

## DIGITALIZACIÓN Elemento de distorsión

# Los 'datos sucios' embarran el desembarco de los algoritmos en las empresas

**Definir una estrategia que separe el grano de la paja en el ingente magma del 'big data' es el gran reto pendiente para el desarrollo sólido de la IA en las compañías**

ADRIÁN ESPALLARGAS

La calidad de los datos es una cuestión fundamental para la correcta ejecución de aplicaciones de inteligencia artificial (IA) generativa. «Tu inteligencia artificial es tan buena como la calidad de sus datos», dice Raúl Bartolomé, managing director de insights & data de Capgemini España, al citar una frase bastante repetida últimamente dentro del sector. «Si los datos no son buenos, la respuesta no será buena», añade Bartolomé para hacer hincapié en la importancia de contar con información de calidad para que la IA trabaje de manera adecuada, por lo que minimizar y eliminar los datos sucios, o 'dirty data' en inglés, es una labor a la que las organizaciones cada vez dan más importancia.

«El 'dirty data' es información desactualizada, errónea, duplicada o incompleta. Puede suponer un desafío para las empresas a la hora de gestionar correctamente todos los datos que reciben», dice Juan Luis Moreno, partner & chief innovation officer de The Valley. Los datos sucios pueden producirse por varias razones. Puede tratarse de información que se ha introducido incorrectamente, puede ser el resultado de un cálculo erróneo o incluso puede deberse a que proceda de dos fuentes distintas, afirma Bartolomé, de Capgemini.

### Disparidades

Dentro de las organizaciones, a veces ocurre que los mismos datos pueden estar almacenados de forma distinta en dos bases de datos diferentes, por ejemplo, el departamento de marketing puede tener el nombre de una empresa con las siglas «SL» al final, mientras que el departamento de ventas está almacenado como «S.L.».

Uno de los primeros pasos para resolver este tipo de problemas es gestionarlos y normalizarlos adecuadamente, es decir, definir qué miden los datos, cómo se introducirán en la base de datos, cómo se calcularán, quién tendrá que registrarlos y quién será el responsable de gestionar esta información dentro de la empresa. Es necesario invertir en herramientas tecnológicas para homogeneizar los datos y establecer una cultura dentro de la organización centrada en cuidar su calidad, explica Bartolomé, de Capgemini.

«Cuando los modelos de inteligencia artificial se entrenan con conjuntos de datos que contienen errores, sesgos o información incorrecta, es probable que produzcan resultados inexactos o 'alucinaciones'. Estas pueden manifestarse como respuestas incorrectas, conclusiones sesgadas o predicciones inexactas», dice Moreno, de The Valley.

Esta es una de las consecuencias más graves de introducir datos erróneos en la IA generativa, que el sistema puede 'alucinar'. Estas 'alucinaciones'

ocurren cuando la IA inventa respuestas basadas en información incorrecta. La IA generativa siempre responderá con cierto grado de confianza, pero si se le proporciona información incorrecta, es probable que genere respuestas incorrectas que pueden parecer verídicas. Esto es lo que se conoce como 'alucinaciones' de la inteligencia artificial, según explica Bartolomé.

«La calidad de los datos es

### EN ALIANZA

**Es clave la inversión en tecnología, pero también mejorar la cultura de la organización**

un proceso continuo. Los problemas relacionados con los datos de baja calidad son comunes, pueden ser complejos de resolver e implicar un gran esfuerzo si no se atajan a tiempo», asegura David Hurtado, director de innovación de Microsoft en España. Hurtado cree que hay una serie de prácticas que pueden llevar a una mala gestión de los datos, como haber implantado procedimientos erróneos de gestión de datos, no actualizar la información, una categorización inadecuada o la falta de

inversión en una infraestructura tecnológica adecuada.

También existen soluciones automáticas, como las que ofrece la empresa Datarmony, que tratan de corregir e identificar errores en los datos antes de que causen problemas graves. El objetivo de estas soluciones es mejorar la fase de recogida de datos para que estos se almacenen correctamente. «La calidad de los datos es un tema del cual nos damos cuenta de su importancia cuando perdemos ventas o clientes», explica Enrique Quintero, CEO de la empresa.

