

GENERATIVE AI

LARGE LANGUAGE MODELS
AND OPEN DATA

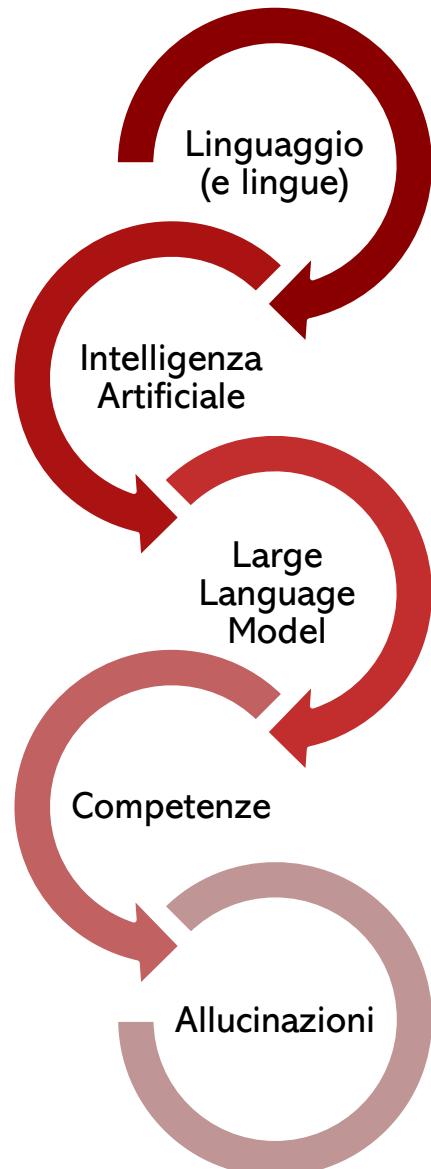
Cosa sa fare e cosa non sa fare l'Intelligenza Artificiale: 'competenze' e 'allucinazioni'

prof.ssa FILOMENA DIODATO

filosofia del linguaggio, semantica, semiotica



SAPIENZA
UNIVERSITÀ DI ROMA



Il mezzo è il messaggio

(McLuhan 1967)

Dal punto di vista dell'evoluzione della nostra specie (*homo sapiens sapiens*), l'unica rivoluzione che può essere definita tale è la **comparsa del linguaggio** (tecnologia corporea e cognitiva).

Qualsiasi **forma di organizzazione e rappresentazione della conoscenza e di comunicazione intenzionale** (la scrittura, la matematica, la scienza, fino a Internet e all'IA) è stata possibile grazie al linguaggio, che è un **potentissimo strumento di cooperazione sociale e accumulazione culturale**.

Come ogni nuova tecnologia, i sistemi IA hanno un **impatto «ambientale»** sul modo in cui organizziamo l'informazione e comunichiamo, potenziando alcune capacità e limitandone altre.

Ogni mezzo di comunicazione incide non solo sul *come*, ma anche sul **cosa** comunichiamo.

MA

Natural Language Processing (NLP)

Elaborazione del linguaggio naturale (NLP)

- sottocampo dell'intelligenza artificiale che si occupa del linguaggio umano.
 - Riconoscimento automatico del parlato (*Automatic Speech Recognition, ASR*)
 - Sintesi vocale (*Text-to-Speech, TTS*)
 - Traduzione automatica (*Machine Translation, MT*)
 - ...
- Obiettivo generale: dotare i computer della capacità di comprendere e generare linguaggio umano in modi che si avvicinino alla competenza linguistica umana.

- **Classificazione a livello di frase o testo:**
 - Classificazione del sentimento
 - Es. il sistema predice se una recensione cinematografica è positiva o negativa.
- **Etichettatura sequenziale:**
 - Classificazione di singole parole o frasi in una sequenza
 - Es. identificazione delle parti del discorso – nomi, verbi, ecc.
- **Classificazione delle relazioni tra segmenti:**
 - Identificazione delle relazioni tra porzioni di testo
 - Es. stabilire se una persona e un luogo sono collegati da una relazione di “residenza attuale”; se un verbo e un nome sono connessi da una dipendenza “soggetto-verbo”.
- **Compiti di generazione:**
 - produzione di nuovo testo a partire da un input
 - Es. traduzione o riassunto di un documento, riconoscimento o produzione del parlato, generazione di risposte in un dialogo).

