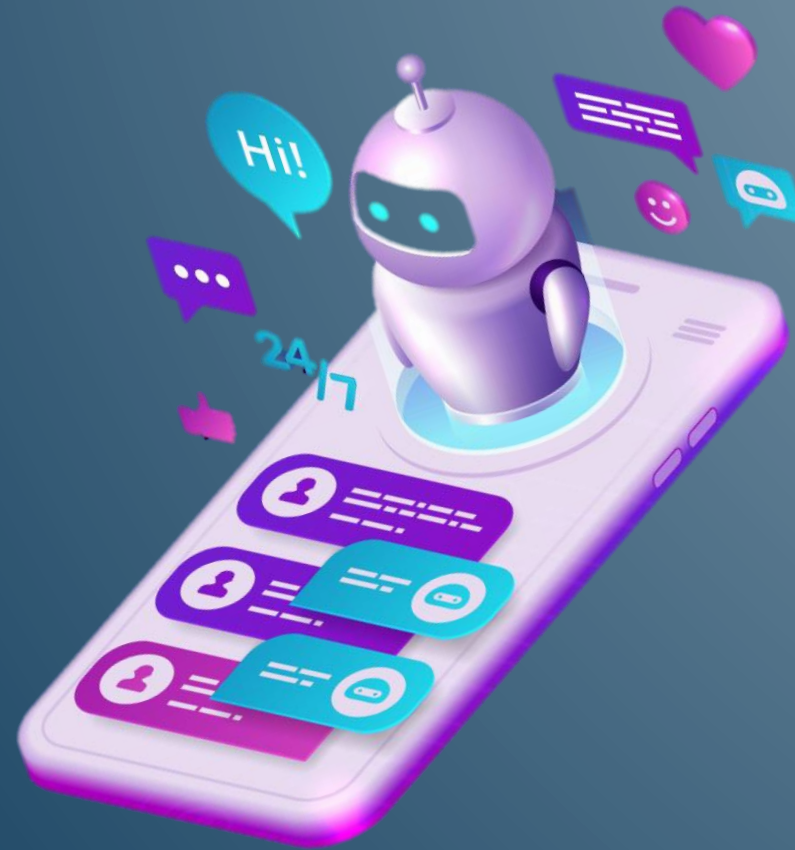


ChatGPT

Esplorando il potere dell'Intelligenza Conversazionale

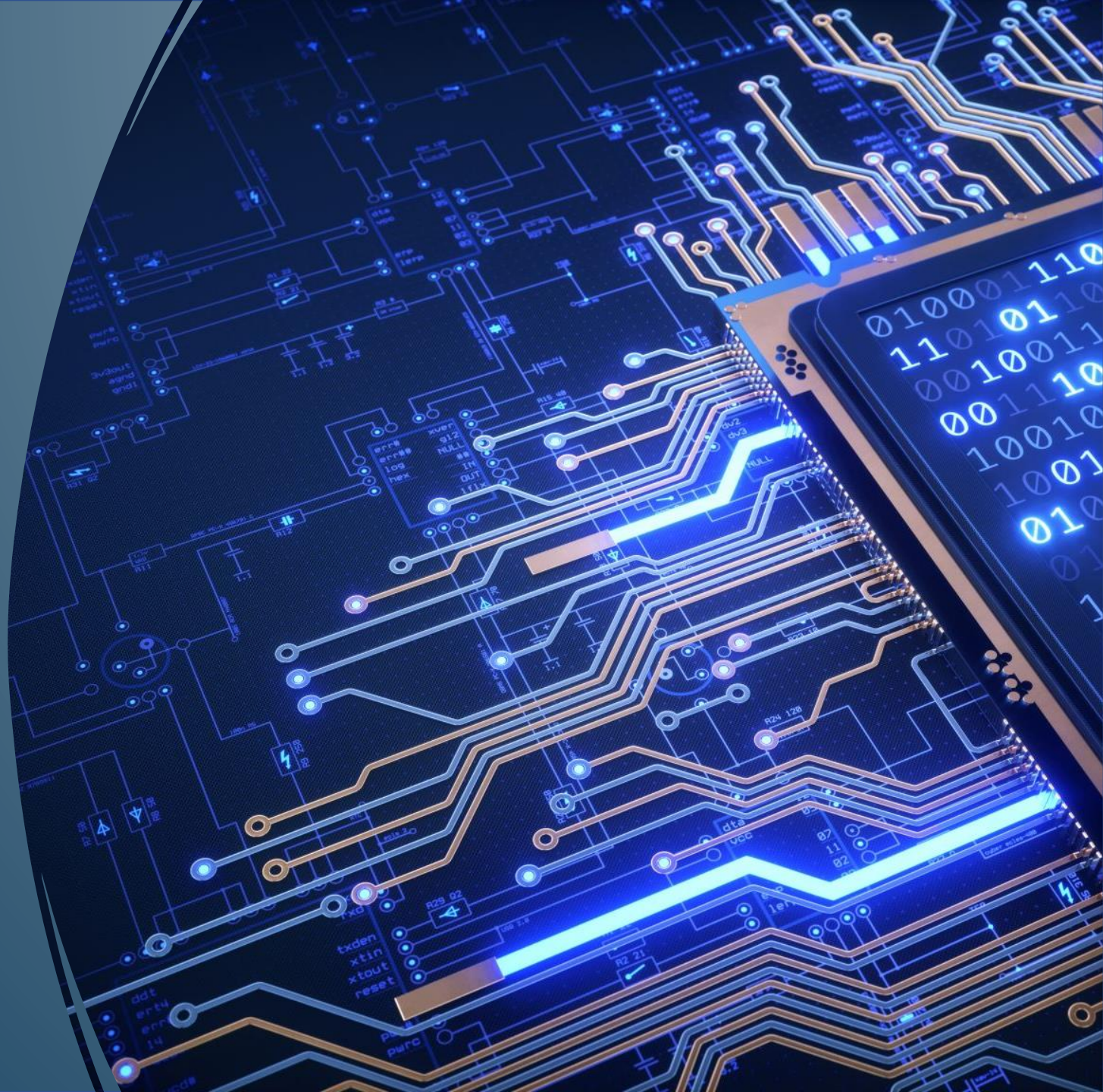


Relatori :
Mattias Caliandro, Mirco Caputo

Cos'è ChatGPT?

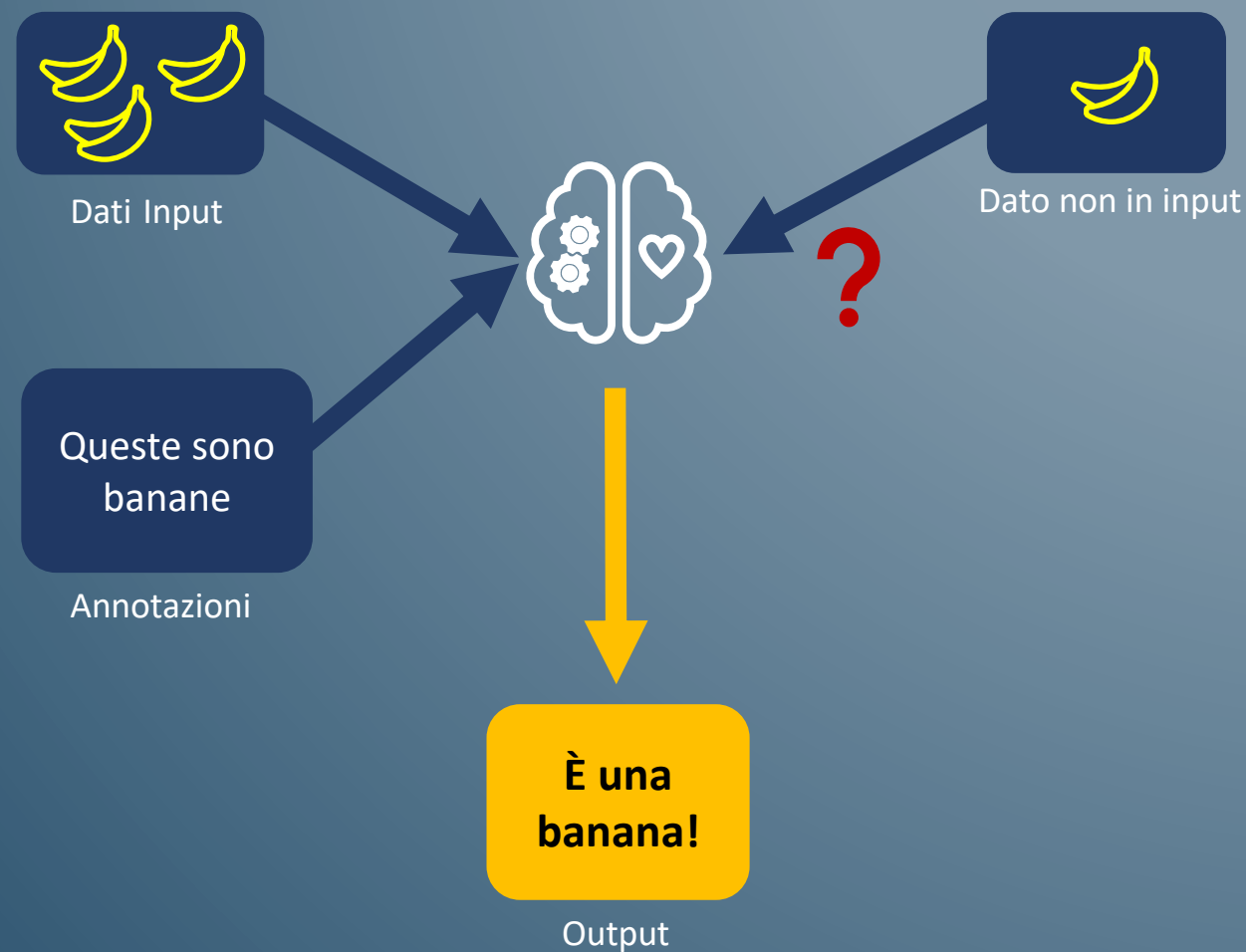
ChatGPT è un chatbot basato su intelligenza artificiale e apprendimento automatico sviluppato da OpenAI specializzato nella conversazione con un utente umano.

