

## Quanta elettricità consuma l'intelligenza artificiale? Infodata 20.07.2024

<https://www.infodata.ilsole24ore.com/2024/07/20/quanta-elettricit%C3%A0-consuma-lintelligenza-artificiale/>

È la domanda a cui i chatbot non sanno rispondere. In modo preciso. E non solo loro, in questo momento i Big dell'AI gen sono alquanto timidi quando si discute di sostenibilità e impronta energetica di questa energia. Tanto che non esistono studi che mettono in relazione Gemini Advanced, Gpt4 o Microsoft Copilot con il loro consumo energetico per singola operazione. Un recente articolo di *The Verge* ha provato a raccogliere quello che sappiamo finora. C'è da dire che in brevissimo tempo i modelli di linguaggio di grandi dimensioni (LLM) sono stati sottoposti a cure dimagranti per diventare meno energivori. E che i nuovi LLM saranno meno "generalisti" e più snelli. Va anche detto l'addestramento di un modello su cui si stanno concentrando numerosi studi è solo una parte del quadro. Dopo la creazione, un sistema viene distribuito ai consumatori che lo utilizzano per generare output, un processo noto come "inferenza". Ciò premesso ecco alcuni numeri estratti dall'articolo.

1.300 MWh

Quanto costa addestrare Gpt-3?

Si stima che l'addestramento di un modello linguistico di grandi dimensioni come GPT-3, ad esempio, utilizzi poco meno di 1.300 megawattora (MWh) di elettricità; circa la stessa energia consumata annualmente da 130 case americane. Per contestualizzare il tutto, lo streaming di un'ora di Netflix richiede circa 0,8 kWh (0,0008 MWh) di elettricità. Ciò significa che dovrete guardare 1.625.000 ore per consumare la stessa quantità di energia necessaria per addestrare GPT-3.

2,907 kWh

Quanto consuma generare le immagini? Secondo un studio pubblicato a dicembre firmato da Sasha Luccioni, ricercatrice presso la società di intelligenza artificiale franco-americana Hugging Face con ricercatori della Carnegie Mellon University il consumo di energia è molto più alto nella produzione di immagini rispetto a quella di testo. Il metodo adottato è stato quello di eseguire test su 88 modelli diversi ripetendo i prompt mille volte e stimando il costo energetico. Il testo richiede 0,002 kWh per classificare e 0,047 kWh per generare testo. Se usiamo come paragone un'ora di streaming di Netflix equivalgono all'energia consumata guardando rispettivamente nove secondi o 3,5 minuti. Le cifre erano notevolmente maggiori per i modelli di

generazione di immagini, che utilizzavano in media 2,907 kWh per 1.000 inferenze. Come osserva il documento, lo smartphone medio utilizza 0,012 kWh per caricarsi, quindi la generazione di un'immagine utilizzando l'intelligenza artificiale può utilizzare quasi la stessa quantità di energia della ricarica dello smartphone.

134 TWh

Quanto consuma il settore dell'intelligenza artificiale? Entro il 2027 l'intera industria dell'intelligenza artificiale **potrebbe consumare tra gli 85 e i 134 terawattora all'anno**. La stima arriva da Alex de Vries, dottorando presso la VU di Amsterdam che su Joule ha calcolato l'impatto in base alle specifiche energetiche dell'hardware di Nvidia (e le relative previsioni di vendita) che rappresenta il 95% del mercato dei chip dell'AI gen. Una seconda stima da prendere in considerazione è quella dell'Agenzia Internazionale dell'Energia che in un suo report ha calcolato la domanda di elettricità nei data center. **L'agenzia afferma che l'attuale consumo energetico dei data center si attesta a circa 460 terawattora nel 2022 e potrebbe aumentare tra i 620 e i 1.050 TWh nel 2026, equivalenti al fabbisogno energetico rispettivamente di Svezia o Germania**. A spingere i consumi le criptovalute e l'intelligenza artificiale.

