

CHATGPT



Dalla fondazione ad oggi



AI is the new (old) web3



Noah Smith



@Noahpinion

2016: AI is going to change the world!

2017: AI is going to change the world!

2018: AI is going to change the world!

2019: AI is going to change the world!

2021: Crypto is going to change the world!

2022: ...um...AI is going to change the world!



Google published attention is all you need

OpenAI founded

GPT

GPT-2

GPT-3

InstructGPT

ChatGGPT

2016

2017

2018

2019

2020

2021

2022

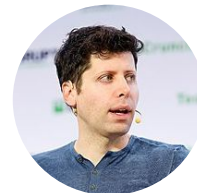
2023

OpenAI founded

Il lancio dell'azienda è
partita già con
1 Miliardo dollari



Elon Musk
TESLA



Sam Altman
OpenAI



Reid Hoffman
LinkedIn



Peter Thiel
PAYPALL



Greg Brockman
Stripe



Jessica Livingston
Ycombinator

2016

2017

2018

2019

2020

2021

2022

2023

OpenAI founded

I primi prodotti che hanno lanciato

Gymnasium 2017

ImageNet più “ampio” per il mondo della ricerca, [link di approfondimento](#)

RoboSumo 2017

robot umanoidi in “meta-apprendimento”, [link di approfondimento](#)

Debate Game 2018

Si entra nel mondo del explainable AI, insegnare alla macchina a discutere di fronte un giudice umano, [link di approfondimento](#)

OpenAI Five 2018

Gioco 5x5 dove un sistema di apprendimento rinforzato permette di giocare contro se stessi per centinaia di volte

2016

2017

2018

2019

2020

2021

2022

2023

Google published attention is all you need

Google pubblica Attention Is All You Need

Introduce:

L'architettura del trasformatore

Rete neurale che pone le basi per molti modelli di linguaggio di grandi dimensioni aprendo la strada a due concetti fondamentali:

- a) **LLM**
Large language models, avere accesso ad una miriade di dati pre-trainedi
- b) **SOTA**
State of the art, scegliere il modello migliore per fare un particolare task



Google published attention is all you need



Explain me Google paper “Attention Is All You Need”



“Attention Is All You Need” is a research paper published by Google in 2017 that introduced a new neural network architecture for sequence-to-sequence learning called the Transformer. The paper proposed a novel approach to natural language processing that eliminated the need for traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs) that were previously used in machine translation tasks.

The Transformer model was introduced as an alternative to the previous sequence-to-sequence models such as the RNN-based models, which were computationally expensive and required a lot of memory. The authors proposed a new model architecture that relies entirely on an attention mechanism for generating the output sequence from the input sequence.

The Transformer architecture consists of two main components: the encoder and the decoder. The encoder takes an input sequence of tokens and produces a sequence of hidden states, while the decoder takes the encoder's output and generates an output sequence.

