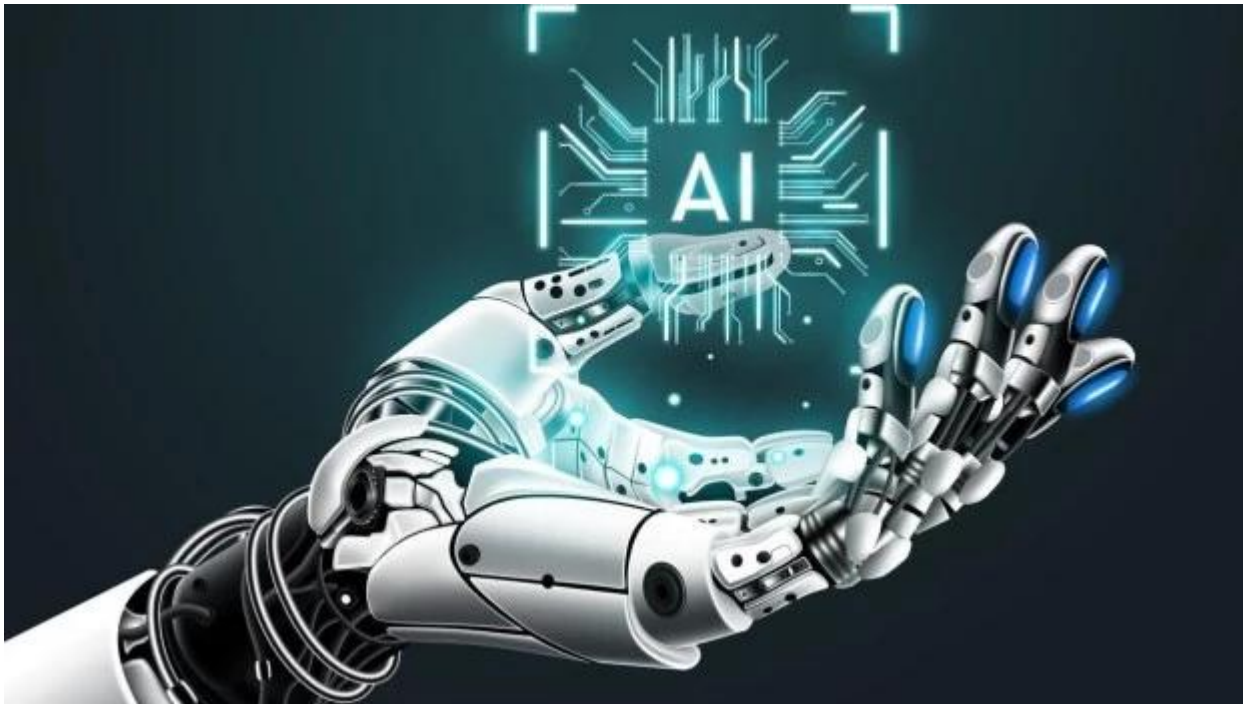


Del problema a la solución: cómo hacer que la IA generativa sea más sustentable

27/03/2025



Para las empresas que dependen de la inteligencia artificial generativa, la sustentabilidad energética ha pasado a ser una necesidad estratégica. La creciente demanda de procesamiento y almacenamiento de datos ha llevado a las organizaciones a replantear sus infraestructuras para reducir su impacto ambiental y optimizar el uso de la energía. Esto no solo implica cambios en el diseño de los centros de datos, sino también la adopción de nuevas tecnologías que permitan un mejor rendimiento con menor consumo.

En el primer artículo de esta serie dedicada al informe TMT Predictions 2025 de Deloitte, analizamos el impacto del consumo energético de la IA generativa y el desafío que representa para las empresas. Ahora, exploramos las soluciones

y estrategias que pueden hacer que los centros de datos sean más eficientes y sustentables.

Uno de los principales desafíos en este sentido es la optimización del hardware. Los procesadores utilizados en la IA generativa han evolucionado rápidamente en el último tiempo, pero su eficiencia energética aún es un punto crítico. Según Deloitte, la densidad de potencia en los racks de servidores ha aumentado velozmente: en 2023 un rack de alto rendimiento consumía aproximadamente 36 kW, pero se espera que esta cifra alcance los 50 kW en 2027. Esta tendencia ha llevado a un cambio en las estrategias de enfriamiento, con una adopción creciente de la refrigeración líquida y el enfriamiento por inmersión para mejorar la disipación térmica sin aumentar el consumo energético.

Además del hardware, el software también juega un papel clave en la eficiencia energética. Optimizar los modelos de IA generativa para que utilicen menos recursos computacionales sin comprometer el rendimiento puede reducir drásticamente el consumo energético. Estrategias como el pruning de modelos (eliminar ciertas conexiones dentro de una red neuronal para reducir su tamaño y mejorar su eficiencia pero sin afectar su rendimiento), la cuantización de estas redes y el uso de técnicas de inferencia más eficientes están ayudando a disminuir la carga de trabajo sin perder precisión.

Otro aspecto esencial es la transición hacia fuentes de energía sustentables. Los centros de datos han comenzado a migrar hacia un modelo de consumo basado en energías renovables, aprovechando fuentes como la solar, eólica e hidroeléctrica. Según el informe de Deloitte, en 2024 más del 60% de los nuevos centros de datos en América del Norte y Europa estuvieron diseñados para operar con electricidad proveniente de fuentes limpias. Algunos gigantes tecnológicos, como Google y Microsoft, ya han establecido objetivos para operar con un 100% de energía libre de carbono en los próximos años.

La ubicación de los centros de datos también influye en su impacto ambiental. Cada vez más empresas están considerando construir estas instalaciones en regiones con climas fríos, como Escandinavia o Canadá, donde las temperaturas más bajas permiten un enfriamiento natural y reducen la necesidad de sistemas de aire acondicionado de alto consumo energético.

Como se ve, el futuro de la IA generativa dependerá de la capacidad de las empresas para equilibrar innovación y eficiencia. Es que la sostenibilidad ha pasado de ser una tendencia entre muchas otras, a un factor clave para la viabilidad y el crecimiento de la industria tecnológica. Adoptar centros de datos sustentables no solo es una medida responsable con el medio ambiente, sino también una ventaja competitiva en un mundo donde la optimización de recursos será cada vez más determinante.