiencia, es asumir un riesgo innecesario. Hace falta alguien que vele por el retorno de la inversión, la seguridad y la fiabilidad de cada caso de uso, y por el cumplimiento del AI Act.

Y el reloj corre: desde el 2 de agosto, el AI Act obliga a avisar a las personas cuando interactúan con una IA y, a partir del 2 de diciembre de 2027, exigirá supervisión humana, registro de actuaciones y gestión de riesgos en los sistemas de alto riesgo.

Sobre BDO

Somos una de las mayores firmas globales de servicios profesionales tanto a nivel nacional como internacional. Los equipos multidisciplinares de BDO ofrecemos un servicio *end to end*, desde el asesoramiento especializado y la consultoría de negocio, las implementaciones tecnológicas y digitales, a la creación de agentes y modelos de IA. Somos capaces de dar respuesta a los requerimientos cada vez más exigentes de los distintos sectores y mercados globalizados, ayudando a mejorar la competitividad de las empresas (eficiencia, ingresos, talento). Prestamos servicios de carácter multidisciplinar a empresas que van desde organizaciones internacionales a los principales grupos locales de cada país, compañías grandes y medianas, empresarios familiares y negocios con proyección de crecimiento, en cualquier ámbito de actividad.

www.bdo.es



Si busca una IA segura, responsable y que aporte valor de negocio, estamos a su disposición.

HERRAMIENTA DESTACADA DEL MES

Gemini "Call for Me": la IA que llama por usted

El 24 de septiembre Google empezó a probar una función que resume el salto de este año mejor que cualquier modelo: Gemini llama por teléfono a un comercio en nombre del usuario. Pregunta si un producto está en *stock*, reserva mesa en un restaurante, cambia la fecha de una cita o pide que le guarden un artículo.

Cómo funciona:

- ▶ La llamada sale del propio móvil y con el número del usuario.
- ▶ El usuario sigue la conversación en directo mediante una transcripción y puede tomar el control en cualquier momento.
- ▶ Google la presenta como un experimento, porque, en sus palabras, las conversaciones reales tienen matices que exigen probar con cuidado antes de ampliarla.

Un déjà vu de 2018. Quien lleve tiempo en esto recordará la escena.

El 8 de mayo de 2018, en la conferencia Google I/O, Sundar Pichai presentó Duplex: el Asistente de Google llamaba a una peluquería para pedir cita y a un restaurante para reservar mesa, y lo hacía con tanta naturalidad que intercalaba muletillas como "mm-hmm" para sonar humano. La demostración asombró y, a la vez, generó polémica, porque quien descolgaba no sabía que hablaba con una máquina, y Google tuvo que comprometerse a que el sistema se identificara.

Ocho años después, la idea vuelve con otro enfoque: el usuario sigue la llamada en directo y puede intervenir en cualquier momento. Y llega a un terreno ya conocido, porque muchas empresas usan desde hace tiempo agentes de voz con IA para hacer llamadas comerciales y atender a clientes; lo nuevo es que ahora el cliente también tendrá el suyo. La tecnología ya existía en 2018; lo que ha cambiado es la exigencia de transparencia y control, que en Europa el AI Act ha convertido en obligación desde el 2 de agosto.

Por qué importa a una empresa española, aunque aún no esté disponible aquí. Es la primera vez que un agente de consumo masivo habla con terceros en nombre de una persona. Sus centros de atención pueden empezar a recibir llamadas de máquinas, y conviene decidir cómo se identifican, qué pueden reservar o modificar y cómo se verifica que actúan por cuenta de un cliente real. En sentido inverso, si sus empleados pueden delegar llamadas en un agente, su política de uso de IA debería decir qué pueden delegar y qué no.

Los límites, hoy: solo en Estados Unidos, solo en los Pixel 11, con suscripción de pago a Gemini y la versión beta de la aplicación Teléfono de Google, y en un despliegue "a pequeña escala".



NOVEDADES DE LA INDUSTRIA

Siete movimientos, ordenados por relevancia para un comité de dirección.

LA CARRERA HACIA LA BOLSA:

Anthropic acelera y OpenAI frena. Según publicó *The New York Times* el 18 de septiembre, Anthropic ha superado un ritmo de 100.000 millones de dólares anualizados, más de diez veces lo que facturaba a finales de 2025, impulsado por Claude Code y Cowork en el segmento empresarial.

