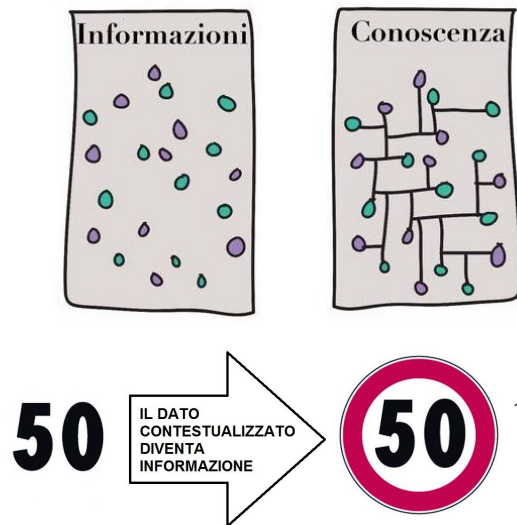
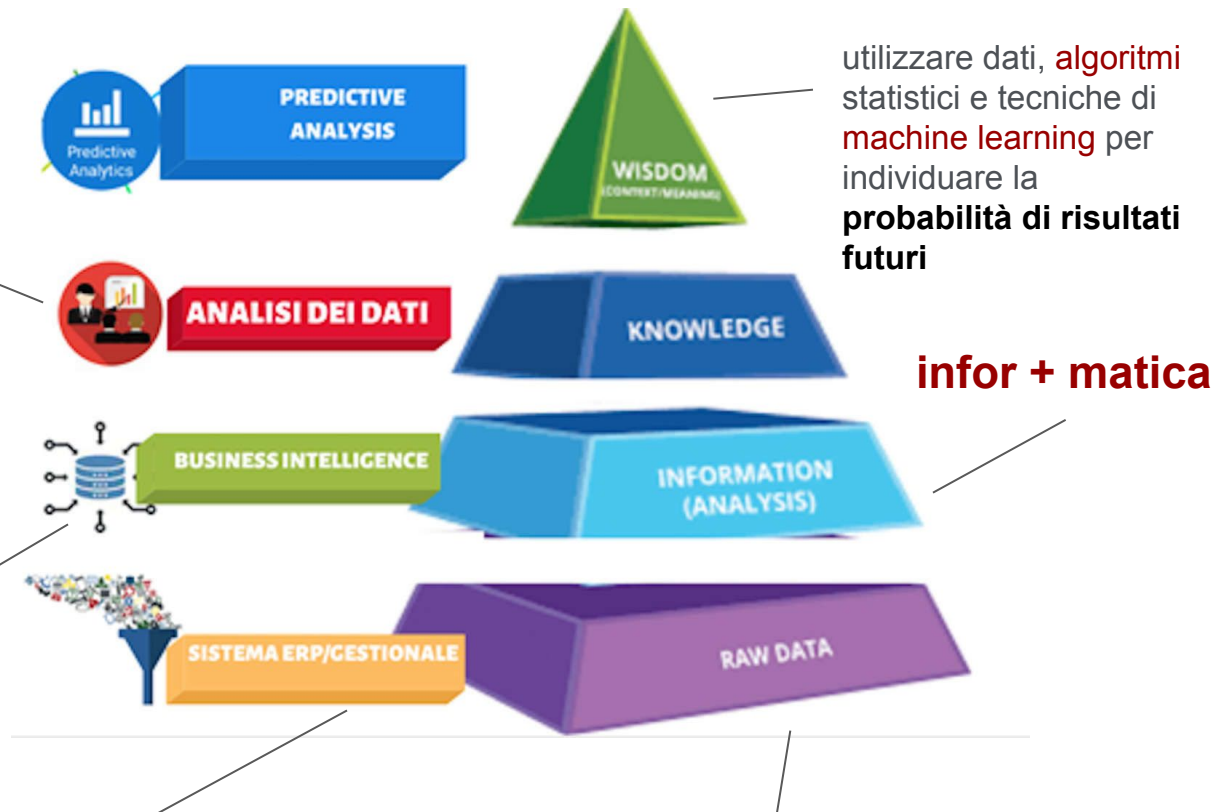


<https://www.nunziocastaldi.it/fad/ai/>

Piramide DIKW



ERP (Enterprise Resource Planning) **software** che aiuta le organizzazioni ad **automatizzare** e **gestire** i processi aziendali principali

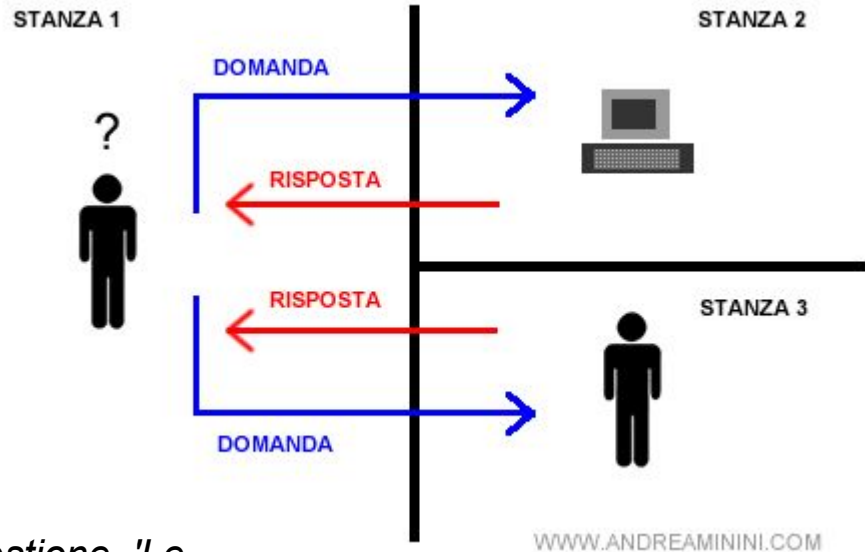


Intelligenza / Test di Turing

Il Test di Turing è concepito per verificare se, attraverso una conversazione in chat, un computer **può convincere** un umano di essere umano.