2016

2017

2018

2019

2020

2021

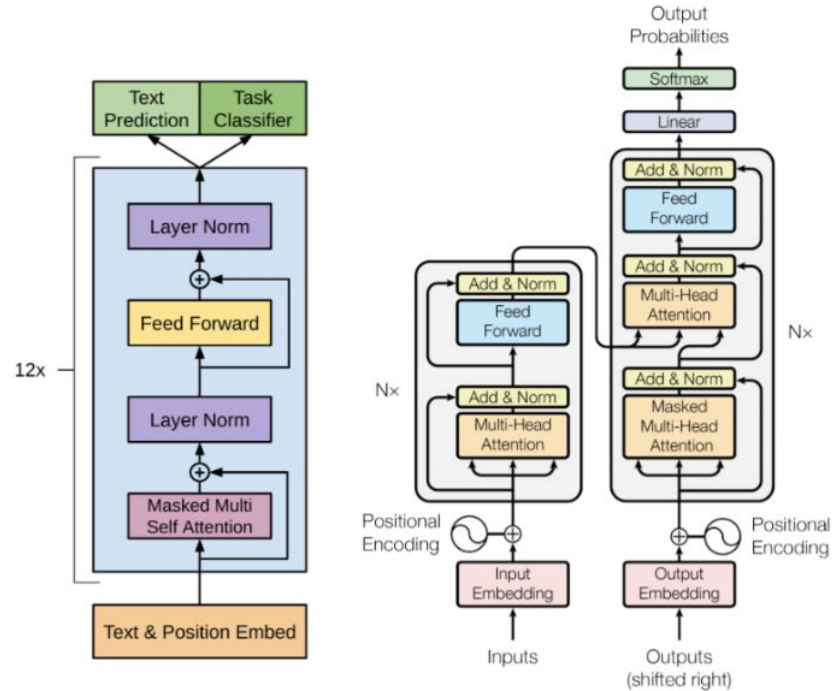
2022

2023

Come funziona

- 1) incorporamento dell'input
- 2) auto-attenzione a più teste
- 3) rete feed-forward
- 4) normalizzazione e collegamenti residui

[Per un maggior approfondimento](#)



Google published attention is all you need

L'attenzione consente al modello di concentrarsi sulle parti più rilevanti dell'input e dell'output imparando la rilevanza o la somiglianza tra gli elementi , che sono solitamente rappresentati da vettori. Se si concentra sulla stessa sequenza, si chiama auto-attenzione



Tom likes to eat apples. He eats them every day.

The diagram shows red curved arrows representing attention weights. A thick red arrow points from the word 'apples' back to 'eat', indicating a strong self-attention connection. Several thinner red arrows point from the word 'eats' to various words in the sentence, including 'eat', 'apples', 'He', 'every', and 'day', illustrating how the model's focus is distributed across the input sequence.

2016

2017

2018

2019

2020

2021

2022

2023

I prodotti che hanno lanciato nel tempo

GPT2 2019

Viene introdotto in Language Models are Unsupervised Multitask Learners, che può eseguire una serie di attività senza supervisione esplicita durante la formazione.

GPT3 2020

Viene introdotto in Language Models are Few-Shot Learners, che può funzionare bene con pochi esempi nel prompt senza messa a punto.

DALL-E 2021

è un algoritmo di intelligenza artificiale capace di generare immagini a partire da descrizioni testuali. Sviluppato da OpenAI

InstructGPT 2022

viene introdotto nei modelli linguistici di addestramento per seguire le istruzioni con feedback umano, che possono seguire meglio le istruzioni dell'utente perfezionando con il feedback umano.

ChatGPT 2022

un fratello di InstructGPT, viene introdotto in ChatGPT: Ottimizzazione dei modelli linguistici per il dialogo . Può interagire con gli umani nelle conversazioni, grazie alla messa a punto con esempi umani e all'apprendimento per rinforzo dal feedback umano (RLHF)