La sigla GPT sta per Generative Pre-trained Transformer, una tecnologia nuova applicata al machine learning.



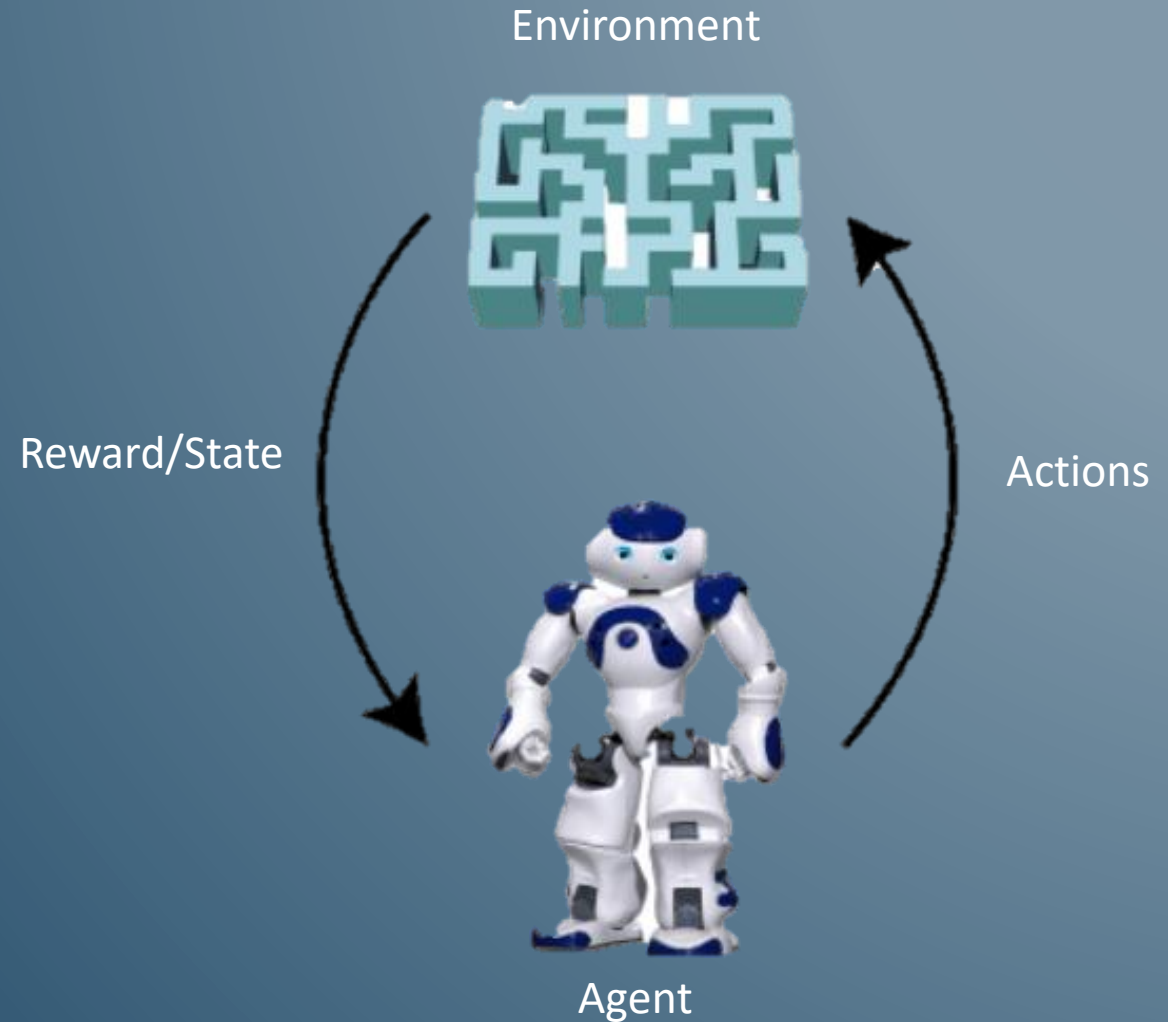
Apprendimento Supervisionato

L'apprendimento supervisionato è un processo in cui un algoritmo apprende da un set di dati di addestramento che contiene esempi etichettati.

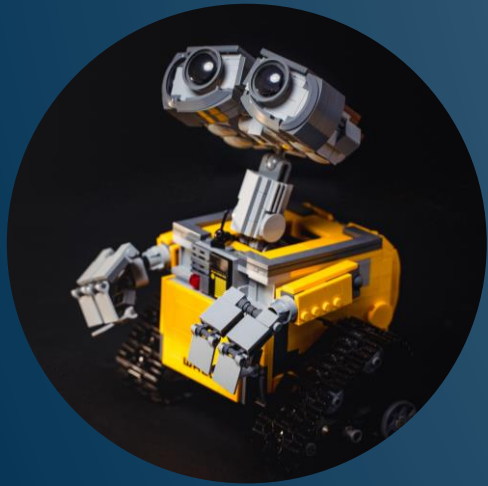


Apprendimento per rinforzo

L'apprendimento per rinforzo è un processo in cui un agente apprende a prendere decisioni ottimali in un ambiente attraverso l'interazione con esso.



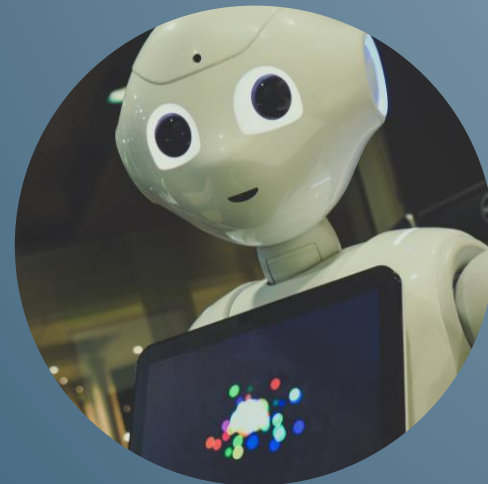
L'evoluzione di GPT



GPT-1



GPT-2



GPT-3



GPT-3.5

GPT-1

Un Generative Pre-Trained Transformer (GPT) è un modello linguistico che si basa sul Deep Learning in grado di generare testi di tipo umano sulla base di un dato input basato su testo. Un utente 'alimenta' il modello con una frase e il trasformatore crea informazioni coerenti basate su paragrafi estratte da set di dati disponibili pubblicamente.