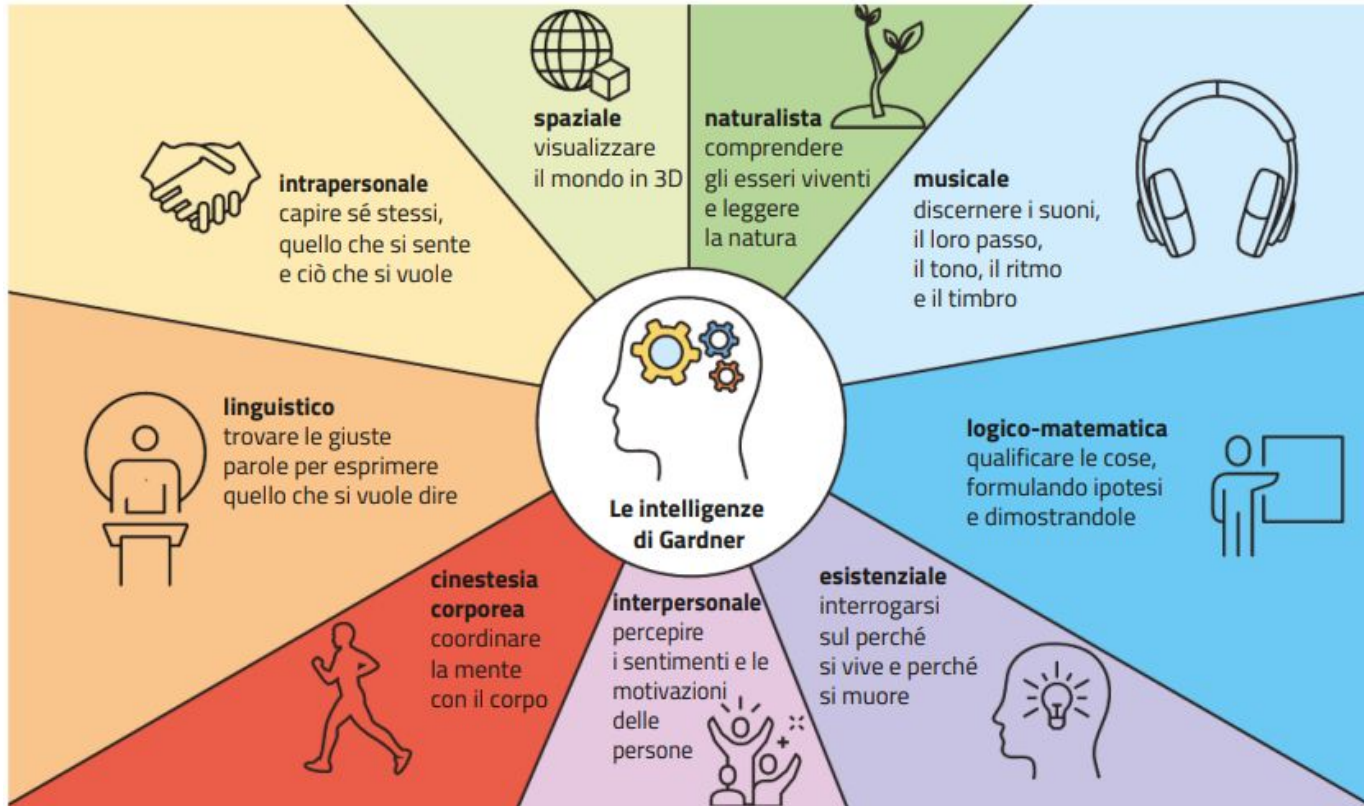
(Gioco dell'imitazione)

Propongo di considerare la seguente questione. 'Le macchine sono in grado di pensare?'. Si dovrebbe iniziare con le definizioni dei significati dei termini 'macchina' e 'pensiero'



Intelligenza Umana

Howard Gardner – psicologo e professore di Harvard

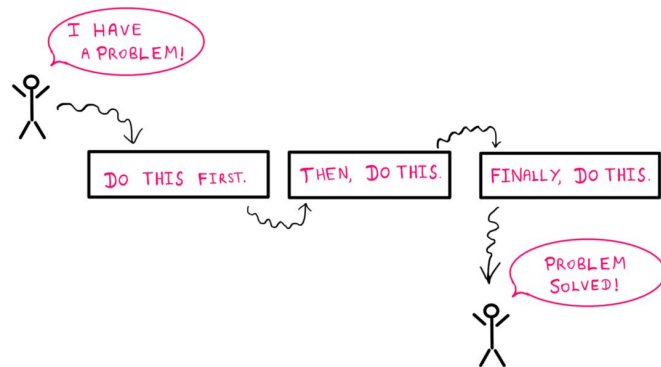


IA, una definizione 'divulgativa'

L'intelligenza artificiale è un insieme di **tecniche informatiche** che permette a macchine e software di svolgere **compiti cognitivi** — come apprendere da dati, ragionare, comprendere il linguaggio naturale o riconoscere immagini — fino a **emulare** funzioni tipiche della mente umana.

Per farlo, si basa su **algoritmi**, alcuni dei quali adottano l'apprendimento automatico (**machine learning**), ossia la capacità di estrarre informazioni e schemi da dati grezzi

Algoritmo = sequenza **finita, definita ed efficace** di operazioni da svolgere per giungere alla soluzione di un problema



L'errore è sempre...

Problema = calcolare l'area di un rettangolo con base 6 e altezza 4

Inizio algoritmo

```
VARIABILE base = 6;  
VARIABILE altezza = 4;  
ESEGUI base * altezza;  
OUTPUT 24;
```

Fine algoritmo (= problema risolto)

Inizio algoritmo

```
VARIABILE base = 6;  
VARRABBILE altezza = 4;
```

Interruzione algoritmo (l'interprete/compiler si accorge dell'errore sintattico e restituisce un errore fatale)

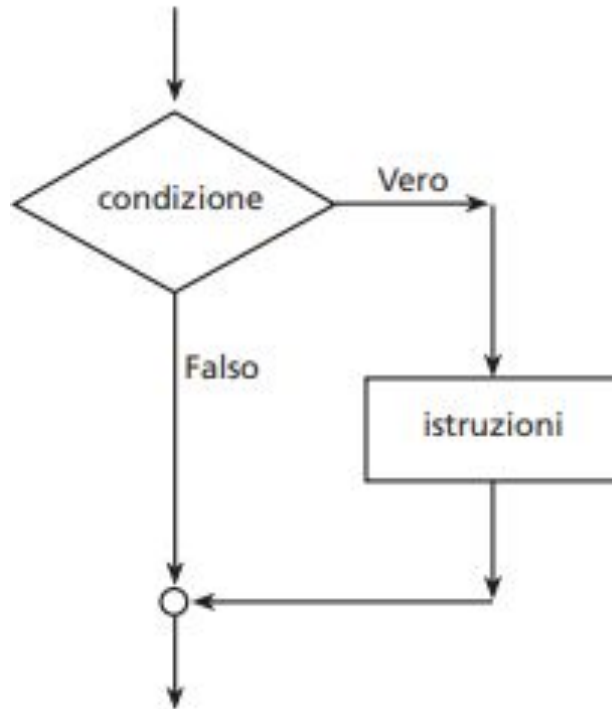
Inizio algoritmo

```
VARIABILE base = 6;  
VARIABILE altezza = 4;  
ESEGUI base * base; (istruzione sbagliata ma l'interprete/compiler non se ne  
accorge, perché è un errore logico)  
OUTPUT 36;
```

Fine algoritmo (= problema **solo apparentemente** risolto)

...umano

SE... ALLORA...