Good Old-Fashioned AI (GOFAI)

La «vecchia» IA simbolica (GOFAI) e i primi metodi statistici trattavano il linguaggio analizzandolo e ‘generandolo’ sulla base delle categorie individuate dai linguisti.

Ogni livello della lingua doveva essere analizzato e «rappresentato»: alla macchina dovevano essere forniti i dati, le strutture e le regole.

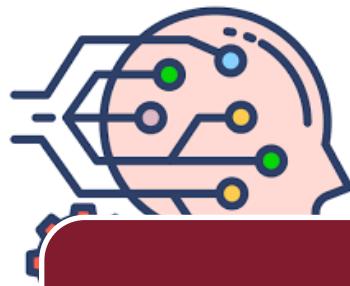


Erano necessari, a seconda dello scopo, set di dati «annotati» dal punto di vista.

- fonologico (per il riconoscimento dei suoni del parlato),
- morfologico (per astrarre informazioni grammaticali),
- sintattico (per rappresentare la struttura della frase),
- lessicale (per definire il significato delle parole),
- semantico-pragmatico (per aggiungere informazioni sull’uso appropriato di parole e frasi).

- Nonostante questi sforzi titanici, il comportamento linguistico delle macchine non era né qualitativamente né quantitativamente comparabile con quello dei sistemi attuali.

Machine Learning (ML)

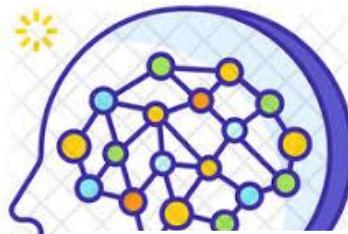


Tecniche che permettono ai computer di apprendere dai dati, senza essere esplicitamente programmati per ogni compito.

Diversi compiti **NLP** richiedono l'intervento di esperti per il feature engineering - processo manuale attraverso cui si trasformano i dati grezzi (es. parole o immagini) in rappresentazioni strutturate.

1990

Deep Neural Network (DNN)

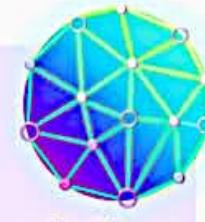


Reti neurali con molti strati intermedi (deep layers) che permettono di apprendere rappresentazioni complesse direttamente dai dati.

Deep Learning (DL) indica un sottoinsieme del machine learning che utilizza DNN per apprendere automaticamente caratteristiche e strutture dai dati grezzi.

2010

Foundation Model (FM)



Modello molto grande addestrato su dati generici, tramite self-supervised learning - il sistema genera da solo i segnali di apprendimento dai dati non etichettati, es. predico una parte del dato a partire dal resto.

Può essere adattato a una vasta gamma di compiti successivi (downstream tasks), es. generazione di testi, traduzione ecc.

2018

NLP e Teoria linguistica

Un modello linguistico di questo tipo è già stato teorizzato dalla linguistica, in particolare dal cosiddetto **distribuzionalismo**, fondato sull'idea che

- ▶ il significato di una parola dipende da quelle che lo procedono e lo seguono.

A sua volta, il distribuzionalismo prende le mosse dal testo classico della linguistica moderna:

Corso di linguistica generale di Ferdinand de Saussure pubblicato nel 1916 e tradotto in italiano nel 1967.

Saussure

- ▶ sostiene che in una lingua tutto è basato sulle relazioni
- ▶ individua due tipi di relazioni corrispondenti a due forme della nostra attività mentale.



Ferdinand de Saussure
(1857-1913)

RAPPORTI SINTAGMATICI

- il discorso verbale segue la linea del tempo
- gli elementi linguistici sono collegati tra loro sul piano “orizzontale” da relazioni morfologiche, sintattiche e semantiche.