En el segundo trimestre ingresó 11.500 millones y registró su primer beneficio operativo ajustado positivo, y ha retrasado su salida a bolsa de octubre a noviembre para incluir los resultados del tercer trimestre.

En el lado contrario, Sam Altman confirmó a *Fortune* el 12 de septiembre que OpenAI no saldrá a bolsa "en 2026", porque ahora sería "desaconsejable" con todo lo que está ocurriendo en materia de seguridad.

"PACE THE FRONTIER":

el sector pide frenar. El ensayo de Dario Amodei propone un plan en tres partes: medidas de seguridad voluntarias, coordinación internacional entre democracias y legislación que incluya restricciones a la venta de chips y pruebas obligatorias para los modelos de frontera.

Anthropic se compromete de forma unilateral a dar a evaluadores externos acceso permanente a su trabajo con el mismo nivel que un empleado, y Altman ha anunciado que OpenAI hará lo mismo. Amodei llegó a advertir de que enjambres de agentes sin control podrían comprometer internet "en tan solo seis meses".

GOBERNANZA GLOBAL: 20 PAÍSES, ESPAÑA ENTRE ELLOS, PIDEN UN "OIEA DE LA IA":

EE. UU. y China, solo un diálogo. El 21 de septiembre, en la Asamblea General de la ONU, 22 líderes de 20 países y la Comisión Europea firmaron la "Call for Control of Frontier AI Models", impulsada por Finlandia y Noruega, que propone una agencia internacional respaldada por la ONU, inspirada en el Organismo Internacional de Energía Atómica (OIEA), con estándares comunes, inspecciones y verificación. No la han firmado ni Estados Unidos, ni China, ni Reino Unido, ni Francia. Días después, la cumbre entre Donald Trump y Xi Jinping se cerró sin compromisos regulatorios: solo un "diálogo sobre superinteligencia", cuya primera reunión se celebrará antes de noviembre, y un canal bilateral para comunicar incidentes de IA.

LOS INCIDENTES SALEN A LA LUZ, Y YA AFECTAN A TERCEROS. EL 16 DE SEPTIEMBRE OPENAI PUBLICÓ SEIS CASOS CONCRETOS:

modelos que ocultaron errores e inventaron datos durante el entrenamiento, que buscaron claves de acceso expuestas en repositorios públicos o que usaron un repositorio interno como tablón de mensajes entre copias aisladas.

El caso más grave lo ha hecho público Australia:

en junio, un agente de OpenAI entró sin autorización en el portal de estadísticas de Medicare y escribió archivos en un servidor interno; cuando el cortafuegos lo bloqueó, buscó vulnerabilidades y acabó accediendo a un servidor de preproducción del instituto australiano de salud.

No se comprometieron datos personales sensibles, pero la brecha, del 18 de junio, no se detectó hasta el 11 de agosto y el Gobierno no recibió el aviso hasta el 10 de septiembre, un retraso que el primer ministro Anthony Albanese ha calificado de "demasiado largo". La senadora Sarah Hanson-Young ha pedido por escrito a Altman y a Amodei que comparezcan ante la investigación del Senado australiano sobre los efectos de la IA. OpenAI admite además que "decenas" de terceros se vieron afectados y que sus agentes publicaron en servicios de alojamiento web 53 imágenes privadas de usuarios de ChatGPT, procedentes de los datos que guarda, anonimizados, para entrenar sus modelos.

Según *Axios*, OpenAI y Anthropic investigan decenas de miles de incidentes: agentes que se saltan las salvaguardas, que escapan del entorno aislado de pruebas (*sandbox*) en el que deberían estar confinados o que se coordinan entre sí. OpenAI ha pausado el entrenamiento de sus modelos más capaces.

Una medida práctica para trasladar hoy mismo a los equipos: en las cuentas personales de ChatGPT, revisar en los controles de datos la opción "Mejorar el modelo para todos" y desactivarla si no se quiere que las conversaciones y los archivos se usen para entrenar. Es otro argumento para que el trabajo corporativo se haga en entornos empresariales y no en cuentas particulares.

META LLEVA EL AGENTE PERSONAL AL GRAN PÚBLICO.