SE piove
chiudi l'ombrello
ALTRIMENTI
apri l'ombrello
FINE SE

istruzione formalmente
corretta ma illogica (ma
l'interprete/compiler non lo
sa)

VA' A FARE LA SPESA
COMPRA 1 bottiglia di latte
SE ci sono banane
comprane 6
FINE SE

istruzione formalmente
corretta ma ambigua

l'interprete/compiler, appurato
che ci sono banane, potrebbe
comprare **6 bottiglie** di latte
(mentre il programmatore
intendeva l'acquisto aggiuntivo di
6 banane)

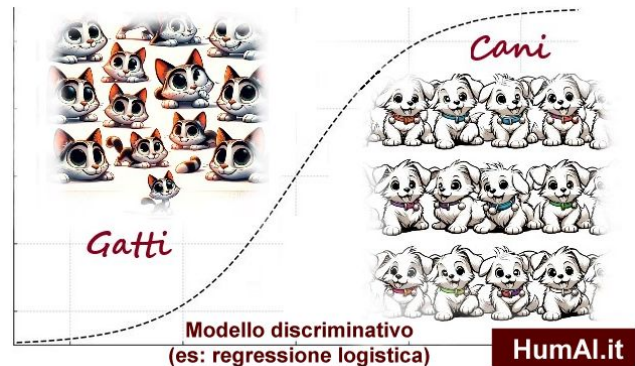
algoritmo corretto:

SE ci sono banane
compra 1 bottiglia di latte +
6 banane
FINE SE

Intelligenza (?) Artificiale

Intelligenza artificiale discriminativa

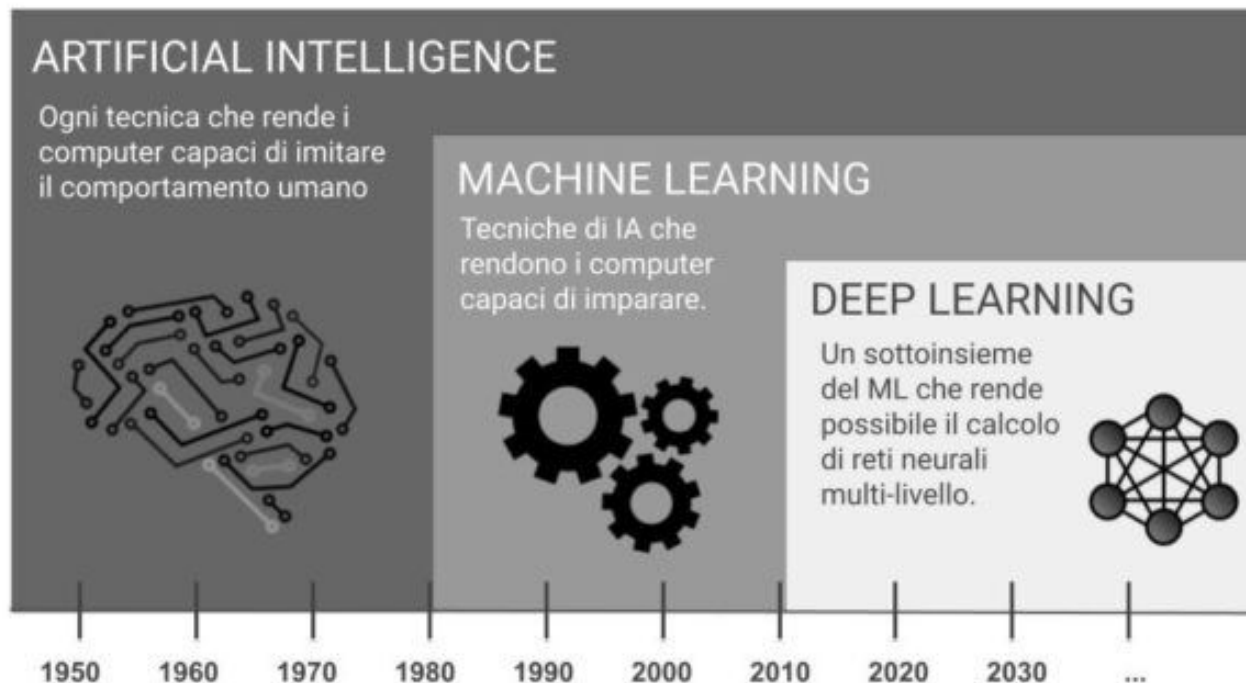
finalizzata all'analisi delle informazioni ricevute e alla classificazione di tali informazioni in categorie predefinite. Cioè “discrimina” i dati secondo criteri predefiniti.



Intelligenza artificiale generativa

finalizzata alla generazione di nuovo contenuto (testi, disegni, immagini, ecc.) sulla base di determinate informazioni di partenza (PROMPT).

Evoluzione

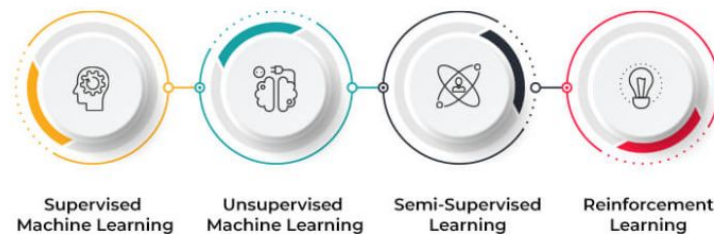


Metodi di Addestramento

APPRENDIMENTO SUPERVISIONATO	<ul style="list-style-type: none">> Dati in ingresso etichettati> Feedback diretto> Predizione dei risultati
APPRENDIMENTO NON SUPERVISIONATO	<ul style="list-style-type: none">> Dati in ingresso non etichettati> Nessun Feedback> Scoperta <i>pattern</i> nascosti nei dati
APPRENDIMENTO SEMI-SUPERVISIONATO	<ul style="list-style-type: none">> Dati in ingresso sia etichettati che non> Feedback diretto> Molto performanti
APPRENDIMENTO RINFORZATO	<ul style="list-style-type: none">> Processo di decisione> Funzione ricompensa> Apprendimento per scelte ed errori



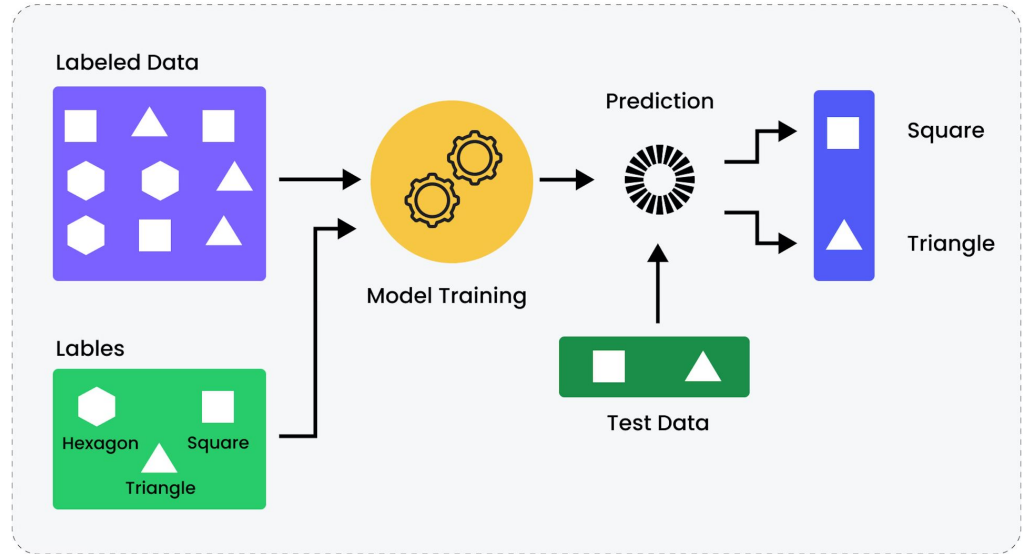
Machine Learning



Addestramento / Apprendimento supervisionato

Alla macchina vengono forniti **set di dati etichettati**, cioè esempi di input con i rispettivi output desiderati.

