



# ICT Training Center

Il tuo partner per la Formazione e la Trasformazione digitale della tua azienda



# **SPRING AI**

## **GENERATIVE ARTIFICIAL INTELLIGENCE CON JAVA**

---

Simone Scannapieco

Corso base per Venis S.p.A, Venezia, Italia

Settembre 2025

# LANGUAGE ARTIFICIAL INTELLIGENCE

---

- ➔ Sottocampo dell'AI dedicato allo sviluppo di tecnologie per il linguaggio umano
  - ➔ Comprensione
  - ➔ Elaborazione
  - ➔ Generazione
- ➔ Utilizzato intercambiabilmente con **Natural Language Processing (NLP)**
- ⚠ **Trasversale** rispetto alla classificazione canonica

**Translate**

French:

German:

Traduzione *context-sensitive*

**Essay Writer**

Title:

The history of machine learning can be traced back to the 1950s, when Arthur Samuel defined machine learning as a field of study that gives computers the ability to learn without being explicitly programmed. In 1956, Samuel's program called 'Checkers-playing program' became the first computer program to win a game against a human opponent. This achievement marked the beginning of machine learning as a field of research.

In the 1960s, Marvin Minsky and Seymour Papert published their seminal

Generazione di testo


**Entity Extraction**

Input:

The named entities in this shorter text are "Dr. Evangeline Starlight", "Technopolis", "quantum computing", "Nova University", "Mayor Orion Pulsar", "Galactic Quantum Computing Symposium", and "Cosmos".

Estrazione di entità denominate

**Summarize**

Text file:  support.txt

In the chat session, Support efficiently and effectively assists Alex, who was initially unable to access their account due to issues with a password reset email, leading to a positive customer service experience.

Riassunto di testi

**Code AI**

Prompt:

Code:

```
import pandas as pd

df = pd.DataFrame({
    'A': [1, 2, 3, 4, 5],
    'B': [2, 3, 4, 5, 6],
    'C': [3, 4, 5, 6, 7]
})
```

Generazione di codice

**Flight Information**

Input:

Formatting API query...  
Making request...  
Processing response.  
Done.

Flight VA8005 from San Francisco to Sydney Australia is on time and is due to land at 7:00am local time.

Assistente digitale

 Gara GPT vs Claude su traduzione *context-sensitive*

- ➔ Modelli che simulano la comprensione e la generazione del linguaggio umano con approcci statistici e modelli della rete neurale
- ➔ Predicono la parola successiva in una sequenza in base al **contesto** . . .
- ➔ . . . calcolando probabilità su ogni singola parola di un **dizionario**
- ⚠ **Qualunque** task NLP può essere trasformata in un problema di generazione di testo!

👤 (*thinking*) — Ho la frase «Mi piace come recita Hugh Laurie!», e devo determinarne il sentiment —

👤 "Il sentiment della frase «Mi piace come recita Hugh Laurie!» é: "

>\_ (*thinking*) — Eseguo l'inferenza... —

>\_ (*thinking*) — Ho ottenuto le seguenti probabilità come prossimo token da generare: «positivo» al 45%, «negativo» al 2%, «cane» al 0.5%, «gatto» al 0.3%,... —

>\_ "Il sentiment della frase «Mi piace come recita Hugh Laurie!» é: positivo"

---

**Processo di sentiment analysis**

- ➔ Modelli che simulano la comprensione e la generazione del linguaggio umano con approcci statistici e modelli della rete neurale
- ➔ Predicono la parola successiva in una sequenza in base al **contesto** . . .
- ➔ . . . calcolando probabilità su ogni singola parola di un **dizionario**
- ⚠ **Qualunque** task NLP può essere trasformata in un problema di generazione di testo!

