

# IA sub-simbolica: il caso ChatGPT

Angelo Montanari

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Università degli Studi di Udine

Udine, 7 aprile, 2024

# IA sub-simbolica: l'esplosione del machine/deep learning

Quali sono le ragioni dell'esplosione della ricerca e delle applicazioni del machine/deep learning?

Enorme **potenza di calcolo** ed enorme disponibilità di **dati** (i dati sono il nuovo **petrolio**).

Il caso di **ChatGPT (Generative Pre-trained Transformer)**: chatbot in grado di dialogare con l'utente in linguaggio naturale

- 1.800.000.000.000 parametri

- dataset di training di 1.000 Terabyte (1 Terabyte = 1.000 Gigabyte).

# IA sub-simbolica: applicazioni

Tecniche di computer vision e machine learning per riconoscimento di corpi estranei (**intruder detection**).

Tecniche di computer vision e deep learning per il riconoscimento di difetti superficiali in materiali compositi e prodotti (**detection of surface defect**).

Rilevamento precoce dei guasti e manutenzione predittiva con tecniche di machine/deep learning (**early failure detection and predictive maintenance**).

Integrazione di **monitoring** e **machine learning** per raffinare l'insieme di condizioni da monitorare (applicabile in una molteplicità di ambiti)

**IA generativa.** Un esempio ben noto: ChatGPT.

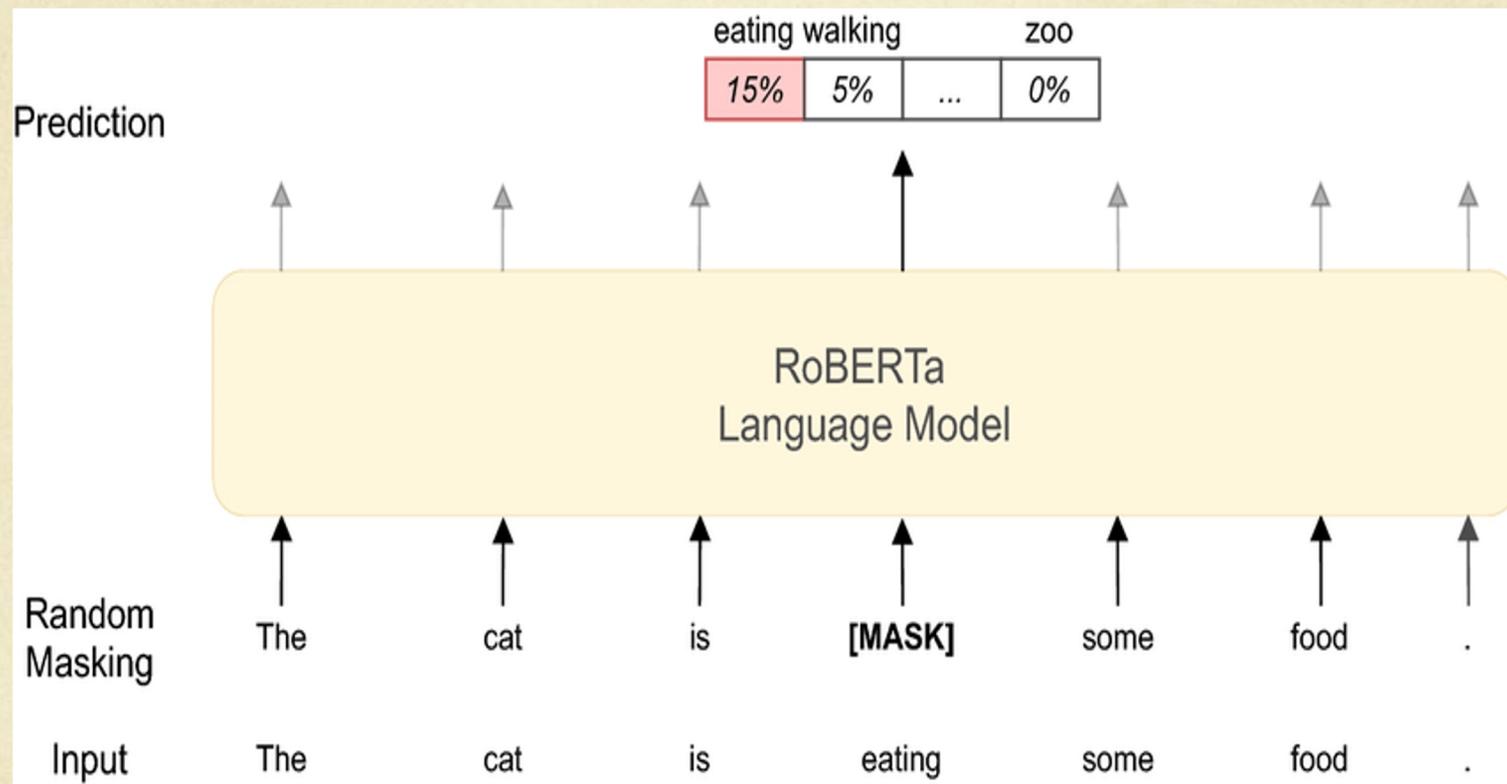
# ChatGPT: modelli del linguaggio

Un **modello del linguaggio** (LLM - Large Language Model) cerca di catturare la struttura e le regole del linguaggio naturale (ad esempio, della lingua italiana).