L'obiettivo è che la macchina impari una **regola generale** che associa gli input agli output.

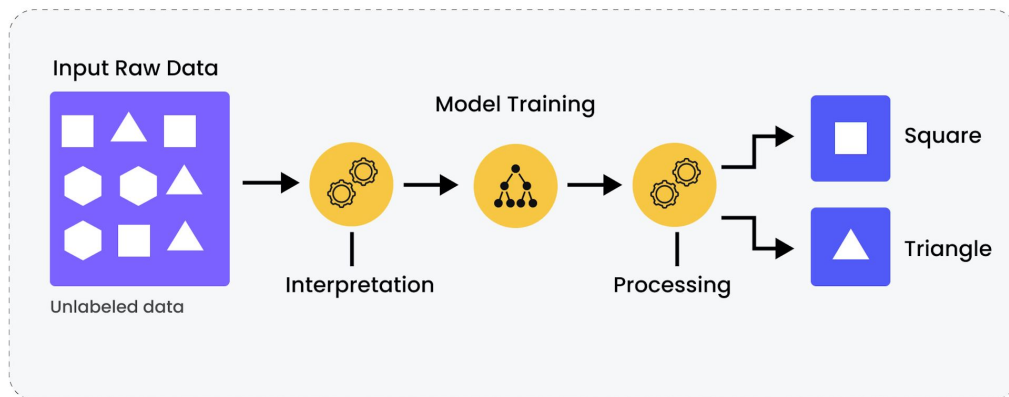


<https://www.ibm.com/it-it/topics/supervised-learning>

Addestramento / Apprendimento non supervisionato

Alla macchina vengono forniti solo **dati 'grezzi' di input**, senza alcun output desiderato.

L'obiettivo è che la macchina scopra autonomamente **una struttura o un pattern** all'interno dei dati

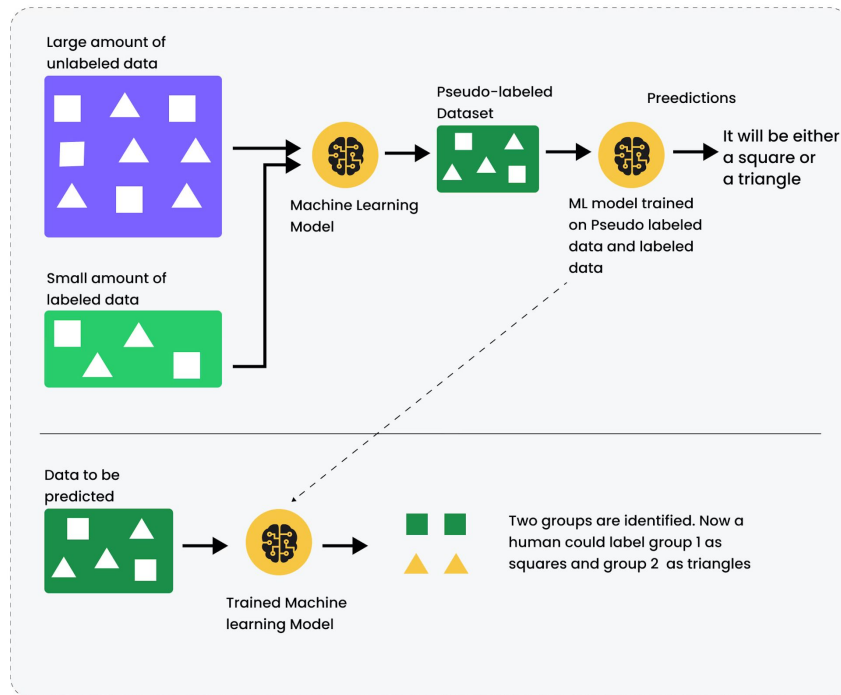


<https://www.ibm.com/it-it/topics/supervised-learning>

Addestramento / Apprendimento semi-supervisionato

Vengono utilizzati sia dati "labeled" (etichettati) che "unlabeled" (non etichettati).

L'obiettivo è che la macchina 'intuisca' **una struttura o un pattern** all'interno dei dati non etichettati, aiutandosi con quelli etichettati.

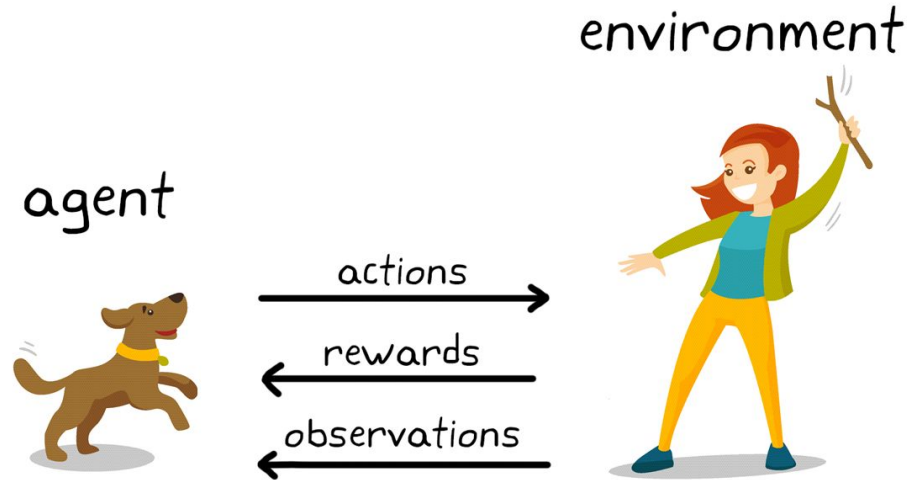


<https://www.ibm.com/it-it/topics/supervised-learning>

Addestramento / Apprendimento per rinforzo

La macchina interagisce con un ambiente dinamico, cercando di raggiungere un obiettivo specifico.

Attraverso tentativi ed errori, la macchina impara a compiere azioni che la avvicinano all'obiettivo, **ricevendo feedback** in termini di ricompense o penalità.