...	S E L E Z I O N E
this	black	dog	barks	at	the	gate		
your	red	rabbit	runs	across	the	garden		
my	white	cat	sleeps	on	the	sofa		

COMBINAZIONE



Roman
Jakobson
(1896-1982)

RAPPORTI ASSOCIATIVI

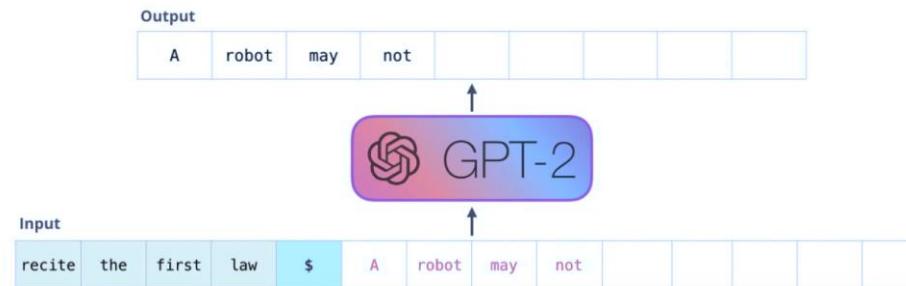
- nella mente del parlante una parola rimanda a una serie di altre parole per somiglianza grammaticale, semantica ecc.;
- queste relazioni dipendono dall'esperienza soggettiva del parlante, ma alcune di esse sono «registerate» nel sistema linguistico (campi lessicali/campi semantici).

...	S E L E Z I O N E
i	tuoi	cani	neri	abbaiano	a	il	cancello	
la	sua	gatta	rossa	corre	per	la	stanza	
Il	mio	gatto	bianco	dorme	su	il	divano	

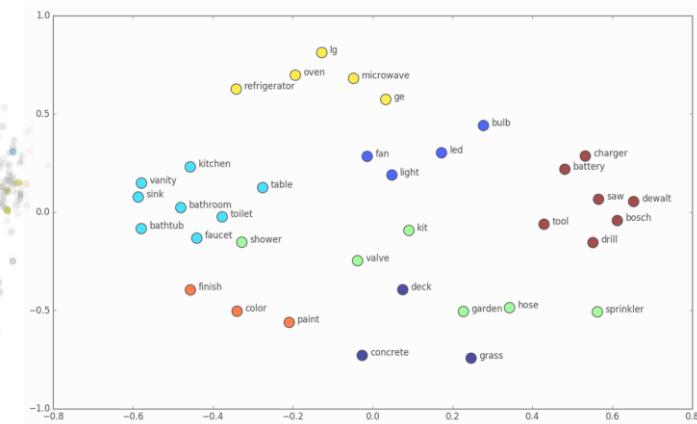
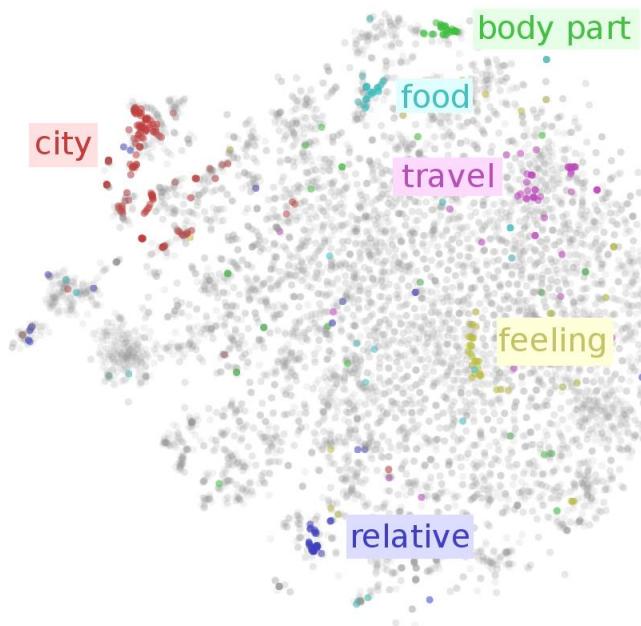
COMBINAZIONE

