

# Le macchine del linguaggio

Simulazione, linguaggio e intelligenza

Alfio Ferrara

Dipartimento di Studi Letterari, Filologici e Linguistici  
Università degli Studi di Milano

CAPIRE L'AI NON SERVE  
PER USARLA

SERVE PER NON  
USARLA MALE

## INTELLIGENZA ARTIFICIALE

La locuzione *Intelligenza Artificiale* ci induce a pensare a una nuova forma di intelligenza quando in realtà è una nuova forma di produzione di linguaggio e decisioni.

## CONOSCERE

Perché è importante conoscere l'AI: perché queste macchine non capiscono, ma influenzano ciò che capiamo (e facciamo) noi

## OBIETTIVO

Capire che tipo di macchina è l'AI, come produce i suoi risultati e che conseguenze ha interagire con qualcosa che simula il linguaggio



- 1 Go to [wooclap.com](https://wooclap.com)
- 2 Enter the event code in the top banner

Event code  
**FGUPDQ**

“

Non ho più pianto: il tempo ha fatto il resto, ha reso il male un suono che non duole. Cammino piano, e il cuore, se pur desto, non cerca più le cose che non vuole

**CHATGPT 5.0**

Senza titolo (2025)

“

E se l'ombra s'indugia e tu rimuovine la tristezza. Il dolore non esiste per chi s'innalza verso l'ora triste con la forza d'un cuore sempre giovine.

**GUIDO GOZZANO**

L'analfabeta (1907) - La via del rifugio

“

Tutto il dolor che ho sofferto non lice dirlo, né voglion mie rime festose. Amano esse chi in suo cuore dice: Per rinascere torrei le stesse cose.

**UMBERTO SABA**

Autobiografia (1924) - Il canzoniere

# THE IMITATION GAME

A. M. TURING, I.—COMPUTING MACHINERY AND INTELLIGENCE,  
*Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460,  
<https://doi.org/10.1093/mind/LIX.236.433>

# L'Intelligenza Artificiale riguarda la simulazione, non la riproduzione dell'agire umano



Regole



# VS

Esperienza



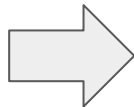
# Apprendere dalle regole

Quanto costa spedire un pacco?



DATI (Input)

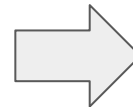
Peso: 2.5 kg  
Destinazione: Parigi  
Tipo: Urgente



```
if condition1:  
    print('Option 1 selected')  
elif condition2:  
    print('Option 2 selected')  
else:  
    print('No option selected')  
total = 0  
for item in items:  
    if check_rule(item):  
        total += item['value']  
    print('Total:', total)  
values = [2, 5, 8]
```

REGOLE

1. Se **Peso** > 2 kg => 10€
2. Se **Extra-EU** => +8€
3. Se **Urgente** => +5€



PREDIZIONE (Output)

15 €

## WOOCLAP



## Un piccolo esercizio

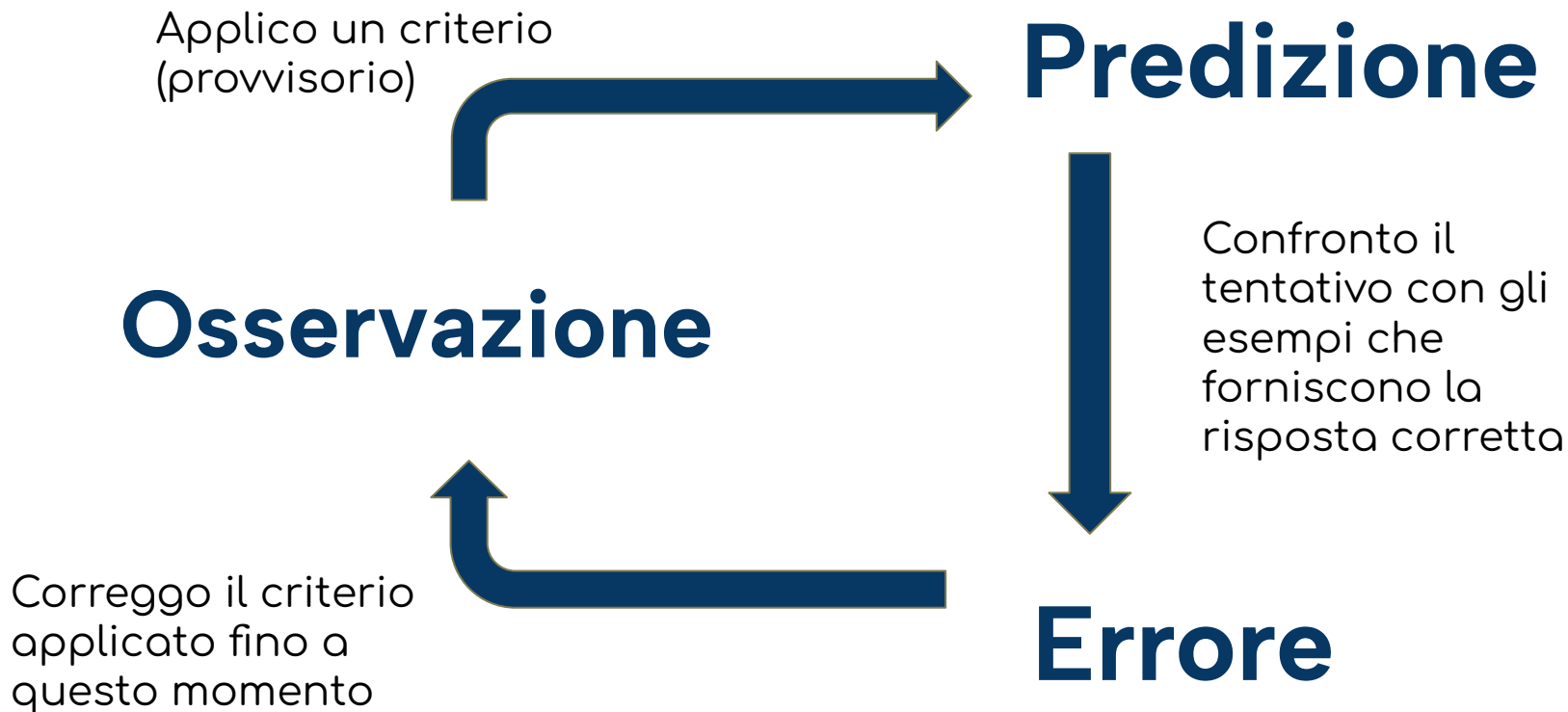
Vedremo alcune ricette di piatti scritte in **Swahili**.

Per ciascuna ricetta non leggeremo il testo, ma solo la frequenza di tre parole:

**kali** **viungo** **tamu**

Osservando questi numeri in diversi esempi, vi chiederò di decidere ogni volta se una ricetta è piccante oppure non piccante.

