



# Logra una mejor rentabilidad y rendimiento mediante la IA híbrida

## AUTOR

**Steven Dickens**

Asesor jefe de tecnología | The Futurum Group

**Ron Westfall**

Director de investigación | The Futurum Group

**AGOSTO DE 2024**

EN COLABORACIÓN CON





## Resumen ejecutivo

Actualmente, los responsables de la toma de decisiones empresariales y de TI (ITBDM) están supervisando la tarea sin precedentes de guiar el viaje hacia la IA de su organización. Las opciones son amplias y la velocidad de la innovación es intensa. Al explorar el mejor camino para el viaje de la IA, vemos que los responsables de la toma de decisiones eligen cada vez más un modelo de IA híbrida como el mejor curso general. ¿Por qué la IA híbrida?

La IA híbrida es la optimización o la extensión de los modelos de IA basados en el aprendizaje automático, el aprendizaje profundo, las redes neuronales con experiencia humana para crear modelos de IA específicos de dominio con una precisión mejorada, o modelos específicos de casos de uso de IA con la máxima precisión o probabilidad de predicción. La IA híbrida es la idea de adherirse a la gravedad de los datos y llevar las soluciones y capacidades de IA allí donde se necesiten: desde dispositivos móviles y de sobremesa, centros de datos privados, computación periférica o incluso la nube. Como tal, el caso de uso y la estrategia de gestión de datos pueden dictar dónde y cómo implementar las capacidades de IA, teniendo en cuenta factores como la seguridad, el coste, la infraestructura de red, la gravedad de los datos y la fiabilidad.

Vemos que este enfoque supera las limitaciones de los métodos de un modelo único o de una infraestructura única, ajustándose a las demandas de los clientes de una gestión de datos flexible y de asistencia a las aplicaciones. La IA híbrida integra la IA de código abierto y las capacidades de código cerrado, como las capacidades ChatGPT de OpenAI, lo que aumenta la productividad, la agilidad, la protección de datos y la democratización de la IA.

Para los principales responsables de la toma de decisiones, incluidos los CIO, los CTO, los directores de datos, los directores de IA/datos y los arquitectos, el paso inicial consiste en definir los problemas, identificar las fuentes de datos y seleccionar los casos de uso, las aplicaciones de IA, con la combinación adecuada de infraestructura y modelos del tamaño adecuado. Este proceso suele incluir el desarrollo de estrategias, la creación de una pila de IA, una fábrica de IA o un centro de excelencia (COE) y la implementación de las soluciones elegidas.

Implementar la IA híbrida requiere evaluar, probar y escalar las tecnologías emergentes con un amplio ecosistema de socios para poder diseñar una solución que tenga la combinación adecuada de técnicas de IA, la arquitectura adecuada y el modelo adecuado para su negocio.

Examinamos por qué la cartera y la visión de Lenovo Smarter AI for All pueden satisfacer las demandas únicas de seleccionar e implementar la IA híbrida en toda la organización. Lenovo Smarter AI for All da prioridad al enfoque basado en los resultados empresariales al proporcionar la combinación y el tamaño adecuados de PC con IA, estaciones de trabajo, soluciones preparadas para IA, experiencia y



capacidades de almacenamiento, computación, infraestructura, software y asociación optimizados para IA con un ecosistema de socios validado. Esto incluye la capacidad de ofrecer soluciones de IA personalizadas, ágiles, escalables y energéticamente eficientes basadas en casos de uso y en el sector, tanto en el ámbito personal como en el industrial (ordenadores, estaciones de trabajo, ubicaciones comerciales), centros de datos (CoLos, locales) y entornos de computación de alto rendimiento (HPC), todo ello con integración/sinergias en la nube pública.

Juntos, descubrimos que Lenovo y NVIDIA crean fábricas de IA que proporcionan entornos empresariales de IA optimizados y habilitados para la IA donde las aplicaciones de IA pueden desarrollarse, implementarse y administrarse de forma segura, privada y a escala.