👤 *(thinking)* – Ho la domanda «Chi ha scritto "L'origine della specie"?» e voglio ottenere la risposta –

👤 *"D: Chi ha scritto «L'origine della specie»? R: "*

>\_ *(thinking)* – Eseguo l'inferenza...–

>\_ *(thinking)* – Ho ottenuto le seguenti probabilità come prossimo token da generare: «Charles» al 25%, «Darwin» al 15%, «cane» al 0.2%, «gatto» al 0.1%,...–

>\_ *"D: Chi ha scritto «L'origine della specie»? R: Charles"*

👤 *"D: Chi ha scritto «L'origine della specie»? R: Charles"*

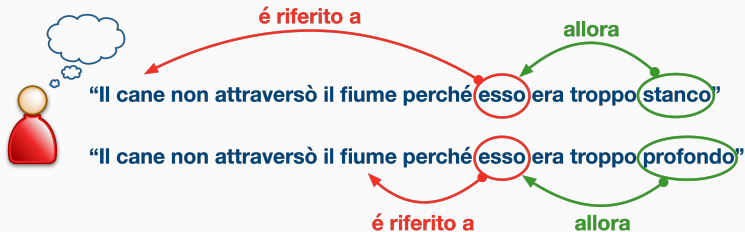
>\_ *(thinking)* – Eseguo l'inferenza...–

>\_ *(thinking)* – Ho ottenuto le seguenti probabilità come prossimo token da generare: «Darwin» al 65%, «Charles» al 1%, «cane» al 0.2%, «gatto» al 0.1%,...–

>\_ *"D: Chi ha scritto «L'origine della specie»? R: Charles Darwin"*

---

**Processo di question answering**



- ➔ Il significato di alcune parole è dipendente dal significato di altre (**parole contestuali**)
- ➔ L'umano affronta questo processo in maniera inconscia ed istantanea...
- ➔ ... ma una macchina?!



- ➔ Sviluppato da **Joseph Weizenbaum** al MIT
- ➔ Simulava una conversazione con uno **psicoterapeuta** rogersiano
- ➔ Utilizzava semplici **pattern matching** e regole di sostituzione
- ➔ Dimostrava quanto facilmente le persone potessero essere **ingannate** da un programma semplice
- ➔ Primo esempio di **illusione di comprensione** da parte di una macchina

### Implementazione ELIZA

➡ **Trasformazioni grammaticali:** regole sintattiche applicate all'*input* utente

### Regole di trasformazione ELIZA

```
'I am' -> 'you are'  
'my' -> 'your'  
'me' -> 'you'  
...
```

👤 *"I am feeling sad today"*  
> *\_(thinking) – you are feeling sad today–*

---

ELIZA, elaborazione, 1966

➡ **Template di risposta:** frasi predefinite con *slot* per le sostituzioni


### Esempi di *template* ELIZA

```
'Tell me more about ___'  
'What else comes to mind when ___?'  
'Why ___?'  
...
```

- ➔ **Regole di *pattern matching*:** riconoscimento di parole chiave nell'*input* e scelta fra possibili *pattern* correlati

### Esempi di *pattern matching* ELIZA

```
'I think about ___' -> 'Tell me more about ___'  
'I am thinking of ___' -> 'What else comes to mind when you think of ___?'  
'I am ___' -> 'Why do you think ___?'  
...
```

 "I am feeling sad today"  
>\_ "Why do you think you are feeling sad today?"

---

**ELIZA, esempio di risposta, 1966**

➡ **Strategia di *fallback*:** formule generiche quando non trova *pattern*

👤 "My whole world is falling apart."  
>\_ "I see."  
(oppure)  
>\_ "Can you elaborate on that?"  
(oppure)  
>\_ "I see."  
(oppure)  
>\_ "What does that suggest to you?"  
(oppure)  
>\_ "Please go on."  
(oppure)  
>\_ "How does that make you feel?"

---

**ELIZA, esempio di risposta, 1966**

- ➔ Menzionato per la prima volta negli anni '50, popolare negli anni 2000
- ➔ Metodo per rappresentare il testo **non strutturato** in formato numerico
- ➔ Il **linguaggio è complicato** per i calcolatori
  - ➔ Il testo perde significato quando rappresentato da 0 e 1
- ➔ **Focus principale:** rappresentare il linguaggio in **modo strutturato** per l'uso da parte dei calcolatori

# BAG-OF-WORDS: COME FUNZIONA

## PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO

- 1 **Tokenizzazione:** processo di divisione delle frasi in parole individuali o sotto-parole (*token*)
  - ➔ Rispetto a un **delimitatore** (specifico o *wildcard*) e una **blacklist**
- 2 **Creazione del vocabolario:** estrazione di entità uniche dai *token*
  - ➔ Applicando estrazione della radice (*stemming*)
  - ⚠ Vettore **ordinato** di radici
- 3 **Conteggio delle parole:** rappresentazione numerica basata sulla **frequenza**