### Sin rumbo

Actualmente, muchas empresas enfrentan problemas de gestión de datos porque desde la aparición del 'big data', empezaron a almacenar todo tipo de información sobre su negocio sin una estrategia y sin saber qué harían con esos datos. «Paralelamente, aparecieron grandes 'cloud providers' (AWS, Azure y Google) que ofrecían de manera económicamente razonable grandes capacidades de almacenamiento. Actualmente en plena

vorágine de la IA, las compañías intentan utilizar sin éxito estos datos para los algoritmos de IA», dice Tony Rodríguez, head of data strategy & data management en NTT Data.

Para Rodríguez, un problema fue que ese almacenamiento se realizó sin ninguna estrategia de calidad ni catalogación sobre qué representa ese dato para la compañía. Ante esta situación, muchas empresas están poniendo en marcha estrategias para procesar la información que tienen de manera adecuada y poder ejecutar herramientas de IA generativa.

El experto de NTT Data considera que las empresas deben comenzar a trabajar en una buena estrategia de gobierno del dato que les permita disponer de un glosario de datos, herramientas para medir la calidad de la información y planes de remediación para responder a errores. Desde Microsoft, Hurtado comenta que otras estrategias eficaces son establecer claramente un conjunto de reglas de calidad o

## UN ROL CADA VEZ MÁS DECISIVO

**El chief data officer (CDO) desempeña un papel fundamental en las organizaciones modernas al gestionar estratégicamente los datos de la empresa. En primer lugar, supervisa la gestión integral de los datos, desde su adquisición hasta su almacenamiento y análisis. Esto implica desarrollar políticas y procedimientos para garantizar la calidad, seguridad y privacidad de los datos. Además, lidera iniciativas para aprovechar el valor de los datos, identificando oportunidades de negocio y aplicando análisis avanzados para obtener información significativa.**

**También colabora estrechamente con otros líderes empresariales para alinear la estrategia de datos con los objetivos organizacionales. El CDO es responsable de garantizar el cumplimiento de las regulaciones de protección de datos y de mitigar los riesgos relacionados con la seguridad de la información. Además, fomenta una cultura empresarial centrada en los datos y fomentando la innovación con el uso estratégico de la información. En resumen, juega un papel crucial en el impulso del valor empresarial a través de la gestión eficaz de los datos.**





realizar una limpieza periódica de la información.

«No quiero dejar de comentar la importancia fundamental de combinar de un modo equilibrado dos elementos: personas y herramientas. Mejorar la calidad de los datos necesita tanto a los expertos como los recursos técnicos para poder desempeñar la tarea», asegura el experto de Microsoft.

«Sin estándares claros, los datos pueden ser inconsistentes y difíciles de interpretar. La introducción manual de datos también es un punto de vulnerabilidad, ya que aumenta el riesgo de errores humanos como la escritura incorrecta o la omisión de información importante. Otro error común es la falta de validación de datos, lo que puede resultar en la inclusión de información incorrecta en las bases de datos», afirma Moreno, de The Valley.

Enrique Serrano, presidente de la comisión de IA y big data de Ametic, considera que en las empresas españolas la gestión del dato empezó a par-

tir de 2018, cuando comenzaron a crearse áreas de científicos de datos para lanzar los primeros proyectos de analítica avanzada. Según los últimos datos del Observatorio Nacional de Tecnología y Sociedad de la Información de 2023, solo el 11,8% de las empresas de más de 10 empleados ya usan IA en sus procesos, «lo cual es un indicador muy bajo, habida cuenta de la facilidad que tienen actualmente todas las empresas, grandes y pequeñas, para adoptar soluciones de mercado en tiempo récord y con presupuestos a la medida de todos: soluciones abiertas, pago por uso o pago completo por producto final», Serrano.

Serrano afirma que la empresa española está fortaleciendo sus capacidades en soluciones de IA y 'big data', pero aún queda por recorrer, ya que todavía se está al inicio de esta tecnología. «Falta ser más disruptivos y perder el miedo a cambiar radicalmente algunos procesos de negocio introduciendo IA generativa y automatizando partes del mismo. Las previsiones que manejamos manifiestan crecimientos previstos de más del 50% en el uso de las IA por las organizaciones en los próximos meses», comenta.