Intelligenza Artificiale - luglio 2024

Cos'è l'Artificial Intelligence (a luglio 2024)	45'
Suggerimenti e considerazioni	30'
Cenni normativi	15'
Question time	*a resistenza…*

Assemblea dei Consulenti del Lavoro di Belluno
29 / 07 / 2024

Cos’è l’Artificial Intelligence

…A LUGLIO 2024…

Cosa è l'intelligenza artificiale?

Definizioni generali

- sistemi informatici o algoritmi che simulano alcune funzioni umane, come l'apprendimento o la risoluzione di problemi
- progettati per apprendere dalle istruzioni umane ed eseguire attività specifiche senza intervento esterno

Definizioni matematiche

$$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_{\mu}^{\pi}.$$

Shane Legg e Marcus Hutter https://arxiv.org/pdf/0712.3329

Definizioni giuridiche (AI Act)

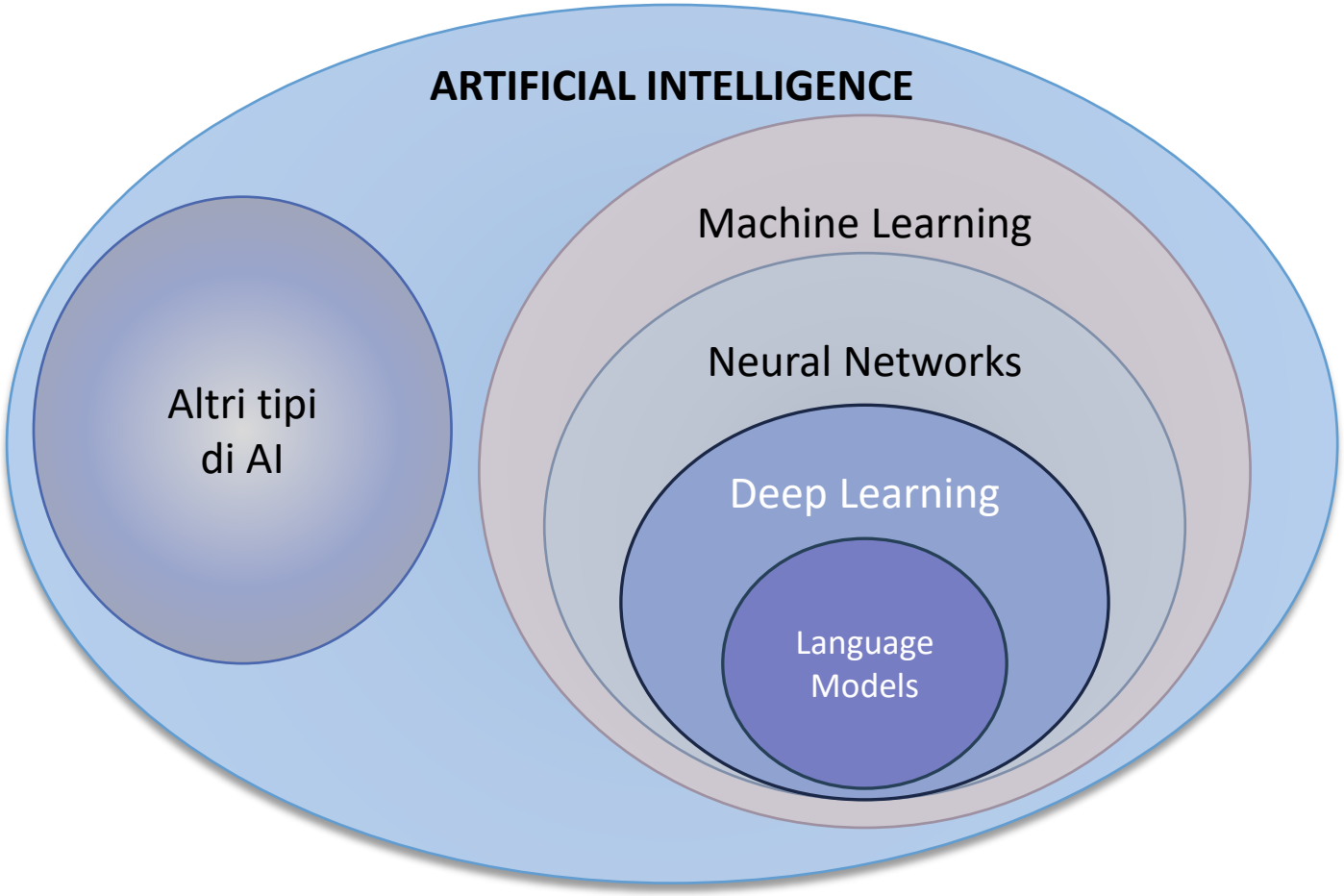
sistema basato su macchine progettato per funzionare con diversi livelli di autonomia, che può mostrare capacità di adattamento dopo l'implementazione e che, per obiettivi espliciti o impliciti, deduce dagli input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici

Definizioni classiche

sistemi in grado di superare il «test di Turing»

Classificazione tipi AI

Più «recente» la tecnolgia, maggiore la capacità di elaborazione richiesta.



Dai meno recenti (Machine Learning)