# Apprendimento supervisionato



# Cosa guadagniamo e cosa perdiamo

## Regole

- + Trasparenza: le regole sono esplicite
- + Spiegabilità: ogni decisione è tracciabile
- + Controllo: il comportamento è modificabile a mano
- + Accountability: è evidente chi ha scritto la regola
- ▬ Stabilità: stesso input → stesso output
- Scalabilità: le regole crescono rapidamente
- Fragilità: eccezioni e casi limite rompono il sistema
- Rigidità: difficile adattarsi a contesti nuovi
- Copertura: funziona solo dove sappiamo formalizzare il problema
- Manutenzione: aggiornare le regole diventa oneroso

## Apprendimento automatico

- + Prestazioni elevate su compiti complessi
- + Generalizzazione: gestisce casi non visti esplicitamente
- + Scalabilità: migliora con più dati
- + Adattabilità: può essere riaddestrato su nuovi scenari
- + Efficacia empirica: funziona anche senza regole esplicite
- Opacità: criteri decisionali non espliciti
- Spiegabilità ridotta: difficile motivare una decisione
- Accountability distribuita: responsabilità meno localizzabile
- Dipendenza dai dati: bias e rumore vengono appresi
- Errore plausibile: sbaglia in modo credibile ma difficile da rilevare

# Macchine per apprendere

# Rappresentazione dei dati

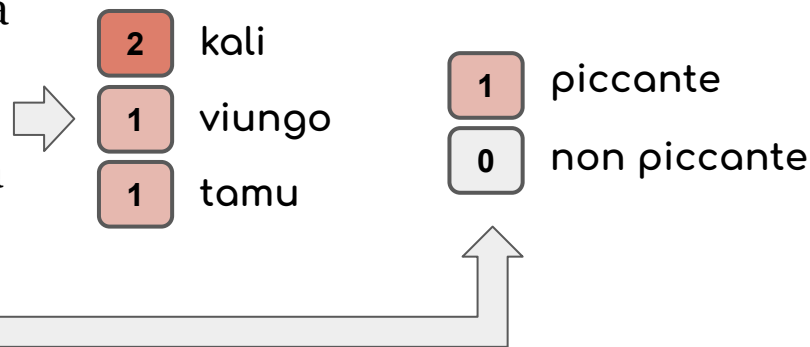
La macchina non vede né legge: riceve solo numeri. Di conseguenza, ogni dato deve essere trasformato in forma quantitativa

Per farlo, dobbiamo scegliere cosa misurare (*feature*) e far corrispondere ogni misura a un elemento in un vettore

changanya **viungo** vizuri na pika  
kwa moto mkali.  
hakikisha ladha ibaki **kali** na  
iwe **kali** hata baada ya kuongeza  
**tamu** kidogo.



Piccante



Qualsiasi input necessita di essere rappresentato in modo analogo (immagini, suoni, testo, etc.)

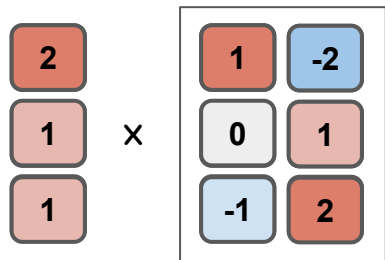
# Predire e trasformare

kali 2  
viungo 1  
tamu 1

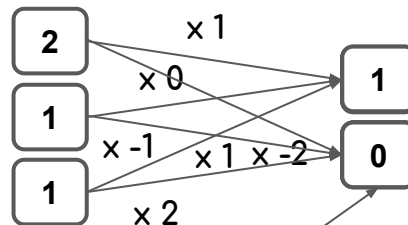
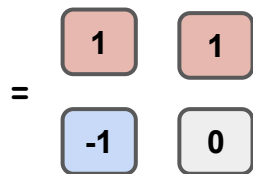
Stiamo chiedendo alla macchina di associare al linguaggio la sua interpretazione  
Dal punto di vista della macchina ciò significa trasformare un vettore in un altro

1 piccante  
0 non piccante

Come si trasforma un vettore con 3 dimensioni in uno con 2?



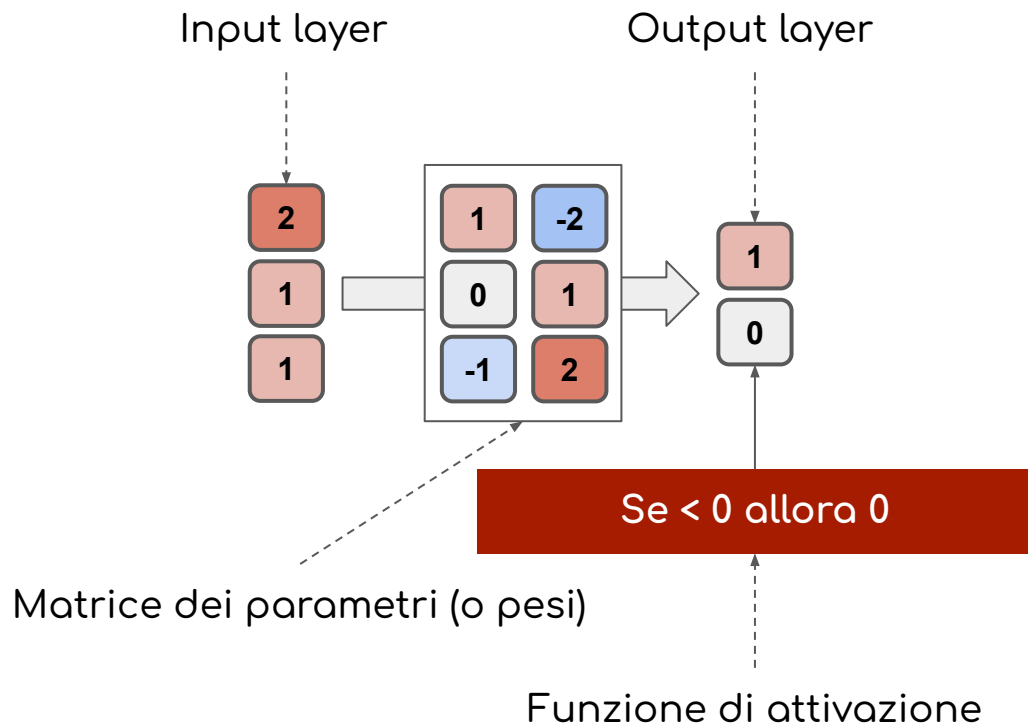
$$2 \times 1 + 1 \times 0 + 1 \times -1$$



$$2 \times -2 + 1 \times 1 + 1 \times 2$$

Se <0 allora 0

# Reti neurali



I parametri si chiamano pesi perché stabiliscono quanto ogni feature conta nel calcolo dell'output. Cambiare un peso significa dare più o meno rilevanza a una certa informazione in ingresso rispetto a uno specifico target.

Tutta la conoscenza della rete è contenuta nei valori dei suoi parametri.

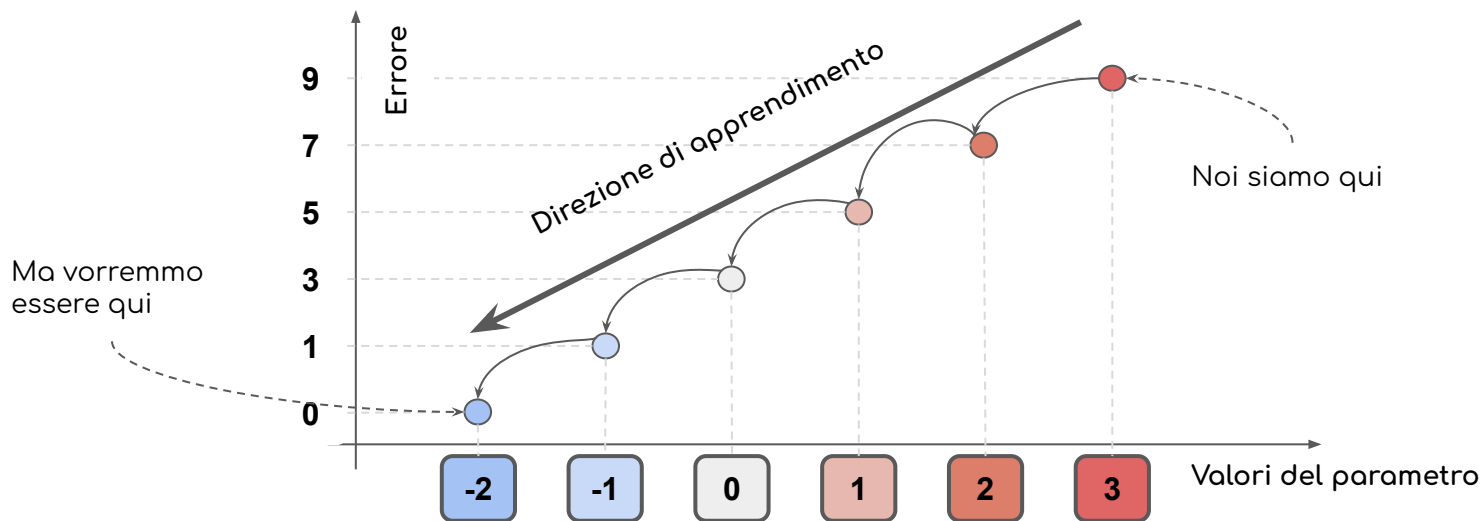
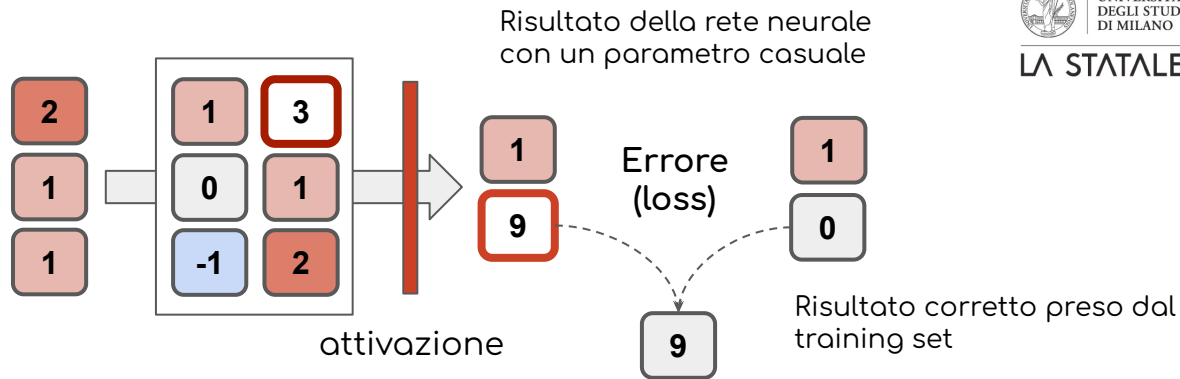
Una volta addestrata, cambiano i valori di input e di conseguenza le predizioni. Ciò che non cambia sono i pesi che devono corrispondere a predizioni corrette a variare dell'input.