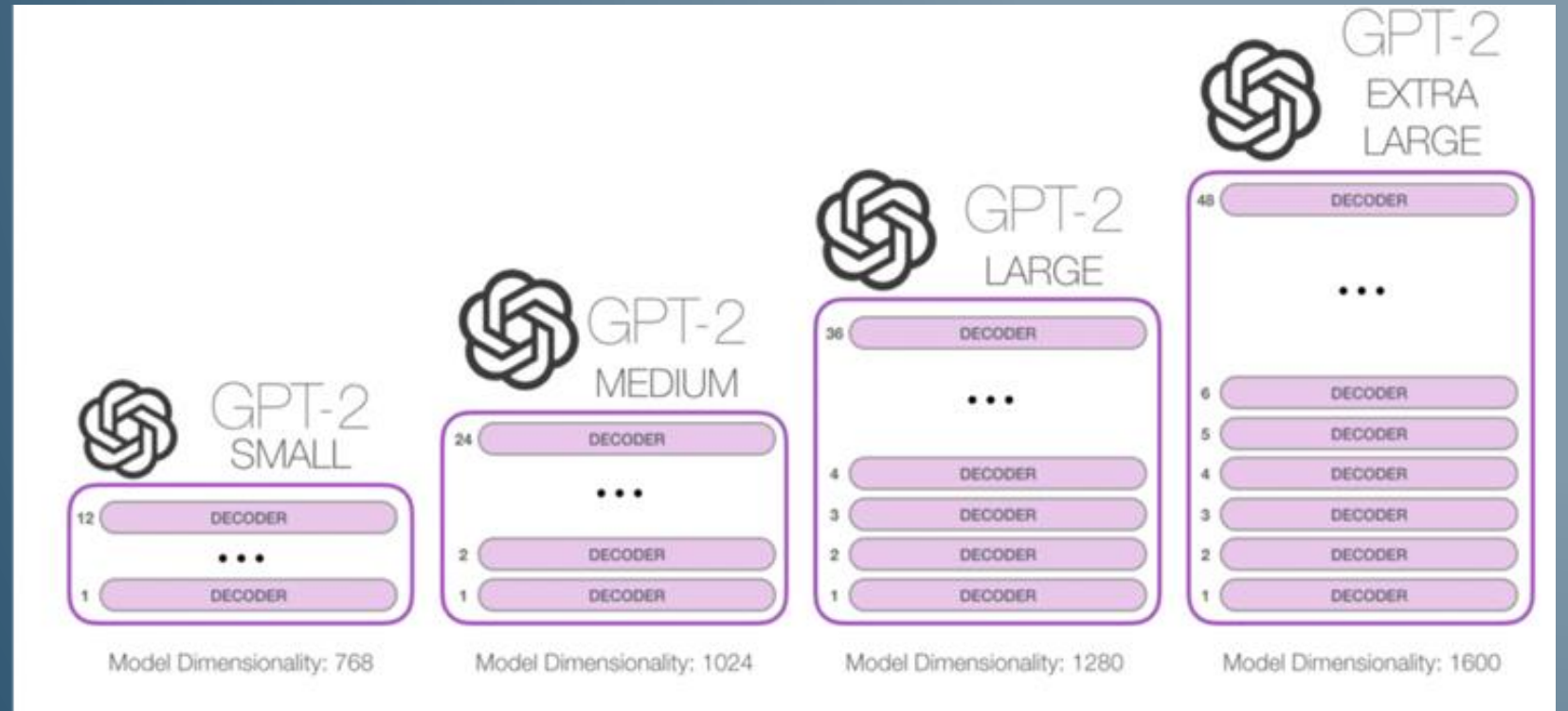


GPT-2

Annunciato nel 2019 da OpenAI, GPT-2 è il successore di GPT. L'obiettivo è : prevedere la parola successiva conoscendo tutte le parole precedenti all'interno di un testo. L'eterogeneità del dataset di training implica che il modello generato si possa adattare a diversi domini applicativi.



GPT-2

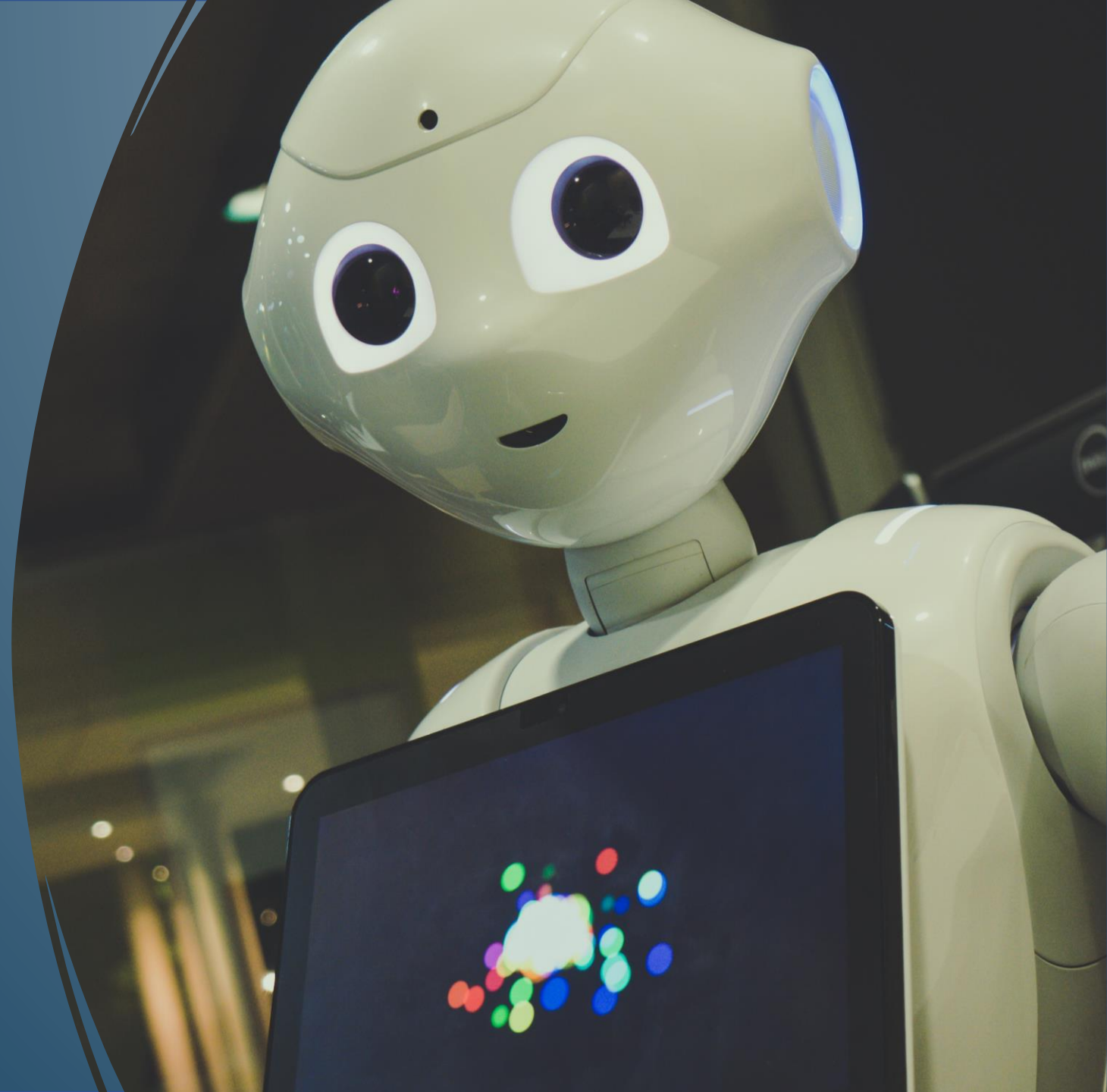


Sono stati rilasciati comunque diversi modelli di GPT2 come mostrato in figura. Questi differiscono a seconda della dimensione:

