POLITECNICO
MILANO 1863

AIRLAB
ARTIFICIAL INTELLIGENCE AND ROBOTICS LAB

Come impara l'intelligenza artificiale?

Andrea Bonarini
andrea.bonarini@polimi.it
https://bonarini.faculty.polimi.it

Artificial Intelligence and Robotics Lab
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
https://airlab.deib.polimi.it

Machine Learning

Dati i successi ottenuti su **un aspetto** dell'intelligenza, l'**apprendimento automatico**, oggi si usa il termine Intelligenza Artificiale soprattutto per indicare la capacità della macchina di **costruire dei modelli** per rispondere ad **esigenze** dei **progettisti** e dei **committenti**.

Ogni modello, a fronte di dati in ingresso deve fornire il risultato per cui è stato costruito: un'interpretazione dei dati, un'azione da fare, una previsione, …

Da cosa impara un sistema di intelligenza artificiale?

Tre modalità principali, che imitano i nostri meccanismi di apprendimento, e che dipendono dall'informazione che abbiamo a disposizione:

- **esempi**: si forniscono alla macchina esempi di cosa ci si aspetta che il modello produca quando riceve certi ingressi. Gli algoritmi di IA **generalizzano** questi esempi producendo un modello che, coerentemente con gli esempi visti, è in grado di dare risposte anche a fronte di dati in ingresso diversi (apprendimento **supervisionato**).

Da cosa impara un sistema di intelligenza artificiale?

Tre modalità principali, che imitano i nostri meccanismi di apprendimento, e che dipendono dall'informazione che abbiamo a disposizione:

- **esempi**: apprendimento **supervisionato**
- **valutazioni**: la macchina propone modelli che sono valutati e, sulla base della valutazione, le parti di modello che portano a buone soluzioni sono favorite a restare, le altre sono gradualmente eliminate (apprendimento per **rinforzo**)

Da cosa impara un sistema di intelligenza artificiale?

Tre modalità principali, che imitano i nostri meccanismi di apprendimento, e che dipendono dall'informazione che abbiamo a disposizione:

- **esempi**: apprendimento **supervisionato**
- **valutazioni**: apprendimento per **rinforzo**
- **dati**: definiti dei criteri di similitudine, la macchina individua le caratteristiche di elementi simili, le regolarità tra i dati (apprendimento **non supervisionato**)

Da cosa impara un sistema di intelligenza artificiale?

Tre modalità principali, che imitano i nostri meccanismi di apprendimento, e che dipendono dall'informazione che abbiamo a disposizione:

- **esempi**: apprendimento **supervisionato**
- **valutazioni**: apprendimento per **rinforzo**
- **dati**: apprendimento **non supervisionato**

In tutti i casi, **l'apprendimento è indirizzato dall'uomo** (progettista e commitente) e la macchina non fa che eseguire operazioni, seguendo quanto il progettista ha determinato.

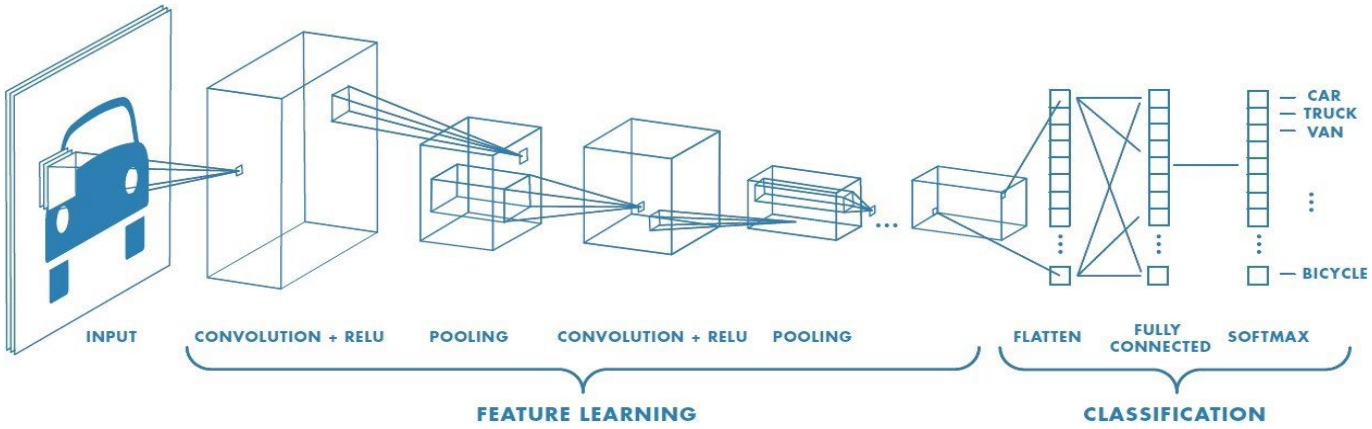
Le operazioni per costruire modelli in molti casi sono così complesse da non essere facilmente **comprensibili**, né in certa misura **governabili**.

