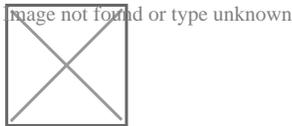


Astronómico costo energético de la inteligencia artificial (Italia no está preparada)

Por: Federico Fubini. 20/03/2024

A todo el mundo le pasa que tiene un momento en el que dice: “¿Cómo no me he dado cuenta de esto antes?”. Me ocurrió en el Foro Económico Mundial de Davos, cuando un hombre llamado Toshihaki Higashihara se levantó en una sala abarrotada. Es el presidente del conglomerado japonés Hitachi, y dijo algo simple y terrible en Davos: citó una estimación según la cual en 2050 los centros de datos informáticos (“centros de datos” o “nubes”) necesitarán “mil veces más electricidad que hoy” debido a su consumo “asociado al funcionamiento de la inteligencia artificial” (IA). Y no es una cuestión remota: está sobre nosotros.

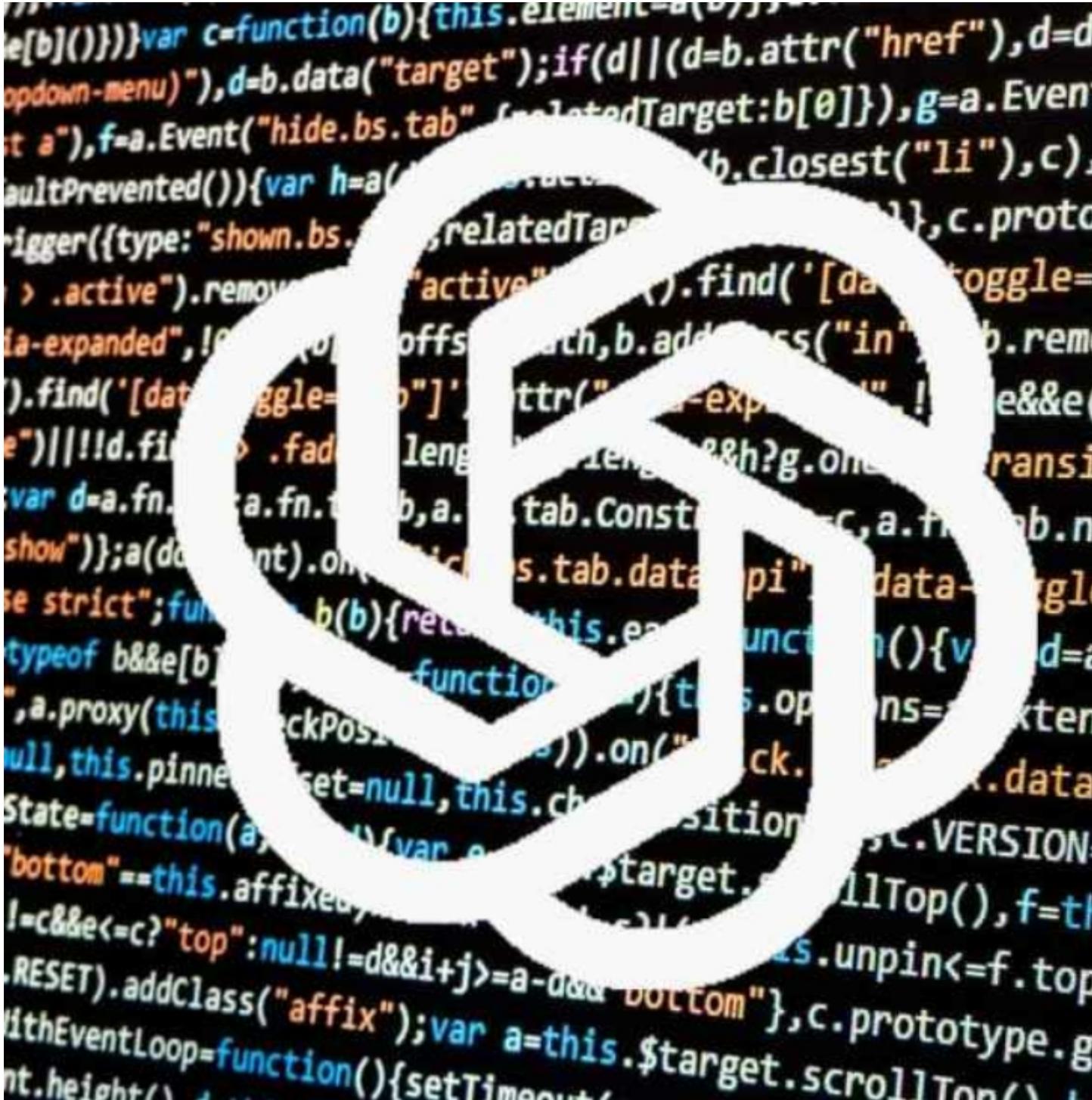


El lunes, Renato Mazzoncini, administrador de la empresa lombarda de servicios de red A2A, dijo que su compañía tendrá que triplicar la potencia eléctrica en Milán para soportar picos de consumo muy superiores. ¿Por qué? Porque una sola búsqueda en ChatGPT, la plataforma de inteligencia artificial de OpenAI, “consume tres veces más energía que Google para dar la misma respuesta”.