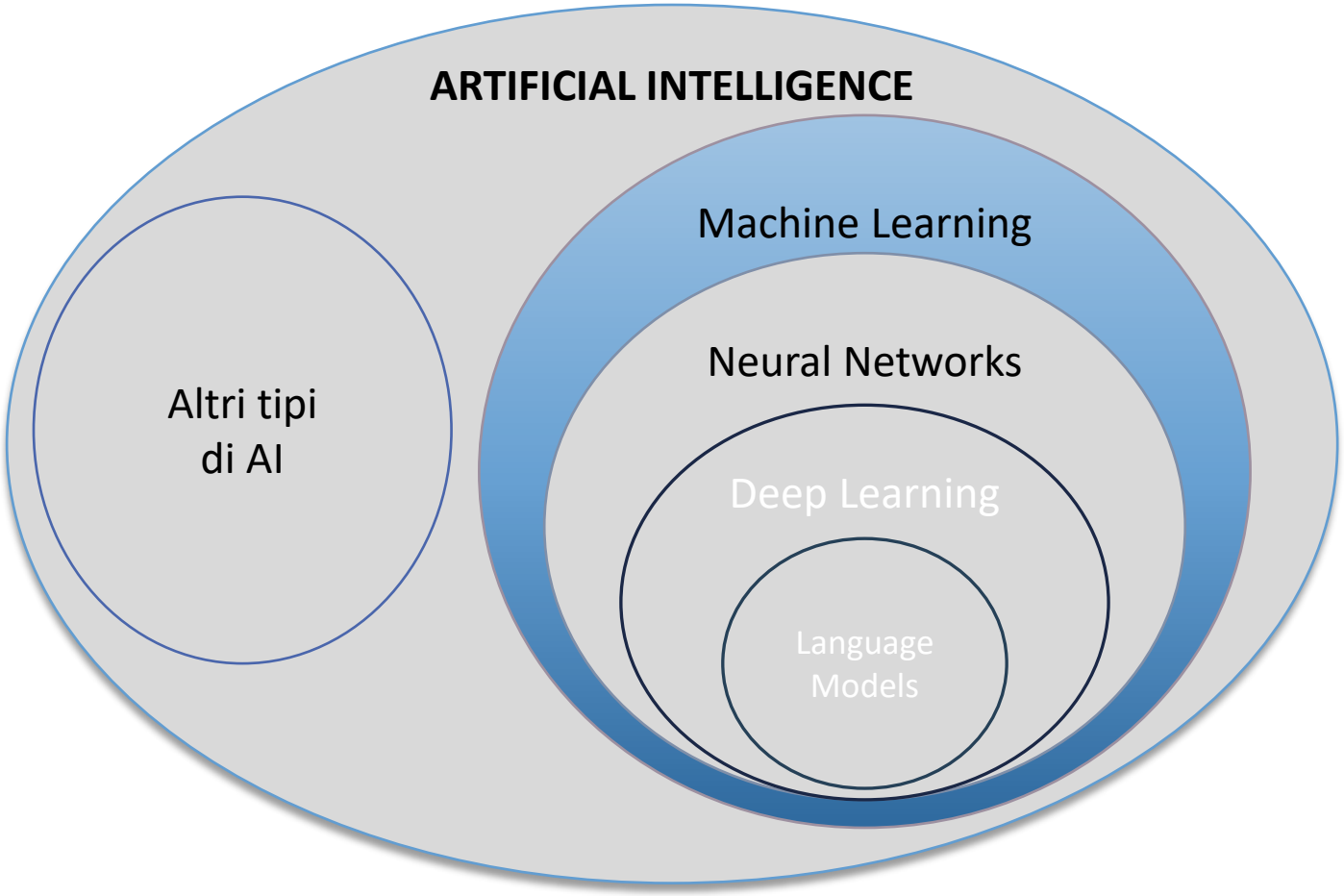
…

verso i più recenti

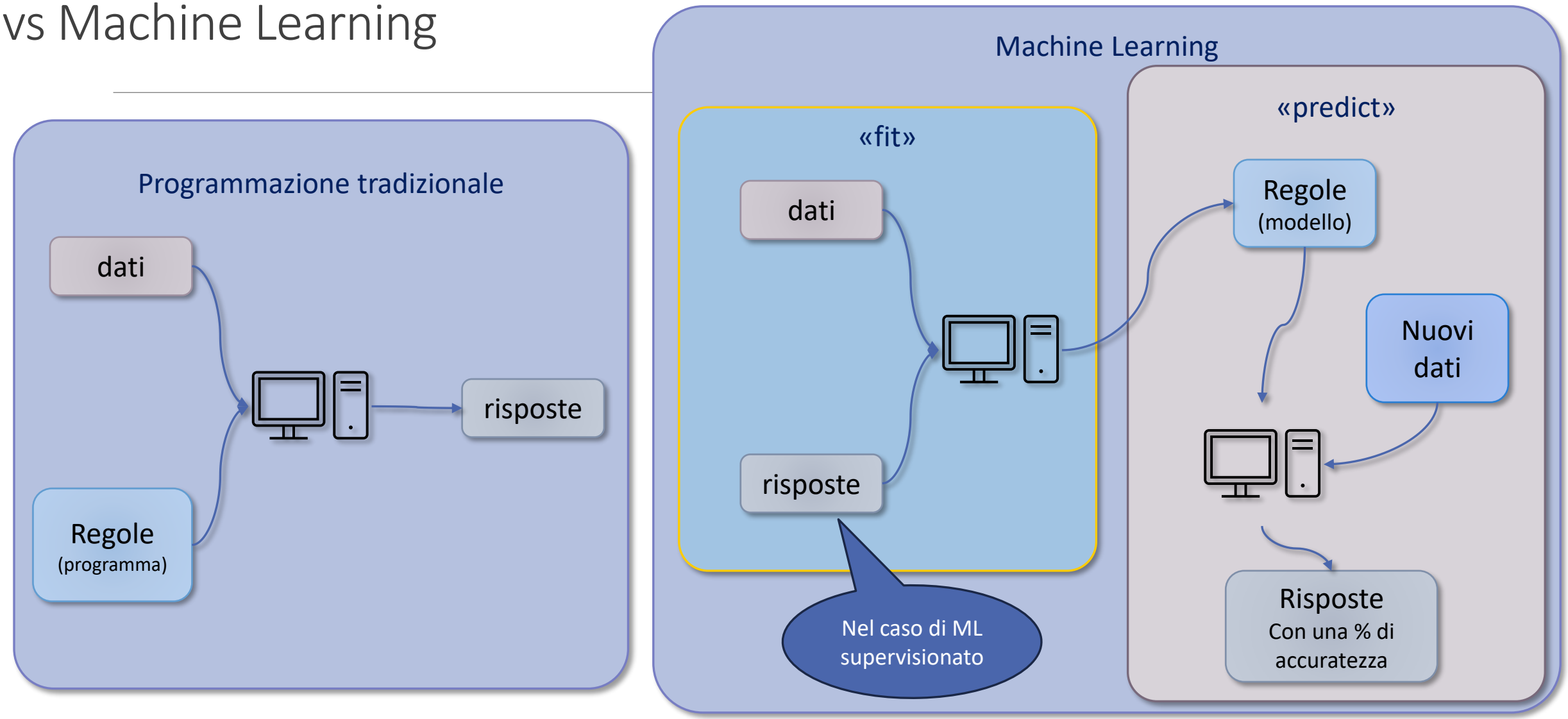
…

(Large) Language Models

Machine Learning



Programmazione tradizionale vs Machine Learning



Natura probabilistica dei modelli

I modelli ML sono probabilistici per natura

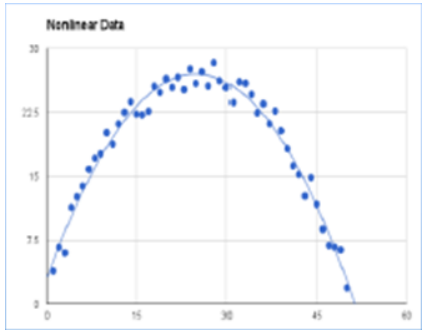
- Prevedono la/le risposte con una certa accuratezza percentuale
- Fondamentale valutare l'affidabilità del modello addestrato
- Si usano tecniche statistiche di misurazione dell'errore
- La «confidenza» dice quanto accurata è la risposta stimata dal modello
- Esempio: «è la foto di un gatto» – «con una probabilità del 90%»

Principali aree applicative del Machine Learning

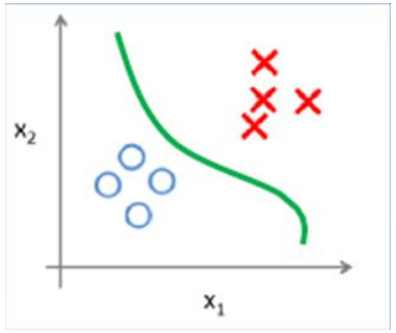
Addestramento e applicazione

- **Apprendimento supervisionato**
  - Regressione
  - Classificazione
  - Serie temporali
- **Apprendimento non supervisionato**
  - Clustering
  - Riduzione dimensionalità
  - Anomaly detection
- **Apprendimento misto**
- **Apprendimento per rinforzo**
  - «premi» o «punizioni» da ambiente esterno