Entriamo più in profondità per cercare di capire…

Apprendimento supervisionato

Migliaia di unità computazionali vagamente ispirate alle **migliaia di miliardi** di neuroni che abbiamo nel cervello ognuno con un **migliaio di connessioni** (160.000 Km di connessioni) per imparare a generare uscite a fronte di dati in ingresso: **reti neurali**.

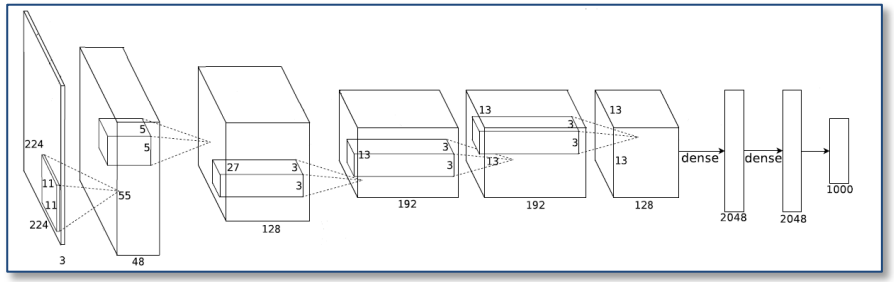
Deep Learning: reti neurali che imparano per livelli gerarchici di complessità.



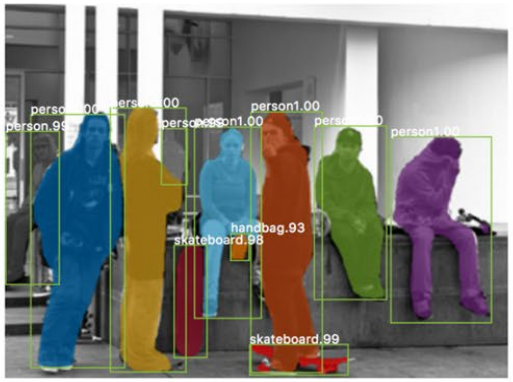
Es.: migliaia di immagini ognuna delle quali etichettata (esempi ingresso/uscita)