### Token

che | bel | cane | che | hai | ma | il | mio | gatto | è | più | bello

- 1 "Che \_bel \_cane \_\_\_\_\_che \_hai!"  
2 "Ma \_il \_\_\_\_\_mio \_gatto \_è \_più \_\_\_\_\_bello"

**Tokenizzazione**

# BAG-OF-WORDS: COME FUNZIONA

## PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO

- 1 **Tokenizzazione:** processo di divisione delle frasi in parole individuali o sotto-parole (*token*)
  - ➔ Rispetto a un **delimitatore** (specifico o *wildcard*) e una **blacklist**
- 2 **Creazione del vocabolario:** estrazione di entità uniche dai *token*
  - ➔ Applicando estrazione della radice (*stemming*)
  - ⚠ Vettore **ordinato** di radici
- 3 **Conteggio delle parole:** rappresentazione numerica basata sulla **frequenza**

### Vocabolario

che		bell-		can-		avere		ma		il		mi-		gatt-		essere		più		bell-
-----	--	-------	--	------	--	-------	--	----	--	----	--	-----	--	-------	--	--------	--	-----	--	-------

1 "Che \_bel \_cane \_\_\_\_\_ che \_hai!"

2 "Ma \_il \_\_\_\_\_ mio \_gatto \_è \_più \_\_\_\_\_ bello"


**Creazione del vocabolario**




# BAG-OF-WORDS: COME FUNZIONA

## PROCESSO DI TOKENIZZAZIONE E CREAZIONE DEL VOCABOLARIO



- 1 **Tokenizzazione:** processo di divisione delle frasi in parole individuali o sotto-parole (*token*)
  - ➔ Rispetto a un **delimitatore** (specifico o *wildcard*) e una **blacklist**
- 2 **Creazione del vocabolario:** estrazione di entità uniche dai *token*
  - ➔ Applicando estrazione della radice (*stemming*)
  - ⚠ Vettore **ordinato** di radici
- 3 **Conteggio delle parole:** rappresentazione numerica basata sulla **frequenza**

Vocabolario												
		che	bell-	can-	avere	ma	il	mi-	gatt-	essere	più	bell-
 	1	2	1	1	1	0	0	0	0	0	0	0
	2	0	0	0	0	1	1	1	1	1	1	1

 1 "Che    bel    cane       , che    hai!"  
 2 "Ma    il        mio    gatto    è    più        bello"

**Vettori bag-of-words**

- ➔ **Perdita dell'ordine:** frasi diametralmente opposte hanno la medesima rappresentazione
- ➔ **Mancanza di semantica:** Non cattura il significato delle parole
- ➔ **Nessuna generalizzazione:** i sinonimi sono trattati come elementi totalmente separati
- ➔ **Alta dimensionalità:** Vocabolari enormi con molti zeri

Vocabolario				
	Simone	mangiare	l'	insalat-
 1	1	1	1	1
 2	1	1	1	1

1 "Simone \_ mangiò \_ l' \_ insalata"

2 "L' \_ insalata \_ mangiò \_ Simone"

**Problema dell'ordine delle parole**

- ➡ **Perdita dell'ordine:** frasi diametralmente opposte hanno la medesima rappresentazione
- ➡ **Mancanza di semantica:** Non cattura il significato delle parole
- ➡ **Nessuna generalizzazione:** i sinonimi sono trattati come elementi totalmente separati
- ➡ **Alta dimensionalità:** Vocabolari enormi con molti zeri

👤 "Concetti come «Re» e «Regina» dovrebbero essere correlati..."

👤 "Sinonimi come «felice» e «gioioso» dovrebbero avere una rappresentazione simile..."

👤 "E con parole come «pitone» e «serpente», una generalizzazione dell'altra?!"