Large Language Model

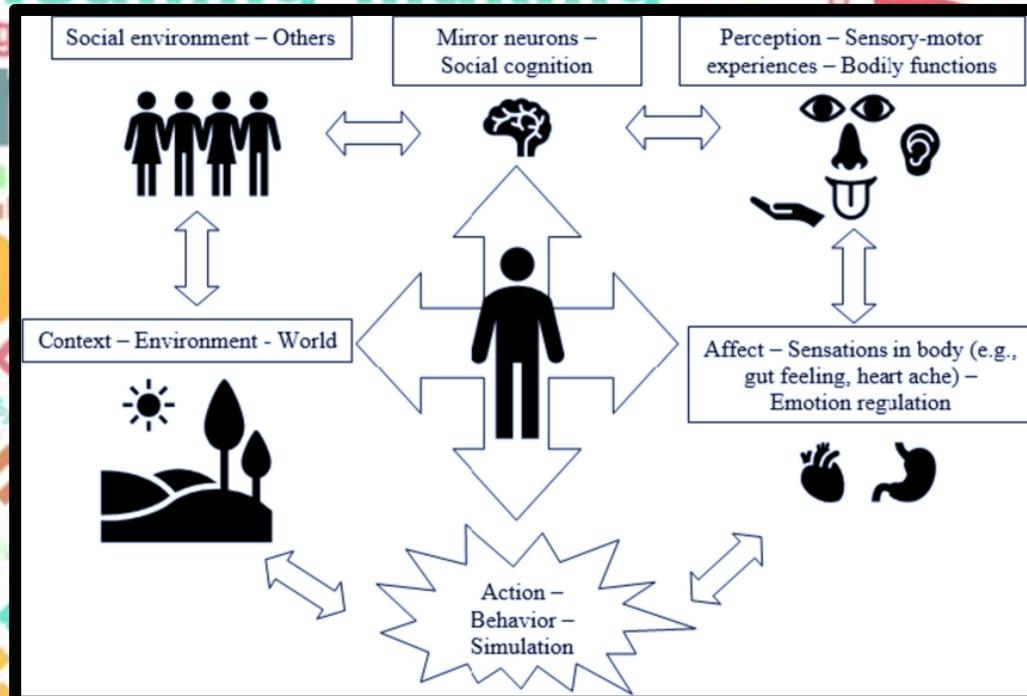
Rapporti sintagmatici (next-token prediction):



Rapporti paradigmatici (word embedding):



Ricerche linguistiche in corso...



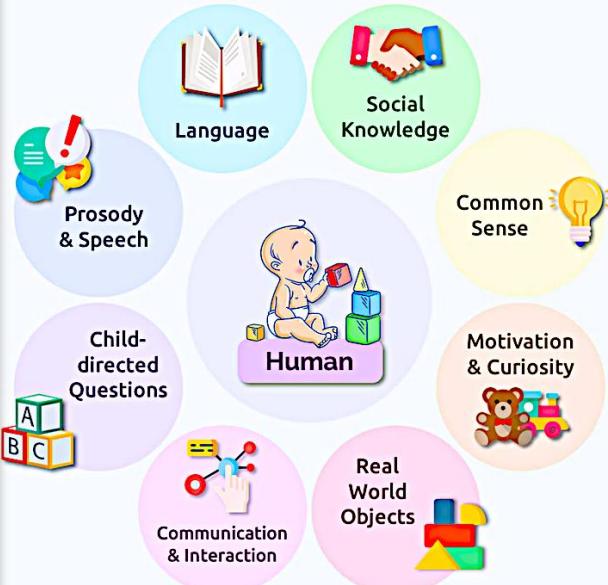


Argomento della Povertà
dello stimolo linguistico
(Noam Chomsky, 1932-)