2016

2017

2018

2019

2020

2021

2022

2023

Da GPT a ChatGPT

Impatto sul mercato
della soluzione

In 5 giorni ChatGPT
raggiunge 1 milione di
utenti

ChatGPT Sprints to One Million Users

Time it took for selected online services
to reach one million users



* one million backers ** one million nights booked *** one million downloads

Source: Company announcements via Business Insider/LinkedIn

2016

2017

2018

2019

2020

2021

2022

2023

Vantaggio competitivo Parametri e costo

GPT2

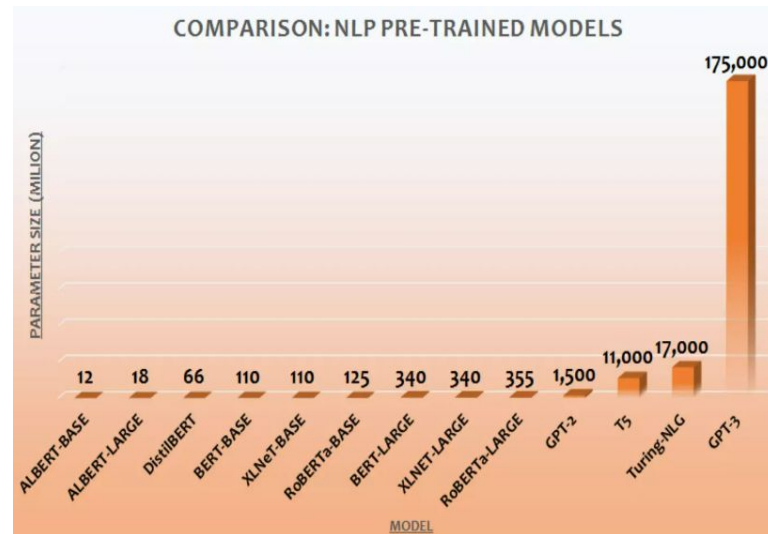
Nella prima versione di GPT era stato usato un pre-addestramento con il database di **BookCorpus** di **7000 libri unici** non pubblicati

GPT-2

GPT-2 aveva aveva **1.5 Miliardi di parametri**

GPT-3

ha **175 Miliardi di parametri** e può imparare a svolgere meglio le sue attività con pochissimi esempi senza supervisione.



2016

2017

2018

2019

2020

2021

2022

2023

Costo giornaliero per mantenere
il servizio up + di 100.000€

2016

2017

2018

2019

2020

2021

2022

2023

Il prezzo cambia a seconda del modello di linguaggio

Più modelli, ognuno con diverse capacità e fasce di prezzo.

I prezzi si valutano in 1.000 token.

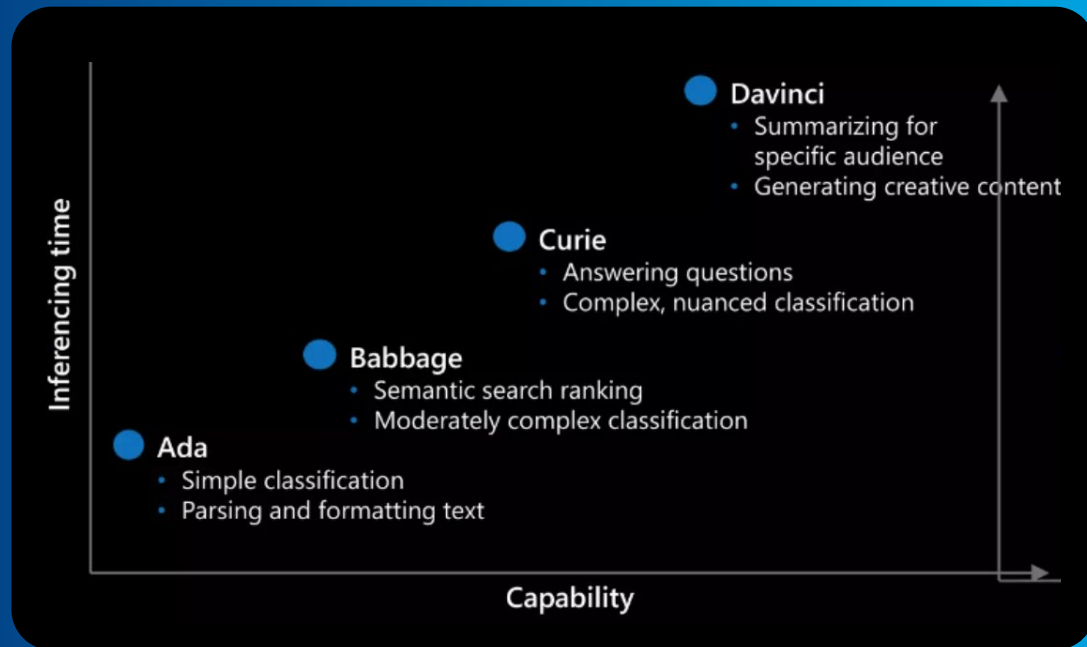
Puoi pensare ai token come pezzi di parole,
dove 1.000 token sono circa 750 parole.

Questo paragrafo per esempio vale 35 gettoni.