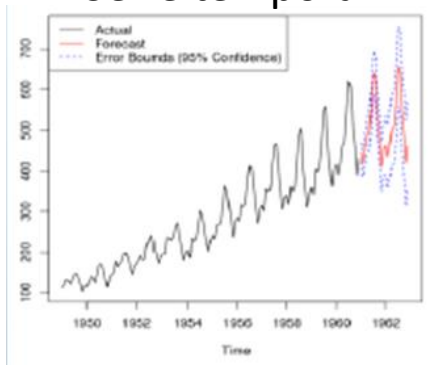
Regressione



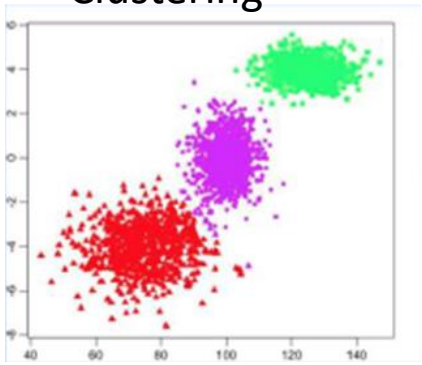
Classificazione



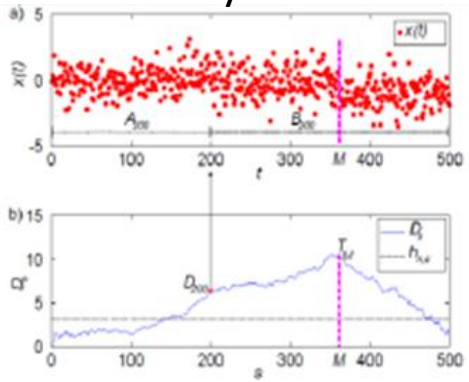
Serie temporali



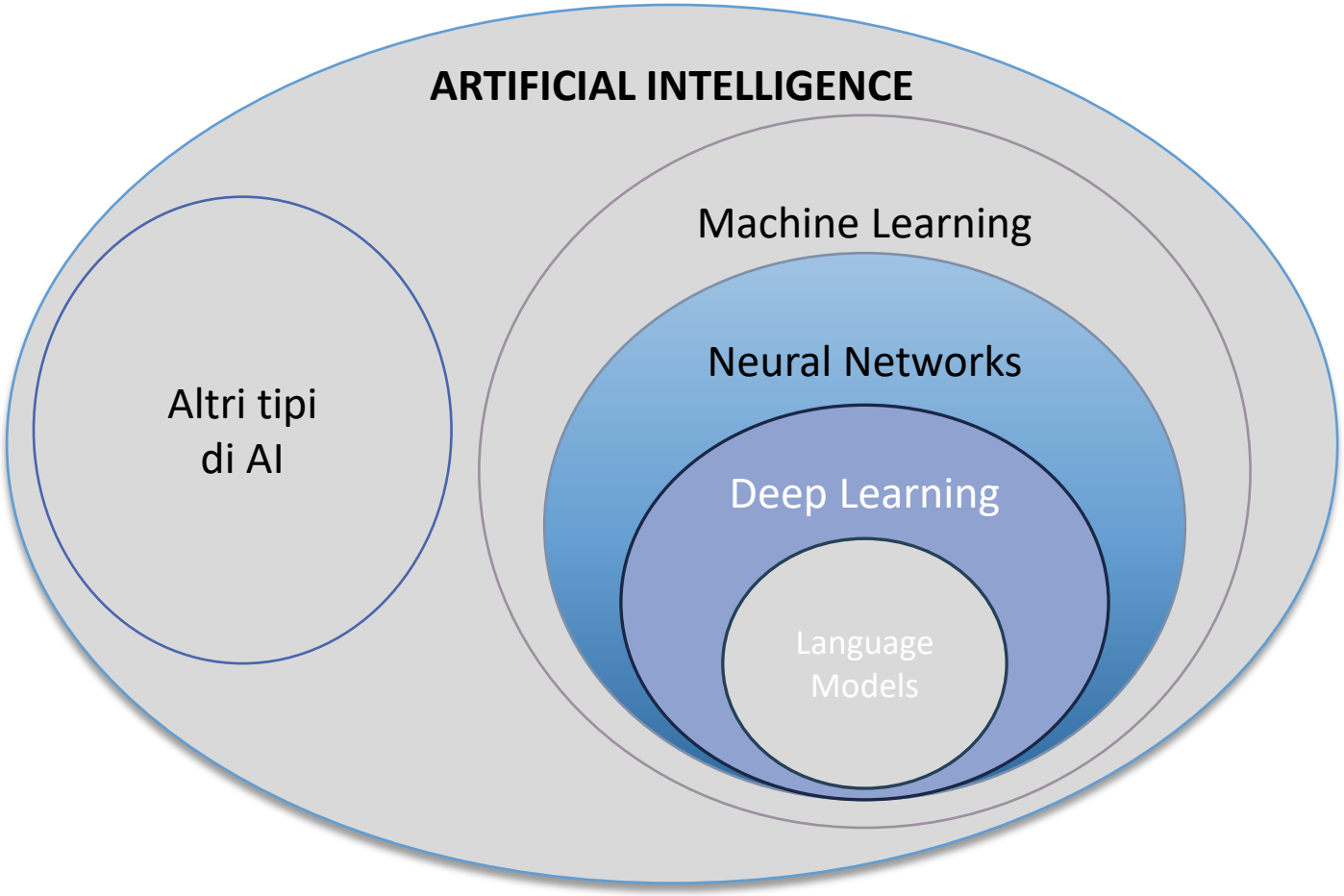
Clustering



Anomaly detection



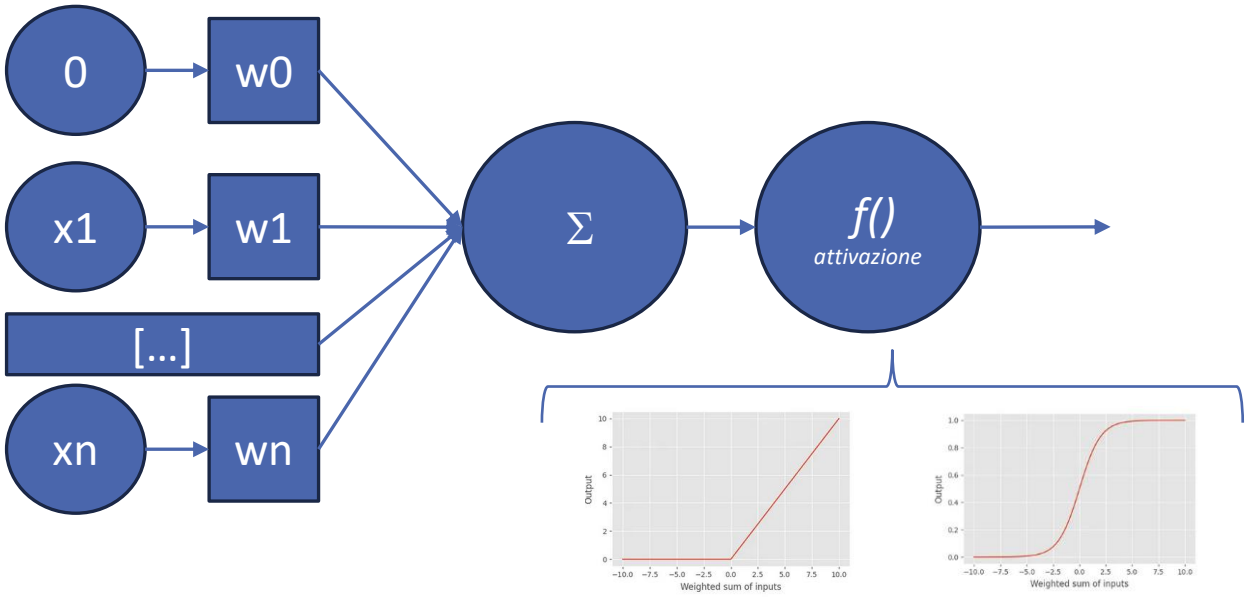
Neural Networks e Deep Learning

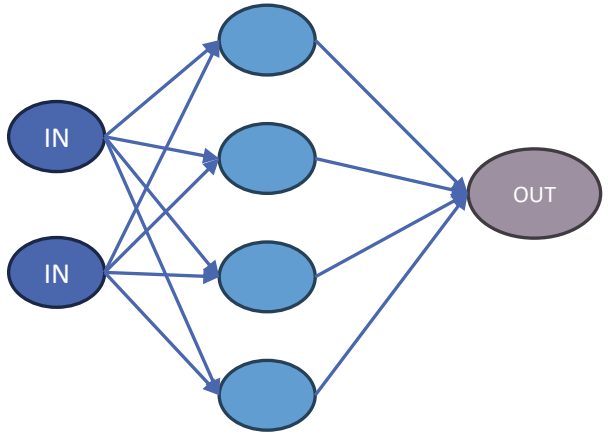


Machine Learning -> **Neural Network**

Architettura che prova a simulare il funzionamento del cervello umano

- Singolo strato di neuroni artificiali = rete neurale semplice (NN, Neural Network)
- Più strati di neuroni artificiali = rete neurale profonda (DNN, Deep Neural Network)
- Ciascun «neurone» è collegato a dati (numerici) di input, e restituisce un valore «pesato» (attivazione) al livello successivo

