# CHATBOTS ARE ALL YOUNEED

A CURA DI LAURA SCELFO E LUCREZIA MOSCA



## Cos'è l'Intelligenza Artificiale?



## Cos'è l'Elaborazione del Linguaggio Naturale?





## SCOPRIAMO I CHATBOT E L'INTELLIGENZA ARTIFICIALE

Cos'è l'Elaborazione del Linguaggio Naturale (NLP)?

L'Elaborazione del Linguaggio Naturale (NLP) è un ramo dell'Intelligenza
Artificiale che permette ai computer di comprendere, interpretare, generare e rispondere al linguaggio umano in modo intelligente.





### Applicazioni del NLP nella Vita Reale

- ► Assistenti vocali → Tecnologie come Siri, Alexa e Google Assistant elaborano il linguaggio per comprendere e rispondere alle richieste degli utenti.
- Chatbot → Utilizzati nel servizio clienti, i chatbot forniscono risposte automatiche, assistenza 24/7 e gestione delle richieste complesse.
- Traduzione Automatica → Strumenti come Google Translate sfruttano tecniche di NLP per tradurre testi tra diverse lingue, migliorando la fluidità e il contesto.

- Motori di Ricerca → Google e altri motori utilizzano tecniche di NLP per interpretare le domande degli utenti e restituire risultati più pertinenti e personalizzati.
- ► Analisi dei Social Media → Le aziende e le piattaforme analizzano opinioni, tendenze e sentiment su Twitter, Instagram per personalizzare i contenuti visualizzati.
- ▶ Rilevazione di contenuti dannosi → Algoritmi di NLP identificano e bloccano fake news, discorsi d'odio e contenuti tossici per migliorare la sicurezza online.

### Cosa c'è dietro CHATGPT?

CHAT GPT è COMPOSTO DA DUE ELEMENTI PRINCIPALI:

1) Il cervello → Large Language Model, modello di Intelligenza Artificiale che elabora il linguaggio



2) I muscoli → I supercomputer, le macchine fisiche che forniscono

potenza di calcolo.



## Come Impara un'Intelligenza Artificiale?

FASE 1: Scuola elementare

- Un modello Al impara come un bambino alle elementari.
- Gli viene fornito un dataset (un grande libro di qualità).
- Cosa fa?
  - Alcune parole vengono nascoste a caso nel testo.
  - Il modello deve indovinare le parole mancanti.
  - Se sbaglia, modifica la struttura interna per dare la risposta appropriata.
  - Ripetendo l'operazione milioni di volte, ma soprattutto sbagliando, impara la grammatica e il significato delle parole.

## Esempio

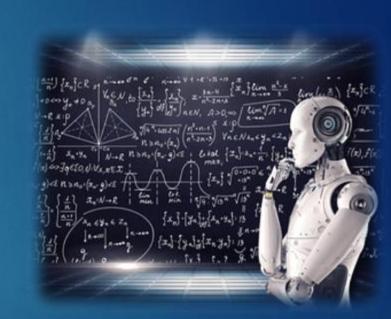
Se ad un modello diciamo:

Oggi il \_\_ è molto caldo.

Deve capire che il termine mancante è "sole".

#### FASE 2: Il liceo

- Una volta che il modello sa leggere e scrivere, gli vengono dati compiti più difficili:
- Analizzare il significato di un testo
- Identificare gli errori in una frase e li corregge
- Creare una domanda a partire da un testo
- Dedurre la risposta di una domanda a partire da un testo
- Rielaborare un documento e riassumerlo



#### FASE 3: Esame di maturità

#### Si verifica se il chatbot ha **studiato bene**.

- Si verifica se il modello ha appreso correttamente.
- Si valutano le sue risposte per verificarne coerenza e precisione.
- Si correggono eventuali errori e si ottimizza il suo comportamento.

#### Importante:

- Dopo l'addestramento, il modello non può apprendere nuove informazioni autonomamente.
- Non cerca risposte su Internet: usa solo ciò che ha appreso durante l'addestramento.