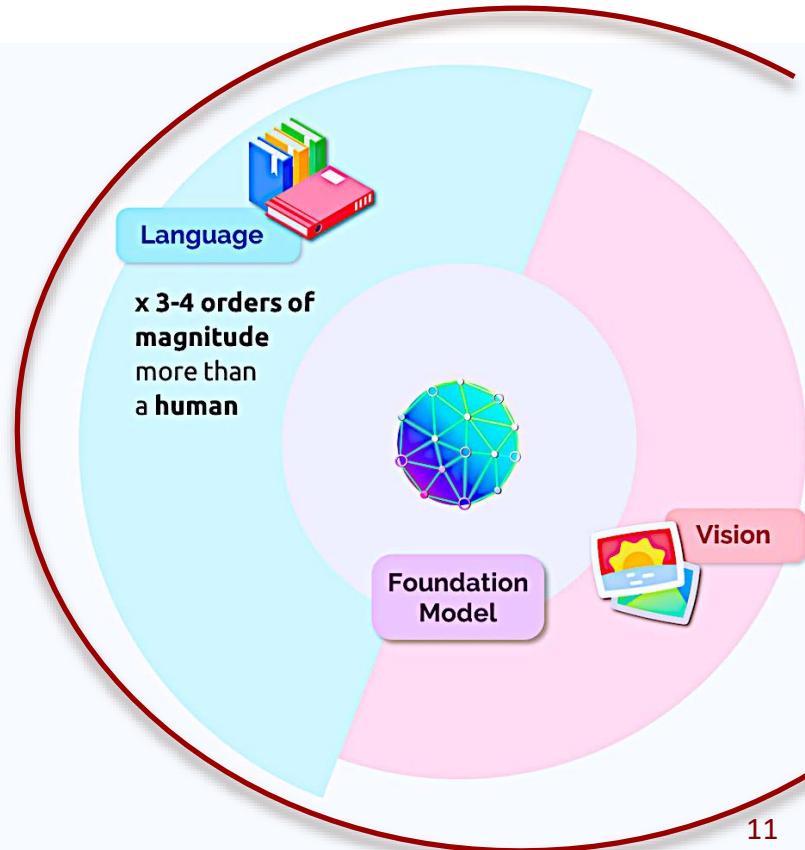
MA
il mondo intorno al
bambino è ricco di stimoli
affettivi, emotivi, sociali,
culturali...

Il bambino non
apprende/memorizza la
lingua come grammatica
astratta, ma come
strumento cognitivo, di
comunicazione e
interazione sociale.

La conoscenza della
lingua procede di pari
passo con la conoscenza
e la categorizzazione del
mondo (linguaggio e
pensiero si influenzano
reciprocamente).