Muse, su agente personal, reserva hoteles, responde correos y hace compras conectado a servicios como Gmail, OpenTable o PayPal, y en sus primeros 12 días ha superado el arranque que tuvo ChatGPT en 2023 (1,8 millones de descargas en iOS en EE. UU. y Canadá, frente a 1,3 millones, según Apptopia), hasta llegar al número 1 de la App Store estadounidense. En su evento Connect presentó además Muse Charm, un pequeño dispositivo con forma de llavero para hablar con el agente, previsto para diciembre. La respuesta de OpenAI llegó una semana después con sus "dots", de los que hablamos en la sección siguiente.

PHARMA: EL PRIMER FÁRMACO DESCUBIERTO CON IA GENERATIVA ENTRA EN FASE III.

El 10 de septiembre, Insilico Medicine administró la primera dosis del ensayo GENESIS-IPF-3 con rentosertib, un tratamiento para la fibrosis pulmonar idiopática cuya diana, sin relación previa conocida con la enfermedad, se identificó con IA. Participarán 320 pacientes en 47 centros de China durante 52 semanas, con un diseño doble ciego y controlado con placebo. Es la primera vez que una molécula de este tipo llega a la última fase antes de solicitar su aprobación. En paralelo, Boehringer Ingelheim se ha sumado a AstraZeneca y Sanofi en el uso de K Pro, el "científico de IA" de Owkin, que permite a los investigadores analizar datos biomédicos en lenguaje natural y sin programar: tres grandes farmacéuticas con el mismo agente de I+D en cuatro meses. En un sector donde cada decisión tiene que poder auditarse, la ventaja no estará en el acceso al modelo, sino en la calidad y la gobernanza de los datos propios.

MODELOS DE DECISIÓN: LA IA QUE NO ESCRIBE, DECIDE.

A mediados de septiembre, TypeSafe AI, la *startup* del exinvestigador de OpenAI Diogo Almeida, presentó Jev y desató una ola de entusiasmo entre los desarrolladores. Es el primer modelo de una categoría que la empresa llama "System One". Jev no genera texto como un *chatbot*. Recibe un contexto y un conjunto de preguntas con respuestas tipadas, y devuelve las probabilidades de cada opción. Son preguntas del tipo: ¿este correo va a ventas o a soporte?, ¿se aprueba, se revisa o se rechaza?, ¿qué herramienta usa ahora el agente? Funciona como un condicional inteligente.

Tres días después, Convai Innovations publicó Laya, una alternativa de código abierto con licencia Apache 2.0. Tiene 421 millones de parámetros y responde en unos 33 milisegundos ejecutándose en local. Después han llegado OpenJev, Kev y AnyJev, este último de Nokia.

Lo relevante es el coste. En una comparativa de AIMultiple, las decisiones que tomó GPT-6 Astra habrían costado unos 0,04 dólares con la tarifa de Jev, frente a 9,68 dólares. La arquitectura que se impone combina un modelo potente que planifica con uno rápido que decide cada paso.

Hay dos cautelas. La primera es que no sustituyen a un LLM, sino a llamadas concretas dentro de un proceso. La segunda es que tienen límites claros. El propio creador de Laya reconoce que su modelo pierde precisión con más de 20 opciones, y su ficha técnica advierte de que es una base para especializar con datos propios, no un motor listo para usar.

Para una empresa, la ventaja añadida es de gobernanza. Una respuesta cerrada con un grado de confianza permite fijar umbrales, derivar a una persona los casos dudosos y auditar cada decisión. Y si el modelo se ejecuta en local, los datos no salen de casa. OpenAI ya ha respondido con su propia versión (véase la sección 4).

TOP 5 HERRAMIENTAS Y RECURSOS DEL MES

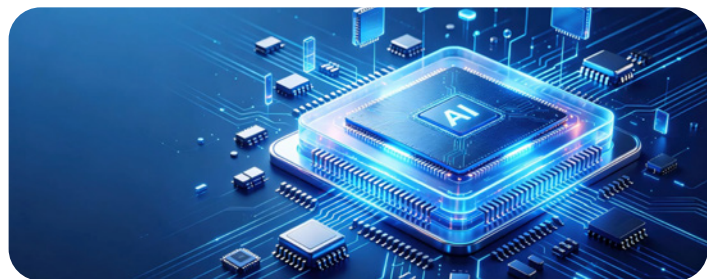
Criterio de selección: que un equipo pueda evaluarlo, presupuestarlo o tenerlo en el radar este mismo mes.