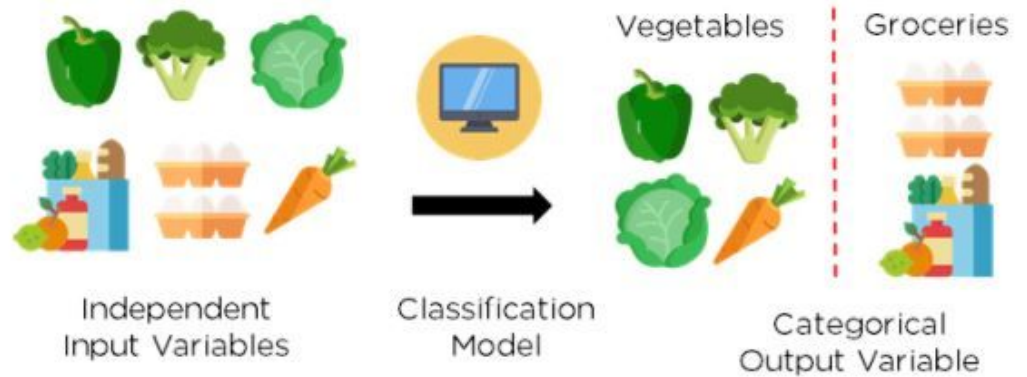


<https://www.ibm.com/it-it/topics/supervised-learning>

Output / Classificazione

L'obiettivo è **dividere gli input in due o più categorie**.

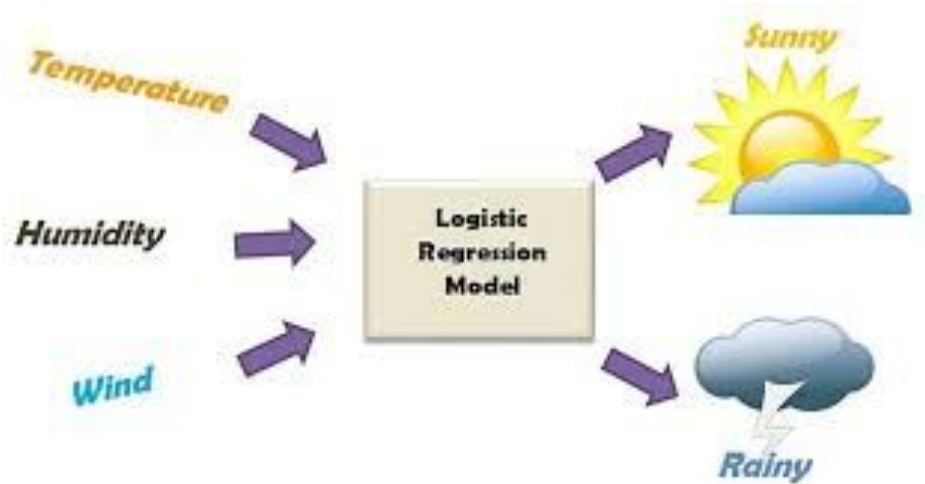
Un esempio è la classificazione di email come spam o non spam.



Output / Regressione

L'obiettivo è **prevedere un valore numerico continuo**.

Un esempio è la previsione del prezzo di un'azione in borsa.



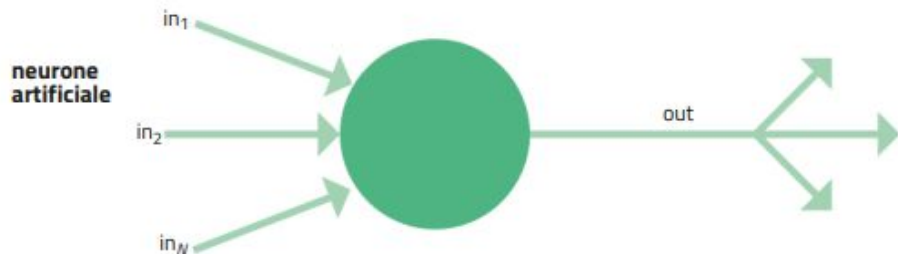
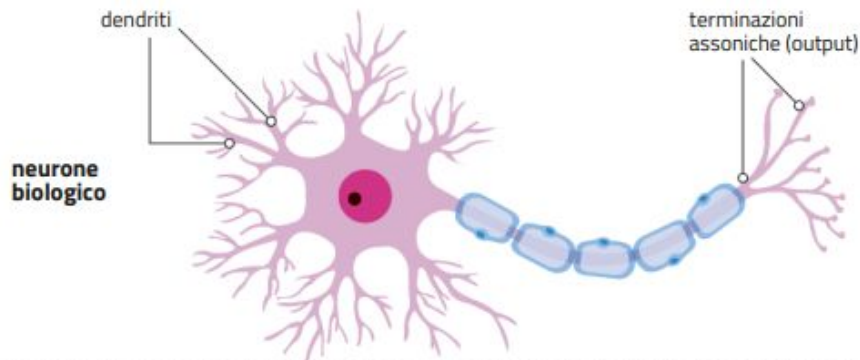
Output / Clustering

L'obiettivo è **raggruppare automaticamente i dati in base alle loro somiglianze**.

Un esempio è la raggruppamento di clienti in base alle loro abitudini di acquisto.



Reti neurali / 1



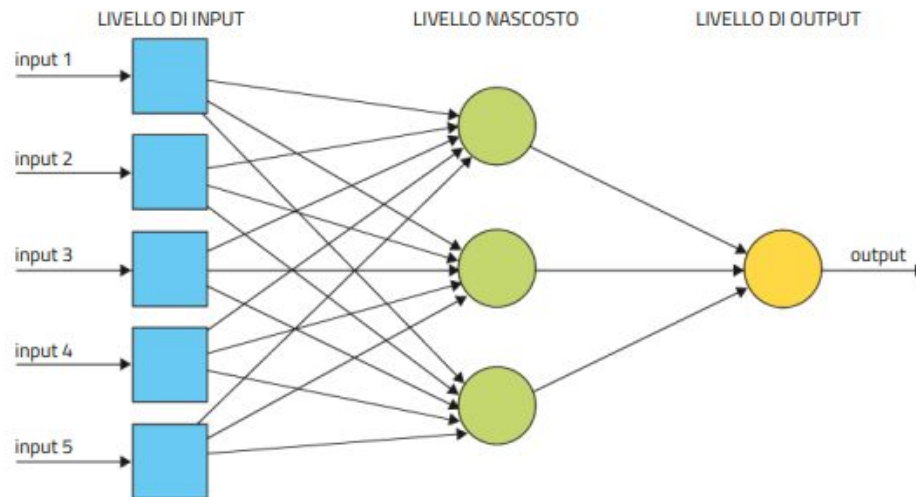
Le reti neurali artificiali (Artificial Neural Networks o ANN) sono **modelli matematici ispirati al funzionamento del cervello umano**. L'idea di base è imitare il modo in cui i neuroni biologici comunicano tra loro attraverso connessioni: ogni neurone riceve segnali, li elabora e li trasmette ad altri neuroni.