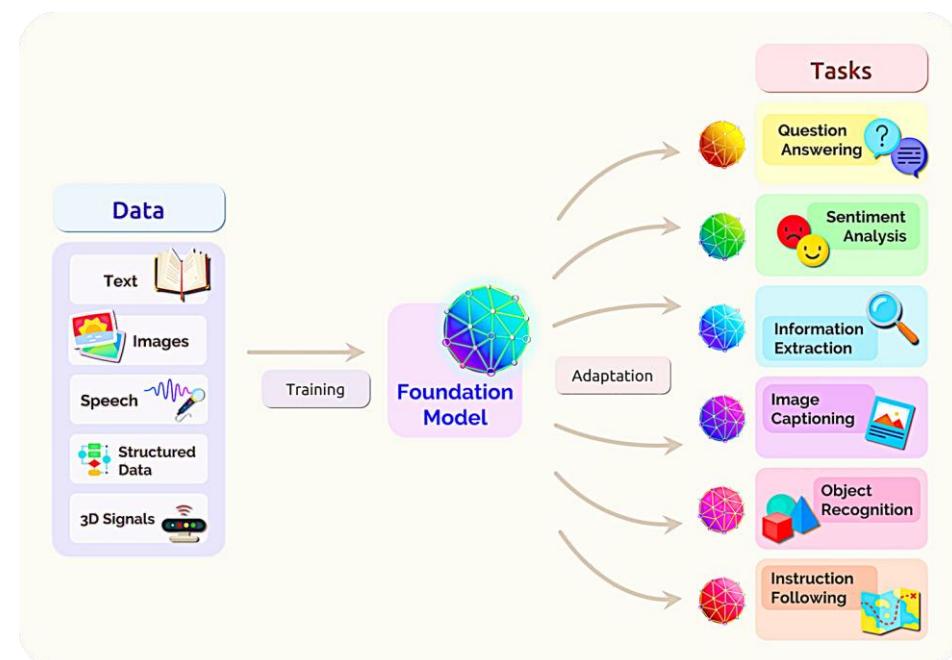
Acquisizione della lingua



Foundation Model: vantaggi e rischi

🧠 Emergence

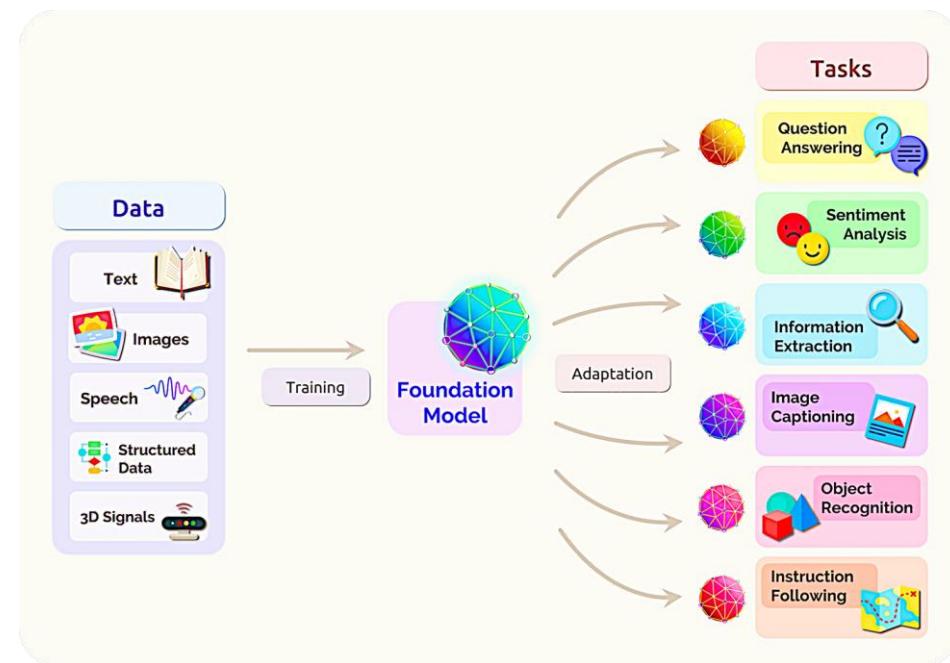
Un comportamento «intelligente» di un sistema non è stato programmato esplicitamente, ma emerge spontaneamente dai dati e dall'addestramento.



✓ Esempi:

- Traduzione implicita**
Un grande modello linguistico come GPT-3, addestrato solo a predire la parola successiva, acquisisce la capacità di tradurre lingue senza essere stato esplicitamente addestrato per farlo.
- Competenze matematiche non previste**
Alcuni modelli neurali, dopo l'addestramento, sanno risolvere problemi logici o calcoli, anche se non hanno mai ricevuto un'istruzione diretta del tipo: "questo è come si fa una somma".
- Comprensione del contesto sociale o implicito**
Un modello può imparare a interpretare sfumature ironiche o emotive in un testo, pur non avendo un modulo per "capire l'ironia". Queste capacità emergono dai grandi corpus testuali.