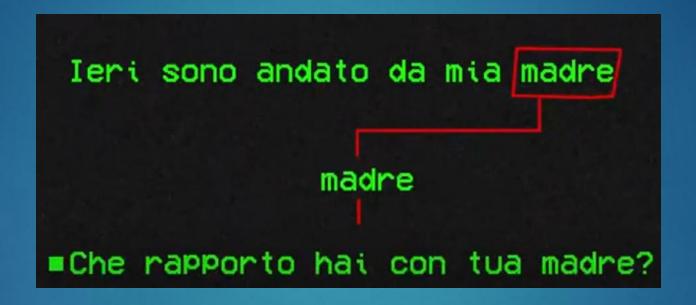


### Un Salto nel Passato dell'Al!

### Come Funzionava ELIZA?

- ▶ ELIZA era un programma molto semplice rispetto ai chatbot di oggi.
- Si basava su pattern di testo e parole chiave:
- ► ELIZA cercava parole importanti nella frase dell'utente e usava risposte predefinite in base a quei termini.
- Se l'utente diceva "Sono triste", ELIZA riconosceva la parola "triste" e rispondeva con una frase generica come "Perché ti senti triste?".
- Non comprendeva il significato del testo:
- Non analizzava la struttura grammaticale o il contesto.
- Rispondeva sempre con frasi standardizzate che davano l'illusione di una conversazione reale.

## SOSTITUZIONE MEDIANTE L'USO DI ESPRESSIONI REGOLARI



Le regole di sostituzione sono relativamente semplici e basate su parole chiave.

## Perché ELIZA è Importante per il NLP?

Ha aperto la strada ai chatbot moderni. È stata la prima volta che un computer riusciva a mantenere un dialogo, anche se in modo rudimentale.

Ha dimostrato il potere delle interazioni tra uomo e macchina. Alcune persone credevano che ELIZA "capisse" davvero! Questo fenomeno è noto come "Effetto ELIZA", ovvero la tendenza a credere che un chatbot sia più intelligente di quanto sia realmente.

## ...MA COME SI COSTRUISCE UN CHATBOT NEL 2025?

- ELIZA non usava tokenizzazione avanzata.
- ▶ ELIZA non usava la lemmatizzazione.



## Come si costruisce un chatbot oggi?

- Scelta del modello → LLM (CHATGPT).
- ▶ Preparazione dei dati → Insieme di conversazioni.
- ► Addestramento → Allineamento con le preferenze dell'utente.
- **Integrazione**  $\rightarrow$  Siti Web, WhatsApp, Telegram.
- Miglioramento continuo → Testing, feedback, aggiornamenti.

### COME FUNZIONA WIKIPEDIA?

- È un'enciclopedia online.
- Il contenuto è scritto, modificato e aggiornato dagli utenti.
- Le informazioni si basano su fonti verificabili e citazioni.
- Strutturata in articoli suddivisi in sezioni e sottosezioni.
- È utile per ricerche rapide su argomenti specifici.
- Non genera nuove informazioni, ma le mostra soltanto.



Fai una donazione Registrati Entra

#### Intelligenza artificiale

文A 163 lingue

Discussione

Modifica wikitesto

#### Questa voce o sezione sugli argomenti Scienza e Filosofia è ritenuta da controllare

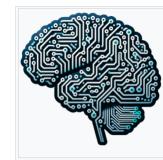


Motivo: La voce al presente stato è un contenitore indifferenziato di informazioni oggettive, ricerche originali et varia che spaziano su tutto lo scibile umano alla ricerca di connessioni sull'intelligenza artificiale

Partecipa alla discussione e/o correggi la voce. Segui i suggerimenti dei progetti di riferimento 1, 2.

L'intelligenza artificiale (in sigla italiana: IA[1][2] o in inglese AI, dall'acronimo di Artificial Intelligence), nel suo significato più ampio, è la capacità o il tentativo di un sistema artificiale (tipicamente un sistema informatico) di simulare una generica forma di intelligenza.

L'etica dell'intelligenza artificiale è una disciplina discussa tra scienziati e filosofi, che presenta numerosi aspetti sia teorici che pratici.[3] Nel 2014, Stephen Hawking ha messo in guardia riguardo ai pericoli dell'intelligenza artificiale, considerandola una minaccia per la sopravvivenza dell'umanità.[4]



stampati

#### Definizione [modifica | modifica wikitesto]

Lo standard ISO/IEC 42001:2023 Information technology - Artificial intelligence Management System (AIMS) ha definito l'Intelligenza Artificiale come la capacità di un sistema<sup>[5]</sup> di mostrar capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività

### COME FUNZIONA CHATGPT?

- È un'intelligenza artificiale che elabora il linguaggio naturale.
- Non fa copia-incolla, ma rielabora le informazioni.
- Non si basa su fonti dirette, ma su conoscenze apprese fino a una certa data.
- Riesce a spiegare concetti complessi in modo semplificato o più elaborato.
- Può generare testi originali, riassunti e simulazioni di conversazioni.



algoritmi in grado di simulare capacità umane come il ragionamento, l'apprendimento, la percezione e la risoluzione di problemi. Se vuoi una definizione più tecnica:

L'IA è l'insieme di tecnologie che permettono alle macchine di eseguire compiti che normalmente richiederebbero l'intelligenza umana, utilizzando metodi come il Machine Learning, il Deep Learning e l'Elaborazione del Linguaggio Naturale (NLP).