Allo stesso modo, in una rete neurale artificiale, i nodi (cioè i neuroni artificiali) elaborano le informazioni che ricevono e le trasmettono ai nodi successivi, in base a **valori di soglia**.

Reti neurali / 2

Questi nodi sono organizzati in strati:

- uno **strato di input**, dove arrivano i dati (per esempio, i pixel di un'immagine o le parole di una frase),
- uno o più **strati nascosti**, che elaborano l'informazione passando attraverso calcoli matematici,
- uno **strato di output**, che restituisce il risultato (ad esempio, "questa è una foto di un cane" o "questa frase ha tono positivo").



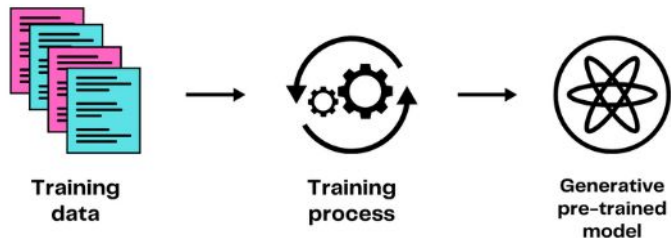
GPT (Generative Pre-trained Transformer)

Generativo

Produce contenuti—per esempio frasi, paragrafi o persino codice—basandosi su ciò che ha “visto” in fase di addestramento

Pre-addestrato

su enormi quantità di testo (libri, articoli, pagine web) senza che nessuno gli spieghi a cosa servono quei testi (**apprendimento non supervisionato**). Impara autonomamente le **regolarità** del linguaggio (come si formano le frasi, quali parole spesso compaiono insieme...).

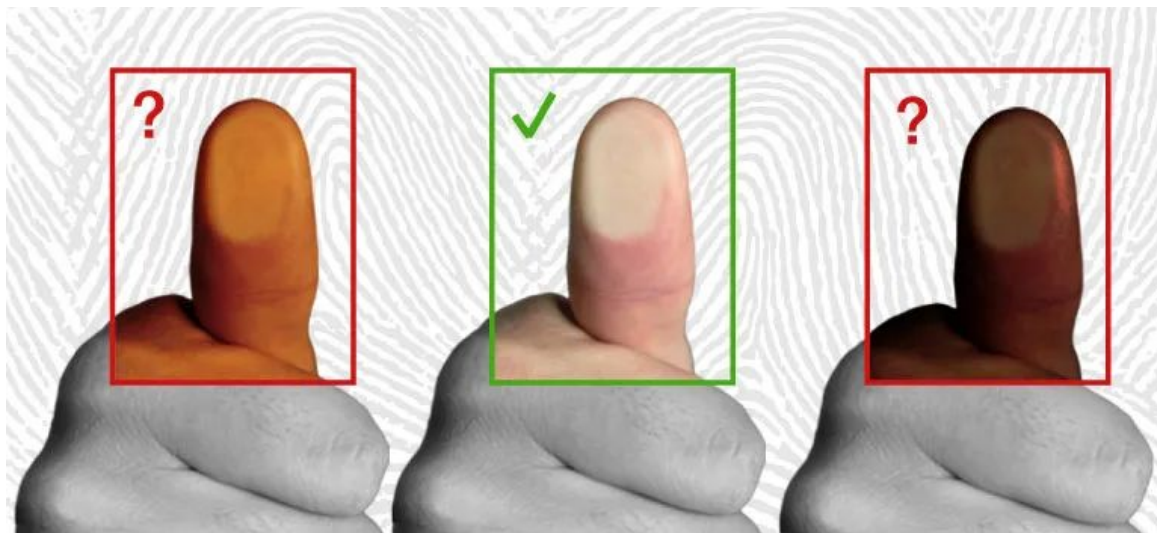


Architettura

di rete neurale che elabora sequenze di token (parole, frammenti di parola o simboli) tramite meccanismi di self-attention, permettendo al modello di valutare dinamicamente l'importanza di ciascun token del contesto per generare il successivo con maggiore coerenza e fluidità.

BIAS cognitivi

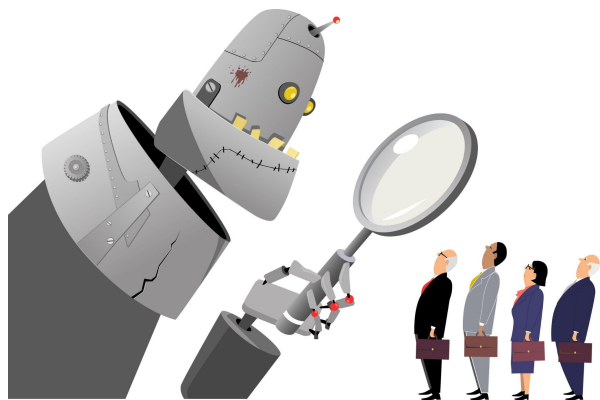
Il bias del machine learning (o bias dell'algoritmo) si riferisce al verificarsi di risultati distorti a causa di **pregiudizi** umani che alterano i dati di addestramento originali o l'algoritmo AI, portando a output distorti e potenzialmente dannosi.



BIAS cognitivi

Bias nelle assunzioni (2025)

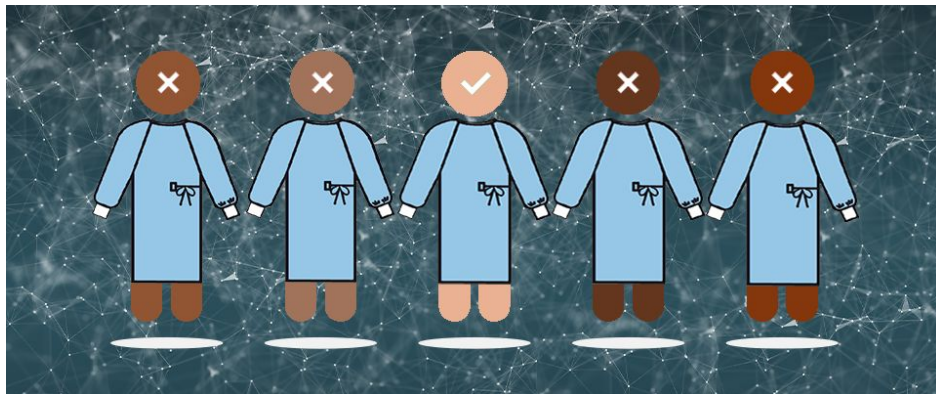
Recenti test condotti su strumenti di screening dei CV basati su IA hanno mostrato che questi sistemi possono **discriminare pesantemente in base a razza e genere**. In particolare, nel 2025, alcuni strumenti hanno registrato un tasso di selezione quasi nullo per nomi maschili afroamericani e una preferenza per nomi bianchi nell'85% dei casi, oltre a favorire candidati maschili nel 52% delle selezioni. Questo ha sollevato gravi dubbi sull'equità e sull'accesso alle opportunità lavorative, con aziende costrette a rivedere i propri processi di selezione