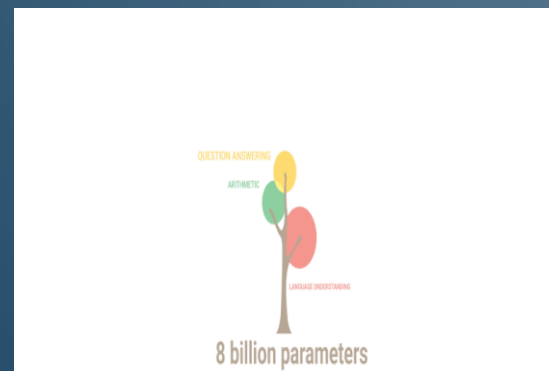
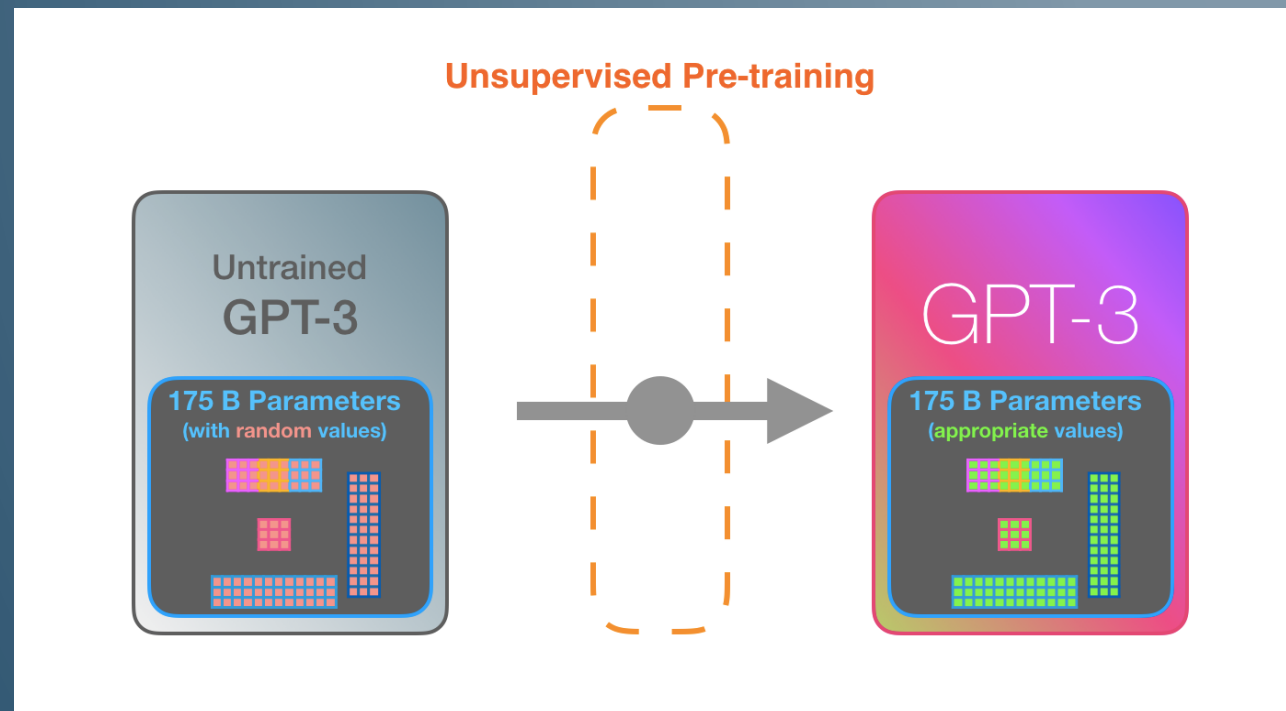
- small (124M parametri)
- medium (355M parametri)
- large (774M parametri)
- extra large (1.5BM parametri).

GPT-3

GPT-3 del 2020 era 100 volte più grande del suo predecessore, mentre i dati del testo di addestramento erano 10 volte più grandi. Il modello ha imparato a tradurre da altre lingue, eseguire operazioni aritmetiche, eseguire semplici programmazioni, ragionare in sequenza e molto altro come risultato dell'espansione della quantità che ha bruscamente aumentato la qualità.

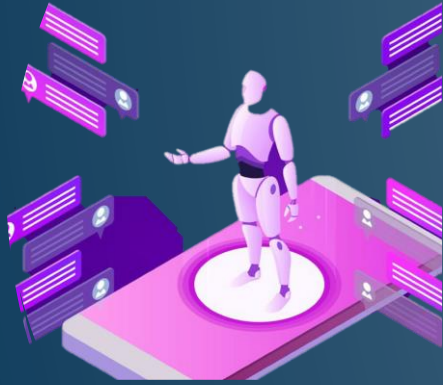


GPT-3



GPT-3 è unicamente un modello di encoder, il che significa che prende in input una sequenza di parole o token e genera una rappresentazione contestuale per ciascuno di essi.

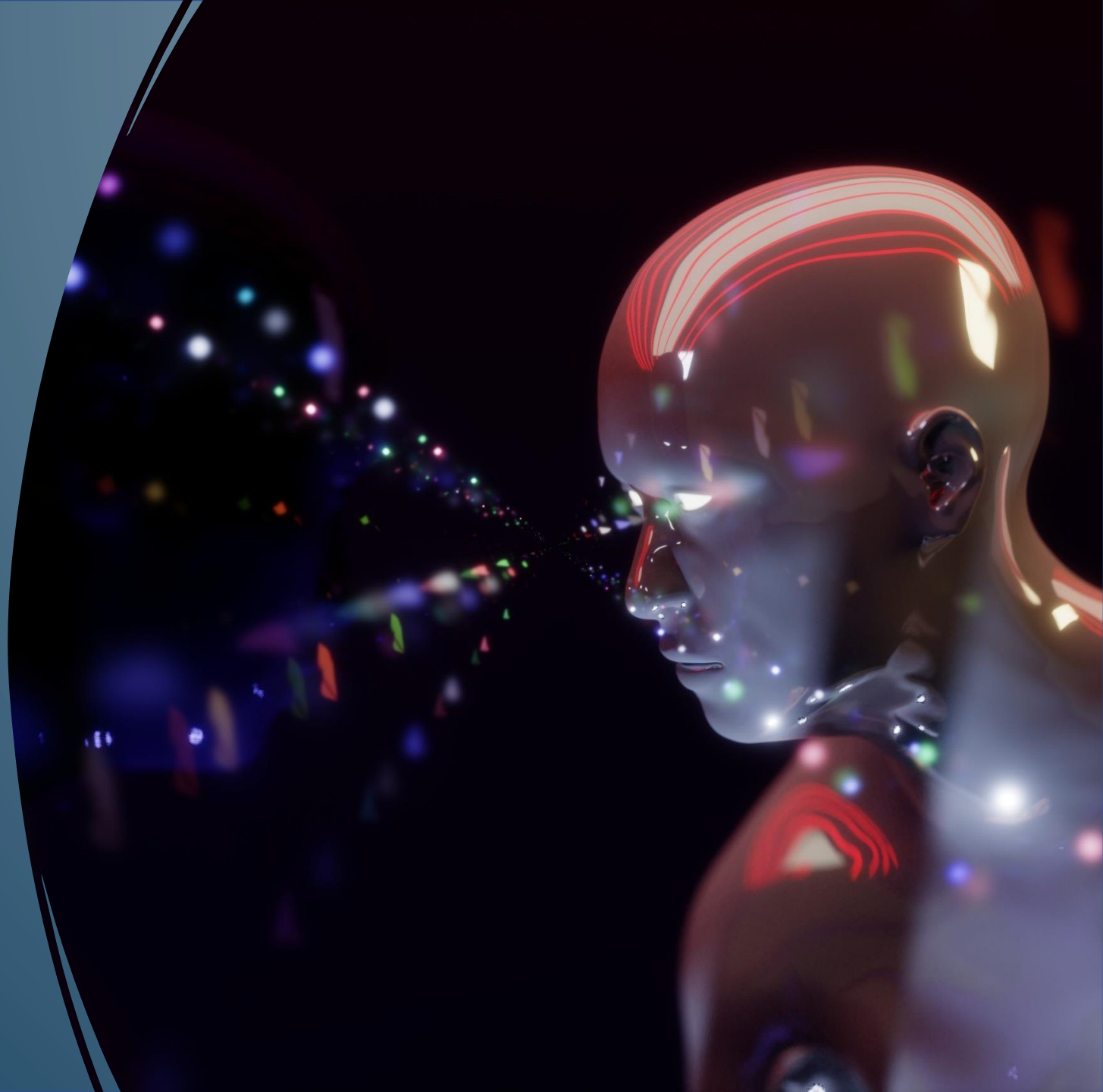
Queste rappresentazioni vengono quindi utilizzate per compiere diverse attività linguistiche come generare testo, completare frasi o rispondere a domande.



GPT-3.5 (InstructGPT)

InstructGPT (o Codex) è stato addestrato su una vasta quantità di testo di programmazione proveniente da diverse fonti, consentendogli di comprendere il contesto e generare codice Python funzionante.

