

El grito de guerra de la inteligencia artificial: aprovecha los datos, pero solo los buenos

Cinco formas de optimizar tus estrategias de inteligencia artificial en torno a la limpieza, la preparación y la gestión de los datos

Las empresas están apostando su futuro a la inteligencia artificial, ya que casi tres cuartas partes de las organizaciones dan prioridad a la inteligencia artificial sobre todas las demás inversiones digitales actuales, según Accenture.¹ Y no se equivocan al apostar todo a esta tecnología emergente, dado que muchos creen que las estrategias impulsadas por la IA acelerarán la innovación, aumentarán la productividad y crearán nuevas y fiables fuentes de ingresos.

Pero ganar con la inteligencia artificial no es cuestión de suerte. El botín será para las empresas que sepan aprovechar el poder de sus datos para sacar el máximo partido de las iniciativas de inteligencia artificial. Lamentablemente, no todos los datos son iguales. Los llamados datos sucios —datos obsoletos, incompletos, mal etiquetados o poco seguros— pueden corromper rápidamente los resultados del algoritmo de aprendizaje automático mejor codificado.

«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, jefe de tecnología de campo y distinguido tecnólogo en 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 los requisitos de rendimiento de las cargas de trabajo de inteligencia artificial 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 nos ofrece cinco pasos concretos para acelerar la implementación de inteligencia artificial a través de mejores operaciones de datos.

¹ [«Reinvención de las operaciones empresariales»](#), Accenture, mayo de 2023

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 dato es este? ¿Cuál es su origen? ¿Se aplica a mi caso de uso empresarial? ¿Qué antigüedad tiene? ¿Dónde debería guardarse?»

No se trata solo de una cuestión de buen mantenimiento. Una gobernanza de datos eficaz garantiza que los datos que impulsan los modelos de inteligencia artificial de una organización no solo estén correctamente definidos, sino que también sean seguros y fiables. 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 regulaciones, como el actual Reglamento General de Protección de Datos (RGPD) de la Unión Europea, pueden exigir además 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 esta, 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 práctica 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 inteligencia artificial», afirma. «Esto significa que las personas que comprenden cómo la organización extraerá valor de la inteligencia artificial deben ser las que establezcan las reglas a 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 inteligencia artificial, 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 un único snapshot que abarca todos los activos de datos 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, pueden ver y acceder rápidamente a los datos que necesitan».

Esta visión panorámica no solo permite crear un mapa de todos los datos existentes, sino que también permite eliminar el desperdicio de datos, dice Matt Hausmann, director de marketing de HPE Ezmeral Software. «Hasta un 30 % de los datos de una empresa nunca se tocan», afirma Hausmann. «Un espacio de nombres global puede hacer que se usen más datos, ya que son visibles y accesibles». Esto, a su vez, puede llevar a que se aprovechen más datos de toda la empresa para generar información impulsada por IA mucho más precisa a partir de todos esos datos.

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 desestructurados, 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 con API abierta, que admita múltiples tipos de datos, ofrece a los desarrolladores el ingrediente crucial que necesitan para triunfar: la flexibilidad. «En el espacio del análisis de datos 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 y preparar los datos que necesitan, independientemente del formato en el que se encuentren», afirma Smykay.

Para Hausmann, adoptar una plataforma de datos API es tanto prepararse para el futuro de la inteligencia artificial como alimentar los modelos actuales. «En el ámbito de la inteligencia artificial, asistimos a un ciclo constante de nuevos avances, desde el aprendizaje profundo y la detección de imágenes hace unos años, hasta el auge de los LLM el año pasado», reflexiona. «Una plataforma de datos unificada compatible con múltiples formatos de archivos, objetos, flujos, tablas y vectores, permite seguir evolucionando la estrategia de datos para satisfacer las necesidades futuras de la empresa».

4.

Aumentar la escalabilidad de los datos

Los datos son el combustible que impulsa el entrenamiento y los resultados de los modelos de inteligencia artificial. 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 inteligencia artificial generativa, se requiere una capacidad de almacenamiento considerable. Para acelerar la recopilación, preparación y despliegue de datos de inteligencia artificial, 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 inteligencia artificial actuales, la tecnología híbrida es una de las estrategias de gestión de datos más asequibles y escalables, explica. «Las grandes soluciones tradicionales de hiperescalador en la nube pública se han vuelto demasiado caras, especialmente para este tipo de cargas de trabajo de inteligencia artificial», dice Smykay. «La nube híbrida puede satisfacer las demandas de escalabilidad distribuyendo los activos dondequiera que se generen o se ingieran 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. Se prevé que la cantidad de datos que se crea 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, según IDC.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 inteligencia artificial que surja en el futuro.

² «Global Datasphere, Data Marketplaces, and Data as a Service» («Esfera de datos mundial, mercados de datos y datos como servicio»), IDC, agosto de 2023

5.

Acelerar los pipelines de datos para alimentar más rápidamente los modelos de inteligencia artificial

En términos de valor, los datos tienen fecha de vencimiento, dice Hausmann. «El valor de los datos suele decaer rápidamente con el tiempo. Empiezan a perder valor en el momento en que se crean y, según cómo se aprovechen, pueden perder la mayor parte de su valor en segundos, horas, días o semanas», explica. Las empresas tienen que garantizar que los datos estén disponibles tan rápido como sea posible para sacarles el máximo partido analítico, aconseja.

Pero a pesar de esta urgente necesidad de velocidad, el 76 % de las organizaciones afirman que sus actuales procesos de gestión de datos no pueden seguir el ritmo de sus necesidades empresariales.³ No obstante, las soluciones de streaming integradas en una plataforma de datos unificada pueden ayudar a acelerar el proceso. «Has creado tus modelos de inteligencia artificial y los has entrenado. Ahora, el desafío es mantenerlos alimentados», afirma Smykay. «La buena noticia es que se puede mantener el flujo de los datos en tiempo real si se cambia la forma en que se ingieren en el back-end con un formato de streaming».

Una ingestión de datos más rápida puede ser muy valiosa para garantizar la calidad de los mismos, señala Smykay, ya que ofrece a los desarrolladores un plazo más corto para la limpieza y preparación de los datos. Aunque están surgiendo herramientas automatizadas que ayudan a preparar los datos para los modelos de inteligencia artificial, 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 inteligencia artificial.

El crecimiento de los datos no se frena. De hecho, a medida que la IA y el análisis de negocios continúan 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 «Being a Data-First Leader Continues to Matter» («Un líder basado en datos sigue siendo importante»), Enterprise Strategy Group, octubre de 2023

Más información en

[HPE.com/AI](https://www.hpe.com/AI)

Visita [HPE GreenLake](#)

 **Iniciar chat**