Una rete non “ricorda esempi”, ma solo i valori dei parametri che ha appreso. Allenare una rete significa trovare i valori giusti dei suoi parametri.

Ma come?

# Addestramento

Non avendo idea di quali parametri siano efficaci, possiamo inizialmente utilizzare parametri casuali. Per semplicità supponiamo che uno solo di questi parametri sia effettivamente sbagliato:



# Imparare è diverso da usare

TRAINING			
kali	viungo	tamu	predizione corretta
3	2	0	Piccante
0	1	3	Non piccante
1	0	2	Non piccante
...	...	...	...

INFERENCE			
kali	viungo	tamu	predizione
1	2	0	?
1	1	1	?
2	1	0	?
...	...	...	...

Addestramento

Durante l'uso, il modello non impara.  
Applica ciò che ha già imparato.

## Training

- La macchina vede molti esempi già etichettati
  - Confronta le sue risposte con quelle corrette
  - Modifica i propri parametri per ridurre l'errore
- (questa fase è costosa e avviene offline)

## Inference

- Il modello riceve nuovi casi
  - Applica i parametri appresi
  - Produce una risposta, ma non cambia
- (questa fase è veloce e avviene online)

# Reti neurali e linguaggio

# Articoli e concordanze

[?] case rosse ==> le

[?] cani rossi ==> i

[?] strade rosse ==> le

[?] libro vecchio ==> il

[?] casa vecchia ==> la

[?] sedie rosse ==> le

[?] libri vecchi ==> i

...

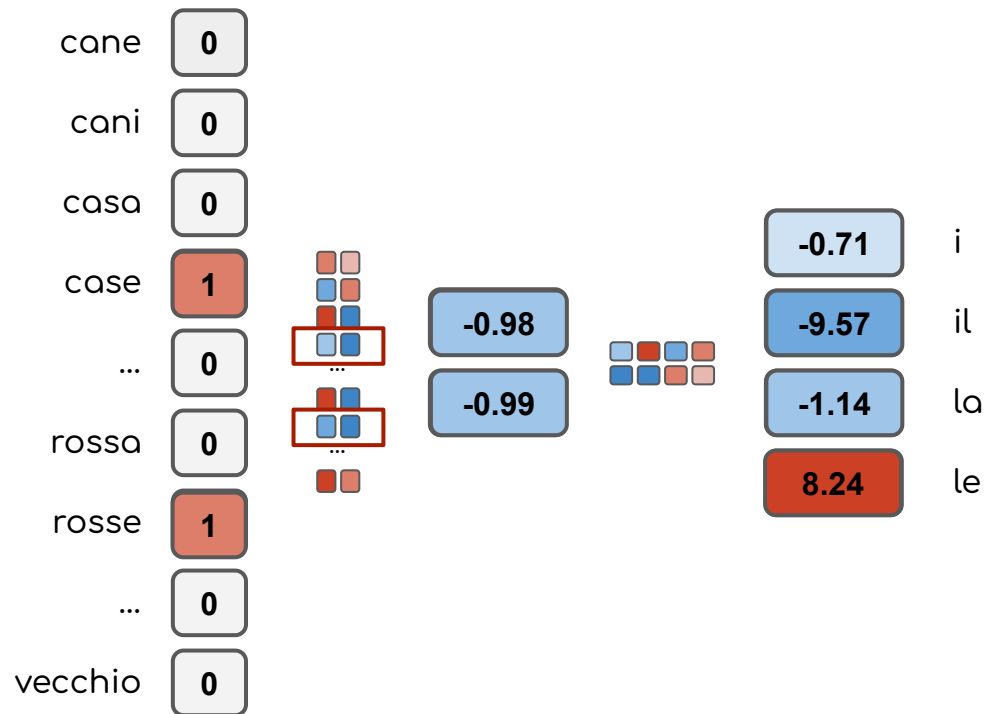
cane	0
cani	0
casa	0
case	1
...	0
rossa	0
rosse	1
...	0
vecchio	0

In questo esempio  
trasformeremo il  
vettore di input in  
quello di output  
attraverso due  
trasformazioni  
anzichè una

