

# DIGITALIZATION & DECARBONIZATION

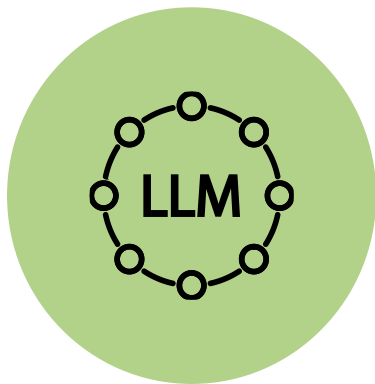
## Report 2024

Le sfide per la  
decarbonizzazione dell'AI:  
consumi, emissioni e strategie  
di mitigazione

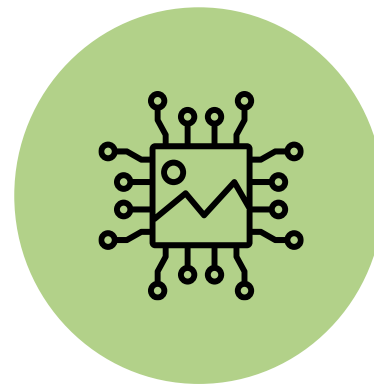
# Introduzione all'AI

## I principali modelli di Generative AI

L'AI generativa è una branca dell'intelligenza artificiale che permette di **creare nuovi contenuti** partendo da dati esistenti.



Large Language Models (LLM)



Multimodal or Image Generative Models

# Il ciclo di vita dei modelli AI

## Addestramento e utilizzo

I modelli di AI generativa nel loro ciclo di vita sono caratterizzati da due fasi principali: il **training** e l'**inference**.

### *Training*

#### **Ottimizzazione dei parametri su enormi dataset**

Il modello analizza grandi quantità di dati aggiustando continuamente i parametri per migliorare le prestazioni.

#### **Processo di lunga durata**

Il processo di training del modello ha una durata variabile (settimane/mesi) in base al tipo di modello e dati.

#### **Consumi e emissioni elevati**

Sono necessari migliaia di kilowattora di elettricità, se generati da fonti fossili causa l'emissione di tonnellate di CO<sub>2</sub>.

### *Inference*

#### **Modello addestrato per generare output**

Il modello risponde a una domanda o genera un testo a partire da un prompt, sfruttando ciò che ha imparato durante il training.

#### **Consumo globale elevato**

L'inferenza richiede meno risorse rispetto all'addestramento, ma i modelli LLM vengono eseguiti milioni di volte al giorno.

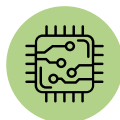
#### **Ottimizzazione dell'inferenza**

Migliorare l'efficienza dei servizi di AI che gestiscono volumi elevati di richieste in tempo reale è cruciale per ridurne i consumi.

# Impatto ambientale dell'AI

## Consumo energetico

Creare **stime accurate dell'uso di energia e delle emissioni di carbonio dei sistemi di AI** nel corso della loro vita è un'operazione **complessa** dal momento che questi calcoli dipendono da molti fattori riportati di seguito.



**Specifiche tecniche dei chip utilizzati**



**Design dei data center**



**Workload**



**Sistemi di raffreddamento**



**Software**



**Fonte dell'elettricità utilizzata**

# Impatto ambientale dell'AI

## Case Study: i consumi e le emissioni per il training

Di seguito vengono riportati i dati relativi alla fase di addestramento di tre modelli di intelligenza artificiale: **GPT-3 (Open AI)**, **Llama 2 (Meta)** e **PaLM (Google)**.

### GPT 3



175 mld di parametri  
355 h di training  
1.287 MWh consumati  
552 t CO<sub>2</sub> emesse

---

1.287 MWh corrispondono a:



Consumo **annuo** di circa  
260 famiglie



~18 viaggi a/r Milano – Roma  
con un Frecciarossa

### Llama 2



70 mld di parametri  
Ore di training **N.D.**  
688 MWh consumati  
291,4 t CO<sub>2</sub> emesse

---

688 MWh corrispondono a:



Illuminazione di un quartiere  
per circa un mese



Consumo **mensile** di un centro  
commerciale

### PaLM



540 mld di parametri  
1.526 h di training  
3.436 MWh consumati  
271,4 t CO<sub>2</sub> emesse

---

3.436 MWh corrispondono a:



Un mese di produzione di  
un grande impianto industriale



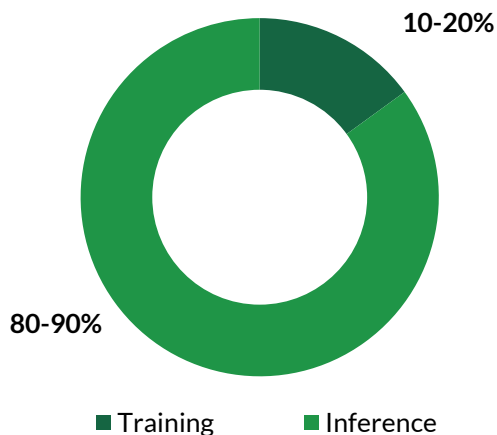
Consumo **mensile** di un quartiere  
di circa 10.000 abitanti