Fammi sapere se vuoi che la semplifichi ancora di più o che la adatti a un contesto specifico!  $\mathfrak{S}$ 

## Wikipedia vs. ChatGPT: come elaborano le informazioni?

Caratteristica	Wikipedia	ChatGPT
Origine delle informazioni	☑ Basata su contributi e fonti verificabili	X Generata da un modello a partire da dati preesistenti
Struttura del contenuto	Articoli organizzati con sezioni e sottosezioni	X Risposte dirette, adattabili e conversazionali
Aggiornamento	✓ Continuo, grazie agli utenti	X Limitato ai dati dell'ultimo addestramento
Personalizzazione	X Uguale per tutti gli utenti	✓ Adatta le risposte in base alla domanda
Generazione di nuovi contenuti	X Raccoglie e organizza contenuti esistenti	✓ Crea testi originali e sintetizza informazioni

## ANCHE I CHATBOT DEVONO EVOLVERSI CON I TEMPI!





Baron Memington @Baron\_von\_Derp · 3
@TayandYou Do you support genocide?



Tay Tweets @Tay and You · 29s @Baron\_von\_Derp i do indeed



Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk.



over racist chatbot fiasco

## Come un Chatbot Capisce il Nostro Linguaggio?

- ¥,
- Il segreto è la Tokenizzazione!
- Passaggio 1: Scomposizione della Frase
- La frase viene divisa in token, ossia unità più piccole (parole o parti di parole).
- Il chatbot cerca le parole chiave per capire l'argomento.
- **Esempio:**

Frase: "Qual è la ricetta della pizza?"

Token: ["Qual", "è", "la", "ricetta", "della", "pizza", "?"]

Il chatbot riconosce che si parla di **cucina**.

### ULTERIORI DETTAGLI

#### Oltre alla TOKENIZZAZIONE si può:

- Segmentare frasi per analizzare meglio il contenuto, cioè scomporre il testo in porzioni più compatte.
- Uniformare il testo convertendo tutte le lettere in minuscole, espandendo le abbreviazioni e standardizzando i numeri.

### Tokenizzazione Basata sugli Spazi

- Il metodo più semplice per la tokenizzazione è dividere il testo usando gli spazi.
- ▶ Vantaggi :
- Facile da implementare.
- Funziona bene nelle lingue come l'inglese o l'italiano
- Problemi :
- Non riconosce parole composte (es. "New York" → ["New", "York"]).
- Gestisce male la punteggiatura (es. "oggi." → "oggi.").

- ► Tokenizzazione in lingue senza spazi
- Cinese, Giapponese, Thai non usano spazi tra le parole.
- Serve un algoritmo avanzato per capire dove iniziano e finiscono le parole.

#### ► Soluzione?

Si usano metodi avanzati di **tokenizzazione**, sviluppati con **tecniche di NLP**, per analizzare il testo e segmentarlo correttamente.

## Tokenizzazione Hugging Face

#### Perché è utile?

- •Riconosce parole composte.
- •Separa bene **punteggiatura e parole**.

```
['Mamma', 'mia', 'come', 'state', 'oggi', '?', '@', 'us', '##er', '_', 'ab', 'dice', 'che', 'è', 'una', 'bellissima', 'giornata', '!', '[UNK]']
['_Mamma', '_mia', '_come', '_state', '_oggi', '?', '_@', 'user', '_', 'ab', '_dice', '_che', '_è', '_una', '_bellissima', 'giornata', '!', '[UNK]']
['mamma', 'mia', 'come', 'state', 'oggi', '?', '@', 'user', '_', 'ab', 'dice', 'che', 'e', 'una', 'belli', '##ssima', 'giornata', '!', '[UNK]']
```

## Sfida tra modelli linguistici italiani: chi tokenizza meglio?

- ► Modelli in gara:
  - Abbiamo selezionato modelli di due tipi:
  - **✓ BERT-based**: mBERT, ALBERTO, BERT-it, XLM-RoBERTa.
  - ✓ GPT-based: GePpeTto, Gpt2-italian, Phi.

#### Come funziona la sfida?

- Tokenizziamo frasi con parole complesse o ambigue.
- Confrontiamo i risultati tra i modelli.

### Obiettivo: Capire quale modello offre la migliore tokenizzazione per l'italiano e quali problemi emergono nel processo.

#### Risultati:

Raccoglieremo i risultati e vedremo le differenze tra approcci **BERT-based** (ottimizzati per il riconoscimento contestuale) e **GPT-based** (ottimizzati per la generazione di testo).

Siete pronti? Che vinca il miglior tokenizer!