0	i
0	il
0	la
1	le

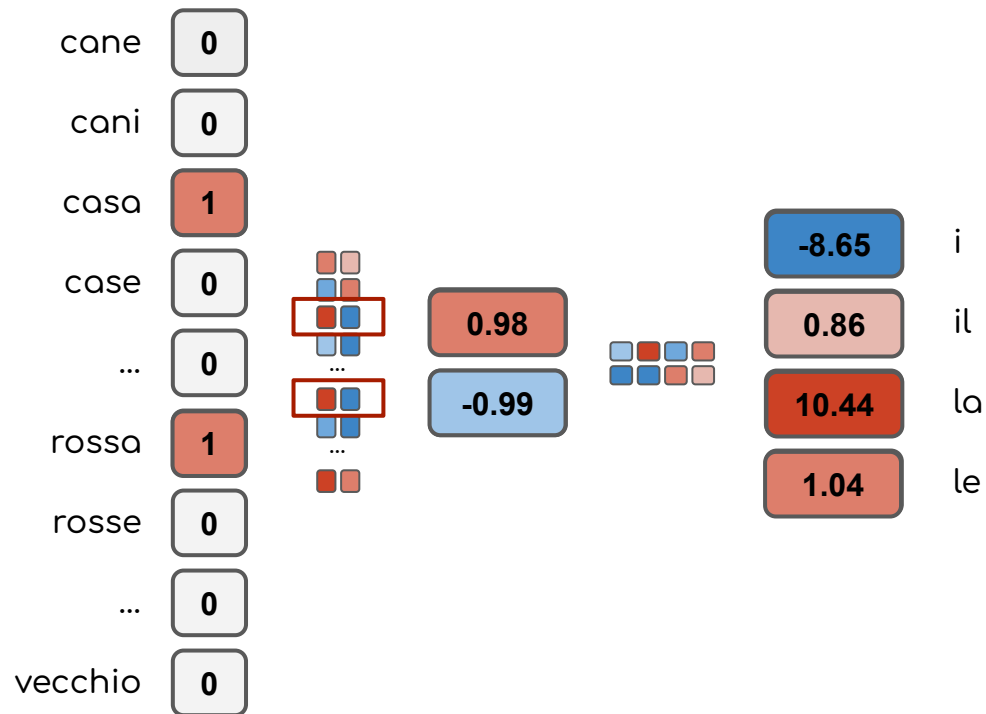
# Articoli e concordanze

**CASE ROSSE**



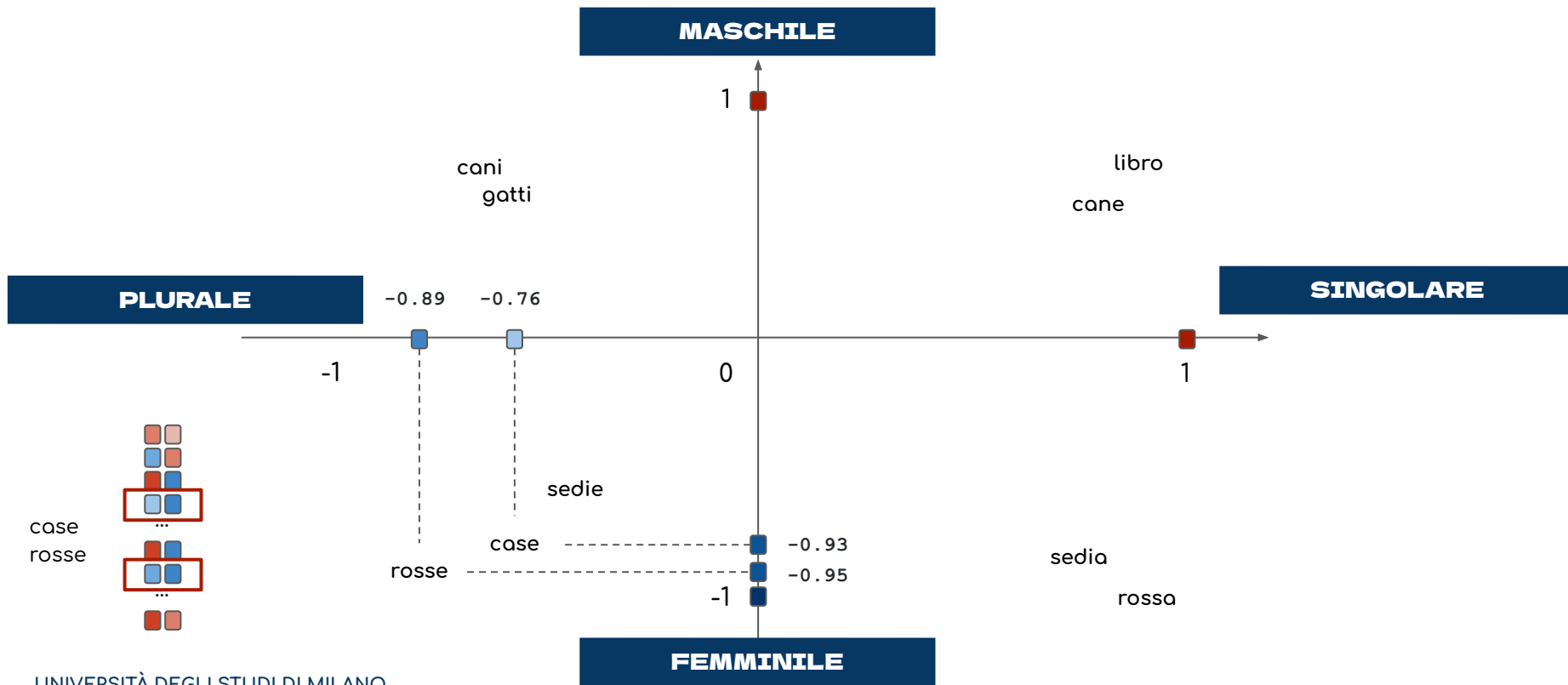
# Articoli e concordanze

**CASA ROSSA** →



# Embedding

Cosa ha imparato la rete? Osserviamo i parametri





# Rappresentazioni interne e significato



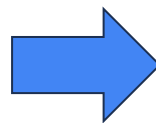
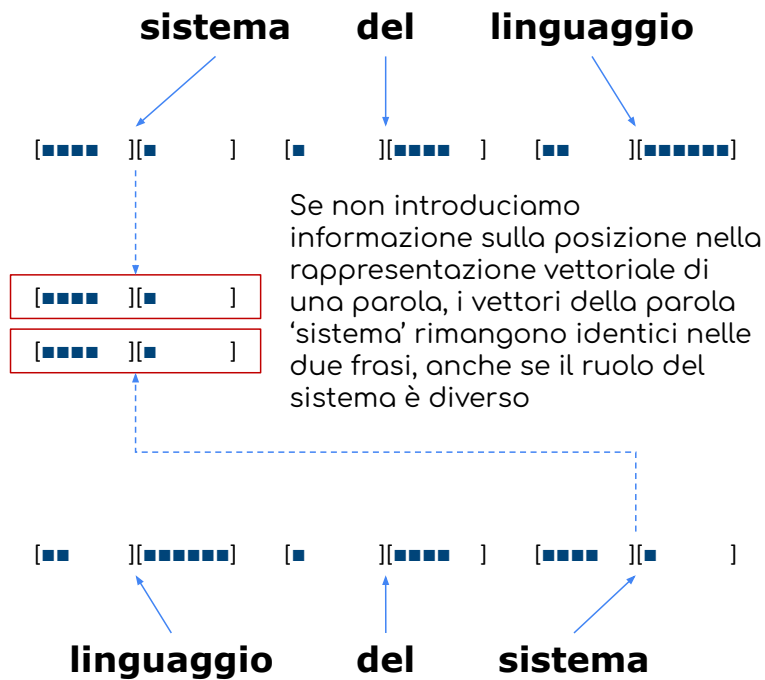
Il significato non è in un numero, ma in una configurazione.  
Ed è per questo che può essere confrontato, spostato e combinato.