## El panorama de la IA híbrida evoluciona rápidamente

Las organizaciones y sus empleados están ansiosos por aprovechar el poder transformador de la IA generativa (GenAI), yendo más allá de los experimentos iniciales y los pilotos de IA/ML para aprovechar sus propios datos. Sin embargo, la adopción generalizada a menudo se ve obstaculizada por problemas legales, de cumplimiento y de gobernanza, lagunas de conocimientos, silos de datos, problemas de seguridad, una infraestructura de IA inadecuada y la complejidad del ecosistema. Las limitaciones presupuestarias complican aún más las cosas. Aunque el 96 % de los CIO anticipan un aumento de las inversiones en IA el próximo año, solo el 20 % espera que los presupuestos generales de TI crezcan más de un 10 % (CIO Study 2024).

La IA híbrida está emergiendo como un componente crucial de las estrategias empresariales de IA, permitiendo a las organizaciones desarrollar soluciones de IA y datos que se alinean con sus objetivos y acercan la IA a las fuentes de datos, las interacciones en tiempo real y los puntos de creación de experiencias. Aunque las herramientas gratuitas de IA, los modelos fundacionales públicos y las suscripciones a la nube existentes facilitan los pilotos iniciales, muchas organizaciones se encuentran atrapadas en ciclos interminables de prueba de concepto y costes crecientes. Esto suele deberse a la falta de una estrategia sólida de IA híbrida que integre la IA en entornos de nube privada y pública, a la ausencia de modelos de IA personalizados, al acceso limitado a datos organizativos de calidad y al tamaño y la planificación informática de la IA inadecuados.

Desde nuestro punto de vista, ejecutar GenAI e IA localmente en el perímetro, incluidos los dispositivos, se está convirtiendo en algo crítico para las organizaciones con datos distribuidos y ubicaciones en el perímetro (por ejemplo, comercio minorista, fabricación, hospitales, lugares de trabajo híbridos), aquellas que requieren análisis y acceso en tiempo real, y entidades con necesidades de computación desconectadas (por ejemplo, operaciones remotas de petróleo y gas o entornos altamente protegidos).

Las ubicaciones y los dispositivos perimetrales generan grandes cantidades de datos, la mayoría de los cuales no se encuentran homologados. La IA y la GenAI presentan una oportunidad importante para aprovechar este recurso sin explotar. Con aplicaciones de IA desarrolladas para casos de uso perimetrales, modelos más pequeños cada vez más eficientes y dispositivos habilitados para IA como los PC con IA, la computación perimetral se convierte en una realidad. Esto elimina la necesidad de transferir todos los datos o modelos de IA a nubes públicas o centros de datos centralizados para su análisis en tiempo real, abordando las limitaciones de seguridad, latencia y ancho de banda.

El procesamiento local de GenAI permite analizar los datos en el punto de creación, lo que reduce la latencia y mejora la eficiencia, algo crucial para aplicaciones en tiempo real, como los procesos de fabricación y la robótica. Además, por motivos de privacidad, a menudo se necesita un procesamiento local para garantizar que los datos confidenciales permanezcan en el dispositivo.

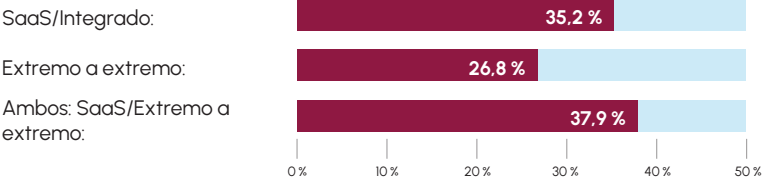
Los modelos lingüísticos de gran tamaño (LLM) basados en la nube presentan desafíos que las organizaciones intentan evitar cada vez más. Estos incluyen el riesgo de alucinaciones, resultados lógicamente incoherentes pero plausibles, y la propagación de imprecisiones a partir de los datos de entrenamiento. Además, las organizaciones tienen un control y una visibilidad limitados sobre la infraestructura y la implementación de los servicios gestionados en la nube, lo que plantea problemas de seguridad.