---

**Mancanza di relazioni semantiche**

- ➡ **Perdita dell'ordine:** frasi diametralmente opposte hanno la medesima rappresentazione
- ➡ **Mancanza di semantica:** Non cattura il significato delle parole
- ➡ **Nessuna generalizzazione:** i sinonimi sono trattati come elementi totalmente separati
- ➡ **Alta dimensionalità:** Vocabolari enormi con molti zeri





👤 *"Fornisci una stima del vocabolario usato in Wikipedia."*  
>\_ *"[...] English Wikipedia: ordine di grandezza  $10^5-10^7$  [...]. Tutte le Wikipedie insieme (tutte le lingue, forme di parola): ordine di grandezza  $10^5-10^8$  [...]."*  
👤 *"Quindi un documento Wikipedia in Bag-Of-Words sarebbe rappresentato da un vettore di dimensione almeno  $10^5$ ?!"*

---

**Problema della dimensionalità**

- ➔ **Word2Vec**: primo tentativo di successo per catturare il significato del testo negli *embeddings*
- ➔ **Embeddings**: rappresentazioni vettoriali di dati che tentano di catturarne il significato
- ➔ Addestrato su **enormi quantità** di dati testuali
  - ➔ British National Corpus
  - ➔ English Wikipedia
  - ➔ Google News
  - ➔ English Gigaword
  - ➔ ...
- ➔ Utilizza **reti neurali** per generare rappresentazioni semantiche
- ⚠ **Grandezza della rappresentazione vettoriale limitata a priori**

- ➔ **Principio fondamentale:** parole che appaiono in contesti simili tendono ad avere significati simili
- ➔ **Addestramento:** predire se due parole sono vicine in una frase
- ➔ **Risultato:** parole con significati simili hanno *embeddings* vicini nello spazio

 <sub>1</sub> "[...] Il mio cane ama dormire nella sua cuccia [...]"  
 <sub>2</sub> "[...] Il **veterinario** ha deciso di sterilizzare il gatto [...]"  
 <sub>3</sub> "[...] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario [...]"  
 <sub>4</sub> "[...] Ho comprato una cuccia per il mio gatto, ma continua a preferire il divano! [...]"  
>\_ (*thinking*) – Devo capire quali termini condividono gli stessi contesti linguistici... –

---

**Principio alla base di Word2Vec**

*You shall know a word by the company it keeps.*

---

**Ipotesi distributiva, John Rupert Firth, 1957**

- ➔ **Principio fondamentale:** parole che appaiono in contesti simili tendono ad avere significati simili
- ➔ **Addestramento:** predire se due parole sono vicine in una frase
- ➔ **Risultato:** parole con significati simili hanno *embeddings* vicini nello spazio

1 "[...] Il mio cane ama dormire nella sua cuccia [...]"  
2 "[...] Il veterinario ha preferito sterilizzare il gatto [...]"  
3 "[...] Mentre giocava, il mio cane si è fatto male e l'ho dovuto portare dal veterinario [...]"  
4 "[...] Ho comprato una cuccia per il mio gatto, ma continua a preferire il divano! [...]"  
>\_ (thinking) – Devo avvicinare le rappresentazioni di «cane» e «gatto» nello spazio perché condividono spesso gli stessi vicini «veterinario» e «cuccia»! –

---

**Addestramento di Word2Vec**

- ➔ **Principio fondamentale:** parole che appaiono in contesti simili tendono ad avere significati simili
- ➔ **Addestramento:** predire se due parole sono vicine in una frase
- ➔ **Risultato:** parole con significati simili hanno *embeddings* vicini nello spazio

>\_ "«Computer» è simile a «laptop», «pc», «desktop», «workstation», ..."  
>\_ "«Re» sta a «uomo» come «Regina» sta a «donna»!"  
>\_ "«Lunedì», «martedì», «mercoledì», ...formano un unico gruppo."

---

**Proprietà emergenti: sinonimie, classificazioni, relazionalità**

 [Portale Word2Vec](#)



- ➔ **Polisemia:** confusione con parole utilizzabili con diversi significati in contesti diversi
- ➔ **Contesti insufficienti:** parole rare potrebbero non avere abbastanza esempi

>\_ (*thinking*) – Ma...«cannonata» in senso bellico...O calcistico? –

---

**Problema della polisemia**

- ➔ **Polisemia:** confusione con parole utilizzabili con diversi significati in contesti diversi
- ➔ **Contesti insufficienti:** parole rare potrebbero non avere abbastanza esempi

>\_ (*thinking*) — *Non ho molto capito il senso delle parole «supercazzola», «scappellamento» e «Antani»... —*

---

**Amici miei, 1975**