Las implicaciones para el medio ambiente o para las opciones ante las que tal revolución pone a cada país son gigantescas: ¿debemos aumentar exponencialmente nuestro consumo de electricidad -con todo lo que ello conlleva para las fuentes de energía y el clima- o es mejor aceptar un (más o menos) feliz decrecimiento en el que perdemos terreno frente a países que utilizan más la tecnología y son, por tanto, más productivos, más rápidos, más ricos, más capaces de atraer incluso a nuestros jóvenes mejor formados? ¿Nos estamos dando cuenta de que nos encontramos en esa encrucijada?

Confieso que no soy especialista en IA, así que me apoyo en quienes saben infinitamente más. En una conferencia del Aspen Institute Italia celebrada en

Venecia hace dos días, Alberto Sangiovanni-Vincentelli, profesor de Ingeniería Eléctrica e Informática de la Universidad de California en Berkeley, resumió lo que está ocurriendo con una comparación: “En perspectiva”, dijo, “el uso de servidores para entrenar redes neuronales, que son sólo una parte de los sistemas de inteligencia artificial, consume tanta electricidad como Suiza. Puede que en el futuro tengamos nuevas arquitecturas que consuman menos”, observó Sangiovanni-Vincentelli, “pero ese es el panorama actual”.



Jensen Huang, consejero delegado del gigante estadounidense de los microchips Nvidia, predice que la revolución de la IA en los próximos cinco años hará que se duplique el valor de las inversiones en centros de datos en todo el mundo, pasando