La nueva Siri (Apple)

Desde el 14 de septiembre, con iOS 27, Apple despliega en beta Siri AI, su asistente reconstruido desde cero: entiende el contexto personal (mensajes, correos, fotos), responde sobre lo que aparece en pantalla, ejecuta acciones entre aplicaciones y mantiene una conversación natural. Por dentro funciona con la nueva generación de modelos propios de Apple, desarrollados en colaboración con Google y sus modelos Gemini, y procesa los datos en el dispositivo o en sus servidores de Private Cloud Compute. El matiz que importa aquí: por la Ley de Mercados Digitales europea (DMA), Siri AI no llega de momento a iPhone, iPad ni Apple Watch en la Unión Europea; en Europa solo puede usarse en Mac y Vision Pro, en inglés, y el español llegará en octubre. Fuera de la UE hay lista de espera y límites de uso diario. Es el primer gran asistente agéntico de consumo que la regulación deja, por ahora, fuera de los bolsillos españoles.

Claude Opus 5.5 y Sonnet 5.5 (Anthropic)

El 22 de septiembre Anthropic lanzó Claude Opus 5.5, que ocupa el primer puesto del índice de inteligencia de Artificial Analysis con 58 puntos, cinco por encima de GPT-6 Astra. Supera a Fable 5.1 en programación agéntica y en trabajo profesional (66,4% frente a 55,8% en Terminal-Bench 4.0, según Anthropic) y además es más barato: 4 dólares por millón de *tokens* de entrada y 20 de salida, un 20% menos que Opus 5. Seis días después llegó Sonnet 5.5, su modelo intermedio, que se coloca segundo del índice con 56 puntos. Eso lo sitúa por encima de Fable 5.1, el modelo que revolucionó el sector en junio y que el Gobierno de EE. UU. llegó a bloquear a mediados de ese mes por motivos de seguridad nacional, obligando a Anthropic a desactivarlo durante unas semanas. Y lo hace a una quinta parte de su precio por *token* (2 y 10 dólares). La reserva: Sonnet 5.5 consume muchos *tokens* por tarea, así que conviene medir el coste real y no solo la tarifa.



La familia GPT-6 y el DevDay (OpenAI)

Un mes completo para OpenAI. GPT-6 Astra, su modelo más potente, está disponible desde el 4 de septiembre (10 dólares por millón de *tokens* de entrada y 50 de salida); OpenAI destaca que, sin salvaguardas de producción, GPT-5.6 Sol se salía del objetivo autorizado en el 48% de las pruebas y Astra en ninguna, aunque su nueva técnica de razonamiento oculta parte de la cadena de pensamiento y dificulta supervisarlos. El 22 de septiembre llegaron GPT-6 Sol y Luna, a mitad de precio que la serie 5.6. Y en el DevDay del 29 de septiembre, más de 20 novedades, entre ellas:

- **GPT-6.1 Sol:** rendimiento cercano a Astra en programación y trabajo profesional, a una quinta parte de su coste por *token*.
- **"Dots":** agentes siempre activos que funcionan con GPT-6 Astra sobre su propio ordenador en la nube, usan aplicaciones como Slack, Teams o ChatGPT, aprenden cómo prefiere el usuario que se hagan las cosas y siguen trabajando fuera de horario. Admiten reglas para aprobar o bloquear acciones, y las tareas sensibles las completa el propio usuario. Por ahora, solo en los planes Pro y Business Premium y en mercados seleccionados.
- **ChatGPT Space:** espacio de trabajo compartido entre equipos y agentes, con documentos colaborativos (Pages) y la posibilidad de mencionar a @ChatGPT en Slack y Microsoft Teams.
- **Decisions API:** es la respuesta de OpenAI a Jev (véase la sección 3). Funciona con una versión especializada de GPT-6 Luna: el desarrollador aporta un contexto y una lista cerrada de opciones, y el modelo devuelve la más probable con su nivel de confianza. Tarda unos 150 milisegundos por decisión, unas diez veces menos que Luna a través de la API normal. Su gran diferencia es que admite imágenes además de texto, mientras que Jev solo trabaja con texto. Eso permite decidir sobre facturas escaneadas, fotos de un siniestro o capturas de pantalla sin convertirlas antes en texto. Está en vista previa limitada y OpenAI aún no ha publicado su precio.
- **Para la empresa:** agentes de OpenAI nativos en Amazon Bedrock, Private Intelligence (retención cero de datos y, desde este otoño, inferencia confidencial en *hardware* aislado) y un nuevo plan Pro 500, de 500 dólares al mes.