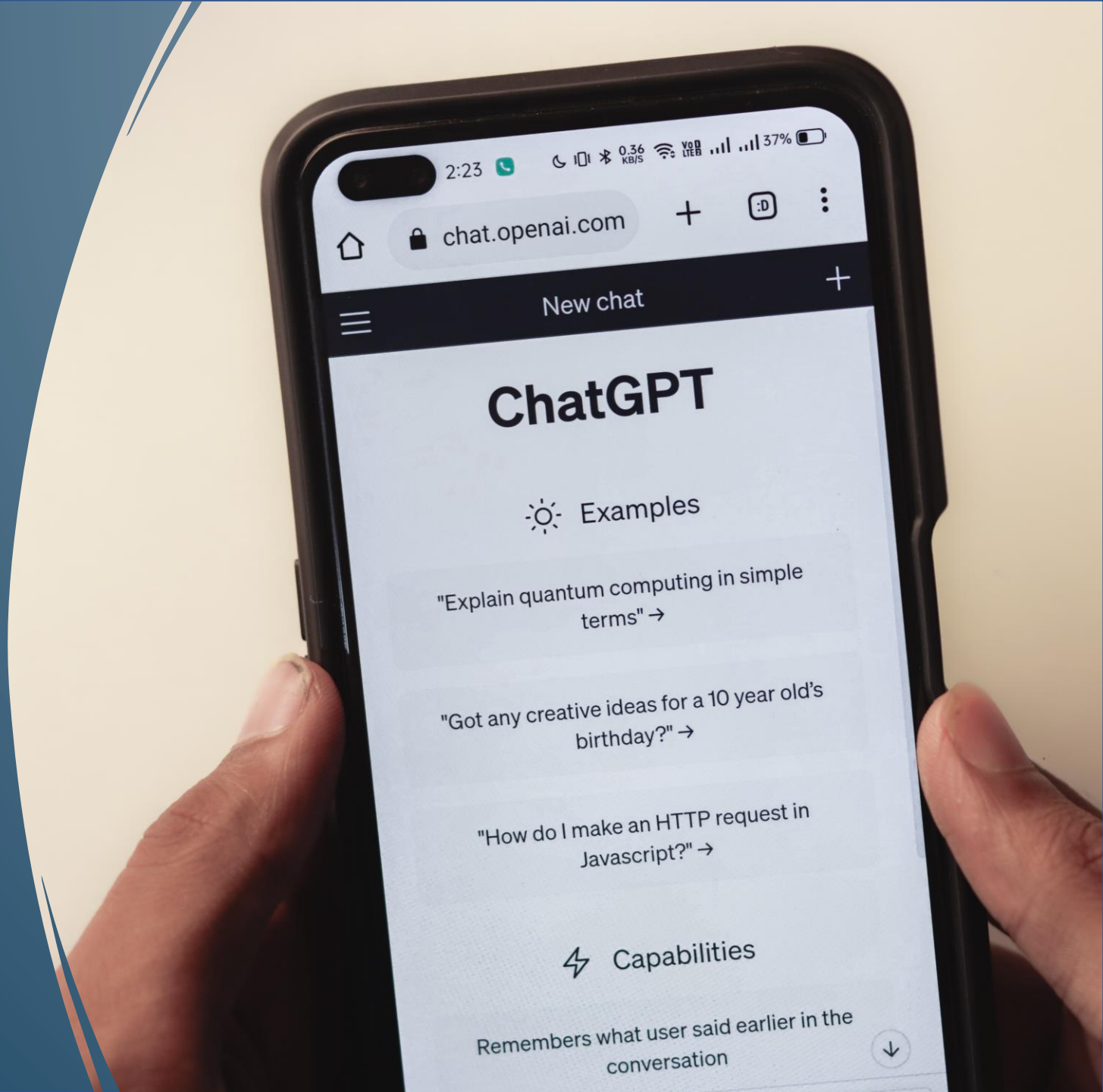
Può essere utilizzato per assistere gli sviluppatori nella scrittura di codice, fornendo suggerimenti, completamenti automatici e persino scrivendo intere funzioni o blocchi di codice.





ChatGPT

La chiave del successo di ChatGPT è la nuova interfaccia user-friendly. ChatGPT utilizzando la "finestra di dialogo" dei messenger è diventato uno strumento disponibile per tutti contemporaneamente.



ChatGPT

Per generare testi con ChatGPT, è essenziale fornire al modello un "prompt". Il prompt è l'input testuale che viene utilizzato per generare il testo di risposta. Può essere qualsiasi cosa, come una frase di avvio per scrivere un articolo o una storia, oppure una domanda per la quale si desidera ottenere una risposta. In generale, maggiore è la precisione e la specificità del prompt, maggiore sarà la precisione e la specificità del testo generato da ChatGPT.

I prompt degli utenti vengono presentati al valutatore di prompt di ChatGPT. Se il valutatore risponde con "no", viene restituito un messaggio di errore all'utente. Se il valutatore risponde con "sì", il prompt viene passato a ChatGPT per generare la risposta desiderata.



Creazione di una ChatBot

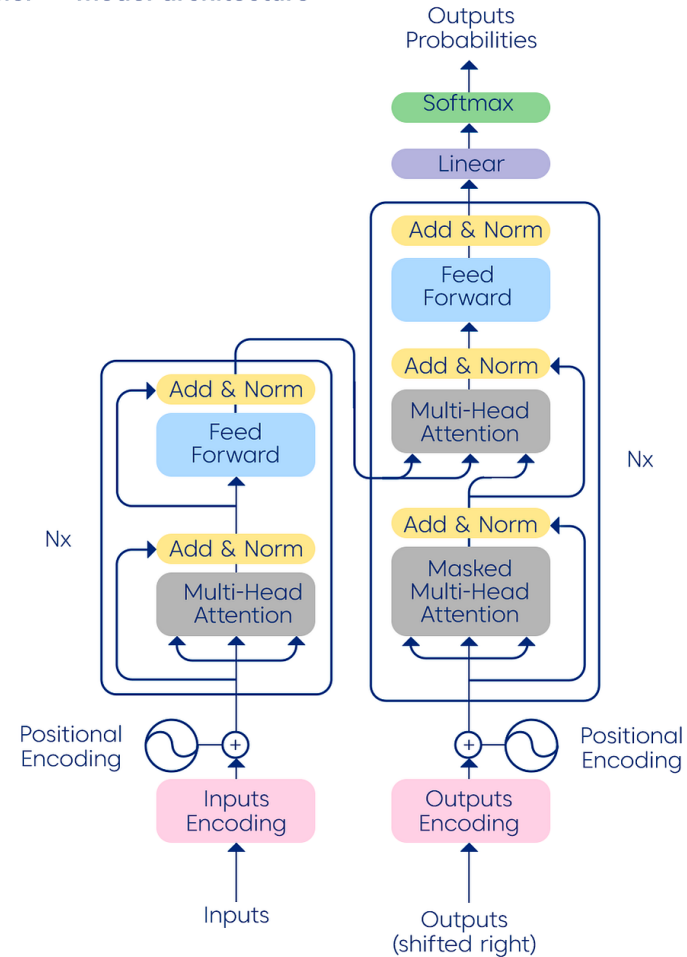
Grazie al framework Gradio e le API di OPENAI abbiamo creato una ChatBot in codice Python.