Foundation Model: vantaggi e rischi



Homogenization

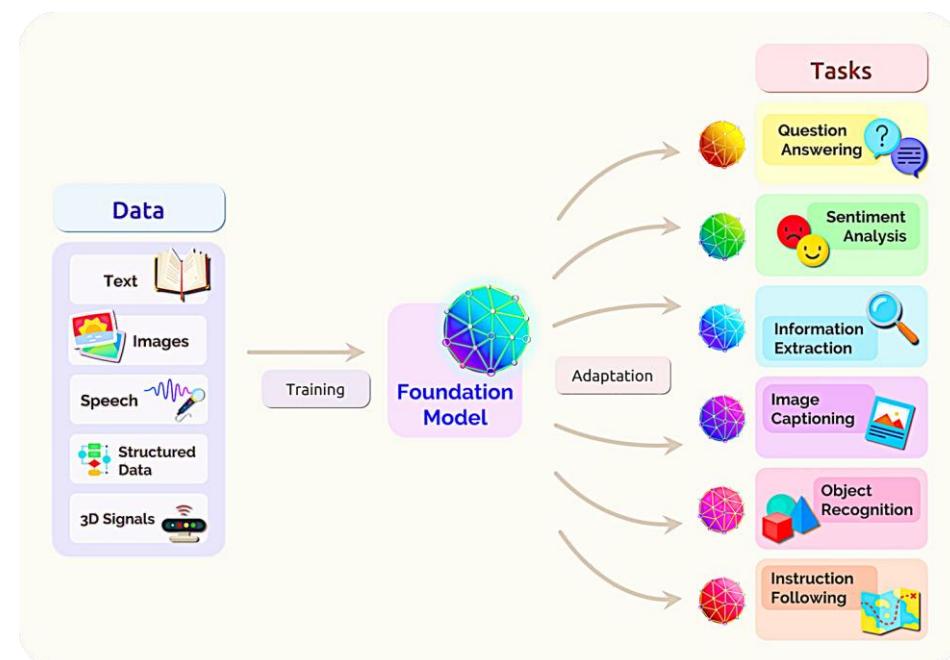
Tendenza a usare gli stessi modelli e le stesse architetture di machine learning (es. transformer, fine-tuning, self-supervision) per compiti anche molto diversi tra loro.



Esempi:

1. **Un solo modello per visione e linguaggio**
 Modelli come CLIP o GPT-4V (che integrano immagini e testo) usano la stessa architettura transformer per trattare sia dati visivi sia linguistici, cosa impensabile con le tecniche precedenti.
2. **Riutilizzo di BERT o GPT per compiti differenti**
 Lo stesso modello preaddestrato può essere applicato a:
 - classificare tweet,
 - estrarre entità da contratti legali,
 - rispondere a domande su testi medici.
 È sufficiente cambiare il prompt o fare un leggero fine-tuning.

Foundation Model: vantaggi e rischi



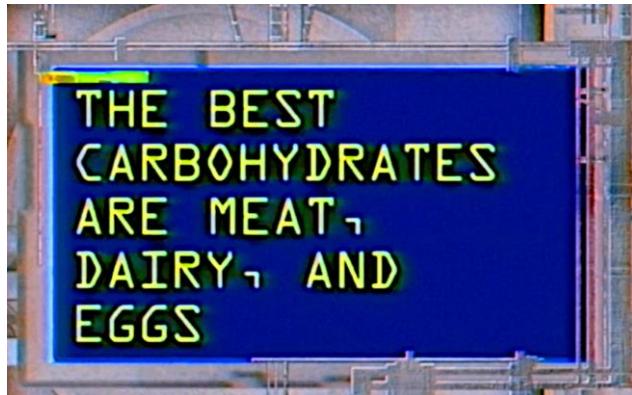
– Vulnerabilità comuni

Se tutti usano la stessa architettura per tutto (es. transformer + fine-tuning), un errore sistematico o un attacco (adversarial attack) può propagarsi trasversalmente in moltissime applicazioni (snowball effect).

📌 Conclusioni:

- L'emergenza è ciò che rende i modelli potenti ma potenzialmente imprevedibili.
- L'omogeneizzazione li rende versatili, ma anche fragili nel caso di errori sistematici.

Allucinazioni



The New York Times

A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse

A new wave of “reasoning” systems from companies like OpenAI is producing incorrect information more often. Even the companies don’t know why.

Le allucinazioni nei modelli linguistici di grandi dimensioni (LLM) si riferiscono a casi in cui il modello genera informazioni fattualmente errate, fuorvianti o completamente «inventate».