Un ejemplo notable se produjo en marzo de 2023 cuando ChatGPT de OpenAI experimentó una interrupción debido a una vulnerabilidad en una biblioteca de código abierto, lo que podría exponer la información de pago de los clientes. Este incidente subraya la importancia de realizar rigurosas pruebas de seguridad de LLM y de contar con mecanismos sólidos de defensa frente a ciberamenazas cuando se utilizan servicios de IA basados en la nube.

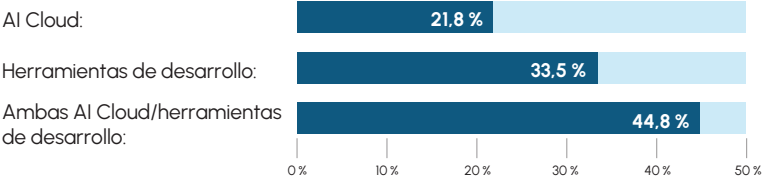
Además, vemos que las organizaciones están adoptando un enfoque equilibrado sobre cómo implementan los servicios de IA:



Externalización



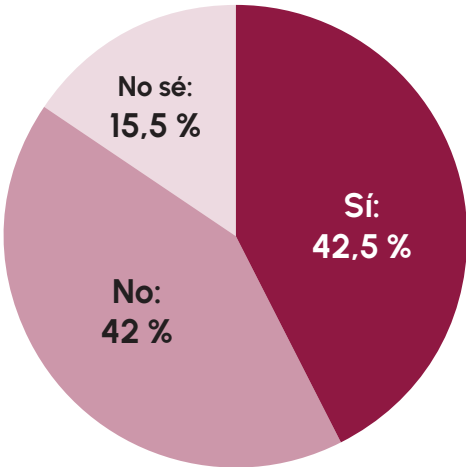
Internalización



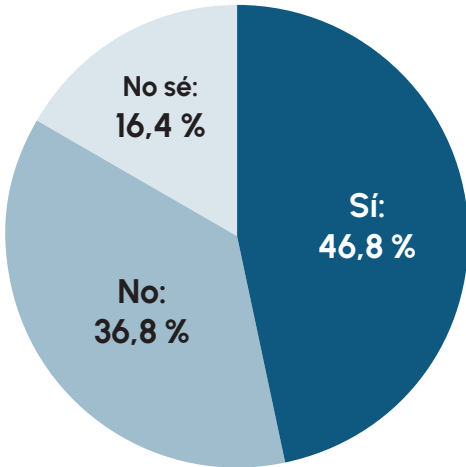
Fuente: The Futurum Group

El interés cada vez mayor en la IA híbrida está creando nuevas dinámicas competitivas a medida que las organizaciones exploran cada vez más planes para cambiar o agregar nuevos proveedores (2023-24):