# Impatto ambientale dell'AI

## Consumo energetico dell'inferenza

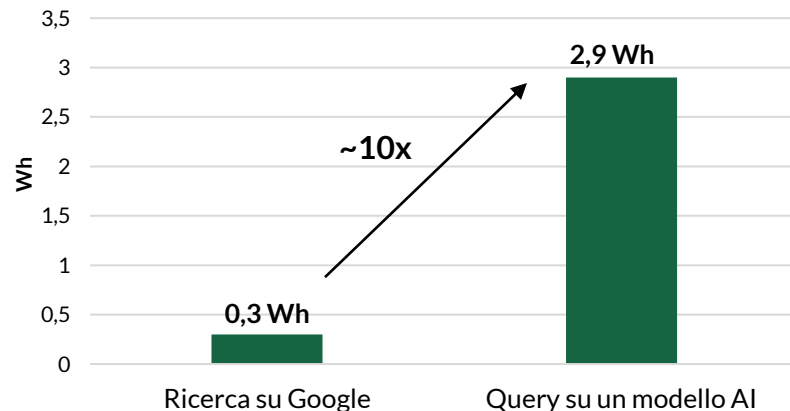
Studi dimostrano che i costi energetici dell'AI derivano principalmente dall'utilizzo dei modelli, ossia dall'inferenza, che rappresenta l'80-90% del workload nei data center.

Energia consumata dai modelli AI durante la loro vita



Fonte: Centre for Data Innovation, S&P Global.

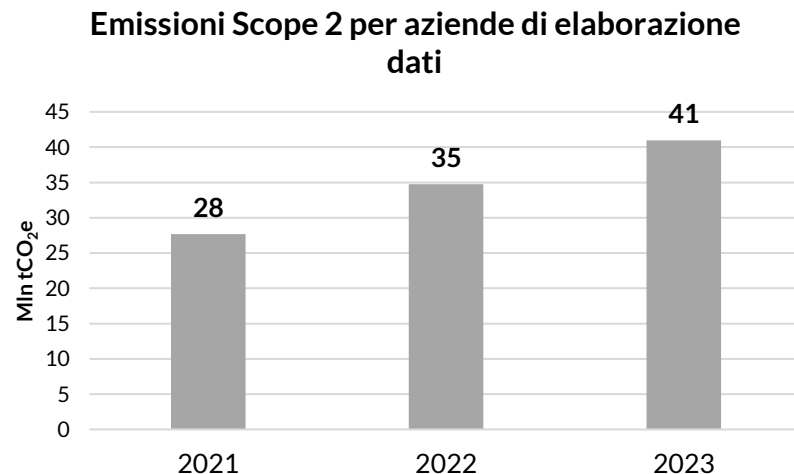
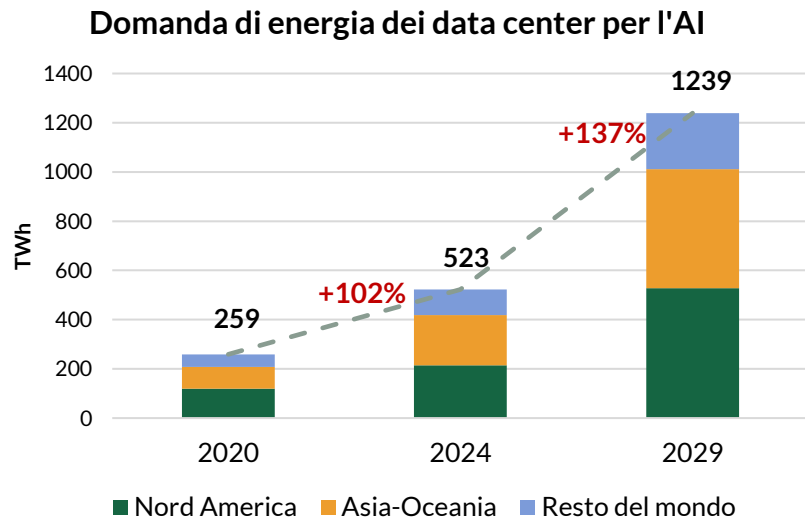
Confronto tra l'energia richiesta per una ricerca sul web e un'interazione con un modello AI



# Impatto ambientale dell'AI

## Consumo energetico ed emissioni

I dati storici indicano che la **domanda di energia dei data center è raddoppiata tra il 2020 e il 2024** e si prevede possa **più che raddoppiare (+137%) entro il 2029**. Negli ultimi anni, le emissioni di CO<sub>2</sub> indirette classificate come **Scope 2** delle aziende che gestiscono o affittano data center sono aumentate del **48% tra il 2021 e il 2023**.