Intelligenza artificiale: l'energia e l'acqua di cui ha bisogno di David Berreby  
21 Febbraio 2024 <https://www.rivistaenergia.it/2024/02/intelligenza-artificiale-lenergia-e-lacqua-di-cui-ha-bisogno/>

In attesa del numero di marzo di ENERGIA in cui parleremo, tra le altre cose, di Intelligenza Artificiale, pubblichiamo un'analisi sul suo consumo di energia ed acqua. **L'intelligenza artificiale generativa utilizza enormi quantità di energia per il calcolo e l'archiviazione dei dati e milioni di litri d'acqua per raffreddare le apparecchiature nei data center**. I legislatori e i regolatori negli Stati Uniti e nell'Unione Europea stanno iniziando a chiedere responsabilità. ... **1.000 terawatt, il consumo di elettricità previsto dei data center nel 2026, pari circa al consumo totale del Giappone**

L'IA può funzionare su molti tipi dispositivi. Una sua versione semplice in grado di correggere l'ortografia dei messaggi funzionerà sugli smartphone. Ma **il tipo di IA che la maggior parte delle persone desidera utilizzare è troppo grande per la maggior parte dei dispositivi personali**, afferma Dodge. "I

modelli in grado di scrivere una poesia o un'e-mail sono molto grandi", afferma. "Le dimensioni sono vitali affinché abbiano queste capacità".

Le grandi IA devono eseguire un numero enorme di calcoli molto rapidamente, di solito su unità di elaborazione grafica specializzate: processori originariamente progettati per calcoli intensivi che elaborino grafiche sugli schermi dei computer. Rispetto ad altri chip, le GPU (graphics processing unit) sono più efficienti dal punto di vista energetico per l'intelligenza artificiale e lo sono ancora di più quando vengono eseguite in grandi "data center cloud", edifici appositi pieni di computer dotati di tali chip. Quanto più grande è il data center, tanto più efficiente può essere dal punto di vista energetico. I miglioramenti nell'efficienza energetica dell'intelligenza artificiale negli ultimi anni sono in parte dovuti alla costruzione di "hyperscale data center", che contengono molti più computer e possono espandersi rapidamente. Laddove un tipico data center cloud occupa circa 100.000 piedi quadrati, un centro su vasta scala può essere di 1 o addirittura 2 milioni di piedi quadrati.

Le stime del numero di data center cloud in tutto il mondo vanno da circa 9.000 a quasi 11.000. Altri sono in costruzione. L'Agenzia internazionale per l'energia (IEA) prevede che il consumo di elettricità dei data center nel 2026 sarà il doppio di quello del 2022: 1.000 terawatt, più o meno equivalente all'attuale consumo totale del Giappone.

Tuttavia, come esempio di problema nel modo in cui vengono misurati gli impatti dell'IA, la stima dell'AIE include tutta l'attività del data center, che va oltre la sola IA includendo molti aspetti della vita moderna. Anche l'esecuzione dell'interfaccia di Amazon, la distribuzione dei video di Apple TV, l'archiviazione delle e-mail di milioni di persone su Gmail e il "mining" di Bitcoin vengono eseguiti dai data center. (Altri rapporti IEA escludono le operazioni crittografiche, ma raggruppano comunque tutte le altre attività dei data center.)

La maggior parte delle aziende che gestiscono data center non rivela quale percentuale dei loro consumi energetici sia imputabile all'IA. L'eccezione è Google, che afferma che il "machine learning" – la base per un'intelligenza artificiale simile a quella umana – rappresenta poco meno del 15% del consumo energetico dei suoi data center.

Un'altra complicazione è il fatto che l'intelligenza artificiale, a differenza del mining di Bitcoin o dello shopping online, può essere utilizzata per ridurre

l'impatto dell'umanità. L'IA può migliorare i modelli climatici, trovare modi più efficienti per realizzare tecnologie digitali, ridurre gli sprechi nei trasporti e tagliare in altro modo l'uso di carbonio e acqua. Ad esempio, una stima ha rilevato che le case smart connesse all'IA potrebbero ridurre il consumo di CO<sub>2</sub> delle famiglie fino al 40%. E un recente progetto di Google ha scoperto che un'IA in grado di elaborare rapidamente i dati atmosferici consentirebbe ai piloti di aerei di percorrere traiettorie di volo che lasceranno il minor numero di scie. ...