Externalización



Internalización



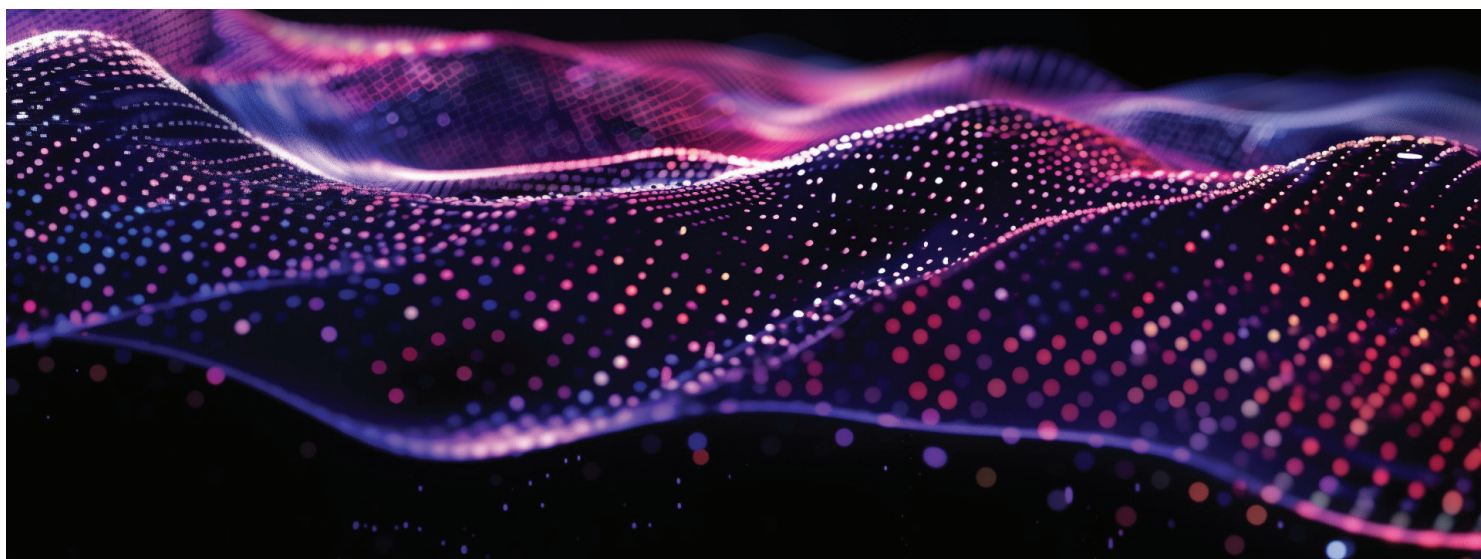
Fuente: The Futurum Group

El éxito de la IA híbrida depende del papel fundamental que desempeñan los ecosistemas en la selección de soluciones de IA adaptadas a las necesidades de la organización. Los ecosistemas reúnen a una amplia gama de partes interesadas de la cadena de valor, lo que permite la integración perfecta de los sistemas de IA en entornos híbridos y la creación de soluciones de IA hechas a medida. Las soluciones de IA suelen requerir una combinación de tecnologías y conocimientos, y dependen de los ecosistemas para proporcionar acceso a la amplia gama de software, aplicaciones, herramientas y capacidades para desarrollar soluciones que puedan abordar directamente los desafíos empresariales a medida que se desarrolla el viaje de la IA.

Además, la flexibilidad inherente de la IA híbrida es esencial para satisfacer las diversas necesidades de las distintas personas implicadas en la IA dentro de las empresas. Los desarrolladores de aplicaciones de IA y los científicos de datos prefieren usar herramientas y marcos de IA conocidos, independientemente de su distribución. En la actualidad, sectores como el comercio minorista y la industria manufacturera, esperan y necesitan respuestas en tiempo real de los sistemas de IA, además de la capacidad de procesar datos con rapidez.

Sin embargo, el uso efectivo de la IA híbrida requiere una planificación y ejecución meticulosas, teniendo en cuenta la tecnología, las personas, los procesos y los enfoques de la IA responsable. Las empresas se enfrentan a iniciativas aisladas, opciones de soluciones de IA y presupuestos desalineados, mientras que los equipos de TI luchan por modernizar la infraestructura y satisfacer las diversas demandas de computación de IA, el cumplimiento, los estándares de seguridad y la necesidad de ejecutar una operación híbrida. A medida que la gestión de datos y el uso de la nube pública se vuelven incontrolables, la seguridad, la gobernanza y la escalada de los costes de la nube surgen como preocupaciones acuciantes. Estos desafíos deben sopesarse con las habilidades y la gestión del cambio necesarias para lograr resultados empresariales significativos con la IA.

Por ejemplo, los científicos de datos y los ingenieros de aprendizaje automático no deberían dedicar la mayor parte de su tiempo a la calidad de los datos, las regulaciones y el cumplimiento. Las empresas deben involucrar a los equipos de TI y operaciones en las decisiones informáticas relacionadas con la IA. Del mismo modo, un enfoque ascendente por parte de los equipos de TI para introducir la IA en la empresa puede dar lugar a un aprovisionamiento excesivo o insuficiente sin una comprensión profunda de las necesidades empresariales y los casos de uso.