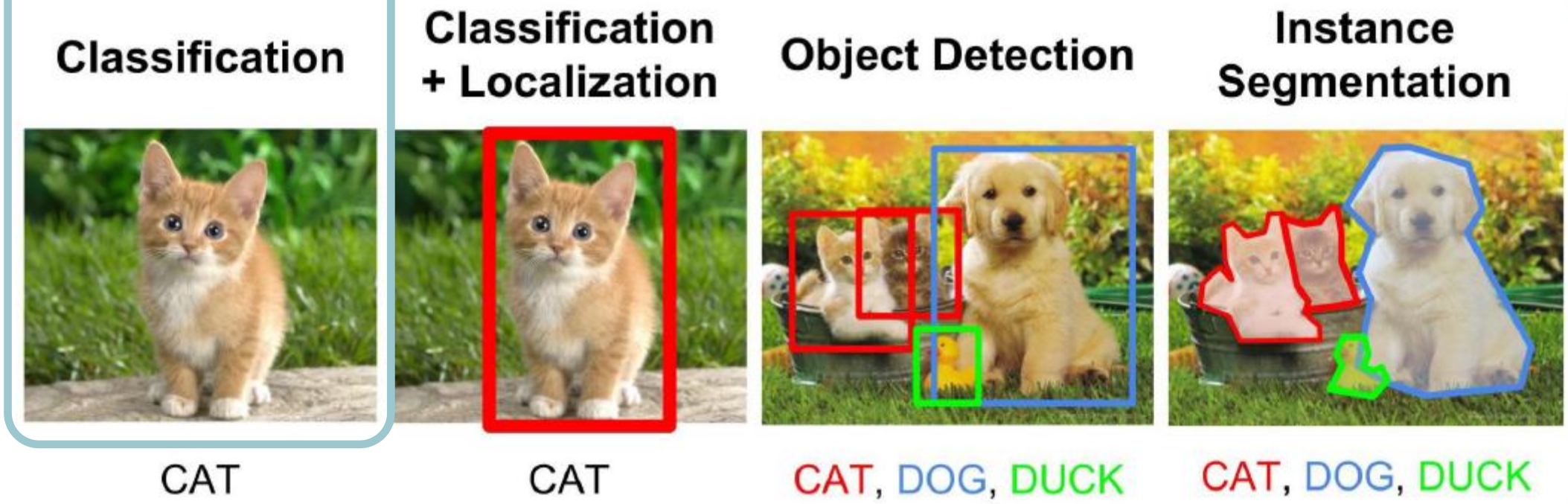




Esempi: relu, sigmoid, …

Esempio: tipologie di riconoscimento immagini

esperimento



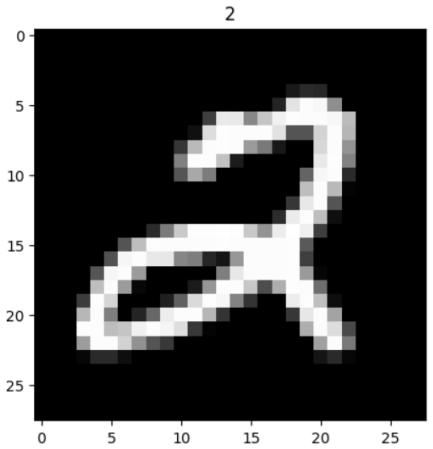
Oggetti singoli

Oggetti multipli

Esperimento didattico di classificazione da immagine 1/2

Riconoscimento numerico con keras e tensorflow su dataset mnist: train set e struttura NN

```
import tensorflow as tf
import tensorflow_datasets as tfds
[…]
train_ds, test_ds = tfds.load( 'mnist’, split = ['train', 'test’], as_supervised = True, batch_size = 8)
[…]
model = Sequential([
        Flatten(input_shape=(28, 28, 1)),
        Dense(20, activation='relu’),
        Dropout(0.5),
        Dense(10, activation='softmax’)
        ])
```



Una delle immagini di addestramento
28 x 28 px, 1 livello colore
28 x 28 x 1 = 784 parametri
X 20 neuroni = 15’680 + 20 bias = 15’700
Strato di risposta: 10 classi x 20 neuroni
= 200 + 10 = 210 parametri

Totale parametri da addestrare: **15’700**

Back propagation + metodo ottimizzazione
(es: discesa del gradiente)

Esperimento didattico di classificazione da immagine 2/2

Addestramento e valutazione modello

```
report = model.fit(train_ds, epochs=5)
[…]
model.evaluate(test_ds)
[…]
my_image = load_image('my3.png’)
[…]
probs = model.predict(my_batch)
print( np.argmax(probs) )
3
probs
```

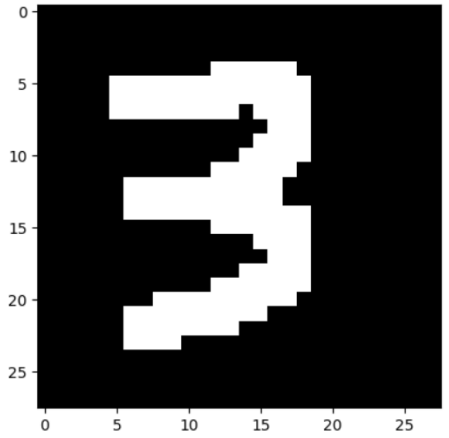


Immagine «nuova», mai vista dalla rete in precedenza

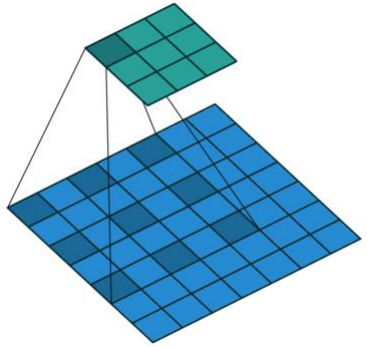
Vettore delle probabilità di appartenenza ad una delle 10 classi. Scelgo la risposta con probabilità maggiore, qui con confidenza al 99%

```
array([[3.4815632e-16, 1.0379343e-12, 2.5142677e-07, 9.9996901e-01, 1.9259736e-18, 3.0697793e-05, 1.2065921e-17,
1.6817488e-11, 6.5785342e-14, 1.5508711e-13]], dtype=float32)
```