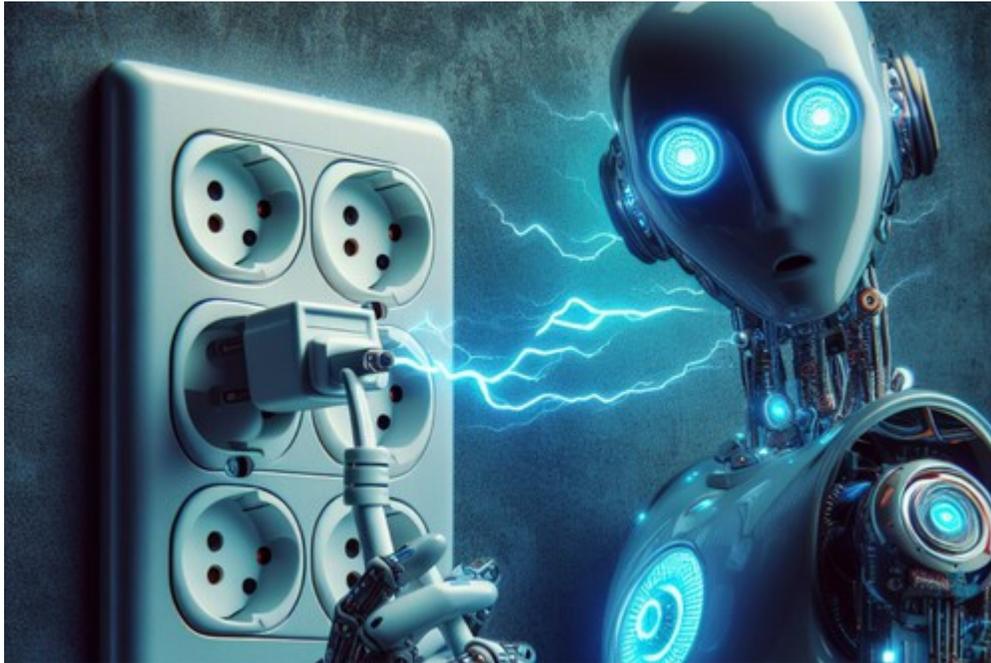
de 1 billón de dólares a 2 billones. Un centro de datos es un gran centro informático formado por miles de ordenadores que reciben, almacenan y procesan los datos producidos o procesados por nuestros ordenadores: desde el más pequeño del portátil de un adolescente hasta los sistemas digitales de grandes empresas, gobiernos o centros de investigación. La inteligencia artificial utiliza los centros de datos con especial intensidad porque sus llamados “grandes modelos lingüísticos”, básicamente los sistemas con los que las plataformas de IA predicen las respuestas a las preguntas que reciben, se forman sobre cantidades colosales de datos para hacer “inferencias” sobre los resultados que estadísticamente parecen más probables. El mero hecho de crear esa inteligencia artificial “generativa” consume mucha energía.

Así, cuando le pides, por ejemplo, a ChatGPT que te escriba un correo electrónico para solicitar un empleo, establecer un contacto comercial o le pides que escriba tu próximo trabajo de italiano sobre Leopardi, esa petición desencadena un nuevo y enorme consumo de energía: a través de los centros de datos, se remonta a millones o miles de millones de ejemplos que ya están en la red y te dará la respuesta que parezca matemáticamente más probable.

Los centros de datos de inteligencia artificial funcionan con semiconductores cada vez más pequeños desde hace décadas. Hoy en día, los transistores más avanzados de un chip tienen un diámetro de dos nanómetros, es decir, dos milmillonésimas partes de un metro: no se ven a simple vista. La llamada “ley de Gordon Moore” -de uno de los fundadores de la industria estadounidense de semiconductores en los años 60- predice que el progreso en la potencia de los microchips, es decir, en su eficiencia y su capacidad para funcionar con menos energía, será exponencial: el doble en dos años. Cada vez cabrán más transistores en un circuito integrado. Pero la revolución de la IA está rompiendo esa “ley”, o más bien inutilizándola involuntariamente.

La brutal avidez de la IA por la infraestructura digital y la energía está aumentando las necesidades de los centros de datos y el consumo de energía de forma aparentemente incontrolable. Giorgio Metta, director científico del Instituto Italiano de Tecnología, me dice: “Con la misma capacidad de cálculo, el consumo de los microchips disminuye gracias a la reducción del tamaño de los transistores: cuanto más pequeños son, menos consumen. Pero es evidente que hemos entrado en una fase en la que el aumento de las necesidades informáticas se acelera más deprisa de lo que podemos hacer chips más pequeños y eficientes desde el punto de vista

del consumo energético”.