## Cartera de productos diseñada conjuntamente por Lenovo y NVIDIA

El crecimiento de la industria de la IA está impulsado por múltiples factores, siendo primordial la creciente demanda de automatización y optimización en todos los sectores. Como se indica en nuestro documento, las empresas están aprovechando los avances en la potencia informática y la infraestructura de la nube para implementar aplicaciones de IA más eficientes. Este cambio ha provocado un aumento de las solicitudes de soluciones de IA personalizadas, incluidos los servicios de consultoría, diseño, implementación y mantenimiento que pueden ayudarles a simplificar ese proceso.

La cartera de Lenovo Smarter AI for All aborda estas necesidades en evolución con dispositivos habilitados para IA, infraestructura inteligente y soluciones sólidas en la nube, lo que facilita la integración perfecta de la IA en varias plataformas. La asociación estratégica entre Lenovo y NVIDIA está acelerando la innovación en el sector de la IA. Su cartera de productos de ingeniería conjunta simplifica la adopción de la IA, ofrece resultados más rápidos y mejora el rendimiento y la eficiencia en entornos de IA híbrida, impulsando aún más el crecimiento del sector.

# Información importante sobre este informe

## CONTRIBUYENTES

### Steven Dickens

Asesor jefe de tecnología | The Futurum Group

### Ron Westfall

Director de investigación | The Futurum Group

## EDITOR

### Daniel Newman

CEO | The Futurum Group

## CONSULTAS

Ponte en contacto con nosotros si quieres hablar sobre este informe y The Futurum Group responderá con prontitud.

## CITAS

La prensa y los analistas acreditados pueden citar este documento, pero deben hacerlo en su contexto, indicando el nombre del autor, su título y «The Futurum Group». Los medios de comunicación y los no analistas deben recibir autorización previa por escrito de The Futurum Group para cualquier cita.

## LICENCIAS

Este documento, incluido cualquier material de apoyo, es propiedad de The Futurum Group. Esta publicación no puede reproducirse, distribuirse ni compartirse de ninguna forma sin el permiso previo por escrito de The Futurum Group.

## DIVULGACIÓN

The Futurum Group presta servicios de investigación, análisis, asesoramiento y consultoría a muchas empresas de alta tecnología, incluidas las mencionadas en este documento. Ningún empleado de la empresa posee acciones de ninguna empresa citada en este documento.



## ACERCA DE LENOVO Y NVIDIA

Lenovo pone al alcance de todos la nueva era de la innovación basada en la IA. Nuestra cartera completa de productos ofrece soluciones de IA potentes, flexibles y responsables para transformar las industrias y empoderar a las personas. Creamos un futuro de IA más inteligente para todos. En Lenovo, creemos que el futuro de la IA pasa por la coexistencia de la IA pública y la empresarial. Lenovo te acerca la IA a ti y a tus datos.

En colaboración con NVIDIA, las soluciones de IA híbrida se han creado específicamente mediante la colaboración en materia de ingeniería para llevar la IA de manera eficiente a los datos de los clientes, donde y cuando los usuarios más la necesiten, lo que hace avanzar la visión de Lenovo de habilitar la IA para todos y ofrecer soporte en tiempo de comercialización de tecnologías y arquitectura de vanguardia para la próxima generación de IA generativa. Las soluciones híbridas de Lenovo, ya optimizadas para ejecutar el software NVIDIA AI Enterprise para una IA de producción segura, compatible y estable, también proporcionan a los desarrolladores acceso a los microservicios de NVIDIA, incluidos NVIDIA NIM y NeMo Retriever.



## ACERCA DE THE FUTURUM GROUP

[The Futurum Group](#) es una empresa independiente de investigación, análisis y asesoramiento, centrada en la innovación digital y las tecnologías y tendencias que alteran el mercado. Cada día, nuestros análisis, investigadores y asesores ayudan a los líderes empresariales de todo el mundo a anticipar los cambios tectónicos en sus sectores y a aprovechar la innovación disruptiva para obtener o mantener una ventaja competitiva en sus mercados.



## INFORMACIÓN DE CONTACTO

The Futurum Group LLC | [futurumgroup.com](https://futurumgroup.com) | (833) 722-5337 |

© 2024 The Futurum Group. Todos los derechos reservados.