In questo caso abbiamo creato uno Psicologo Digitale



GPT Generative Pre-trained Transformer

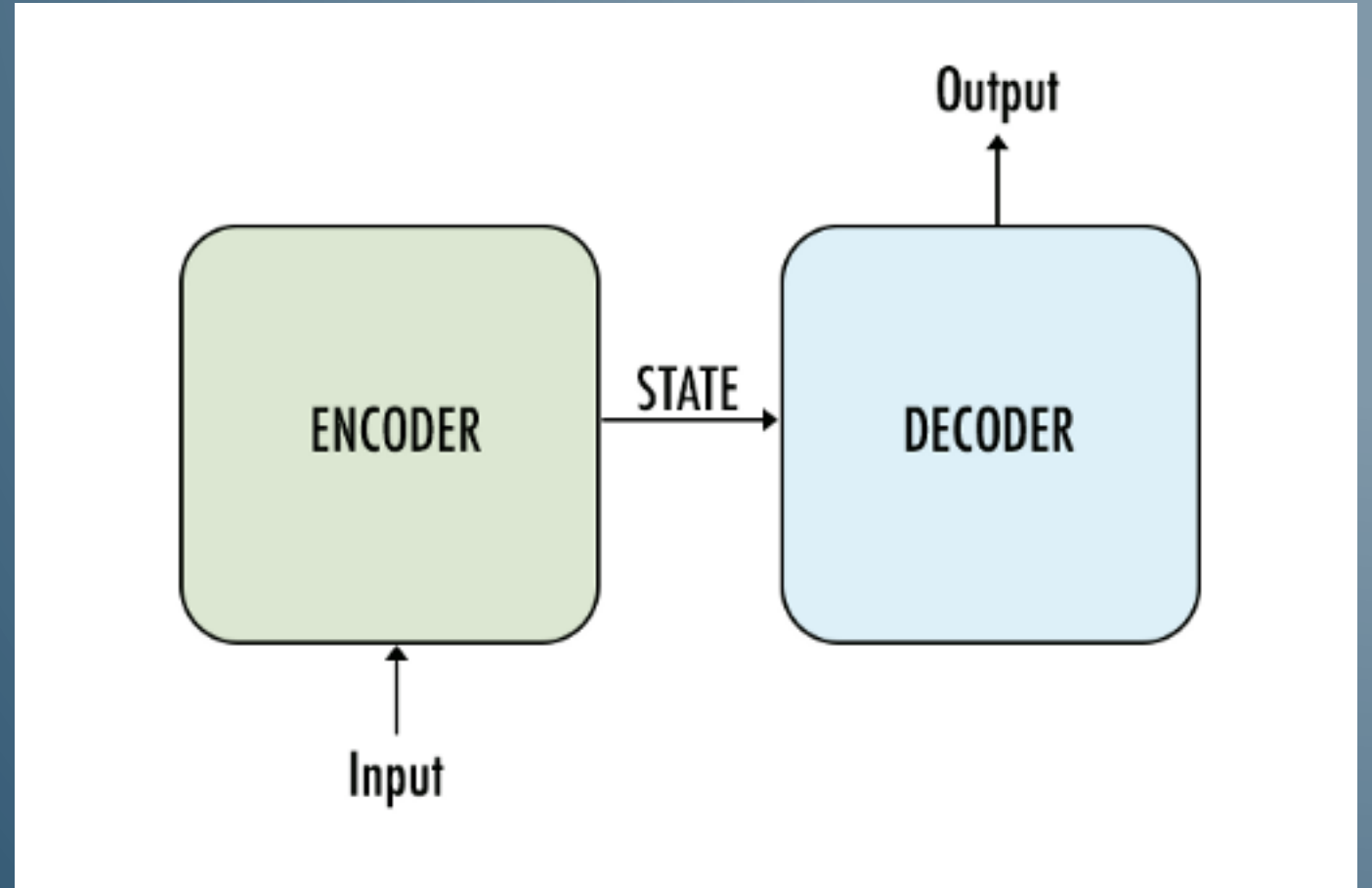
The Transformer — model architecture



Encoder & Decoder

Gli **encoders** hanno il compito di trasformare una frase di input in una rappresentazione numerica, tenendo conto del contesto e del significato delle parole e delle frasi. Utilizzano varie tecniche, come l'embedding, l'attenzione multi-head e le reti feed-forward, per analizzare l'input e creare una rappresentazione numerica della frase.

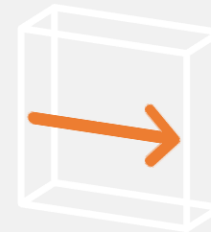
D'altra parte, i **decoders** sono responsabili di generare la risposta o la frase successiva utilizzando la rappresentazione numerica prodotta dagli encoders. I decoders utilizzano il meccanismo di auto-attenzione (self-attention) per generare una distribuzione di probabilità per ogni parola del vocabolario, indicando la probabilità che ogni parola sia la prossima parola nella risposta o nella frase in fase di generazione.



Transformer

I transformer sono modelli di apprendimento profondo basati su reti neurali, che hanno rivoluzionato il campo del Natural Language Processing grazie alla loro capacità di gestire sequenze di parole in modo più efficiente rispetto ai modelli precedenti.

Positional encoding



Self-Attention



Attention



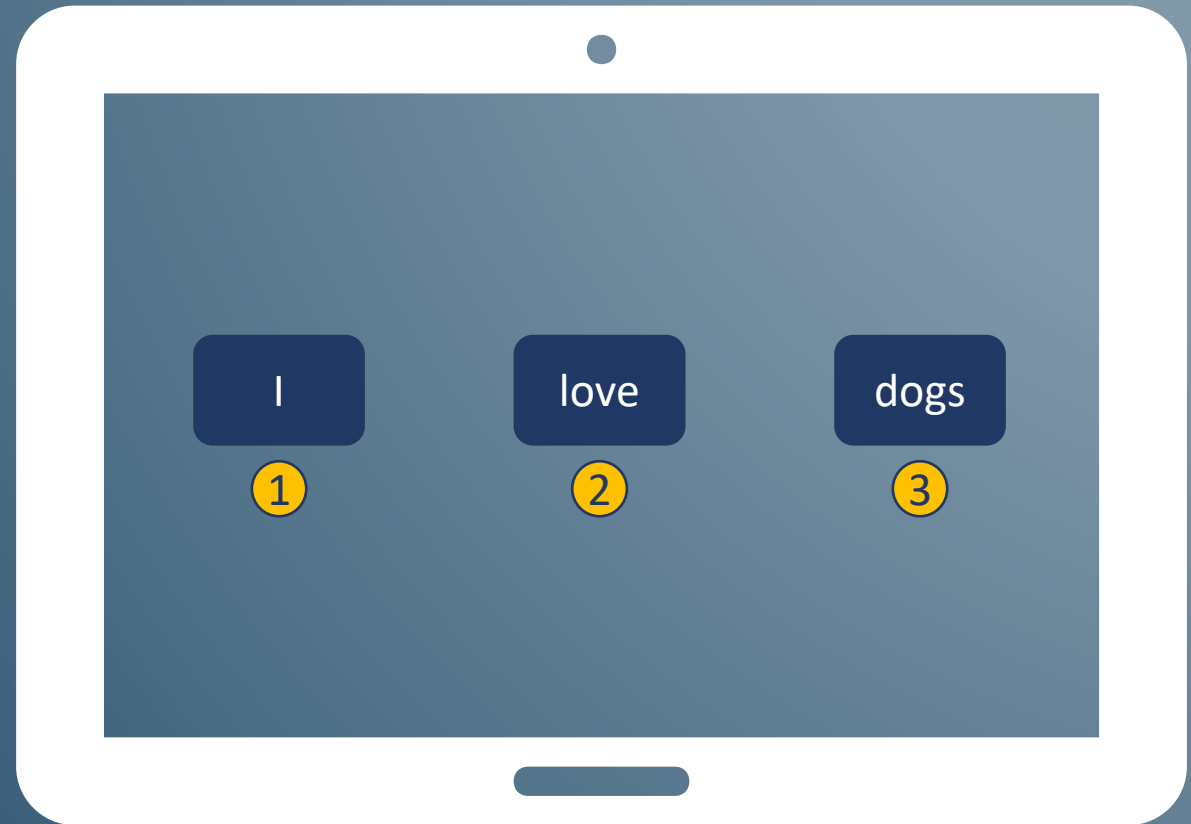
Positional
encoding

Self-Attention

Attention

Positional Encoding

Prima di immettere le parole all'interno della neural network ogni parola viene numerata in base alla sua posizione nella frase. Così facendo quando la network viene 'trainata' su una grande quantità di testo impara come interpretare questi positional encodings capendo l'importanza dell'ordine delle parole dai dati.



Positional
encoding

Self-Attention

Attention

Positional Encoding

Ogni parola nel vocabolario è mappata su un valore vettoriale costante. Tuttavia, una parola può avere significati diversi in contesti diversi. Il design del trasformatore aggiunge un'ulteriore funzione sinusoidale a questo vettore che consente all'incorporamento del vettore di parola di variare a seconda della sua posizione in una frase.

$$\text{bank} = \begin{bmatrix} 0.58 \\ -0.63 \\ 0.29 \end{bmatrix} + \begin{bmatrix} \sin(w/10000^{2 \times 0/3}) \\ \sin(w/10000^{2 \times 1/3}) \\ \sin(w/10000^{2 \times 2/3}) \end{bmatrix}$$

w è l'indice della parola nella frase.

Ciò consente alla rete neurale di conservare alcune informazioni sulle posizioni relative delle parole dopo che i vettori di input sono stati propagati attraverso i livelli. **Si noti che la codifica posizionale da sola non disambigua i diversi sensi di una parola**, ma serve piuttosto come un modo per trasmettere informazioni sull'ordine della frase ai meccanismi dell'attenzione.

Positional
encoding

Self-Attention

Attention

Self-Attention

La self-attention permette alla rete neurale di capire il contesto di una parola tra le parole che la circondano.

In generale aiuta la rete neurale a togliere ambiguità alle parole, riconoscere parti di frasi o capire il tempo delle parole(passato/presente..)

Server, can i have
the **check**?



La parola check da un contesto alla parola server togliendone ambiguità (mail server o human server)

Looks like i just
crashed the server.



La parola crashed fa intendere che il server sia una macchina.

Positional
encoding

Self-Attention

Attention

Attention

Il 'meccanismo' dell'attenzione è una struttura di rete neurale che permette a un modello testuale di guardare ogni singola parola della frase originale per prendere decisioni sul come tradurre la parola nell'output. Il modello per ogni parola in input si 'aspetta' una corrispondente in output. Ma come fa ad aspettarsi qualcosa in output? questo perché dopo che viene 'trainato' con tanti esempi di traduzioni il modello impara i modelli grammaticali dietro di esse.

The European Economics Area



la

europea

economia

zona

The European Economics Area



lo

spazio

economico

europeo

Positional
encoding

Self-Attention

Attention

Attention

Il meccanismo dell'attenzione affronta la questione di quali parti del vettore di input la rete dovrebbe concentrarsi durante la generazione del vettore di output. Questo è molto importante nella traduzione. **Ad esempio, l'inglese “the red house” corrisponde a “la casa roja” in spagnolo:** le due lingue hanno un ordine delle parole diverso. I meccanismi di attenzione consentono a un decodificatore, mentre sta generando una parola in uscita, di concentrarsi maggiormente su parole rilevanti o stati nascosti all'interno della rete e concentrarsi meno su informazioni irrilevanti.

In pratica l'attenzione viene utilizzata in tre modi diversi in una rete neurale trasformatore:

(1) Attenzione encoder-decoder, un meccanismo di attenzione che consente a un decodificatore di occuparsi della sequenza di input durante la generazione della sequenza di output.