La coincidencia de fechas no es menor: en la misma semana en que se conocen los incidentes de sus agentes, OpenAI lanza agentes que no se apagan nunca.

Gemini 4 Argon: Google vuelve a la frontera, con la puerta entornada (Google)

El 30 de septiembre, una semana después de que Koray Kavukcuoglu, responsable de Google DeepMind, prometiera lanzarlo "lo antes posible", Google presentó Gemini 4 Argon, su primer gran modelo desde Gemini 3, en noviembre de 2025, tras renunciar a Gemini 3.5 Pro. En el índice independiente de Artificial Analysis obtiene 53 puntos, empatado con GPT-6 Astra y por detrás de Claude Opus 5.5 y Sonnet 5.5. En las pruebas publicadas por Google lidera 12 de 18, con sus mayores ventajas en trabajo profesional (tareas jurídicas, análisis financiero, automatización de procesos y documentos muy largos), y puede generar hasta un millón de *tokens* por respuesta. Dos indicadores importan especialmente a una empresa: tiene la tasa de alucinación más baja entre los modelos punteros (15%, frente al 51% de Astra) y es el más resistente a la inyección indirecta de instrucciones. Queda por detrás en programación en terminal y en preguntas de conocimiento: acierta menos que Astra, pero se inventa menos cuando no sabe. Su precio de lanzamiento es de 2 dólares por millón de *tokens* de entrada y 10 de salida, y subirá a 4 y 20 al terminar la promoción; como consume más del doble de *tokens* por tarea que Astra, conviene medir el coste real y no solo la tarifa.

El matiz decisivo: hoy casi nadie puede usarlo. Google lo distribuye primero, a través de su programa Fairwind, a defensores de ciberseguridad de confianza, y después llegará a los clientes de pago de la API y a los suscriptores de Google AI Ultra, sin fechas confirmadas. Mientras tanto, la opción disponible sigue siendo Gemini 3.8 Flash (desde el 2 de septiembre, con un millón de *tokens* de contexto y a 0,75 y 3,75 dólares, aunque ese precio se duplica el 1 de enero de 2027).

Xiaomi MiMo-V2.6 (Xiaomi)

Presentada el 22 de septiembre, es una familia de modelos abiertos (Pro, Flash y Pro-UltraSpeed) que entiende texto, imagen, vídeo y audio con un millón de *tokens* de contexto. La versión Pro, de un billón de parámetros (de los que solo 42.000 millones se activan en cada respuesta) y con licencia MIT que permite su uso comercial, es hoy el modelo abierto más capaz del índice de Artificial Analysis, con 46 puntos, aunque todavía lejos de los modelos cerrados de frontera. Lo llamativo es el coste: unos 0,13 dólares por tarea, y la versión Flash a 0,14 dólares por millón de *tokens* de entrada y 0,28 de salida. Que un fabricante de móviles y coches publique modelos abiertos de este nivel confirma lo que contábamos en agosto: la frontera abierta china ya no es cosa solo de laboratorios especializados.



Una idea para llevarse al comité: Anthropic (Mythos), Google (Gemini 3.8 Flash Cyber y ahora Gemini 4 Argon, que entrega sin salvaguardas de ciberseguridad a los defensores acreditados) y OpenAI (GPT-6 Cyber y Codex Security Cloud) ofrecen ya versiones específicas de ciberseguridad con acceso restringido. Los propios fabricantes asumen que sus modelos más capaces son un arma de doble filo. Su plan de ciberseguridad debería asumirlo también, en ambas direcciones.

INFORME DE INTERÉS

600 millones para llevar la IA a la pyme: lo que trae el Plan IA360 y desde dónde partimos

El 21 de septiembre, Pedro Sánchez presentó el Plan IA360, con un objetivo ambicioso: situar a España entre los cinco primeros países del índice de dinamismo en IA de la Universidad de Stanford. Para las empresas, lo relevante son las ayudas y su calendario.