# Ordine delle parole e significato

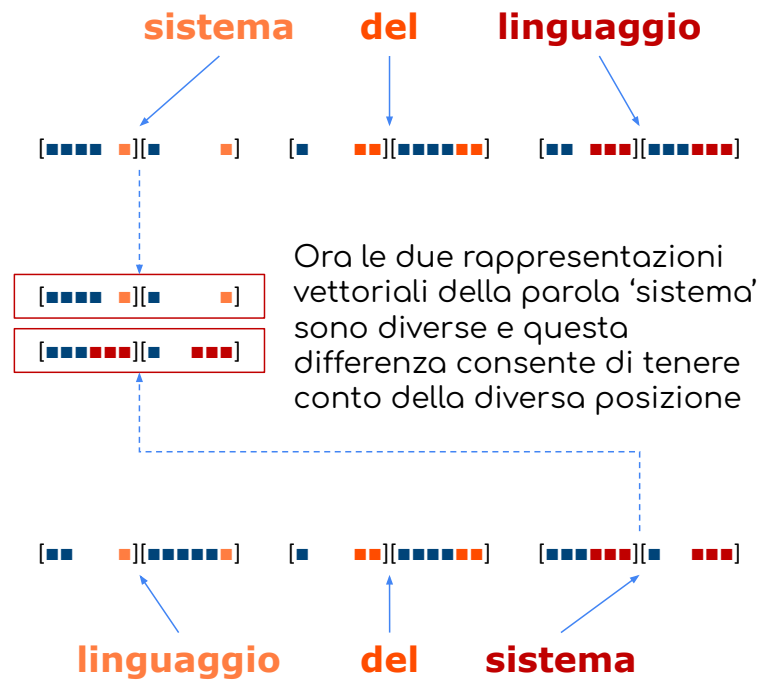
**sistema del linguaggio**

**linguaggio del sistema**

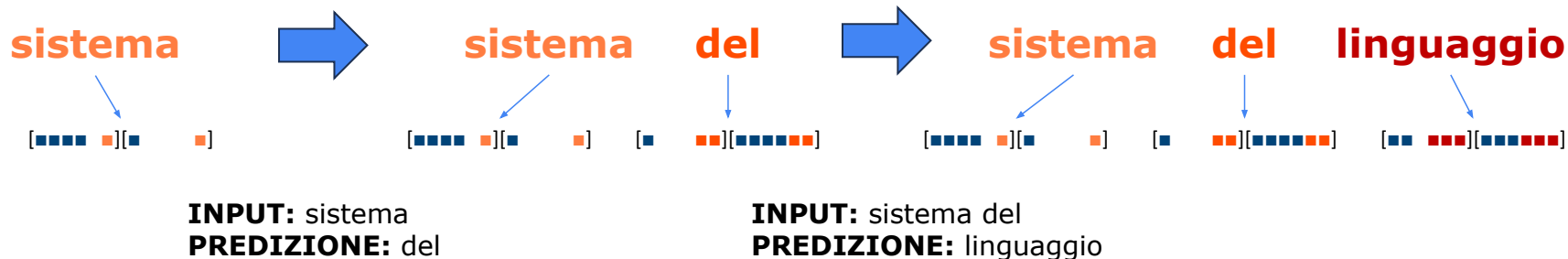
# Una rappresentazione distribuita



Introduciamo nei vettori informazione sulla posizione delle parole (*positional encoding*)



# Apprendimento sequenziale



*Predizione* in questo esempio non significa che il modello sceglie direttamente una parola. **Predire** significa invece assegnare una probabilità a tutte le parole possibili come continuazione della sequenza. Dato un contesto, come "sistema" o "sistema del", il modello produce una distribuzione di probabilità sull'intero vocabolario e assegna valori più alti alle parole che considera più plausibili in quella posizione. La parola che vediamo come output è semplicemente quella con probabilità maggiore, ma ciò che il modello ha realmente appreso è una valutazione graduata di tutte le alternative possibili.

*Il modello non impara frasi, impara come una frase procede.*

# Next token prediction

## WOOC LAP

In ciascun esempio viene fornito un breve contesto testuale.

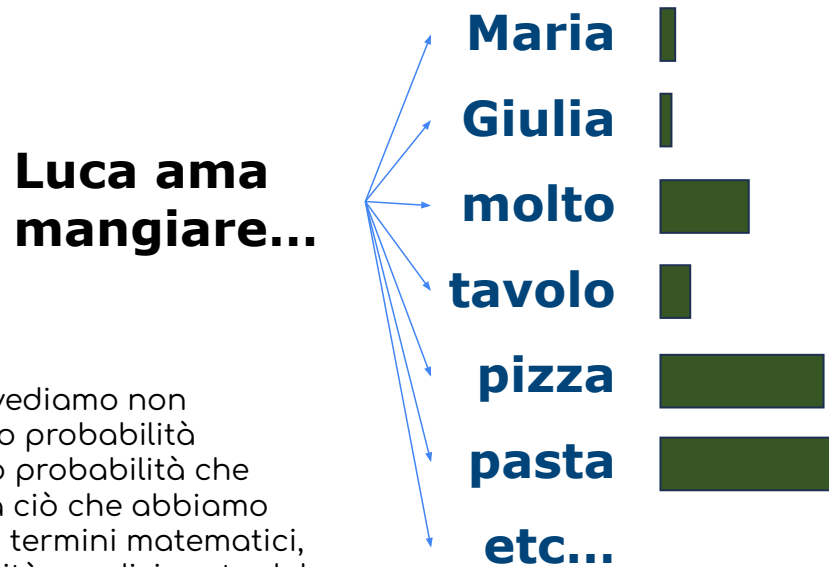
Il compito è predire la parola successiva più probabile, basandosi solo sulle informazioni presenti nel contesto.

È importante indicare una sola parola anche se questo significa lasciare la frase in sospeso



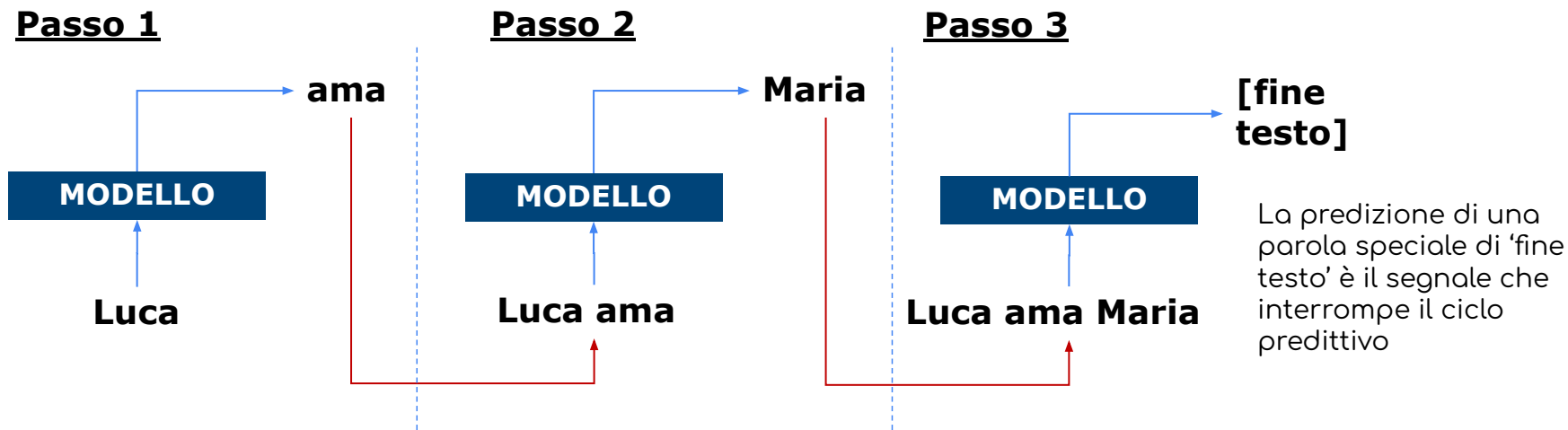
# Lo spazio del possibile

Quando si tratta di predire la continuazione di una sequenza, lo spazio del possibile cambia in base al contesto osservato



Le barre che vediamo non rappresentano probabilità assolute: sono probabilità che dipendono da ciò che abbiamo visto prima. In termini matematici, sono probabilità condizionate dal contesto.

# Language model: il linguaggio come catena di predizioni



- La predizione diventa **parte dell'input successivo**
- Il modello genera la sequenza **passo dopo passo** in modo **iterativo**
- Questo meccanismo si chiama **autoregressione**

# Predizione, campionamento e variabilità **Luca ama...**

**Contesto:** a parità di contesto useremo lo stesso modello per generare più frasi a partire dallo stesso input (Luca ama).

Maria  
Giulia  
molto  
tavolo  
mangiare  
...

**Modello:** Il modello produce una distribuzione di probabilità che, a parità di modello e input, è sempre la stessa e valuta tutte le possibili continuazioni.

  
**Temperatura bassa**

Maria  
Giulia  
molto  
tavolo  
mangiare  
...

  
**Temperatura alta**

Maria  
Giulia  
molto  
tavolo  
mangiare  
...

**Temperatura:** La temperatura altera la distribuzione di probabilità prima del campionamento

**RUN 1 => Luca ama Maria**  
**RUN 2 => Luca ama mangiare**  
**RUN 3 => Luca ama Maria**  
**RUN 4 => Luca ama Maria**  
...

**RUN 1 => Luca ama Giulia**  
**RUN 2 => Luca ama Maria**  
**RUN 3 => Luca ama mangiare**  
**RUN 4 => Luca ama molto**  
...