Model	Training	Usage
Ada	\$0.0004 / 1K tokens	\$0.0016 / 1K tokens
Babbage	\$0.0006 / 1K tokens	\$0.0024 / 1K tokens
Curie	\$0.0030 / 1K tokens	\$0.0120 / 1K tokens
Davinci	\$0.0300 / 1K tokens	\$0.1200 / 1K tokens

Quale modello scegliere?

Se sei indeciso
usa il diagramma!



Come funziona ChatGPT

The image shows a screenshot of the ChatGPT web interface with five numbered callouts:

- 1**: Points to the '+ New chat' button in the top left sidebar.
- 2**: Points to the user's input prompt: 'Explain me Google paper "Attention Is All You Need"'.
- 3**: Points to the OpenAI logo in the left sidebar.
- 4**: Points to the first paragraph of the AI's response.
- 5**: Points to the 'Regenerate response' button at the bottom of the response box.

ChatGPT's Key Features

Explain me Google paper "Attention Is All You Need"

"Attention Is All You Need" is a research paper published by Google in 2017 that introduced a new neural network architecture for sequence-to-sequence learning called the Transformer. The paper proposed a novel approach to natural language processing that eliminated the need for traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs) that were previously used in machine translation tasks.

The Transformer model was introduced as an alternative to the previous sequence-to-sequence models such as the RNN-based models, which were computationally expensive and required a lot of memory. The authors proposed a new model architecture that relies entirely on an attention mechanism for generating the output sequence from the input sequence.

Regenerate response



Mi scrivi il codice Javascript per calcolare la lunghezza dell'Italia?



Mi dispiace, ma non è possibile calcolare la lunghezza dell'Italia con il codice Javascript. La lunghezza di una nazione si riferisce alla sua estensione geografica, che è una grandezza fisica e geografica.

Tuttavia, è possibile calcolare la distanza tra due punti sulla superficie terrestre utilizzando le coordinate geografiche. Se mi fornisci i punti di inizio e fine dell'Italia, posso aiutarti a calcolare la distanza tra di essi.

Ecco un esempio di codice Javascript che calcola la distanza tra due punti sulla Terra, dato il loro latitudine e longitudine in gradi decimali:

```
function distance(lat1, lon1, lat2, lon2) {
  const R = 6371; // Raggio della Terra in chilometri
  const dLat = degToRad(lat2 - lat1);
  const dLon = degToRad(lon2 - lon1);

  const a =
    Math.sin(dLat / 2) * Math.sin(dLat / 2) +
    Math.cos(degToRad(lat1)) * Math.cos(degToRad(lat2)) *
    Math.sin(dLon / 2) * Math.sin(dLon / 2);

  const c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1 - a));
  const d = R * c; // Distanza in chilometri
  return d;
}

// esempio di utilizzo
const ItaliaStart = {
  lat: 41.88837,
  lon: 12.508378
};

const ItaliaEnd = {
  lat: 37.090000,
  lon: 15.500000
};

const distanzaItalia = distance(ItaliaStart.lat, ItaliaStart.lon, ItaliaEnd.lat, ItaliaEnd.lon);

console.log('La distanza tra i due punti è di ' + distanzaItalia + ' km');
```

Nota che questo codice calcola la distanza tra due punti sulla Terra in linea d'aria, senza tenere conto di eventuali ostacoli o deviazioni nel percorso, quindi i risultati potrebbero non essere precisi al 100%. Inoltre, la lunghezza dell'Italia può variare a seconda della definizione di "lunghezza" utilizzata, ad esempio, se si tiene conto solo della costa o anche dell'entroterra.