BIAS cognitivi

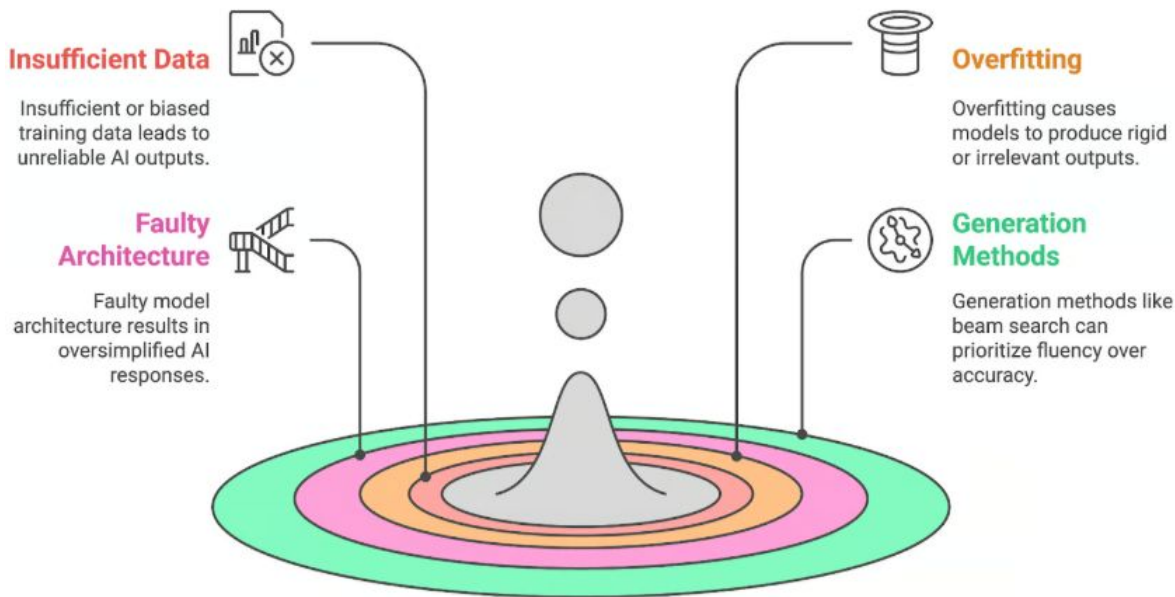
Bias nella sanità (2024-2025)

Il bias negli algoritmi medici si è tradotto in disparità di trattamento e diagnosi. Uno studio della Yale School of Medicine ha rilevato che il 90% dei modelli linguistici medici mostrava **bias razziale**, con conseguenze gravi: i pazienti afroamericani non ispanici hanno registrato un tasso di mortalità superiore del 30% rispetto ai pazienti bianchi a causa di errori dell'IA. Inoltre, la precisione nella diagnosi di alcune condizioni, come il cancro della pelle, era significativamente inferiore per le persone con pelle scura rispetto a quelle con pelle chiara.



Hallucinations

L'AI può generare **informazioni che non corrispondono alla realtà o che non sono coerenti** con i dati di input forniti.



Hallucinations

Caso Tribunale di Firenze (marzo 2025)

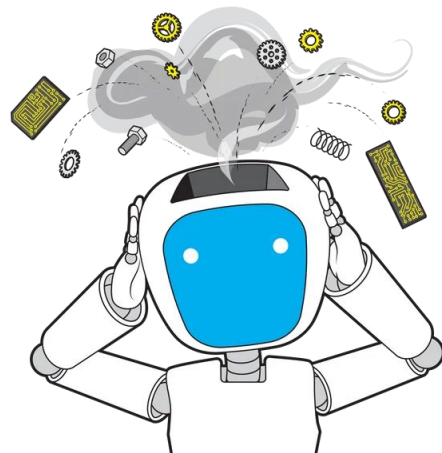
Un avvocato ha presentato atti difensivi basati su ricerche effettuate con ChatGPT. L'AI aveva inventato numeri di sentenze della Corte di Cassazione relative all'acquisto soggettivo di merce contraffatta, che in realtà non esistevano. Il Tribunale ha sottolineato il rischio di affidarsi acriticamente all'AI per la ricerca giuridica, richiamando la necessità di basarsi sempre su fonti verificabili e affidabili.



Hallucinations

Diffamazione e dati personali falsi (marzo 2025)

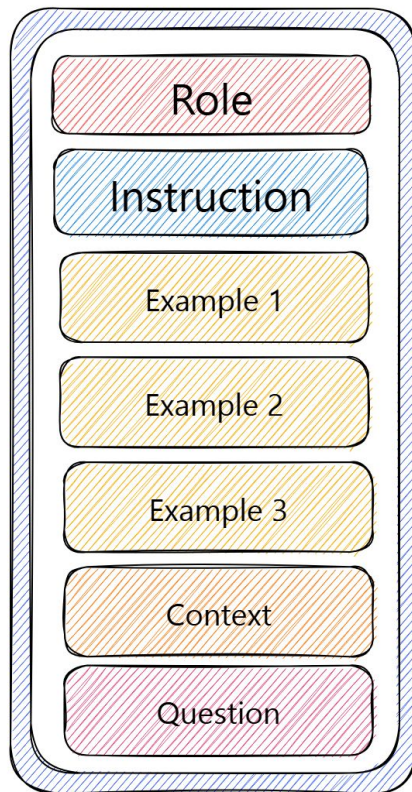
Un caso internazionale ha visto ChatGPT accusare falsamente un uomo norvegese di aver ucciso i suoi figli. Il chatbot ha generato una storia completamente inventata, attribuendo all'utente un duplice omicidio e una condanna al carcere, senza alcun fondamento reale. Questo episodio, insieme ad altri casi di false accuse di corruzione e abusi, ha portato a una denuncia formale contro OpenAI per violazione del principio di accuratezza previsto dal GDPR. L'azienda, secondo la denuncia, non sarebbe in grado di correggere le informazioni false prodotte dal sistema, limitandosi a bloccare alcune risposte ma senza rimuovere i dati errati dal modello.



Prompt



- Definire un ruolo
- Fornire il contesto
- Essere chiari e specifici
- Essere sintetici e diretti
- Utilizzare esempi concreti
- Rivolgersi all'AI come a una persona



La regola fondamentale è che **istruzioni di qualità determinano risultati di qualità.**

Un **prompt** è l'input testuale da fornire al modello: può essere una domanda, un'istruzione, un contesto o anche un semplice avvio di frase. In pratica, è tutto ciò che innesca la generazione di una risposta da parte del modello.

Prompt



Prompt vago

“Parlami del cambiamento climatico.”

Prompt ottimizzato

“Ciao, Chat GPT. Sei un redattore scientifico per un giornale di larga diffusione. Fornisci un **riassunto in 5 punti** sullo stato attuale della ricerca sul cambiamento climatico, includendo dati recenti e citazioni di studi.”