Fonte: S&P Global

# Impatto ambientale dell'AI

## Le azioni di mitigazione

Efficientare il consumo energetico dei modelli di intelligenza artificiale è fondamentale, dato il loro crescente utilizzo. Diverse **tecniche algoritmiche** possono migliorare l'**efficienza energetica** mantenendo prestazioni elevate.

### Possibili interventi per l'efficientamento di un modello AI

 Pruning	 Hyperparameter Tuning	 Hardware Ottimizzato	 Efficienza dei Dati
 Quantizzazione	 Architetture Efficienti	 Gradient Checkpointing	 Federated Learning e Edge Computing
 Knowledge Distillation	 Adaptive Computation	 Batch Size e Mixed-Precision Training	 Green AI e Carbon-aware Scheduling



# Il consumo e le emissioni dell'AI

## Un modello per stimare consumi e emissioni

Il modello per la stima dei consumi e delle emissioni dell'inferenza di un'AI generativa si basa sulla simulazione di un caso d'uso reale.



Classificazione del testo  
0,002 Wh



Classificazione delle immagini  
0,007 Wh



Rilevamento degli oggetti  
0,038 Wh



Generazione di testo  
0,047 Wh



Riassunto del testo  
0,049 Wh



Generazione di immagini  
2,907 Wh

Fonte: Centre for Data Innovation.

# Il consumo e le emissioni dell'AI

## I risultati del modello

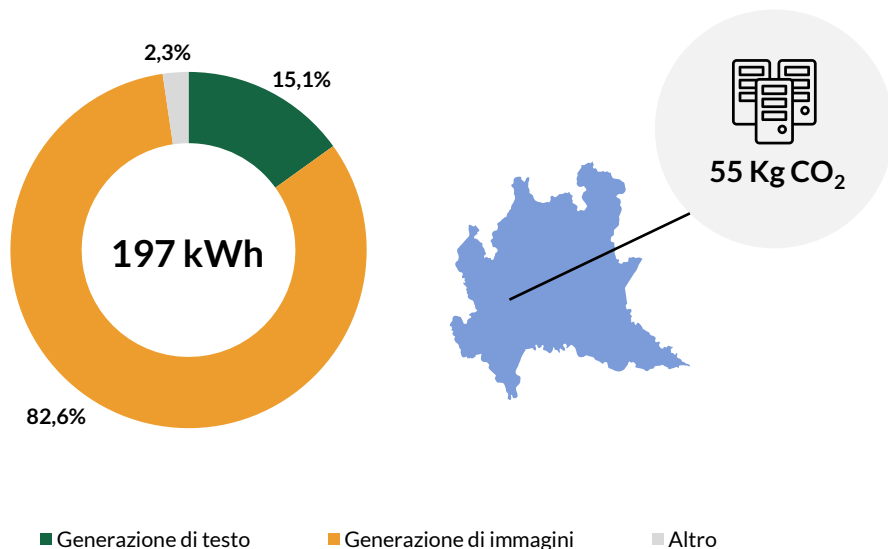
La stima dei consumi energetici e delle emissioni di CO<sub>2</sub> associati all'utilizzo di un modello AI per la stesura di 1.000 report evidenzia un impatto significativo.

L'energia totale consumata è di **197 kWh**, con emissioni pari a **55 Kg di CO<sub>2</sub>** se il data center fosse situato in Lombardia.

La maggior parte dei consumi è attribuibile alla **generazione di immagini**, seguita dalla **generazione di testo**, mentre altre attività come classificazioni, riassunto del testo e rilevamento di oggetti incidono in misura minore.

Per confronto, l'energia necessaria equivale a quella utilizzata da un'auto elettrica per percorrere circa mille chilometri.

Consumi ed emissioni associati all'utilizzo di un modello di AI generativa per la stesura di 1000 report



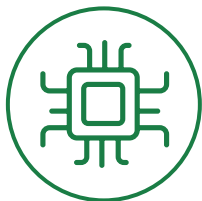
# Messaggi Chiave



L'intelligenza artificiale richiede una notevole quantità di elettricità. **Circa l'80-90% del consumo deriva dall'utilizzo dei modelli di AI generativa**, eseguiti milioni di volte al giorno. **Una singola query su un modello AI può consumare dieci volte più energia rispetto a una normale ricerca sul web.**



I consumi energetici associati ai modelli AI si traducono in emissioni di CO<sub>2</sub> che dipendono fortemente dalla fonte energetica utilizzata nei data center e dall'efficienza complessiva del modello. La stima di consumi ed emissioni di CO<sub>2</sub> nella fase di inferenza evidenzia come **l'utilizzo su larga scala di queste tecnologie comporti un impatto significativo in termini di consumi elettrici ed emissioni.**



**Per bilanciare i benefici dell'intelligenza artificiale con il suo impatto energetico, sono necessarie strategie di mitigazione su più livelli.** A livello algoritmico, riducendo i calcoli richiesti senza compromettere la precisione, e a livello hardware, utilizzando architetture **più efficienti e dispositivi ottimizzati.**