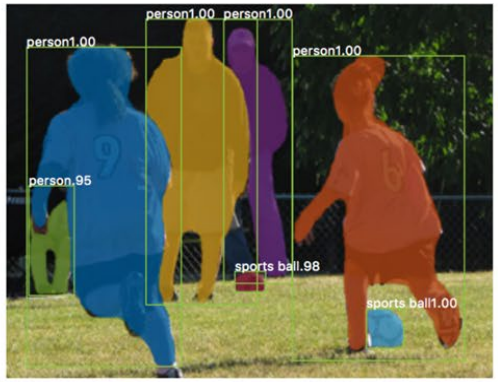
ImageNet

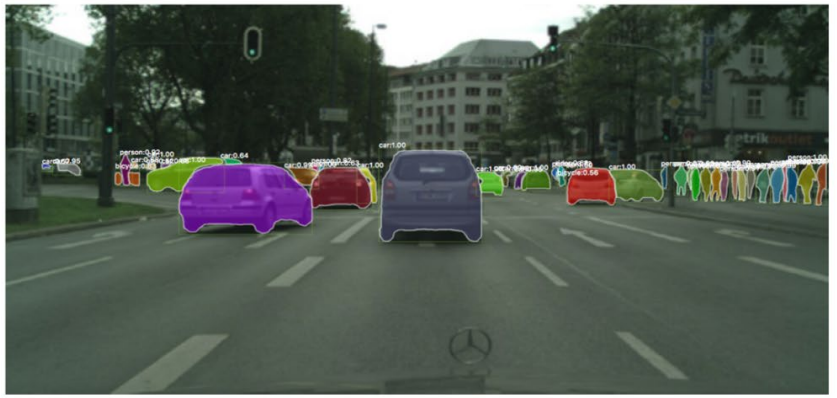
21.000 categorie,
14 milioni di immagini etichettate