Utilizza il precedente codice e calcolami la distanza tra Prato e Firenze



Certo, ecco il codice per calcolare la distanza tra Prato e Firenze in Javascript utilizzando la funzione 'distance' definita nell'etichetta [Regenerate response](#)

```
function distance(lat1, lon1, lat2, lon2) {
  const R = 6371; // Raggio della Terra in chilometri
  const dLat = degToRad(lat2 - lat1);
  const dLon = degToRad(lon2 - lon1);

  const a =
    Math.sin(dLat / 2) * Math.sin(dLat / 2) +
    Math.cos(degToRad(lat1)) * Math.cos(degToRad(lat2)) *
    Math.sin(dLon / 2) * Math.sin(dLon / 2);

  const c = 2 * Math.atan2(Math.sqrt(a), Math.sqrt(1 - a));
  const d = R * c; // Distanza in chilometri
  return d;
}

// esempio di utilizzo
const PratoStart = {
  lat: 43.775000,
  lon: 11.145000
};

const FirenzeEnd = {
  lat: 43.775000,
  lon: 11.145000
};

const distanzaItalia = distance(PratoStart.lat, PratoStart.lon, FirenzeEnd.lat, FirenzeEnd.lon);

console.log('La distanza tra i due punti è di ' + distanzaItalia + ' km');
```

Questo codice restituirà la distanza in chilometri tra Prato e Firenze, calcolata come la distanza in linea d'aria tra le coordinate geografiche dei due comuni.



Quindi quanti KM sono?



La distanza in linea d'aria tra Prato e Firenze è di circa 16.5 chilometri, quindi il codice che ho fornito dovrebbe restituire un valore intorno a questo numero. Tuttavia, tieni presente che il risultato è solo un'approssimazione della distanza effettiva tra i due punti sulla superficie terrestre e potrebbe non riflettere la distanza reale percorsa su strada o su altre vie di comunicazione.



Arafin Siddique Nabil

30 anni



Founder di Taak srls → Media company



Partner di Hastega → Software house



CEO di Awhy srl → Customer care platform AI based

Amante della corsetta all'aperto, appassionato di progetti web in particolari soluzioni SaaS.

Link utili

<https://bootcamp.uxdesign.cc/how-chatgpt-really-works-explained-for-non-technical-people-71efb078a5c9>

<https://research.aimultiple.com/large-language-model-training/>

<https://research.aimultiple.com/gpt/#what-is-gpt-3::~:~:text=A%20training%20run%20is%20estimated%20to%20cost%20%244.6%20million%20and%20it%20takes%20numerous%20training%20runs%20to%20fine%20tune%20the%20training%20process.>

<https://research.aimultiple.com/large-language-models/>

[https://smartstrategy.eu/intelligenza-artificiale/cosa-sono-i-transformer-e-come-vengono-utilizzati-nellelaborazione-del-linguaggio-naturale/#::~:~:text=In%20sintesi%2C%20i%20transformers%20rappresentano.del%20linguaggio%20naturale%20\(NLP\).](https://smartstrategy.eu/intelligenza-artificiale/cosa-sono-i-transformer-e-come-vengono-utilizzati-nellelaborazione-del-linguaggio-naturale/#::~:~:text=In%20sintesi%2C%20i%20transformers%20rappresentano.del%20linguaggio%20naturale%20(NLP).)

https://smartstrategy.eu/intelligenza-artificiale/chat-gpt-3-come-accedere-e-testare-il-nuovo-modello-di-gpt-3/?_gl=1*14e0tj4*_up*MQ..*_ga*MTc0MTc4ODc0NC4xNjc4MDEzODA5*_ga_SSJSEWQ4Z6*MTY3ODAxMzgWOC4xLjAuMTY3ODAxMzgWOC4wLjAuMA..*_ga_FZH5PHKM53*MTY3ODAxMzgWOC4xLjAuMTY3ODAxMzgWOC4wLjAuMA..

https://smartstrategy.eu/intelligenza-artificiale/gpt-3-di-open-ai-cose-e-come-funziona-la-piu-potente-ai-in-circolazione/?_gl=1*14e0tj4*_up*MQ..*_ga*MTc0MTc4ODc0NC4xNjc4MDEzODA5*_ga_SSJSEWQ4Z6*MTY3ODAxMzgWOC4xLjAuMTY3ODAxMzgWOC4wLjAuMA..*_ga_FZH5PHKM53*MTY3ODAxMzgWOC4xLjAuMTY3ODAxMzgWOC4wLjAuMA..

<https://www.gptprompter.com/how-to-use-it>