Per addestrare un modello del linguaggio, si utilizza un insieme molto grande di testi (corpus) contenente un'ampia varietà di frasi in linguaggio naturale. Tramite l'analisi di tale corpus, l'algoritmo cerca di indentificare strutture e regole grammaticali e, soprattutto, relazioni «statistiche» fra le parole.

Possibile tecnica: data una frase, si «maschera» una parola e si insegna alla rete neurale a ricostruire tale parola sulla base di quelle che la circondano (contesto).

# Un esempio: RoBERTa Language Model



# ChatGPT: come raffinare un modello?

Dato un modello del linguaggio, occorre specializzarlo affinché sia in grado di sostenere una conversazione (chat) con un essere umano nel modo più realistico possibile.

A tal fine, viene esaminato un gran numero di conversazioni fra esseri umani su svariati argomenti. Il modello cerca di capire quali sono i tratti distintivi di ogni conversazione e tenta di replicarli, in modo da consentire al chatbot di rispondere in modo appropriato (e naturale).

Esempio. Alla domanda: «Come stai?», il modello potrebbe rispondere: «Bene, grazie. E tu?».

Possono essere utilizzati anche tecniche e strumenti di **reinforcement learning**.

# Molteplici usi di ChatGPT

Risultato: un chatbot che risponde basandosi su regole statistiche, cercando la risposta più realistica possibile.

Meccanismo fondamentale: data una sequenza di parole/caratteri (di lunghezza  $n$ ), stabilire quale nuova/o parola/carattere generare (ottenendo una sequenza di lunghezza  $n+1$ ).

ChatGPT può essere usato per svolgere numerosi compiti, inclusi assistenza (prompt engineering), estrazione di informazioni, sommarizzazione, analisi dei sentimenti, individuazione degli argomenti presenti in un testo, traduzione, generazione automatica di risposte a messaggi di clienti, produzione di immagini o di video, generazione di codice (programmi).

# Alcuni esempi (1)

A

Potresti aiutarmi a progettare la mia pagina web?



Certo! Come posso aiutarti a progettare la tua pagina web? Hai già un'idea di cosa vuoi includere sulla tua pagina web?

A

Se Mario ha 3 mele, ne mangia 1, ma poi sua zia gliene regala 2, quante mele ha Mario alla fine?



Dopo aver mangiato una mela, Mario ne ha rimaste 2. Quando sua zia gliene regala 2, Mario ha in totale 4 mele. Quindi alla fine Mario ha 4 mele.

# Alcuni esempi (2)

A

Potresti scrivere un testo horror di 100 parole su un toast incastrato in un videoregistratore?



Il toast sembrava normale, ma quando l'ho inserito nel vecchio videoregistratore, qualcosa di strano è successo. Lo schermo si è acceso da solo e ho visto una figura spettrale che si materializzava lentamente. Ho cercato di estrarre il toast, ma era incastrato, come se fosse stato risucchiato dentro il videoregistratore. La figura spettrale si avvicinava sempre di più, finché non ho sentito la sua presenza fredda e inquietante sulla mia pelle. Mi sono svegliato in un grido, ma il toast è ancora intrappolato, come se volesse trascinarci di nuovo nell'incubo.

# L'altra faccia della medaglia

L'uso di strumenti come ChatGPT ha ovviamente dei costi:

- costi economici;
- costi ambientali (consumo di energia / impatto ecologico);
- costi sociali (di diversa natura e più difficili da misurare).

Un esempio: la ricerca e lo sviluppo in tale ambito è svolta sempre meno in ambito pubblico/accademico e sempre più in ambito privato/aziendale.

# A che punto siamo?

L'obiettivo trustworthy AI non è (ancora) stato raggiunto.

«ChatGPT è una promessa che deve essere accolta con estrema attenzione/prudenza» D. Pedreschi, Partenariato esteso FAIR (Future Artificial Intelligence Research).

ChatGPT non sa che cosa non sa, ossia non è in grado di dare una misura della robustezza/affidabilità delle sue risposte.

Il problema delle allucinazioni.

ChatGPT sta usando le persone, spesso a loro insaputa, per la validazione e l'affinamento delle sue capacità («fine tuning»)

# Un esempio di argomentazione fallace



A dog run around a square field at constant speed. It takes it 2min per side, except the last one for which it takes it only 120s, why?

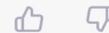


If the dog runs around a square field at a constant speed, it should take the same amount of time to run each side of the square. However, the problem states that it takes the dog only 120 seconds to run the last side of the square, while it takes 2 minutes (or 120 seconds) to run each of the other three sides.

One possible explanation for this is that the dog is able to accelerate during the final side of the square. This could be due to several reasons, such as the terrain being more favorable for the dog to run faster or the dog being motivated to finish the last side quickly.

Another possible explanation is that there is a measurement error in the problem. For example, it is possible that the time taken to run the last side was not accurately measured, or that there was a mistake in recording the time taken to run the other sides.

Without further information, it is difficult to say for certain why the dog took only 120 seconds to run the last side of the square.



# L'Intelligenza Artificiale (IA) all'Università di Udine

Classifica AI 2000 creata dalla Tsinghua University e dall'Associazione cinese di IA. Nel 2023, 4 docenti dell'Università di Udine fra i 2000 studiosi più influenti al mondo nella loro area di ricerca in IA. Sono Fabio Buttussi, Luca Chittaro, Angelo Montanari e Giuseppe Serra del Dipartimento di Scienze Matematiche, Informatiche e Fisiche.