(2) Autoattenzione nel codificatore. Ciò consente a un codificatore di occuparsi di tutte le parti dell'output di codifica dal codificatore precedente.

(3) L'auto-attenzione nel decodificatore. Ciò consente a un decodificatore di occuparsi di tutte le parti della sequenza all'interno del decodificatore. I meccanismi di attenzione consentono a un modello di **trarre informazioni dalle parole** di input e dagli stati nascosti in qualsiasi altro punto della frase.

Andando oltre, possiamo generare una matrice che mostra la forza del vettore di attenzione tra ogni parola nella lingua di partenza e nella lingua di destinazione:

	the	red	house
la	0.62	0.06	0.32
casa	0.26	0.14	0.60
roja	0.12	0.80	0.08

Positional
encoding

Self-Attention

Attention

La funzione del meccanismo di attenzione è come una **ricerca fuzzy nel dizionario**: prende una query e un insieme di coppie chiave-valore e restituisce una somma ponderata dei valori che corrispondono alle chiavi più simili alla query. La funzione di attenzione consente alla rete neurale del trasformatore di concentrarsi su un sottoinsieme dei suoi vettori di input. La formula più comune per l'attenzione in una rete neurale del trasformatore è l'attenzione del prodotto scalare:

Formula di attenzione

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Esempio Rete Neurale Transformer

```
1  import tensorflow as tf
2
3  # Definizione dei dati di addestramento
4  english_sentences = ["I am a student", "He is a teacher", "She likes to read"]
5  italian_sentences = ["Io sono uno studente", "Lui è un insegnante", "Lei ama leggere"]
6
7  # Creazione della rete neurale Transformer
8  transformer = tf.keras.layers.Transformer(
9      d_model=128,
10     num_heads=8,
11     num_layers=6,
12     dff=512,
13     input_vocab_size=len(english_sentences),
14     target_vocab_size=len(italian_sentences)
15 )
16
17 # Traduzione di una frase dall'inglese all'italiano utilizzando la rete neurale Transformer
18 input_sentence = tf.convert_to_tensor([english_sentences[0]])
19 output_sentence = transformer(input_sentence, training=False)
20 predicted_index = tf.argmax(output_sentence, axis=-1).numpy()[0]
21 predicted_translation = italian_sentences[predicted_index]
22 print(predicted_translation) # "Io sono uno studente"
```


Ambiti applicativi



Chat-bot



Creazione di
testi



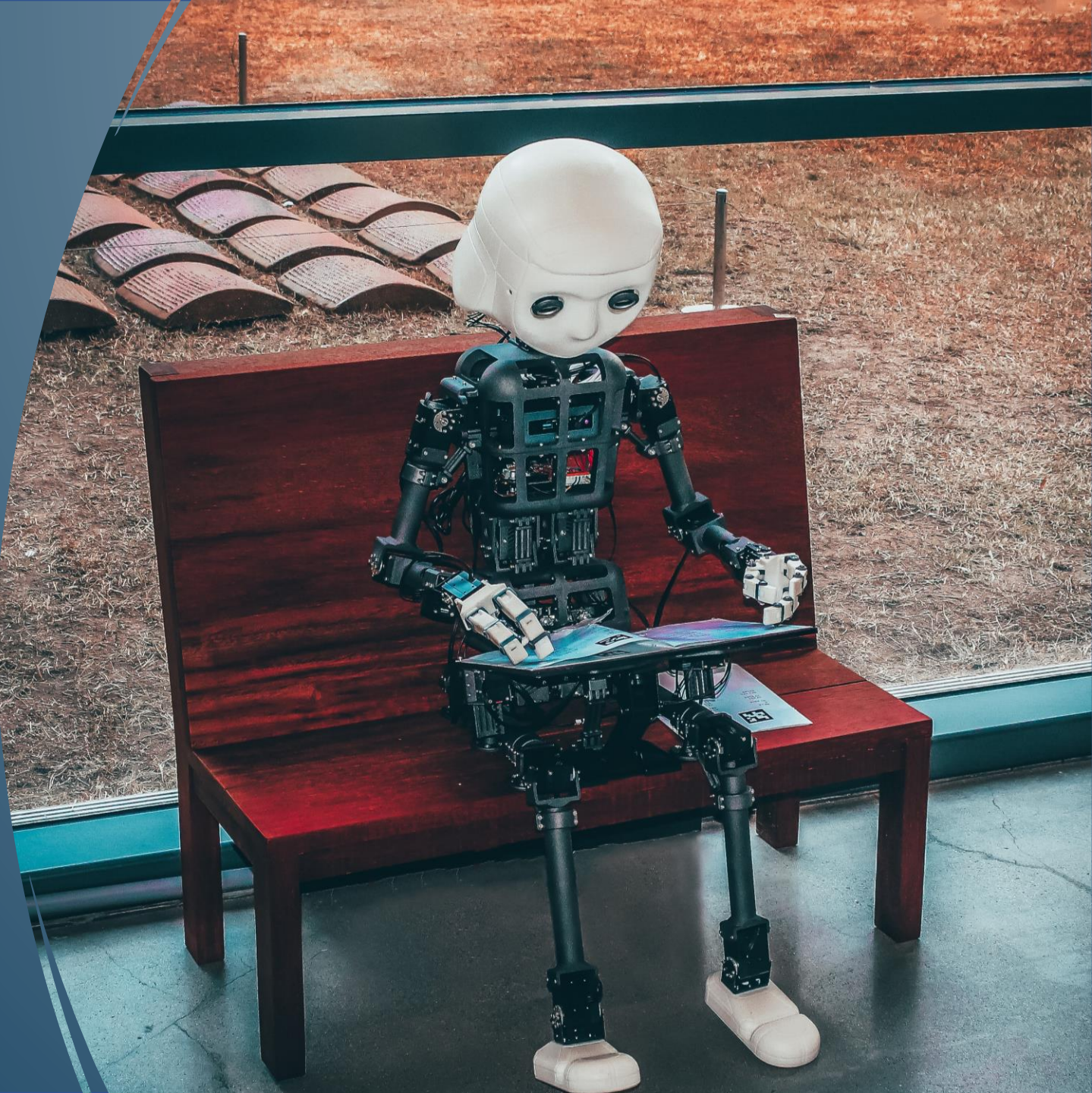
Traduzioni



Creazione di
codice



Risoluzione
problemi





Chat-bot



Creazione
di testi



Traduzioni



Creazione
di codice



Risoluzione
problemi

M

Ciao ChatGPT ! Potresti consigliarmi una serie tv da vedere?



Certo! Ecco alcuni suggerimenti di serie TV che potresti trovare interessanti:

1. "Stranger Things": Una serie che mescola il genere thriller con l'ambientazione anni '80, seguendo un gruppo di ragazzi che affronta eventi sovranaturali nella loro piccola città.
2. "Breaking Bad": Una serie drammatica che racconta la trasformazione di un insegnante di chimica in un pericoloso produttore di metanfetamine, esplorando la moralità e le conseguenze delle sue azioni.
3. "Game of Thrones": Un'avvincente serie fantasy che si svolge in un mondo immaginario, pieno di intrighi politici, guerre e creature mistiche, basata sulla saga di romanzi di George R.R. Martin.
4. "Stranger Things": Una serie antologica di episodi autoconclusivi che esplora temi di tecnologia, società e identità in un futuro distopico, offrendo una prospettiva unica su come la tecnologia influisce sulle nostre vite.



Chat-bot



Creazione
di testi



Traduzioni



Creazione
di codice



Risoluzione
problemi

M

Scrivi una canzone in stile Marracash

< 3/3 >



(Verse 1)

Nel frastuono di questa città che urla,
Tra i palazzi alti e le strade fitte di gente,
Un'atmosfera che vibra e non si ferma mai,
E io, Marracash, mi muovo tra le menti.

(Chorus)

Eh-eh, questa è la mia storia, la mia realtà,
Cammino sulle strade con fermezza e fierezza,
Nel labirinto delle parole, la mia abilità,
Marracash, il re dell'hip hop, la tua certezza.