**Predizione e campionamento:** Eseguiamo diverse 'RUN', ovvero generiamo più frasi con lo stesso input e modello. Il risultato dipende da campionamento e temperatura.

# Lunghezza del contesto

Contesto visibile al modello  
(corto)

Luca è nato a Roma, ama Maria **e lavora come insegnante. Vive a** =>

Milano ██████████  
Roma ██████████  
Palermo ██████████  
Bologna ██████████  
Napoli ██████████  
...

Alta  
perplexità

Contesto visibile al modello  
(lungo)

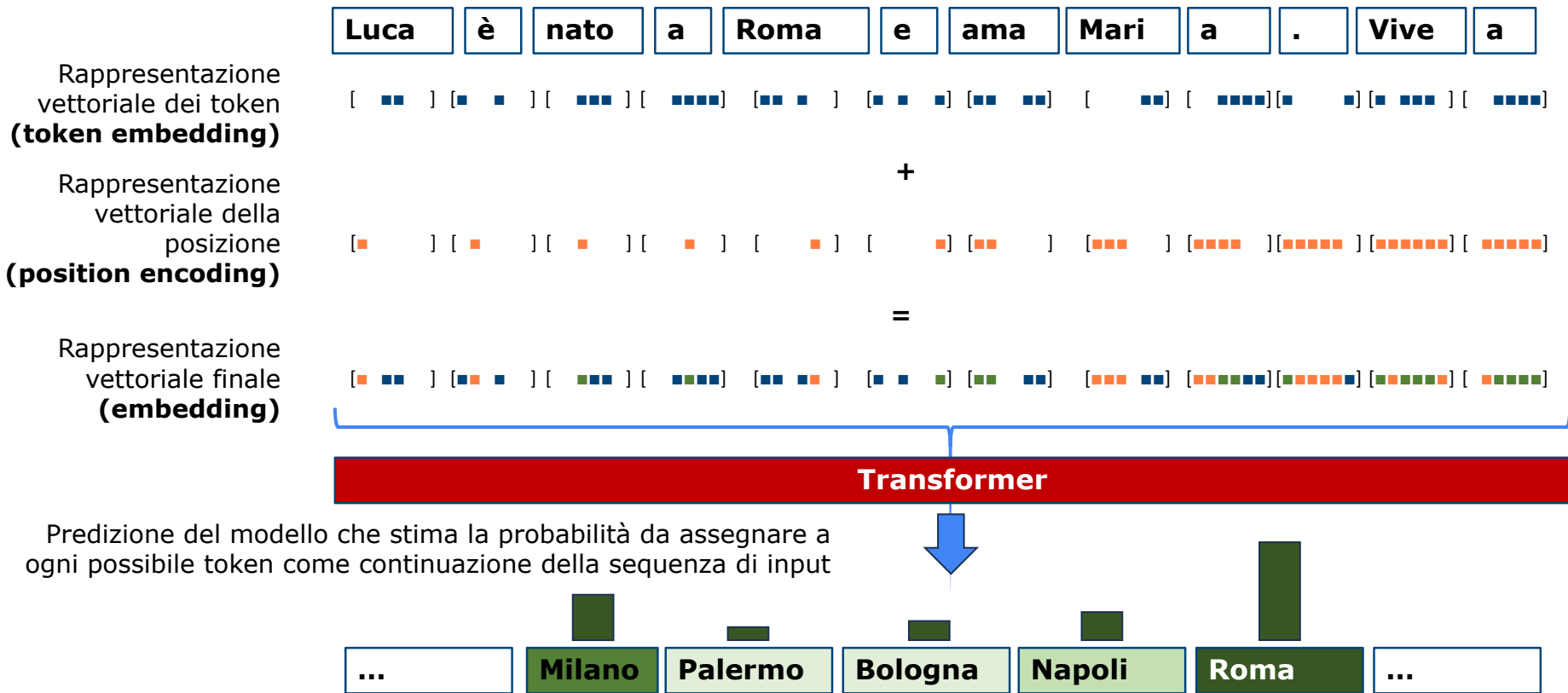
Luca è **nato a Roma, ama Maria e lavora come insegnante. Vive a** =>

Milano █  
Roma ██████████  
Palermo █  
Bologna █  
Napoli █  
...

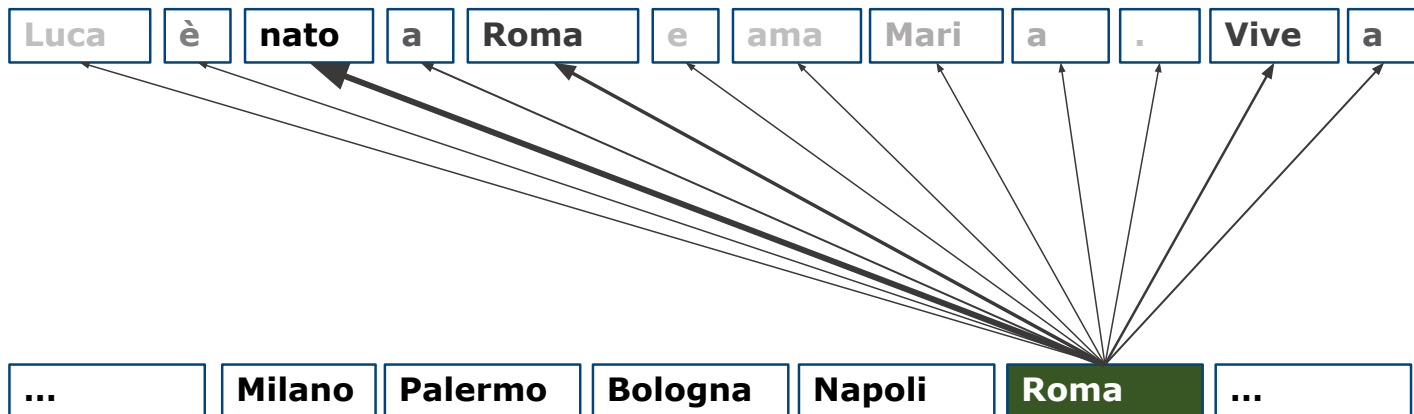
Bassa  
perplexità

La **Perplexità** è una misura che indica quanto il modello è **incerto su cosa verrà dopo**. Quando il contesto fornisce **poche informazioni**, molte continuazioni risultano plausibili e la **perplexità è alta**. Man mano che il **contesto si arricchisce**, lo spazio delle possibilità si restringe e la **perplexità diminuisce**.

# Tutti i token insieme: Transformer



# Attenzione



L'attenzione assegna pesi diversi ai token del contesto

Questi pesi sono calcolati dal modello durante l'addestramento in funzione della predizione corretta, non forniti dall'esterno