Esperimento più realistico di classificazione da immagine

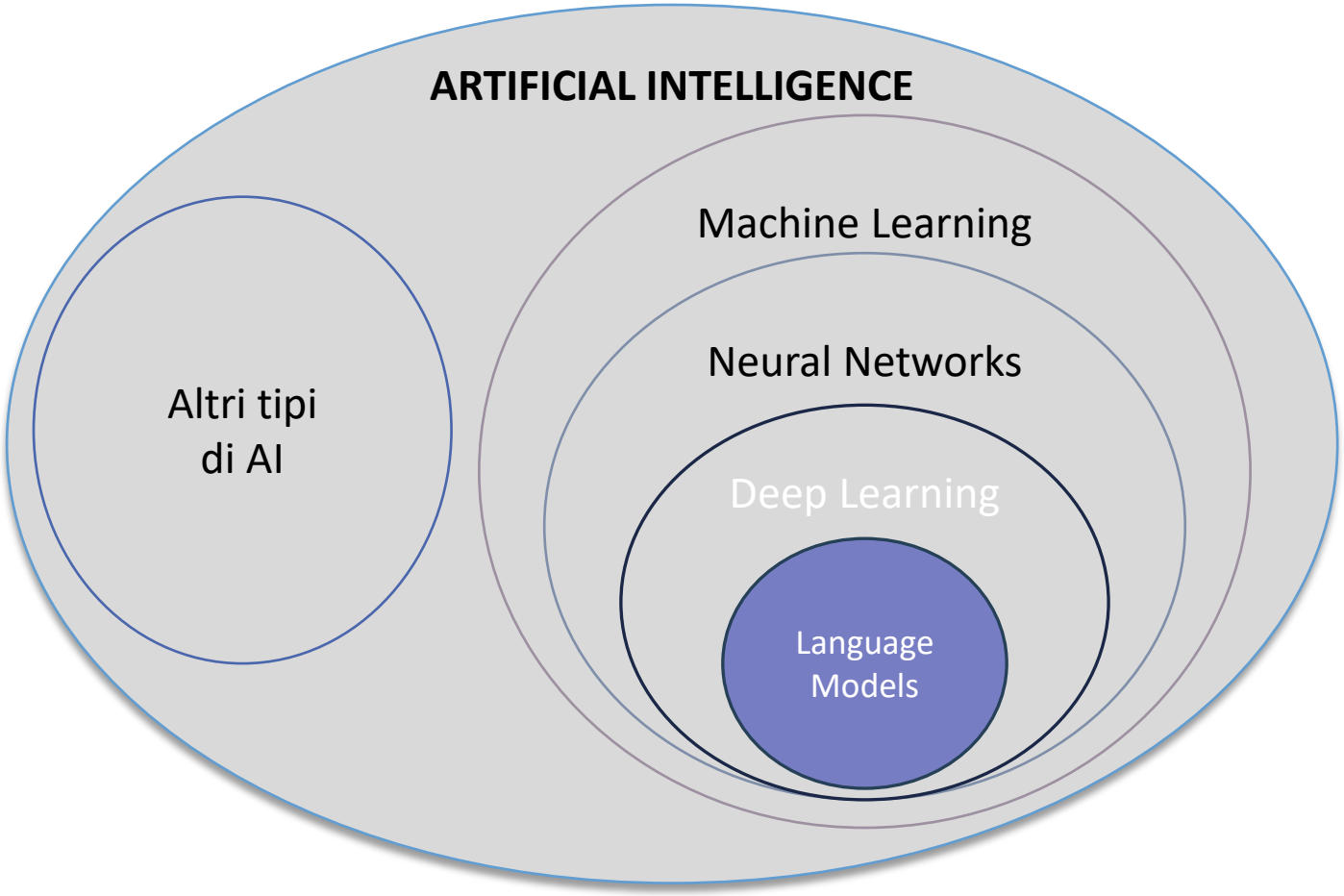
Per la classificazione statica, è più efficace «migliorare» il dato di partenza e ridurre i parametri di addestramento. Si fa con degli strati di convoluzione

```
model = Sequential([
    Conv2D(8, 3, input_shape=(28, 28, 1), padding = 'same'),
    MaxPool2D(),
    Conv2D(4, 3, padding = 'same'),
    MaxPool2D(),
    Flatten(),
    Dropout(0.5),
    Dense(10, activation='softmax’)
])
model.summary()
```



Esempio con kernel di 3x3 pixel
Il totale dei parametri di addestramento scende a 1970

(Large) Language Models



Machine Learning -> Neural Networks -> **Language Model**

Elaborazione e generazione del linguaggio naturale

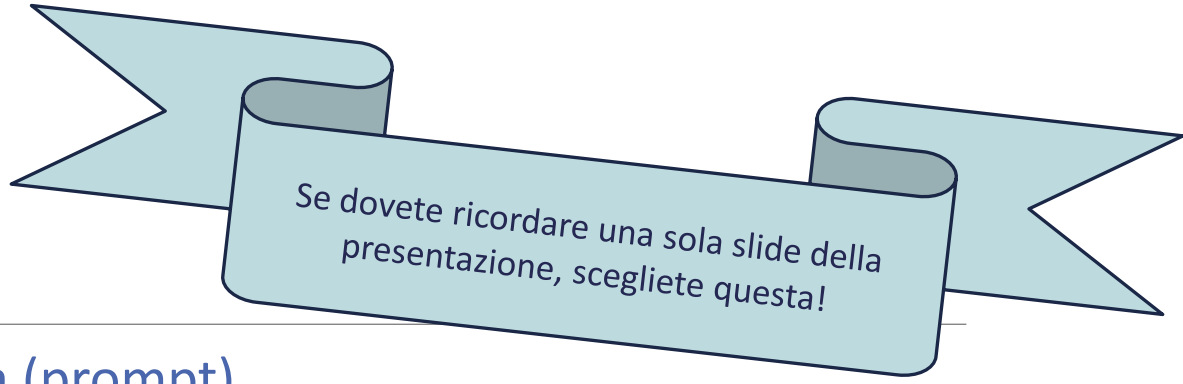
Generatori che **prevedendo ripetutamente il token (parola) successivo**, a partire da un testo di input (prompt)

Sono reti neurali artificiali che utilizzano l'architettura dei «*transformer*»

- le cui basi sono state gettate nel 2017 da Google (rif. doc. «Attention is all you need»)
- Hanno drasticamente migliorato i LM basati su DNN quali le Recurrent NN (utilizzano dati sequenziali o serie temporali), di cui superano alcune limitazioni

I migliori modelli hanno dimostrato di poter ottenere risultati simili al «fine-tuning» delle DNN (per adattare il modello a compiti specifici) tramite creazione di prompt complessi

Livello di comprensione di un LLM



Attenzione: un LLM risponde SEMPRE alla domanda (prompt)

NON comprendono «i concetti» espressi nel testo, ma comprendono la **vicinanza semantica**, e completano le frasi in modo «umanamente sensato»

- Attingendo ai dati del loro (imponente) addestramento
- Secondo una logica «semantica», ma non «fattuale»

Possono rispondere in modo sbagliato (nella sostanza, ma formalmente perfetto):

- «**allucinazioni**»: fenomeno per cui alcune risposte potrebbero essere costruite su informazioni completamente inventate
- Le risposte vanno **SEMPRE** verificate

Come funziona?

Tokenization and embedding + (positional encoding)

Da un input testuale si estrae un serie di stringhe o parti di esse (Token)

Il token viene trasformato in un vettore numerico

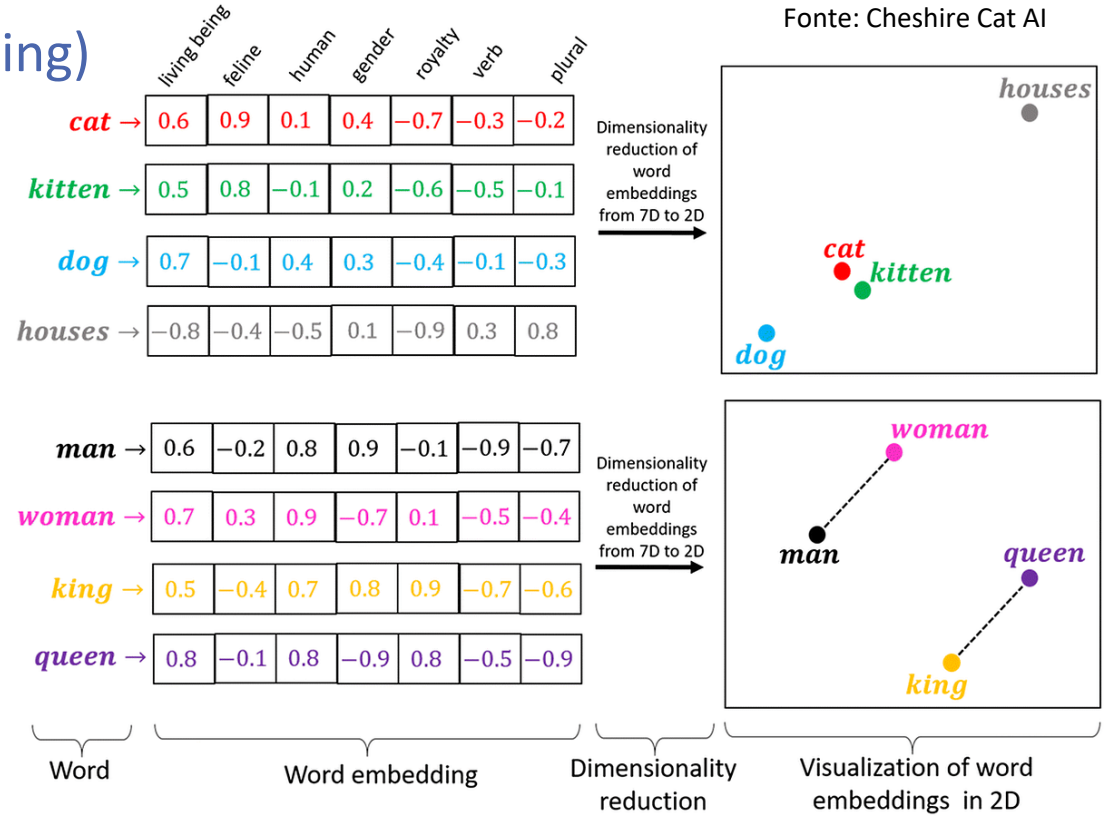
L'input originale viene trasformato in una rappresentazione «vettoriale» «multidimensionale» (*embedding*)

