



**Hewlett Packard
Enterprise**

Para ganar con la IA es necesario poner en orden los datos

Tu estrategia de IA debe empezar con una sólida estrategia de datos. Te presentamos cinco formas de garantizar que tus datos estén preparados para la IA.



Las empresas se están desviviendo por invertir dinero en IA, por lo que se prevé que el gasto mundial en aplicaciones, infraestructura y servicios de IA se duplicará en los próximos cuatro años y alcanzará los 632 000 millones de dólares en 2028, según IDC.¹ Lamentablemente, puede que muchas de estas organizaciones se apresuren para implementar soluciones de IA antes de haber sentado las bases necesarias. A menudo se pasa por alto un primer paso fundamental: garantizar que los datos necesarios para impulsar estas cargas de trabajo de IA estén a la altura.

«Cuando la calidad de los datos es mala, los resultados son malos, por muy bueno que sea el modelo de aprendizaje automático», afirma Chad Smykay, director de tecnología de inteligencia artificial para sectores verticales del sector y distinguido tecnólogo de Hewlett Packard Enterprise. Smykay dice que el secreto de la buena calidad de los datos no es ningún secreto: los datos deben estar preparados, limpios y bien organizados para que los científicos de datos y los desarrolladores de toda la empresa puedan utilizarlos. Sin embargo, la producción de grandes volúmenes de datos limpios para satisfacer las necesidades de rendimiento de las cargas de trabajo de IA con uso intensivo de datos que tenemos actualmente requiere mucho tiempo y trabajo, enormes desafíos en un mundo en el que escasea el talento para los datos y la velocidad es esencial.

¿Cómo superar estos desafíos y optimizar la calidad de los datos? Smykay ofrece cinco pasos concretos para acelerar la implementación de IA a través de unas mejores operaciones de datos.

¹ «Se prevé que el gasto mundial en inteligencia artificial alcance los 632 000 millones de dólares en 2028, según una nueva guía de gasto de IDC», IDC, agosto de 2024

1. Establecer políticas inteligentes de gobernanza de datos

Todas las empresas necesitan una base sólida de gobernanza de datos que garantice que los datos están bien definidos y organizados mientras se recopilan, antes de que se introduzcan en un algoritmo o se utilicen para entrenar un gran modelo de lenguaje (LLM). Estas políticas requieren que las empresas se hagan preguntas como: «¿Qué tipo de datos son estos? ¿De dónde proceden? ¿Se aplican a mi caso de uso empresarial? ¿Qué antigüedad tienen? ¿Dónde deben almacenarse?».

No se trata solo de una cuestión de buen mantenimiento. Una gobernanza de datos eficaz garantiza que los datos que impulsan los modelos de IA de una organización no solo estén correctamente definidos, sino que también sean seguros y fiables. Ello también ayuda a preparar a la organización para el futuro, en cuanto a leyes de privacidad y otras normativas legales que probablemente surjan en los próximos años. Estas normativas, como el Reglamento General de Protección de Datos (GDPR) de la Unión Europea, pueden exigir a las empresas que muestren las tareas realizadas sobre el origen de los datos y cómo se han utilizado para entrenar algoritmos.

Si bien los nuevos productos de software prometen automatizar la gobernanza de datos, Smykay dice que ésta, al menos al principio, debe ser un proceso de personas alineado con la tecnología, gestionado por un director de tecnología, un director de datos o un director de datos e inteligencia artificial. ¿Por qué? Lo más probable es que el software no sea capaz de discernir qué datos serán realmente útiles para satisfacer las necesidades y alcanzar los objetivos empresariales en materia de inteligencia artificial.

«Para obtener información de la inteligencia artificial que te ayude a tomar decisiones basadas en los

datos, es fundamental que los datos que clasifiques, limpies y almacenes sean relevantes para tus casos de uso empresarial de la IA», afirma. «Esto significa que las personas que comprenden el modo en que la organización extraerá valor de la IA deben ser las que establezcan las reglas que se deberán seguir».

2. Simplificar la gestión de datos con un espacio de nombres global

Cuando se trata de optimizar los datos para implementar con éxito la IA, es importante trabajar para simplificar la gestión global de los datos. Para lograr la simplicidad, Smykay recomienda que las empresas creen una estructura de datos con un espacio de nombres global que gestione todos los datos de sus organizaciones. Piensa en ello como una única snapshot que abarca todos los activos de datos a través del entorno informático distribuido e híbrido, claramente organizado con un sistema de nombres unificado y coherente para los diferentes conjuntos de datos que los desarrolladores de inteligencia artificial puedan querer utilizar.