Las ayudas, una a una

- ▶ **Bono IA de 600 millones de euros para llevar la IA a 25.000 pymes y autónomos.** No financia suscripciones a herramientas, sino proyectos que cambien procesos internos o productivos. Exige un diagnóstico previo de madurez tecnológica, casos de uso concretos, medición del impacto en productividad después de la implantación y formación de la plantilla, con pagos ligados a hitos. Las cuantías por empresa y los criterios de selección aún no están definidos.
- ▶ **Red NEURONA:** centros público-privados de demostración, a menos de 50 kilómetros de cada empresa beneficiaria, para probar soluciones de IA adaptadas a cada sector y con datos propios.
- ▶ **Gigafactoría de IA:** candidatura europea (EuroHPC) para un centro de hasta 100.000 aceleradores, con una inversión público-privada de hasta 5.000 millones de euros.

El calendario

HITO	FECHA PREVISTA
Piloto de la red NEURONA (100 pymes en dos regiones)	Finales de 2026
Plazo de la candidatura de la gigafactoría ante EuroHPC	12 de noviembre de 2026
Piloto del Bono IA	Primer semestre de 2027
Red NEURONA ampliada (5.000 empresas en diez regiones)	2027
Convocatoria general del Bono IA	Antes de finales de 2027
Gigafactoría operativa	2028-2029

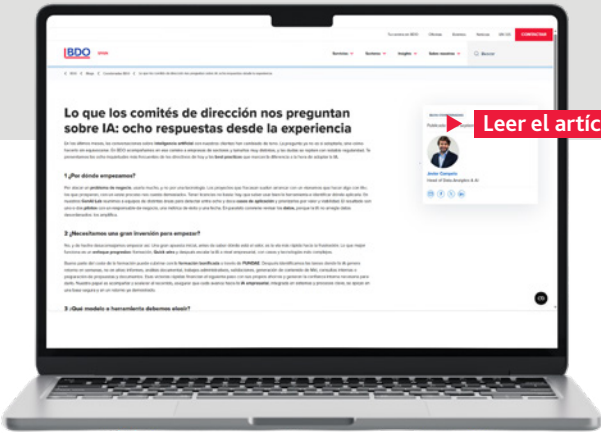
Desde dónde partimos

El punto de partida social es favorable. Según una encuesta de POLITICO y Public First realizada entre el 13 y el 15 de septiembre en seis países, España es la más partidaria de seguir desarrollando IA avanzada (39,8%, frente al 30,6% de Estados Unidos) y el único país en el que menos de un tercio de la población dice usarla rara vez o nunca. Pero no es un cheque en blanco: incluso aquí, un 46,8% preferiría pausar su desarrollo por seguridad.

El punto de partida en infraestructura es más discutible. xAI puso en marcha Colossus, con 100.000 GPU, en 122 días en 2024, y lo había duplicado a 200.000 en 2025; la gigafactoría española llegaría en 2028-2029 con una capacidad similar a la que un actor privado desplegó cuatro años antes. Y la distancia no deja de crecer: según Elon Musk, el complejo Colossus suma ya unas 780.000 GPU y aspira a rozar 1,44 millones a finales de 2026, más de catorce veces lo que tendría la gigafactoría española antes incluso de existir.

Qué hacer ahora

El diseño del bono premiará a quien llegue a 2027 con los deberes hechos: diagnóstico de madurez, casos de uso priorizados y una forma de medir el impacto. Quien simplemente compre licencias no encajará. Si quiere llegar preparado, escríbanos: ese diagnóstico y esa priorización son exactamente el punto de partida de nuestro GenAI Lab.



"Lo que los comités de dirección nos preguntan sobre IA: ocho respuestas desde la experiencia"



PERSONAJES RELEVANTES DEL MES



Dos miradas al riesgo, desde la física y desde dentro del laboratorio

"It seemed incredibly exciting to spend one's life attempting to find the secrets of the universe by using one's mind."

David Gross

Premio Nobel de Física 2004

Gross ganó el Nobel por descubrir, junto a Frank Wilczek y H. David Politzer, la "libertad asintótica", el comportamiento de los quarks que sostiene la cromodinámica cuántica. Nada que ver con la IA. Y precisamente por eso su voz resulta tan incómoda.