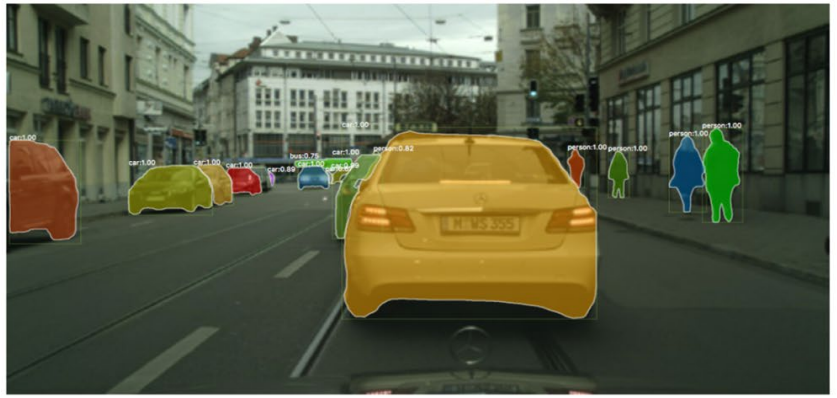


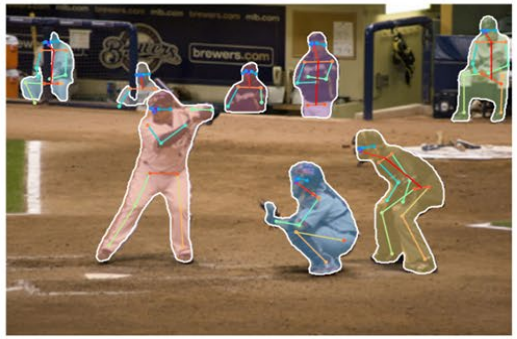
Riconoscimento di oggetti e azioni