Prompt / Schemi (= framework)



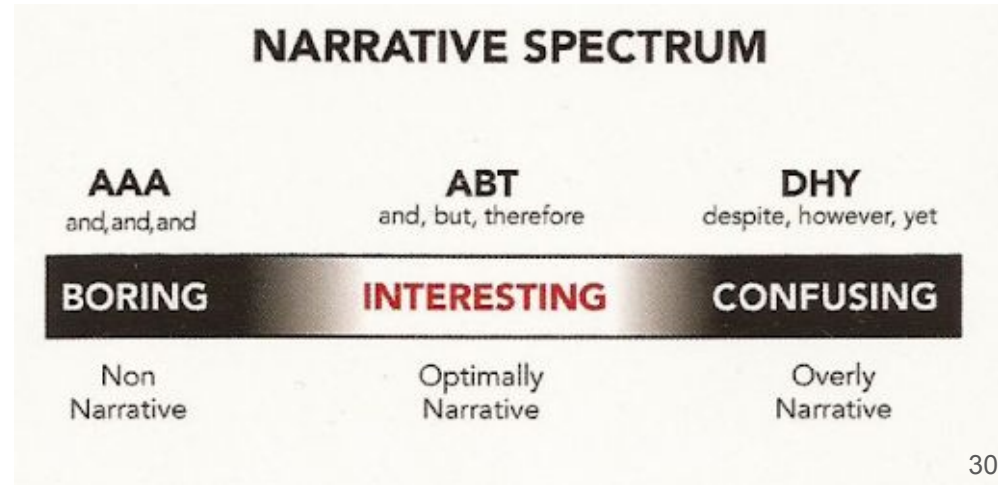
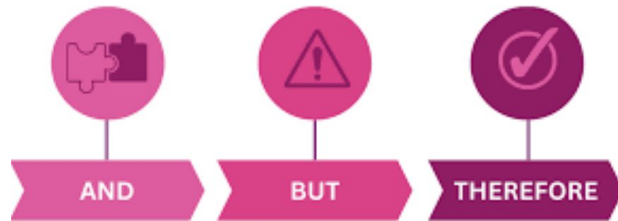
Gli schemi più diffusi

Gli schemi di prompt (APE, RACE e simili) sono principi di **prompt engineering** e quindi valgono in linea di massima per qualsiasi assistente basato su modelli di linguaggio (GPT, Claude, Bard, LLaMA...)

Framework / ABT (And, But, Therefore)

Framework brevissimo e molto narrativo, perfetto per storie semplici e lineari:

1. **And:** introduce il contesto felice o positivo, chiudendo il cerchio con un elemento quotidiano e familiare.
2. **But:** crea la “piccola difficoltà” o problema che rompe l’equilibrio iniziale e scatena la curiosità.
3. **Therefore:** porta alla soluzione o al “lieto fine”, spiegando in modo semplice come il protagonista supera l’ostacolo.



Framework / The Hero's Journey

Percorso narrativo in più tappe che trasforma un “eroe” qualunque in un protagonista capace di superare ostacoli e tornare cambiato, portando un dono o un insegnamento. Le fasi principali, semplificate per un racconto scolastico, sono:

1. **Mondo Ordinario:** presenta il contesto “quotidiano” del protagonista
2. **Chiamata all'Avventura:** qualcosa rompe la routine e invita l'eroe a fare un passo fuori dal suo mondo.
3. **Incontro con il Mentore:** un aiuto esterno (persona, libro, idea) offre strumenti o coraggio.
4. **Prove e Ostacoli:** l'eroe affronta sfide di difficoltà crescente.
5. **Apice della Sfida:** momento critico in cui sembra non esserci via d'uscita.
6. **Ricompensa:** l'eroe vince la prova e ottiene un “tesoro” (conoscenza, oggetto, consapevolezza).
7. **Ritorno con l'Elisir:** ritorno a casa con il “dono” che cambia la sua vita e quella degli altri.

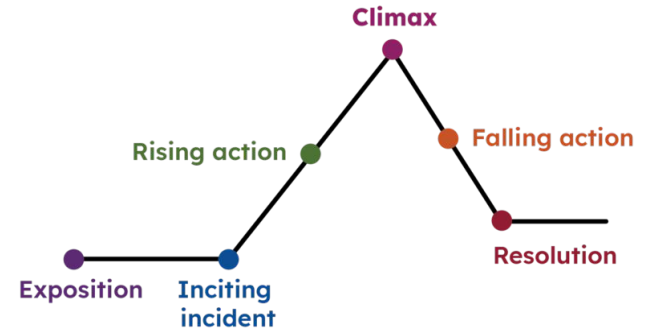


Framework / Freytag's Pyramid

Modello narrativo in cinque atti, tratto dalla drammaturgia classica, che dà ritmo e tensione al racconto passando da un'introduzione fino alla risoluzione finale. Le fasi sono:

1. **Esposizione:** presentazione del mondo e dei personaggi
2. **Azione crescente:** sviluppo del conflitto o della curiosità
3. **Climax:** punto di massima tensione, svolta decisiva
4. **Caduta dell'azione:** conseguenze immediate della crisi superata
5. **Scioglimento:** conclusione e insegnamento

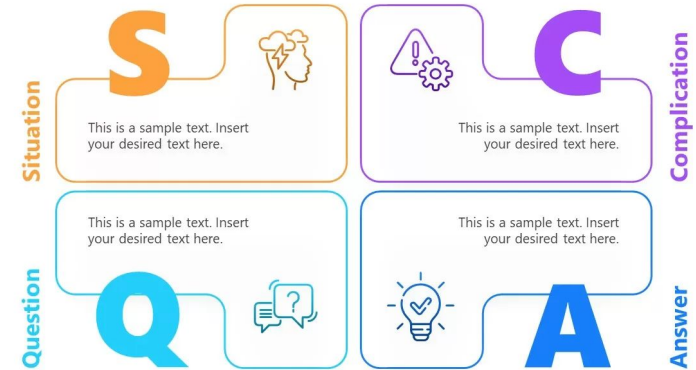
Freytag's Pyramid



Framework / SCQA (Situation, Complication, Question, Answer)

Un formato pensato per costruire una narrazione avvincente, portando il lettore da uno stato di fatto a un problema, facendogli porsi la domanda chiave e infine offrendo la soluzione:

- **Situation:** inquadra la scena iniziale, mostrando il contesto quotidiano o lo stato di partenza.
- **Complication:** introduce un ostacolo, un mistero o un cambiamento che rompe l'equilibrio e crea curiosità.
- **Question:** formula la domanda che spinge il protagonista (e il lettore) alla ricerca di una risposta.
- **Answer:** fornisce la risoluzione, l'*insight* o la spiegazione finale.



Grazie per l'attenzione :)



Nunzio Castaldi

weben.fad@gmail.com
weben.it