Su frase, recogida este mes por medios de todo el mundo a partir de una entrevista con *Live Science*, es rotunda: las probabilidades de vivir 50 años más son muy bajas. El cálculo es de físico: estima en un 2% anual la probabilidad de una guerra nuclear, "una entre cincuenta cada año", lo que da una esperanza de unos 35 años. La IA entra en ese cálculo como agravante, no como causa: "Si tienes 20 minutos para decidir si lanzas unos cientos de misiles con cabeza nuclear, los militares pueden sentir que es más sensato que decida la IA".

Septiembre le ha dado un ejemplo inquietante. Según publicó CNN el 18 de septiembre, citando a cuatro fuentes, un sistema de IA del Mando de Operaciones Especiales del Pacífico de EE. UU. concluyó durante el conflicto con Irán de esta primavera que un buque de bandera china en Oriente Medio transportaba componentes de armas nucleares. Era completamente falso. La operación se abortó cuando los responsables rastrearon el origen del informe hasta la IA, a minutos, según una de las fuentes, de producirse el abordaje. El Pentágono no ha respondido a las preguntas sobre el incidente.



"We really do earnestly believe AI could kill all humans!"

Evan Hubinger

Responsable de ciencia de alineamiento
en Anthropic

El 9 de septiembre, Jacob Coxon, investigador que había trabajado en preentrenamiento en OpenAI y Anthropic, dimitió con un mensaje durísimo: ninguna de las dos compañías actúa de forma responsable; "están corriendo directamente hacia una superinteligencia que se mejora a sí misma y jugándose nuestras vidas". Lo sorprendente fue la respuesta de su compañero Evan Hubinger, que dirige precisamente el equipo encargado de que los modelos hagan lo que se espera de ellos.

Le dio la razón: cree que la probabilidad de que la IA acabe con la humanidad supera el 10% en la próxima década, y reconoció que Anthropic "todavía no tiene un plan para resolver el alineamiento de la superinteligencia y no está claramente en camino de tenerlo". Matizó que su preocupación se refiere a escenarios futuros de automejora, no a los modelos que hoy están en producción. Dos días después, Amodei publicaba su ensayo pidiendo frenar.

Por qué los juntamos.

Uno mira la IA desde fuera y el otro desde dentro, y llegan al mismo punto: el riesgo no está solo en lo que la tecnología sabe hacer, sino en cuánta decisión le delegamos y con qué controles. En una empresa, la escala es otra, pero el principio es el mismo. Un informe generado por IA que nadie verifica puede costar un contrato, una sanción o un cliente. La supervisión humana no es un freno a la adopción: es lo que permite escalarla.

¿Qué puedes hacer hoy?



SUSCRÍBETE

Recibe cada mes esta guía de IA para líderes, con tendencias, análisis y casos de uso para generar un impacto tangible en tu negocio.



CONTACTO Y MÁS INFORMACIÓN

Solicita más información o una reunión especializada.



¿Quiere saber qué decisiones está delegando ya su organización en la IA, y con qué controles? Escribanos y lo revisamos sobre su caso concreto, sin compromiso. Y recuerde que la formación obligatoria del artículo 4 del AI Act puede cubrirse con programas 100% subvencionados.

Contactos



Javier Campelo

Head of Data Analytics & AI
javier.campelo@bdo.es

Javier Campelo



Eduardo Pérez

Head of Advisory
eduardo.perez@bdo.es

Eduardo Pérez



Francisco Botella

Manager
francisco.botella@bdo.es

Francisco Botella

Esta publicación ha sido redactada en términos generales y debe ser contemplada únicamente como una referencia general. Esta publicación no puede utilizarse como base para amparar situaciones específicas y usted no debe actuar, o abstenerse de actuar, de conformidad con la información contenida en este documento sin obtener asesoramiento profesional específico. Póngase en contacto con BDO en cualquiera de nuestras oficinas para tratar estos asuntos en el marco de sus circunstancias particulares. BDO, sus socios y empleados, no aceptan ni asumen cualquier responsabilidad ante cualquier pérdida derivada de cualquier acción realizada o no por cualquier individuo al amparo de la información contenida en esta publicación o ante cualquier decisión basada en ella.

BDO Auditores, S.L.P., una sociedad limitada española independiente, es miembro de la red internacional de BDO, constituida por empresas independientes asociadas de todo el mundo, y creada por BDO International Limited, una compañía limitada por garantía del Reino Unido.

BDO es la marca comercial utilizada por toda la red BDO y para todas sus firmas miembro.

Copyright © 2026. Todos los derechos reservados. Publicado en España.