«Un espacio de nombres global es una solución que ofrece a los desarrolladores la manera más fácil de encontrar datos almacenados en múltiples ubicaciones, ya sea en varias nubes públicas, repartidos entre centros de datos o en el extremo», afirma Smykay. «Donde sea que estén, ellos pueden ver y acceder rápidamente a los datos que necesitan».

El llamado «desperdicio de datos» es otra preocupación que un espacio de nombres global puede ayudar a abordar. «Nunca se accede a una gran parte de los datos empresariales», afirma Smykay. «Con un espacio de nombres global, puedes acceder a estos datos y finalmente darles un buen uso, especialmente para cargas de trabajo de IA».



3. Crear una plataforma de datos flexible, abierta y portátil

En el complejo entorno empresarial actual, los datos son diversos y se encuentran dispersos entre dónde se originan, dónde se almacenan e incluso qué forma adoptan. Ya sean estructurados o no estructurados, generados internamente o extraídos de Internet, los datos actuales se almacenan en un sinnúmero de formatos y, para acceder a ellos, hay que utilizar múltiples interfaces de programación de aplicaciones (API) que pueden no comunicarse entre sí.

«Los datos dispersos y diseminados pueden generar problemas; por ello, la consolidación es la mejor solución», dice Smykay. Reunir todos los activos de datos diversos y dispersos de una empresa en una estructura de datos unificada es un paso fundamental para poner datos de calidad en manos de los desarrolladores. Una única plataforma de acceso a los datos de API abierta que admita múltiples tipos de datos ofrece a los desarrolladores el ingrediente crucial que necesitan para triunfar: la flexibilidad. «En el espacio de los análisis y la inteligencia artificial, los desarrolladores quieren una plataforma de datos que pueda aprovechar múltiples API porque les ofrece la forma más flexible de acceder a los datos que necesitan y prepararlos, independientemente del formato en el que se encuentren», afirma Smykay.

4. Aumentar la escalabilidad de los datos

Los datos son el combustible que impulsa el entrenamiento de los modelos y los resultados de IA. Para hacer frente a los desafíos de rendimiento de las cargas de trabajo con uso intensivo de datos, como los enormes volúmenes de datos necesarios para entrenar los LLM de IA generativa, se requiere una capacidad de almacenamiento considerable. «Para acelerar la recopilación, preparación y despliegue de datos de IA, las empresas necesitan una solución de gestión de datos escalable que abarque la nube híbrida», dice Smykay.

Para los exabytes de datos que se necesitan para impulsar los modelos de IA actuales, la tecnología híbrida es una de las estrategias de gestión de datos más asequible y escalable, explica. «Las grandes soluciones tradicionales de hiperescala en la nube pública se han vuelto demasiado



caras, especialmente para este tipo de cargas de trabajo de IA», dice Smykay. «La nube híbrida puede satisfacer las demandas de escalabilidad distribuyendo los activos dondequiera que se generen o se consuman los datos, desde una nube privada local hasta el extremo».

La gestión de estos datos se está convirtiendo en un enorme desafío, ya que el volumen de datos que se genera se acelera a la velocidad del rayo. Según IDC, se prevé que la cantidad de datos que se crean anualmente se duplique con creces en 2026 con respecto a los niveles de 2022, y que la cantidad de datos empresariales crezca el doble de rápido que los datos de los consumidores.²

Smykay añade que la estrategia del plano de datos de la nube híbrida también es flexible y está orientada al futuro, por lo que está mejor posicionada para dar soporte a cualquier modelo innovador de IA que surja en el futuro.