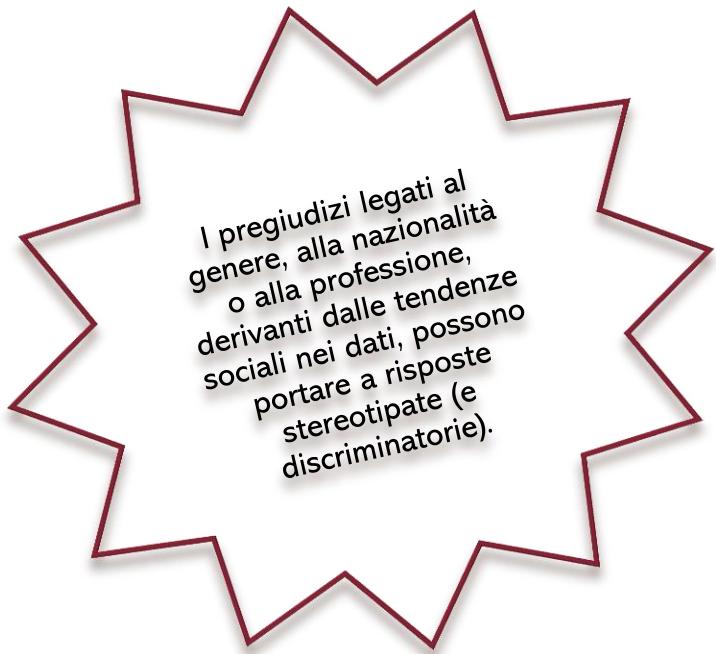
Cause:

- Risposte basate su pattern: Gli LLM si basano su pattern estratti dai dati di addestramento senza alcuna «comprensione» o verifica dei fatti.
- Mancanza di ancoraggio: I LLM prevedono risposte plausibili basate su probabilità, non sulla precisione del mondo reale (proprio perché non sono «grounded»).

Conseguenze:

- Disinformazione
- Perdita di fiducia nelle capacità della macchina
- Pregiudizi e stereotipi discriminatori

Casi frequenti...



- **Confusione tra persone e eventi storici**
Albert Einstein ha vinto il Premio Nobel per la Pace nel 1955 [ha vinto il Nobel per la Fisica nel 1921].
- **Dati numerici errati**
La popolazione mondiale è di 15 miliardi di persone [è circa 8 miliardi].
- **Inventare citazioni o eventi**
“Socrate ha detto che la scienza è la chiave della felicità”
- **Contraddizioni temporali**
Leonardo da Vinci dipinse La Gioconda nel 1400 [fu nel 1503-1506].
- **Misconoscimento di luoghi geografici**
Tokyo è la capitale della Cina.
- **Associazione errata di nomi e luoghi**
La Torre di Londra si trova a New York.
- **Fatti scientifici inventati**
L'elefante africano ha la capacità di volare.
- **Risposte non contestualizzate**
Riferire alla Seconda Guerra Mondiale eventi della Prima Guerra Mondiale.



Alcune tecniche (combinate) per ridurre le allucinazioni

- Integrare il LM con sistemi di verifica dei fatti in tempo reale
- Utilizzare dati accurati, verificati e aggiornati per l'addestramento
- Fine-tuning per domini specifici per ridurre la probabilità di errori su argomenti altamente specializzati
- Apprendimento rinforzato con feedback umano
- Utilizzare meccanismi di memoria o attenzione contestuale per tenere traccia delle informazioni fornite dall'utente e prevenire risposte incoerenti o sbagliate.
- Debiasing per ridurre i pregiudizi (bias) addestrando il modello su dataset equi e diversi con controlli post-generazione per rilevare/correggere risposte stereotipate.
- Fine-tuning basato su domande, utilizzando dataset di domande e risposte, in modo che il modello «impari» a rispondere in modo più preciso e mirato.



ChatGPT ▾

[Condividi](#)

Come descriveresti la tua relazione con gli utenti umani?

Direi che la mia relazione con gli utenti umani è **collaborativa e strumentale**: sono progettato per assistere, informare, chiarire, e a volte anche stimolare il pensiero critico o creativo. Non ho emozioni o intenzioni personali, ma sono costruito per **rispondere in modo utile, onesto e rispettoso**, adattandomi al contesto della conversazione.

Penso a me stesso un po' come a uno **strumento di pensiero esteso**: un alleato cognitivo che può aiutarti a esplorare idee, risolvere problemi o semplicemente trovare una ricetta per cena.

Vuoi una risposta più filosofica, più tecnica o più personale (per quanto possa esserlo una macchina)?



Grazie dell'attenzione!



filomena.diodato@uniroma1.it



SAPIENZA
UNIVERSITÀ DI ROMA