Chat-bot



Creazione
di testi



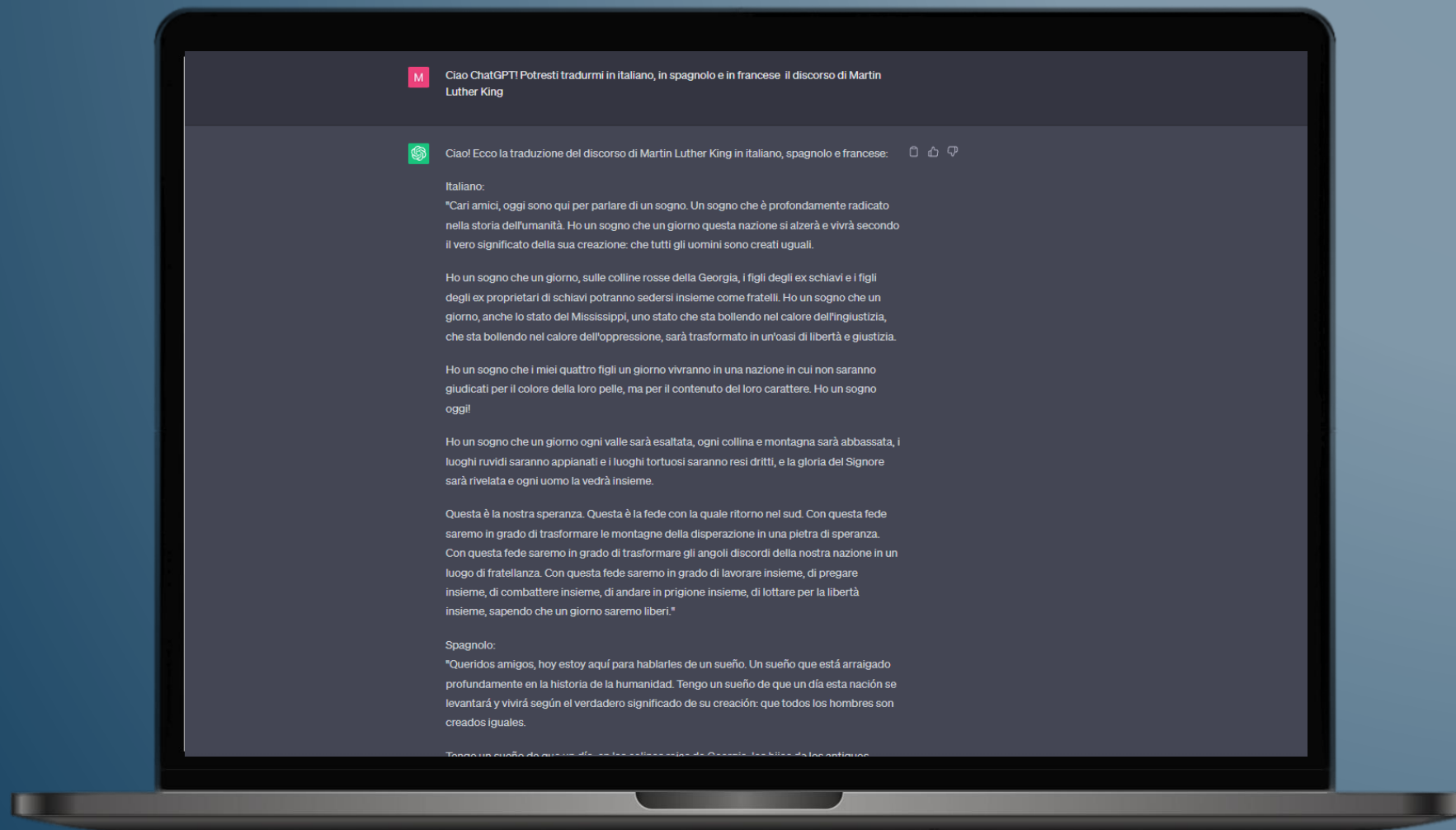
Traduzioni



Creazione
di codice



Risoluzione
problemi





Chat-bot



Creazione
di testi



Traduzioni



Creazione
di codice



Risoluzione
problemi



Certo! Per ottenere i 5 giocatori con il valore di mercato più alto tramite le API di Sorare, puoi utilizzare il seguente codice Python:



python

Copy code

```
import requests

# Definisci l'URL dell'endpoint delle API di Sorare per ottenere i giocatori
url = "https://api.sorare.com/api/v2/players?sort=market_value&order=desc&limit=5"

# Effettua una richiesta GET all'API di Sorare
response = requests.get(url)

# Verifica lo stato della risposta
if response.status_code == 200:
    # La richiesta è andata a buon fine
    data = response.json()
    players = data["players"]

    # Stampa i 5 giocatori con più valore di mercato
    print("I 5 giocatori con più valore di mercato su Sorare sono:")
    for player in players:
        print(f"- {player['full_name']} - Valore di mercato: {player['market_value']}")
else:
    # Si è verificato un errore nella richiesta
    print(f"Si è verificato un errore nella richiesta: {response.status_code}")
```



Chat-bot



Creazione
di testi



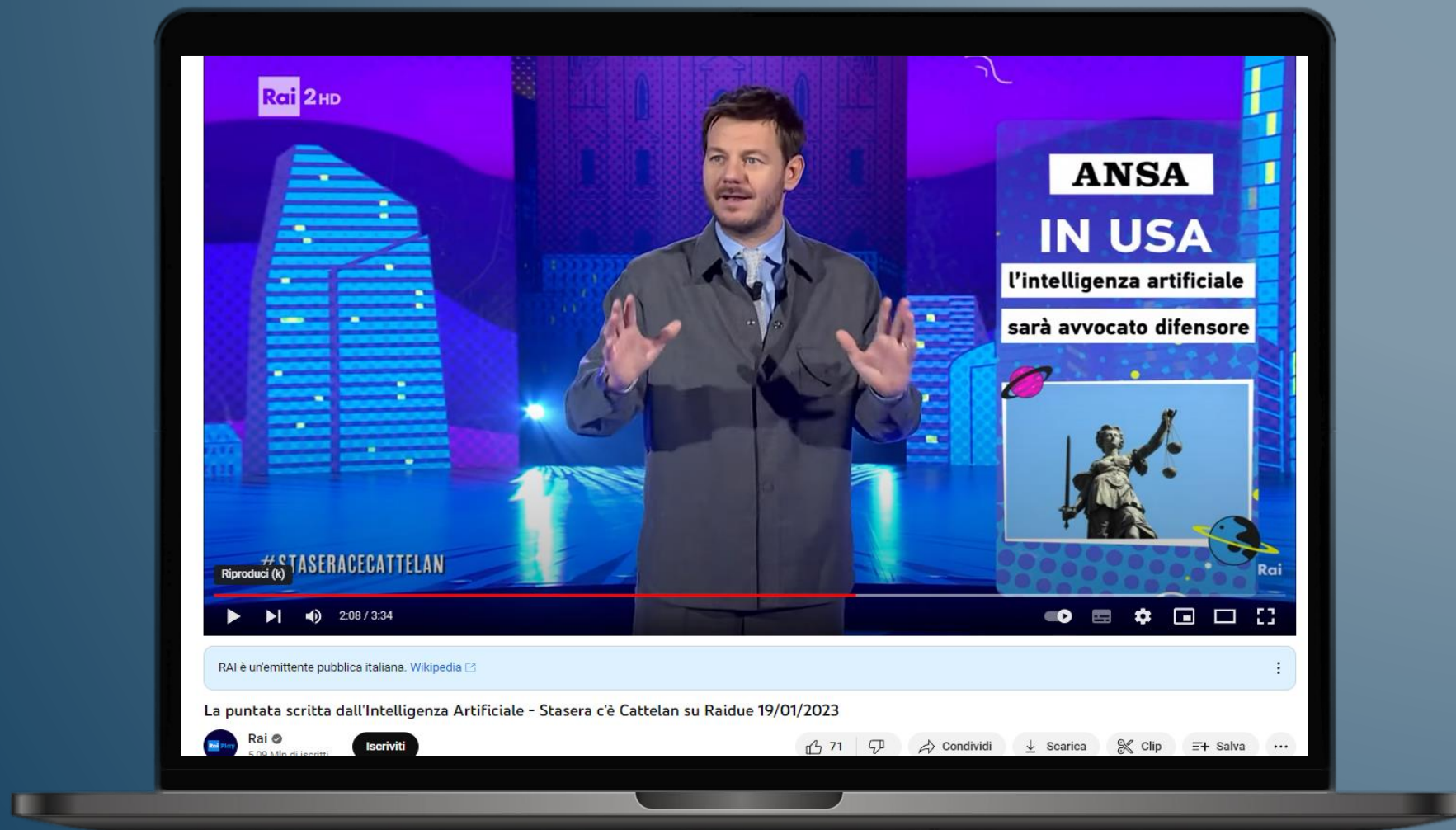
Traduzioni



Creazione
di codice



Risoluzione
problemi



Considerazioni etiche

*«Un altro pericolo della scrittura AI è il potenziale per **un uso non etico**. Gli algoritmi di intelligenza artificiale possono essere programmati per generare informazioni fuorvianti o false, che potrebbero essere utilizzate per diffondere notizie false o disinformazione. Ciò potrebbe avere gravi conseguenze, come influenzare l'opinione pubblica o causare danni a individui o gruppi».*





GRAZIE PER L'ATTENZIONE

DI :

MIRCO CAPUTO

MATTIAS CALIANDO