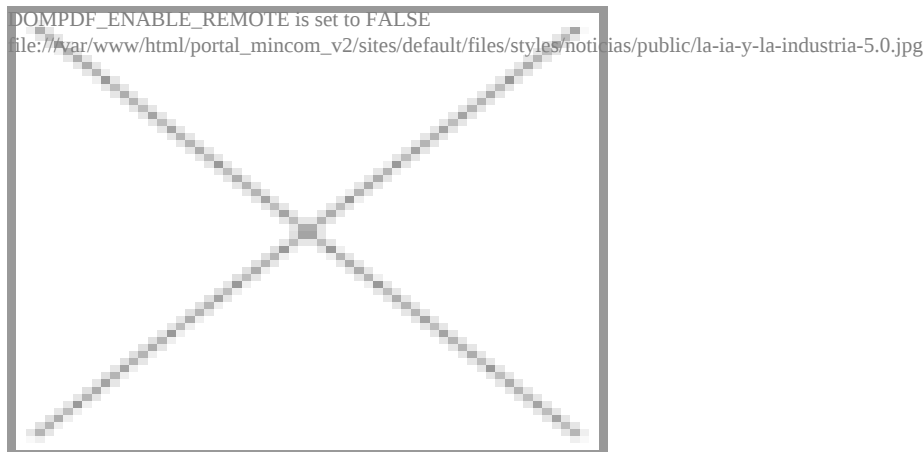




Source:
ComputerWorld

En la feria CES de este año, el concepto de ‘fábrica de IA’ se ha utilizado abundantemente en diferentes contextos. ¿Se trata de una fábrica desde el punto de vista literal? ¿De un centro de datos? ¿De una plataforma de software? ¿O es algo más?

El término ‘fábrica de IA’ ha aparecido en múltiples presentaciones, ruedas de prensa y entrevistas con ejecutivos en la feria Consumer Electronics Show (CES) de este año, que tuvo lugar la pasada semana en Las Vegas.

El concepto parece tener diferentes significados para diversas personas. “El término se utiliza de múltiples maneras, a menudo de forma imprecisa”, afirmó Alex Cordovil, director de investigación de Dell’Oro Group. En conversaciones con operadores, hiperescalares y proveedores de infraestructura, dijo que la forma principal en que se emplea es para referirse a un ‘centro de datos especializado’.

Así es también como el consejero delegado de Nvidia, Jensen Huang, utilizó el término. “La razón por la que no lo llamo centro de datos es que un centro de datos es un lugar donde la gente almacena datos”, afirmó Huang durante una conferencia magistral en el CES.

“Estas fábricas de IA son completamente diferentes de los centros de datos”, indicó Roland Busch, director ejecutivo de Siemens, a los periodistas presentes en la feria. “Consumen más energía, necesitan refrigeración líquida y no se pueden controlar con los controladores normales de los edificios, sino que requieren controles de nivel industrial”.

Bill Vass, director de tecnología de Booz Allen Hamilton, está de acuerdo con esa definición concreta. “Una fábrica de IA es un centro de datos diseñado específicamente para hardware específico de IA”, explicó. Por ejemplo, puede tener hormigón especial para soportar el peso físico de los racks de servidores y utiliza IA, gemelos digitales y simulaciones para optimizar la energía, la refrigeración y la ubicación de los servidores.

Otros, sin embargo, parecen utilizar el término para referirse a algo más pequeño que un centro de datos, refiriéndose más bien a los servidores, el software y otros sistemas utilizados para ejecutar la IA. Por ejemplo, la fábrica de IA de AWS es una combinación de hardware y software que se ejecuta en las instalaciones, pero que es gestionada por la compañía y viene con servicios suyos como Bedrock, redes, almacenamiento y bases de datos y seguridad.

En Lenovo, las fábricas de IA parecen ser servidores empaquetados diseñados para ser utilizados para la IA. “Estamos considerando una arquitectura con un número fijo de racks, que funcionan juntos como un solo diseño”, afirmó Scott Tease, vicepresidente y director general de IA y computación de alto rendimiento del grupo de soluciones de infraestructura de Lenovo. ¿Ese número de racks? Desde un solo rack hasta cientos, según declaró a COMPUTERWORLD. Cada rack es un poco más grande que un frigorífico, viene completamente montado y, a menudo, está totalmente preconfigurado para el caso de uso del cliente. “Una vez que llega a las instalaciones del cliente, nuestro personal de servicio se encarga de conectarlo a la red eléctrica y a la red”, explicó Tease.

Para otros, el concepto de fábrica de IA se centra más en el software. Algunos expertos utilizan la expresión en el sentido de la arquitectura tecnológica de la IA. Thomas H. Davenport y Randy Bean, de MIT Sloan Management Review, la definen como una combinación de “plataformas tecnológicas, métodos, datos y algoritmos desarrollados previamente que permiten crear sistemas de IA de forma rápida y sencilla”. “No poder construir sobre una base establecida hace que sea más caro y lleve más tiempo crear IA a gran escala”, dicen.

La Comisión Europea también tiene una definición amplia del término; prevé ecosistemas que incluyen hardware de TI, talento humano y datos, siendo el producto final modelos y aplicaciones de IA. La institución también habla de “gigafactorías de IA”, gigantescos centros de datos dedicados a entrenar modelos de IA de última generación. En 2025, la UE asignó 20.000 millones de euros para crear hasta cinco gigafactorías de IA.

Por último, Deloitte define una fábrica de IA como una combinación de hardware, software y servicios que permiten el ciclo de vida de la IA. “El producto principal es la inteligencia, medida por el rendimiento de los tokens”, según Nicholas Merizzi, director de Deloitte Consulting y responsable de la oferta de infraestructura de IA de la empresa en Estados Unidos. Al igual que una fábrica tradicional recibe materias primas y genera un producto acabado, una fábrica de IA recibe energía, datos y grandes modelos de lenguaje (LLM) y produce conocimientos, afirma.

Dada la repentina omnipresencia del término, es importante tener claro a qué se refiere un proveedor cuando vende una fábrica de IA. ¿Se refiere al tipo metafórico? ¿A un edificio real? ¿A una serie de racks de servidores? Es el tipo de malentendido que podría costarle miles de millones a una empresa si invierte en el concepto equivocado.

Disponible en:

<https://www.computerworld.es/article/4115478/que-es-exactamente-una-fabr...> [1]

Links

[1] <https://www.computerworld.es/article/4115478/que-es-exactamente-una-fabrica-de-ia.html>