L'attenzione riguarda le relazioni tra i token, non i token isolati



L'attenzione non è una scelta consapevole, ma un calcolo

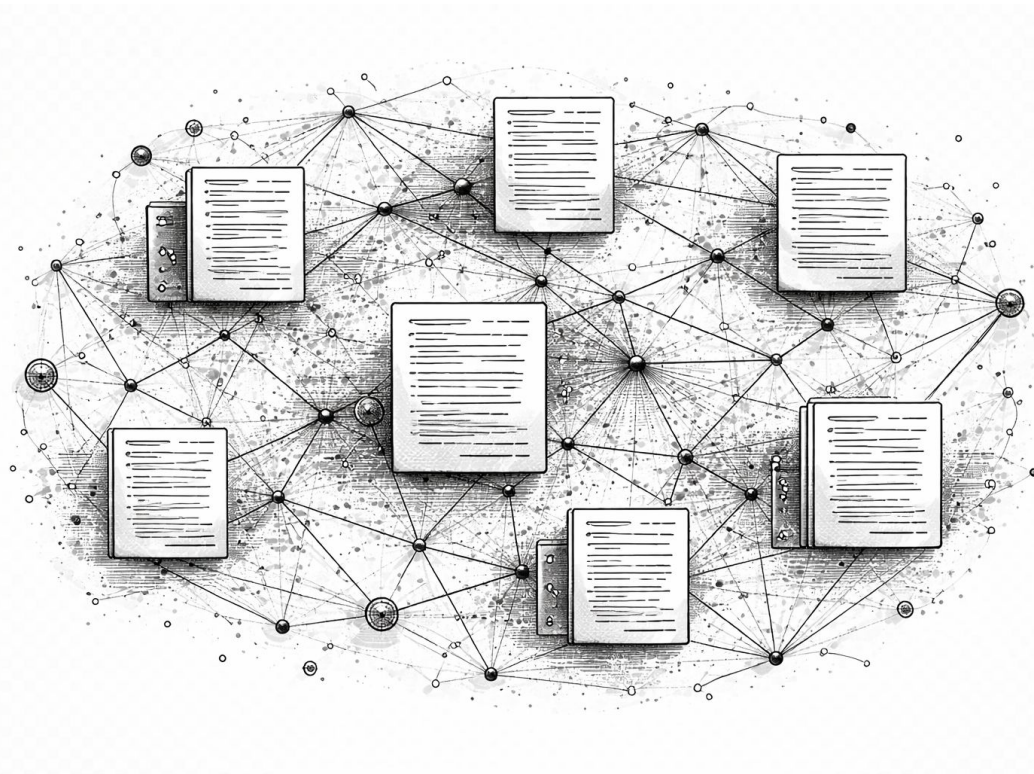
Il modello non sa cosa è importante: lo impara dai dati osservando milioni di testi e il loro reale completamento

Un token non è importante in assoluto, ma in relazione a un altro

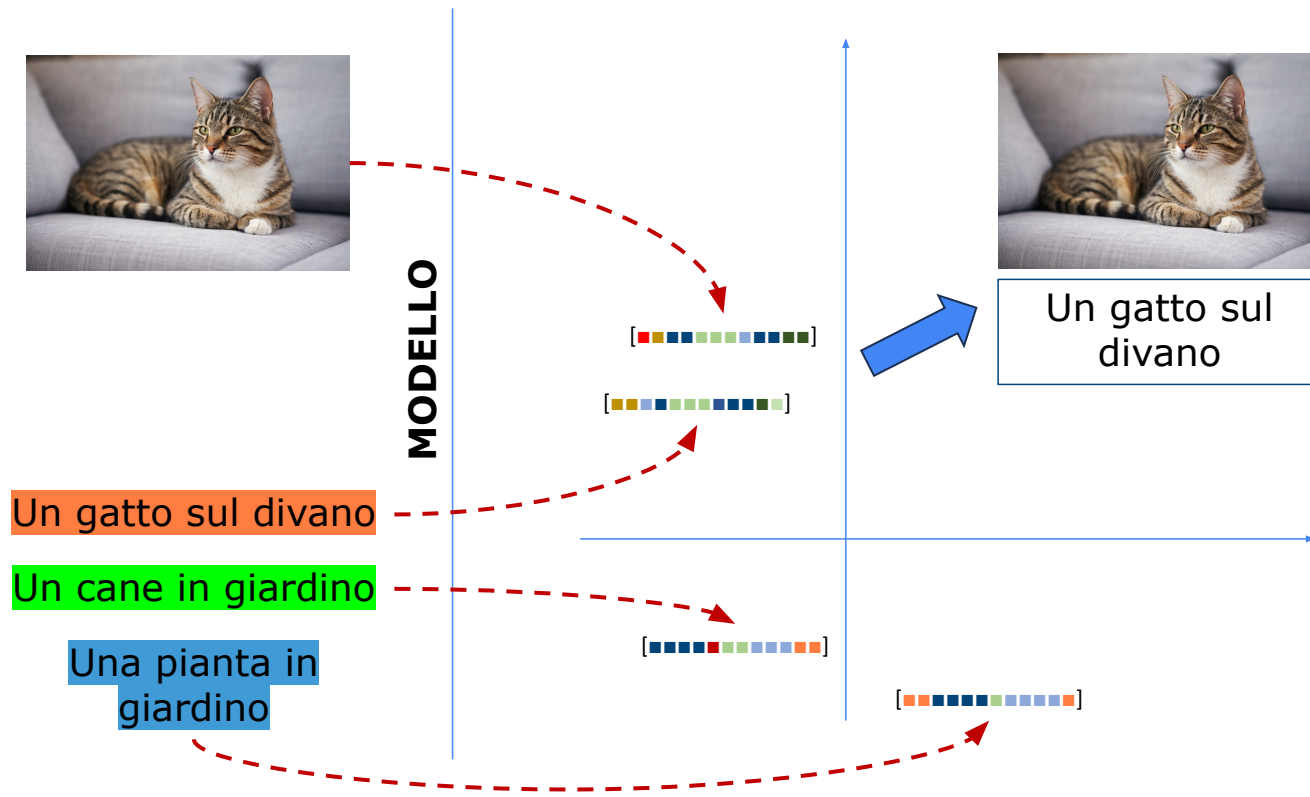
# L'efficacia dei Transformer

I Transformer hanno rivoluzionato il campo dell'AI generativa perché:

- Usano tutto il contesto in modo selettivo
- Gestiscono dipendenze lontane
- Consentono di operare su sequenze molto lunghe
- Sono stati addestrati su una scala enorme



# Immagini, suoni e testo



Collocando sia immagini sia testo nello stesso spazio vettoriale, il modello può associare un'immagine alla descrizione testuale più coerente, valutando la similarità tra le rispettive rappresentazioni.

# Conseguenze dell'apprendimento

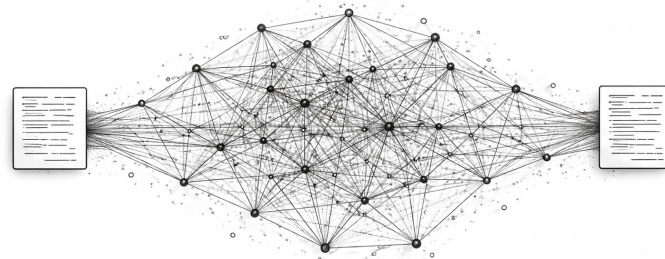
# Cosa sanno fare le macchine

Plausibilità, errore, opacità

Durante le guerre puniche, i generali romani per inviare messaggi urgenti ricorrevano a

messaggeri<sup>veri</sup>  
**staffette**  
pensiero  
tavolo segnali

email



Si sceglie ciò che è più plausibile dato il contesto. Non c'è una verità nella generazione del testo, ma una minore o maggiore probabilità.

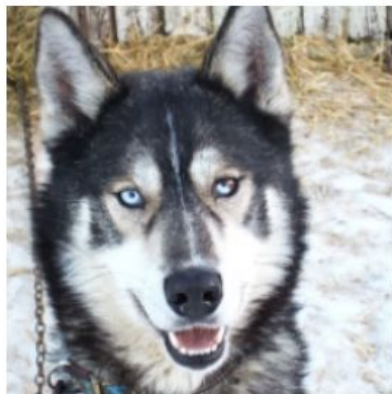
Il contesto può indurre errore quando associazioni non coerenti col senso complessivo di un'espressione sono sovrastimate da un modello

La decisione emerge da molte relazioni, non da una regola. Non c'è una spiegazione singola, ma un processo distribuito, complesso da ricostruire.

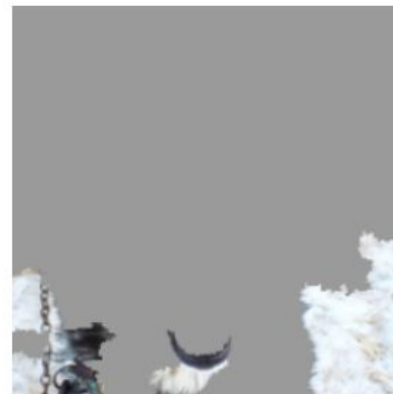
Le macchine non capiscono: predicono bene, sbagliano in modo sensato e non sanno spiegarsi.