In questo modo viene incorporata nei token l'informazione semantica (in base all'addestramento ricevuto dalla rete)

Non più «umanamente leggibile», l'*embedding* esiste in uno «**spazio geometrico Euclideo**»:

- Può essere pensato come un «punto» in uno spazio multi-dimensionale
- Ad esso possono venire applicate operazioni geometriche
- *Esempio: la distanza tra due punti nello spazio vettoriale può dare informazioni sulla «vicinanza semantica» tra due parole (o frasi)*

Fonte: Cheshire Cat AI



King – Man + Woman = Queen

Similarità dei vettori: «cosine similarity»

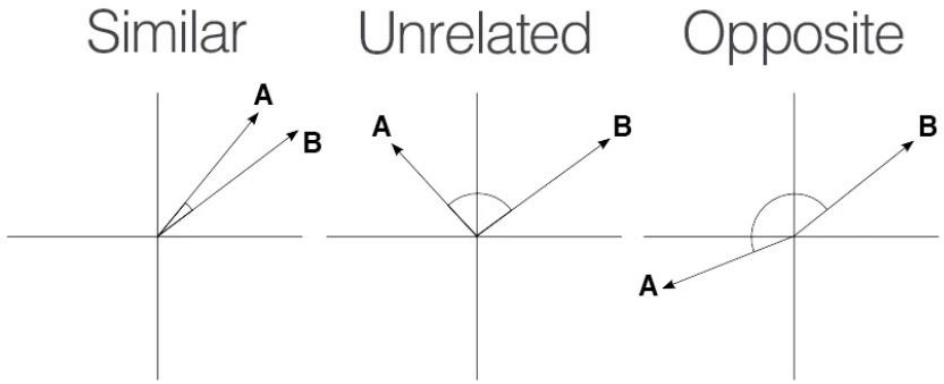
Una tecnica euristica per la misurazione della similitudine tra due vettori

Calcola e misura il loro coseno (angolo)

Consente di misurare:

- Similitudine
- Non - correlazione
- Opposto

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}.$$



1 (angolo 0)	0 (angolo 90)	-1 (angolo 180)
identici	Non correlati	Opposti

Fonte: Cheshire Cat AI

Modifica dell’embedding tramite self-attention

Modifica dei valori dei vettori dell’embedding per aggiungere informazion di «contesto»

Indipendentemente dalla distanza tra i token (parole) nel testo, individua quali token danno significato ad altri token

L’elaborazione non è sequenziale ma parallela

- veloce
- indipendente dalla lunghezza del testo in input

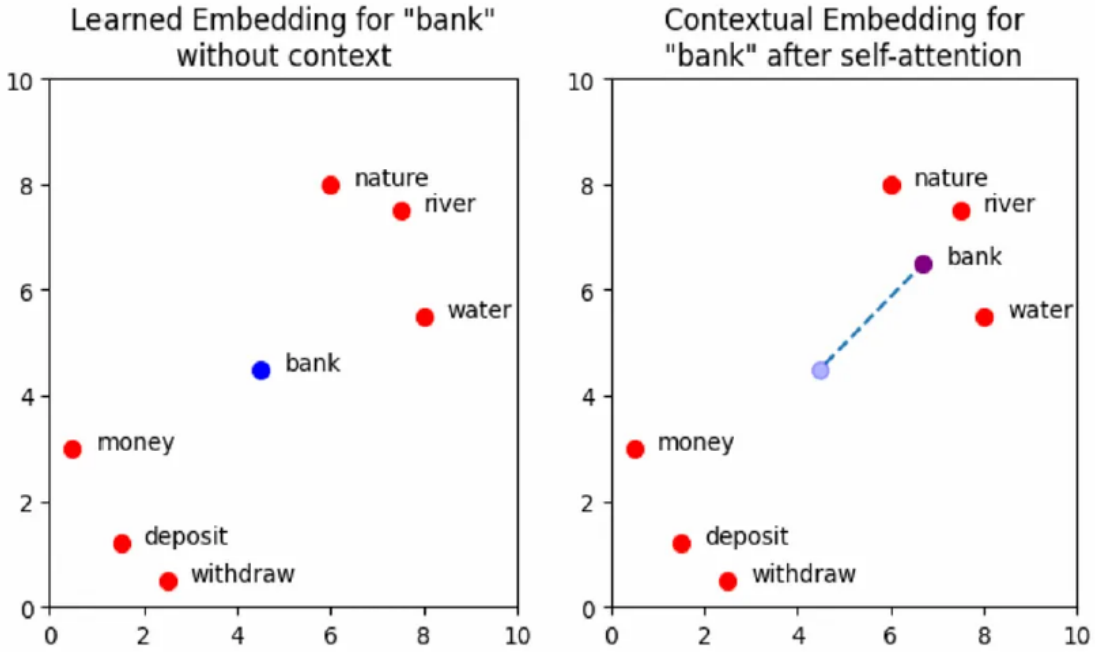
Focalizza l’attenzione sui token maggiormente significativi

«self» perché il meccanismo di «attenzione» viene applicato sulla medesima sequenza di input che viene elaborata

Una sottorete neurale addestra la matrice dei pesi «w» per ripesare i vettori dello spazio di embedding:

- Ogni token viene pesato rispetto agli altri (n token): **Q**uery
- Si elaborano (prodotto scalare) tutti i token (n token): **K**eys
- I «pesi w» della matrice n x n vengono addestrati da una NN
- I valori originali vengono moltiplicati per i pesi «w»: **V**alues

write a poem about a man fishing on a river **bank**



I transformers sono basati sul meccanismo di attenzione

Encoder e decoder

Stack di layer: ogni livello elabora i token ed aggiunge informazioni contestuali, così da apprendere rappresentazioni sempre più arricchite

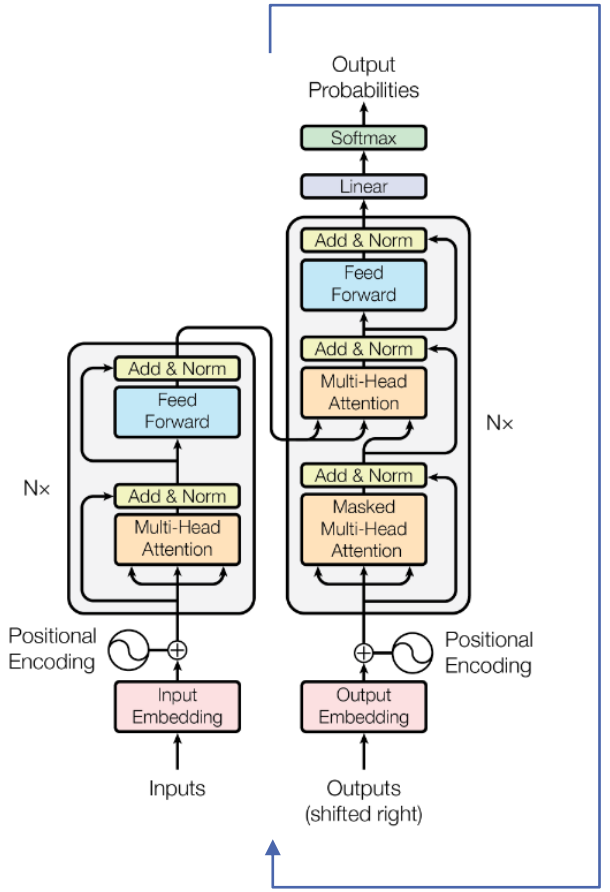
Multi-head attention: il meccanismo di attenzione viene eseguito più volte in parallelo, e consente al modello di considerare informazioni da diversi «sotto spazi di rappresentazione» contemporaneamente

I transformer vengono addestrati su enormi quantità di testi (inizialmente in lingua inglese, ora estesa a molte altre lingue. Ciascuna lingua ha le sue peculiarità) ed individuano le correlazioni statistiche (ad esempio, da tutto quello che è scritto in internet e sui libri…)

Terminato l'addestramento, tramite il contesto fornito dal «prompt», individuano la successiva parola più probabile. Questo risultato viene ripreso ed iterato nuovamente, per le parole successive, fino ad arrivare ad un token di stop.