Hoy en día, un centro de datos típico en Estados Unidos consume diez veces más megavatios que hace diez años, y la necesidad de IA se multiplicará por cinco de aquí a 2025, como peso relativo en el consumo total de los centros de datos. Así pues, para integrar la inteligencia artificial en sus sistemas, en sus empresas, en su sanidad, en su administración, en sus capacidades de defensa o en la gestión de sus bancos, un país necesita un programa asesino. Debe invertir en infraestructuras de centros de datos, de lo contrario se verá obligado a almacenar sus datos más sensibles y estratégicos en centros de otros países: nunca se sentiría realmente seguro.

Tiene que averiguar de dónde sacar nuevas fuentes de energía para alimentar esos centros de datos, y tiene que averiguar cómo hacerlo de una manera que no alimente las emisiones de efecto invernadero: sin duda no podrá abrir nuevas centrales eléctricas de carbón y probablemente tampoco de gas, ni podrá pagar la factura de comprar toda esa energía en el extranjero. También tiene que aceptar que las energías renovables no garantizan la cantidad y estabilidad de flujo necesarias para hacer funcionar cientos, miles de centros de cálculo a flujo continuo.

Luego tiene que pensar en las autopistas por las que viajarán esos datos: las

actuales no resistirían y serían fácilmente atacadas por piratas de la red. Giorgio Metta, director científico del IIT, me dice: “Para conectar entre sí algunos de nuestros centros de investigación, hemos empezado a invertir sumas importantes para equiparnos con fibra óptica dedicada de hasta 400 Gigabits por segundo, porque la conectividad ordinaria no basta. Los datos que hay que mover entre nuestros centros de Génova pueden ser un millón de veces más voluminosos que una película o un vídeo que hay que descargar para un usuario doméstico”.

Algunas personas han empezado a pensar estratégicamente en el futuro, porque la revolución de la IA cambia las estructuras de un sistema. Sam Altman, el CEO de OpenAI que de hecho desencadenó la actual ola de inteligencia artificial “generativa”, se ha dado cuenta de ello: está invirtiendo en la producción directa de microchips, porque no quiere depender de los de Nvidia; y también está invirtiendo en dos start-ups de energía nuclear civil (una de fisión, la otra de fusión) para asegurarse la energía que necesitará su empresa.



Altman afirma: “El mundo necesita más infraestructura de IA -más fábricas de chips, más energía, más centros de datos- de la que tenemos previsto construir hoy”. Francia está poniendo en marcha un programa de catorce nuevas centrales nucleares -que también atrae cientos de millones de euros de start-ups del sector nacidas en Italia- en gran parte para alimentar los centros de datos que está construyendo y competir así en IA. ¿Y nosotros? En Italia podemos decidir que la IA no nos interesa.

Un informe coproducido para el think tank Teha por Microsoft (que es parte interesada, como accionista de referencia de Sam Altman, en OpenAI) predice que la integración de la inteligencia artificial en las empresas, la sanidad y la

administración en Italia supondría un aumento de la capacidad asistencial y un incremento de 312.000 millones en el valor añadido de la economía italiana -con el mismo número de horas trabajadas- de aquí a 2040. Sería una forma de compensar en parte la caída de la capacidad productiva que supondrá la pérdida de casi cuatro millones de trabajadores debido a la demografía.

Pero, efectivamente, podemos decidir que no nos importa. Que queremos seguir con las tecnologías del siglo XX, retrocediendo una marcha o dos. Sería legítimo: elegimos estar entre lo que Jensen Huang, de Nvidia, llama “los que no tienen” de la nueva revolución, frente a los “que tienen”, los ricos que lo tienen. En lugar de eso, queremos tenerlo. Lo queremos todo: este mundo y aquel mundo. Queremos permanecer en las “grandes ligas” del mundo avanzado. Pero, con muy pocas y dispersas excepciones, no estamos haciendo nada al respecto: ni siquiera nos estamos dando cuenta de la visión y el coraje que haría falta.

**Periodista italiano del Corriere della Sera de Milan*

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ](#)

Fotografía: Editores

Fecha de creación

2024/03/20