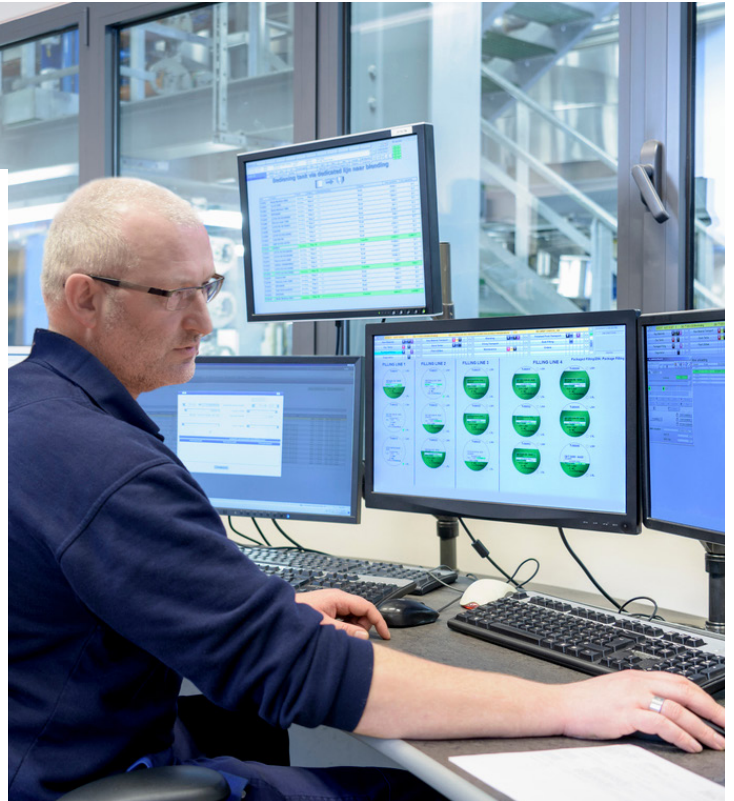
² «Esfera de datos mundial, mercados de datos y datos como servicio», IDC, agosto de 2023

5. Acelerar la canalización de datos para alimentar más rápidamente los modelos de IA

Las cargas de trabajo de IA no solo requieren una gran cantidad de datos, sino también datos precisos para maximizar su utilidad. «Pero la mayoría de los datos tienen una vida útil», dice Smykay, lo que significa que los modelos de IA deben asimilarlos rápidamente mientras aún están actualizados.

Sin embargo, a pesar de esta urgente necesidad de velocidad, el 85 % de las organizaciones del Fortune 500 dicen que no pueden utilizar sus datos de manera efectiva.³ No obstante, las soluciones de streaming integradas en una plataforma de datos unificada pueden ayudar a acelerar el proceso. «Has creado y entrenado tus modelos de IA. Ahora, el desafío es mantenerlos alimentados», afirma Smykay. «La buena noticia es que se puede mantener el flujo de datos en tiempo real si se cambia la forma en que se ingieren en el back-end con un formato de streaming».

Según Smykay, una ingesta de datos más rápida puede ser muy valiosa para garantizar la calidad de los mismos, ya que ofrece a los desarrolladores un plazo más corto para limpiar y preparar los datos. Aunque están surgiendo herramientas automatizadas que ayudan a preparar los datos para los modelos de IA, los desarrolladores y los científicos de datos saben que el control de calidad de los datos es demasiado importante como para externalizarlo a un algoritmo. Pero con las soluciones de datos adecuadas,



las empresas pueden posicionarse con el fin de optimizar los datos de alta calidad para un despliegue rápido, flexible y seguro de la IA.

El crecimiento de los datos no se detiene, sino más bien al contrario: a medida que la IA y los análisis de empresa sigan evolucionando, es probable que el crecimiento de los datos se acelere aún más. Es fundamental prepararse para esta realidad. A la hora de planificar y construir la base de datos de una organización, la estructura de datos es clave. Permitirá incorporar datos que abarquen toda la nube híbrida, admitir una amplia gama de formatos de datos, escalar junto con la empresa y ayudar a la organización a crear una estrategia de gobernanza de datos fiable que garantice la seguridad y la confianza.

³ «Ahogándose en datos por falta de información: ¿Es realmente posible la minimización de datos?», IDC, septiembre de 2024

Más información en

HPE.com/AI/insights

Visita HPE.com

Iniciar chat


**Hewlett Packard
Enterprise**

© Copyright 2024 Hewlett Packard Enterprise Development LP. La información contenida en este documento está sujeta a cambios sin previo aviso. Las únicas garantías de los productos y servicios de Hewlett Packard Enterprise figuran en las declaraciones expresas de garantía incluidas en los mismos. Nada de lo que aquí se indica debe interpretarse como una garantía adicional. Hewlett Packard Enterprise no se responsabilizará de los errores u omisiones técnicos o editoriales que pudiera contener el presente documento.

a50011804ESE