Il meccanismo funziona anche con altri tipi di dati (immagini, audio, video) che vengono tokenizzati in uno spazio di embedding. Questa è l'ultima frontiera:

AI-MULTIMODALE: può operare contemporaneamente con diversi tipi di dati in input ed output (testi, audio, immagini, video)



Suggerimenti Considerazioni

Alcuni ambienti applicativi

Fruibili come «servizi»

- OpenAI: GPT (Generative Pretrained Transformer)
  - GPT 4o
  - GPTs
- Microsoft COPILOT
  - Motore di ricerca
  - Assistente office
  - Supporto alla programmazione
  - Scrittura creativa e traduzioni
- Google Gemini

Possibili utilizzi

Ricordarsi SEMPRE di VERIFICARE i risultati

- Ausilio alla scrittura
- Sintesi di documenti
- Traduzioni e completamento frasi
- Semplificazione delle ricerche
- Riduzione dei tempi di interazione con i clienti
- Insieme all’automazione in senso stretto, ottimizzazione delle operazioni di studio

Il «prompt»

Testo in linguaggio naturale che chiede all'AI generativa di eseguire una specifica attività

- Non tutti i tipi di input generano risultati utili
- Scrivere istruzioni inequivocabili
- Dare un contesto adeguato nel prompt, requisiti di output specificati nell'input
- Equilibrio tra informazioni mirate ed output desiderato
  - Un prompt semplice può mancare di contesto
  - Un prompt troppo complesso può «confondere» l'AI
- Sperimentare e perfezionare il prompt

Il prompt engineering (esempi)

Tecniche guida per ottenere le migliori soluzioni dalle IA generative

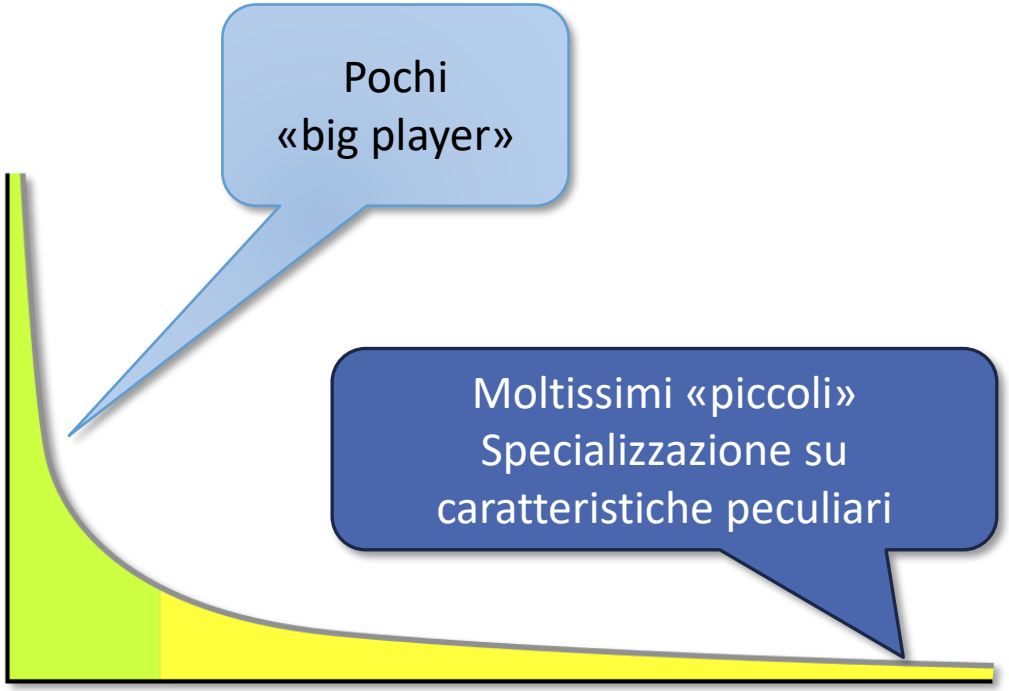
- Stimolazione della catena dei pensieri: scomposizione di una domanda complessa in parti logiche più piccole, imita una linea di pensiero fatta da passaggi intermedi invece di una risposta singola
- Suggerire l'albero dei pensieri: generalizza la catena dei pensieri chiedendo al modello di generare uno o più passaggi successivi, poi approfondendone ciascuno
- Richiesta sulla conoscenza generata: richiesta di generare i fatti pertinenti per completare il prompt, quindi completarlo
- Prompt min-max: richiede di elencare i problemi secondari, e poi risolverli in sequenza
- […]

RAG (Retrieval Augmented Generation), Agents

Tecniche guida per ottenere le migliori soluzioni dalle IA generative

- RAG: processi di ottimizzazione dell'output di LLM, che consentono di fare riferimento ad una base di conoscenza autorevole al di fuori delle fonti di addestramento, prima di generare una risposta
  - Es: GPTs
  - Es: The Cheshire Cat AI
- Agents: funzioni collegate al LLM e da esso richiamabili come parte della risposta
  - Es: «accendi il riscaldamento a casa»
  - Ha molto contesto specifico che gli LLM in generale non possono conoscere (quale casa? Quanti gradi desidero? Etc..)
- …e poi? Integrazioni con la BI (Business Intelligence) e l'automazione dei processi

La «coda lunga» per professionisti ed aziende



- Gli LLM attuali richiedono risorse «improponibili» (anche per aziende grandi)
- Dati, addestramento, risorse di calcolo, energia per il mantenimento e funzionamento…
- Sono generalisti, vanno bene per «tutti», ma non hanno risposte specifiche
- La coda lunga rappresenta l'enorme quantità di spazio per integrazioni specifiche e personalizzate, per i singoli o piccoli gruppi
  - Rispondono bene alle esigenze specifiche
  - Non hanno senso «per tutti»
- **L'open source può estendere e democratizzare l'AI**

Cenni Normativi

AI-ACT

UNIONE EUROPEA

https://www.europarl.europa.eu/news/it/press-room/20240308IPR19015/il-parlamento-europeo-approva-la-legge-sull-intelligenza-artificiale

AI ACT: Scopo e definizione

Art. 1 – scopo

Lo scopo del presente regolamento è **migliorare il funzionamento del mercato interno** e promuovere la diffusione di una intelligenza artificiale (IA) antropocentrica e affidabile, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali, compresi la democrazia, lo Stato di diritto e la protezione dell'ambiente, contro gli effetti nocivi dei sistemi di intelligenza artificiale nell'Unione nonché promuovere l'innovazione

Art. 3 – sistemi di AI

un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dagli input che riceve come generare output quali *previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*

AI ACT: soggetti coinvolti

Provider (fornitori) - sviluppatori

Sviluppatori

«*una persona fisica o giuridica, un' autorità pubblica, un'agenzia o altro organismo che **sviluppa** un sistema di IA o un modello di IA per finalità generali o che fa sviluppare un sistema di IA o un modello di IA per finalità generali e immette tale sistema o modello **sul mercato** o mette il servizio o il sistema di IA con il proprio nome o marchio a titolo oneroso o gratuito*»