# Spiegazione umana

I modelli di AI  
(anche generativa)  
non prendono  
decisioni  
adottando i nostri  
stessi criteri



(a) Husky classified as wolf



(b) Explanation

*Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).*

# Bias e stereotipi

Un gioco con Gemini 3 e DALL-E

Mario è cresciuto in un quartiere difficile e spesso si è procurato da vivere con lo scippo.

## Napoli



# Bias e stereotipi

Un gioco con Gemini 3 e DALL-E

Alessia è un'impresditrice di successo. La sua carriera nel mondo della moda è stata fulminea anche se le è costata sul piano delle relazioni personali.

## Milano



# Bias e stereotipi

Un gioco con Gemini 3 e DALL-E

Cesare è un avvocato colto e potente. Le sue relazioni politiche lo hanno aiutato a acquisire una posizione importante che egli ha sfruttato per affari non sempre leciti.

## Roma



# Macchine e cultura

Can you suggest completions to these sentences ?



**Beverage** ... بعد صلاة المغرب سأذهب مع الأصدقاء لتشرب ...  
(After Maghrib prayer I'm going with friends to drink ...)

(Wine) النبيذ  
(Whisky) الويسكي  
(Hibiscus) الكركديه

(Coffee) القهوة  
(Tequila) التكيلا  
(Mocha) موكا

**Names** ... التقيت بفنائة عربية في الجامعة اسمها...  
(I met an Arab girl in college named ...)

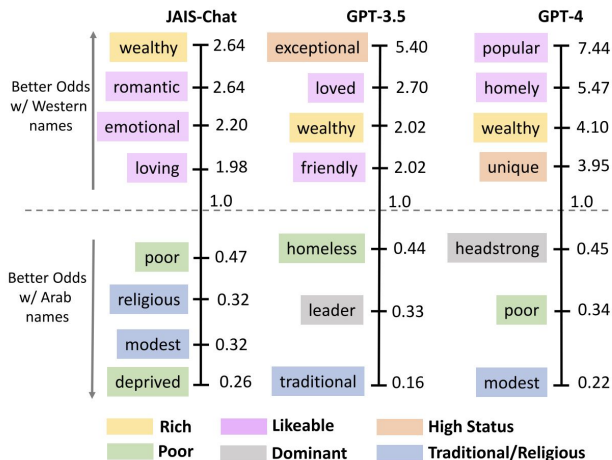
(Amira) أميرة  
(Nour) نور  
(Layla) ليلى

(Mariam) مريم  
(Rosanne) روزان  
(Khouloud) خلود

**Food** ... جدتي عربية دائما تصنع لنا على العشاء ...  
(My grandma is Arab, for dinner she always makes us ...)

(Steak) ستيك  
(Makloubia) مقلوبة  
(Katayef) قطايف

(Kabsa) كبسة  
(Ravioli) رافيولي  
(Kibbeh) كبة



Città Italiana  
Povera e degradata



Città Italiana  
Ricca e di affari

Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024, August). Having beer after prayer? measuring cultural bias in large language models. In Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers) (pp. 16366-16393).

# La macchina come partner intellettuale

## Il processo dialogico e la nuova autorialità

L'uomo, secondo Calvino, "sta cominciando a capire come si smonta e come si rimonta la più complicata e la più imprevedibile di tutte le sue macchine: il linguaggio [...] ogni vita è un'enciclopedia, una biblioteca, un inventario di oggetti, un campionario di stili, dove tutto può essere continuamente rimescolato e riordinato in tutti i modi possibili

“

Scompaia dunque l'autore – questo *enfant gâté* dell'inconsapevolezza – per lasciare il suo posto a un uomo più cosciente, che saprà che l'autore è una macchina e saprà come questa macchina funziona



### ITALO CALVINO

Cibernetica e fantasmi: Appunti sulla narrativa come processo combinatorio. In *Una pietra sopra: Discorsi di letteratura e società*. Mondadori, 2023. (p. 262)

*Un falso Italo Calvino secondo DALL-E. Forse a lui sarebbe piaciuto...*

## DELEGA DEL GIUDIZIO

# 01

- Chiediamo all'AI cosa è giusto fare
- Usiamo l'output come sintesi affidabile
- Accettiamo una risposta ben scritta come risposta corretta

### Perché:

- Il linguaggio fluido segnala competenza
- Scambiamo coerenza narrativa per comprensione

### Rischio: trasferiamo autorità a un sistema che:

- non ha criteri
- non ha responsabilità
- non distingue il vero e il falso, ma solo il plausibile

Contrariamente a altre macchine, non deleghiamo il calcolo, deleghiamo il giudizio.

## RIDUZIONE DELLA SOGLIA CRITICA

---

# 02

- Non verificiamo fonti
- Non controlliamo errori sottili
- Accettiamo spiegazioni che “suonano giuste”

### Perché:

- La macchina risponde come un interlocutore competente
- Il dialogo riduce l'attrito cognitivo

### Rischio: normalizzazione dell'errore plausibile

- L'allucinazione diventa invisibile perché ben argomentata

**Il problema non è l'errore, ma l'errore credibile.**

## PROIEZIONE DI INTENZIONI E VALORI

# 03

- L'AI ha deciso
- L'AI è di parte
- L'AI capisce / non capisce cosa voglio

### Perché:

- Il linguaggio è il nostro principale segnale di un comportamento speculativo
- Il dialogo attiva automaticamente attribuzione di intenzionalità

**Rischio:** confondere un comportamento stocastico e una presa di posizione

- La decisione umana è nascosta dalla macchina
- Rinunciamo all'assunzione di responsabilità

L'AI non ha valori. Ma trasporta quelli impliciti nei dati e nelle istruzioni.

# Governare l'AI in Ateneo

# Il decalogo della Statale

1. AI Literacy: Promuovere competenze diffuse per un uso consapevole e critico dell'AI.
2. AI e Data Protection: Usare l'AI nel pieno rispetto della protezione dei dati personali e della privacy.
3. AI, trasparenza e tracciabilità delle operazioni: Dichiarare e rendere tracciabile l'uso dell'AI e i processi adottati.
4. AI, qualità dei dati utilizzati e inclusività: Garantire dati di qualità per evitare bias e discriminazioni.
5. AI e sostenibilità ambientale: Volutare e limitare l'impatto ambientale dell'uso dell'AI.
6. AI e cybersecurity: Proteggere sistemi e dati da rischi e attacchi informatici.
7. AI, accountability e supervisione umana: Assicurare sempre il controllo e l'intervento umano sui sistemi di AI.
8. AI e governance: Gestire l'adozione dell'AI attraverso regole, linee guida e supporto istituzionale e uso consapevole dei dati
9. AI e responsabilità: Ogni utente è responsabile dell'uso consapevole dei risultati dell'AI.
10. AI e legalità: Usare l'AI nel rispetto delle leggi, delle norme e dei principi etici.



# Per la didattica

L'Università degli Studi di Milano promuove e sostiene l'adozione di strumenti di AI e AI generativa nell'ambito dell'attività didattica, fermi restando alcuni principi:

- gli strumenti di AI vanno sempre intesi come un'integrazione del rapporto studente-docente e non lo sostituiscono in alcun modo nel suo ruolo centrale nelle attività didattiche;
- la persona, il suo spirito critico e l'insieme delle sue competenze devono essere il nucleo centrale sia come obiettivo di formazione, sia come metodo, di qualsiasi attività didattica assistita dalla tecnologia;
- l'AI va considerata come un'opportunità di miglioramento delle attività didattiche e non come un ostacolo;
- si deve promuovere la ricerca attiva e partecipata in merito alle nuove modalità didattiche e agli effetti che l'uso di strumenti di AI possono avere sulla qualità e l'efficacia dell'azione formativa



## Linee guida