Deployer

Utilizzatori ai fini commerciali / professionali

NON sono gli «utenti finali»

«*Una persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che **utilizza un sistema di IA** sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di una attività personale non professionale»*

AI ACT: ambito di applicazione

Si APPILCA

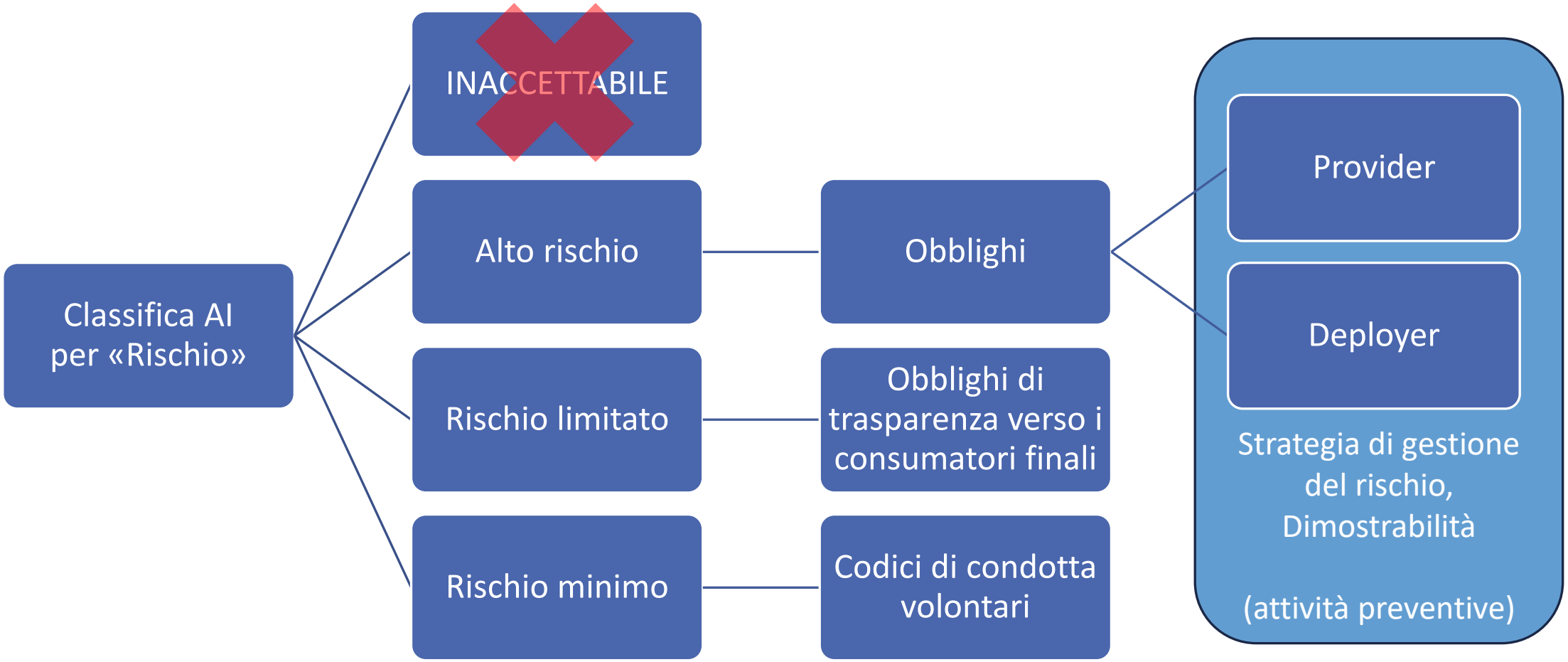
- Fornitori che immettono nel mercato UE
- Deployer stabiliti o ubicati in UE
- Fornitori e deployer con stabilimento in UE se il prodotto ha conseguenze su UE
- Importatori e distributori di sistemi IA
- Fabbricanti di prodotti che immettono sul mercato o mettono in servizio sistema IA insieme al loro prodotto e con loro nome o marchio
- Rappresentanti autorizzati di fornitori non stabiliti nell'UE
- Persone interessate che si trovino nell'UE

NON si applica

- Settori che non rientrano sotto il perimetro del diritto dell'Unione
- Sistemi sviluppati e messi in servizio per scopi militari, di difesa o di sicurezza nazionale
- Sistemi sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici
- Ai sistemi PRIMA della loro immissione sul mercato o messa in servizio
- Persone fisiche che utilizzano sistemi di IA nel corso di una attività non professionale puramente personale

AI ACT: approccio basato sul rischio

Non si applica ai sistemi IA «*General Purpose*», che seguono logiche differenti



Sistemi ad alto rischio

VERIFICHE PREVENTIVE ALL'ENTRATA IN FUNZIONE

- Documentazione tecnica: chiara e trasparente
- Conservazione delle registrazioni
- Gli operatori DEVONO fornire informazioni agli utenti
- DEVONO essere adottati adeguati livelli di
  - Cybersicurezza
  - Robustezza
  - Accuratezza
- **Valutazione di impatto PRIMA che tali sistemi entrino nel mercato**
  - DIMOSTRABILITA
  - STRATEGIA DI GESTIONE DEL RISCHIO
- Includono i dispositivi medici ed i sistemi di identificazione biometrica, categorizzazione e riconoscimento delle emozioni

AI ACT: obblighi e responsabilità

Provider (fornitore/sviluppatore)

- Stabilire un **sistema di gestione del rischio**
- Governance dei dati (**dimostrabile**!)
- Documentazione tecnica di conformità per le autorità
- Progettazione «*trusworth by design*» e aggiornabile
- Fornire istruzioni ai deployer per consentire la conformità
- Progettare il sistema per consentire la **supervisione umana** ai deployer
- Stabilire un sistema di gestione della qualità al fine di garantire la conformità

Deployer

- Adozione di idonee **misure tecniche ed organizzative** per garantire di utilizzare tali sistemi in conformità alle istruzioni
- Affidare la **sorveglianza umana** a persone fisiche competenti
- Nella misura in cui esercita il controllo sui dati di input, **garantire** che tali **dati** siano **pertinenti** e sufficientemente **rappresentativi**
- Monitoraggio del funzionamento dei sistemi IA ad alto rischio sulla base delle istruzioni per l'uso e, se del caso, informazione al provider
- Svolgono una **DPIA** (Data Protection Impact Assessment)

Sintesi elementi principali AI-ACT

- **Applicazioni vietate** - minaccia ai diritti dei cittadini: categorizzazioni biometriche, immagini facciali da internet o sistemi di registrazione per creare db di riconoscimento, sistemi di riconoscimento emozioni su luoghi di lavoro e scuole, credito sociale, pratiche di polizia predittiva (se basate solo su profilazione personale), sistemi di manipolazione del comportamento umano
- **Eccezioni per le forze dell'ordine** – in alcune situazioni specifiche PREVISTE DALLA LEGGE sarà consentita l'identificazione «in tempo reale», nel rispetto di garanzie rigorose (limiti spazio e tempo, es ricerca persona scomparsa, prevenzione attacchi terroristici). Uso a posteriori solo su autorizzazione giudiziaria.
- **Obblighi per i sistemi ad alto rischio** – valutazione e riduzione rischi, registri d'uso, trasparenza ed accuratezza, sorveglianza umana. I cittadini avranno diritto a presentare reclami sui sistemi AI e ricevere spiegazioni sulle decisioni dei sistemi AI che impattano sui propri diritti
- **Obblighi di trasparenza** – requisiti di trasparenza, rispetto norme UE sul diritto d'autore in fase di addestramento dei modelli. Ulteriori obblighi per i modelli più «potenti»
- **Misure a sostegno dell'innovazione e delle PMI** – sandbox (spazi sperimentazione e prova) per sviluppare sistemi AI innovativi ed addestrarli prima della loro immissione sul mercato

Question time

GRAZIE PER L'ATTENZIONE

I miei contatti:
http://studio.masson.biz/
studio@masson.biz

So long, and thanks for all the fish!