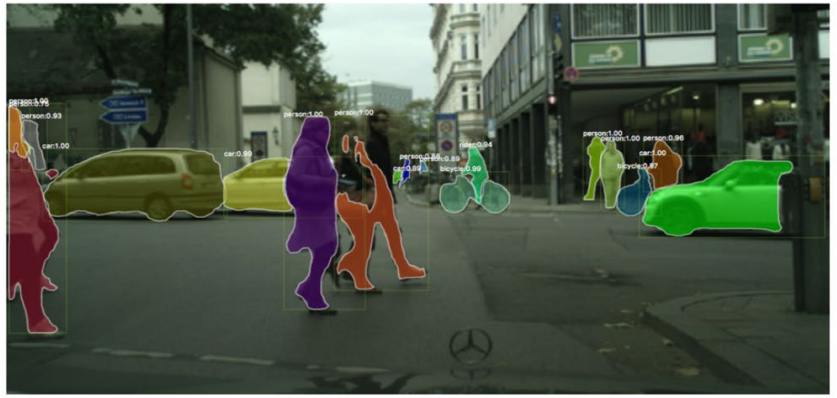






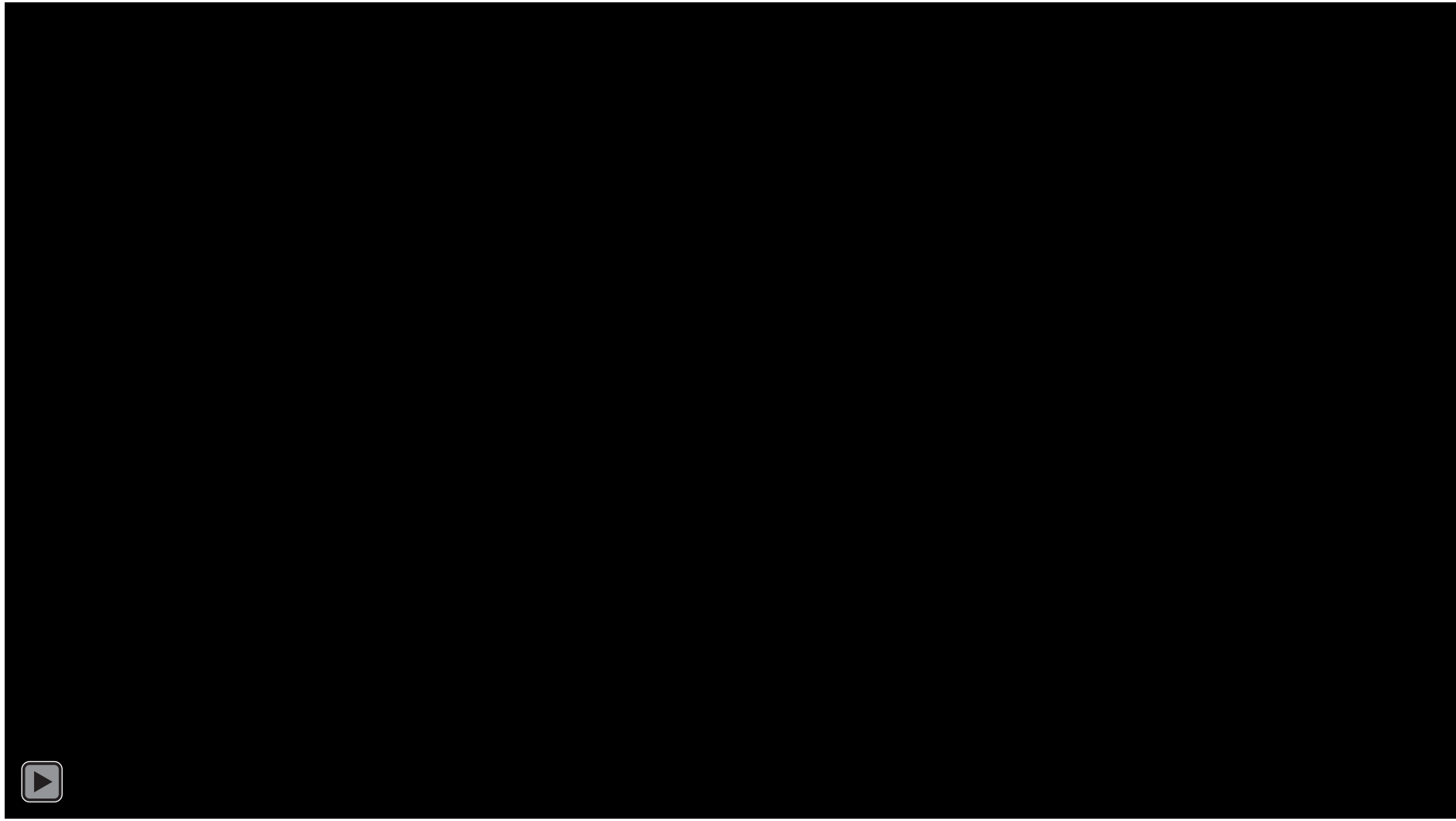




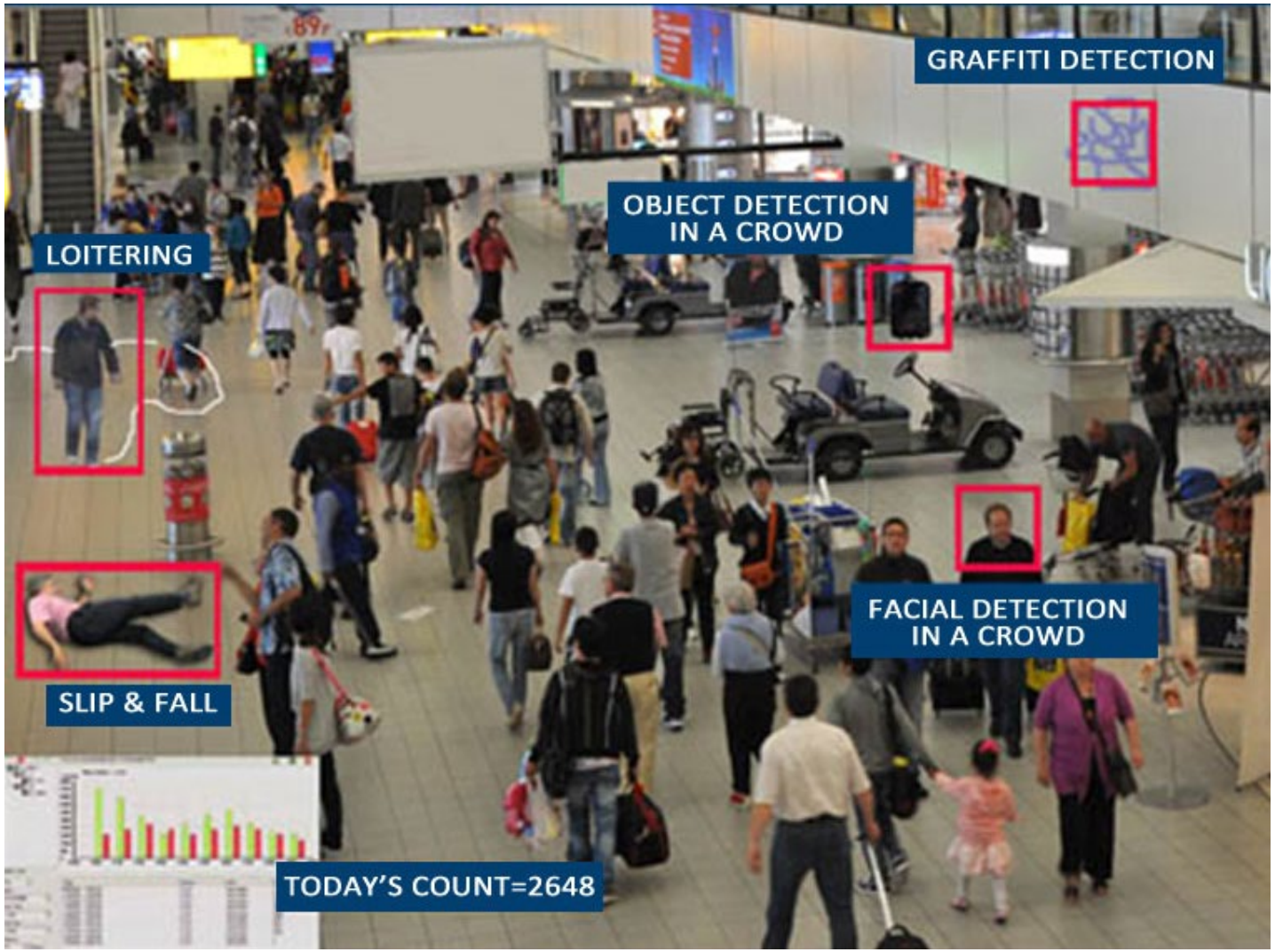


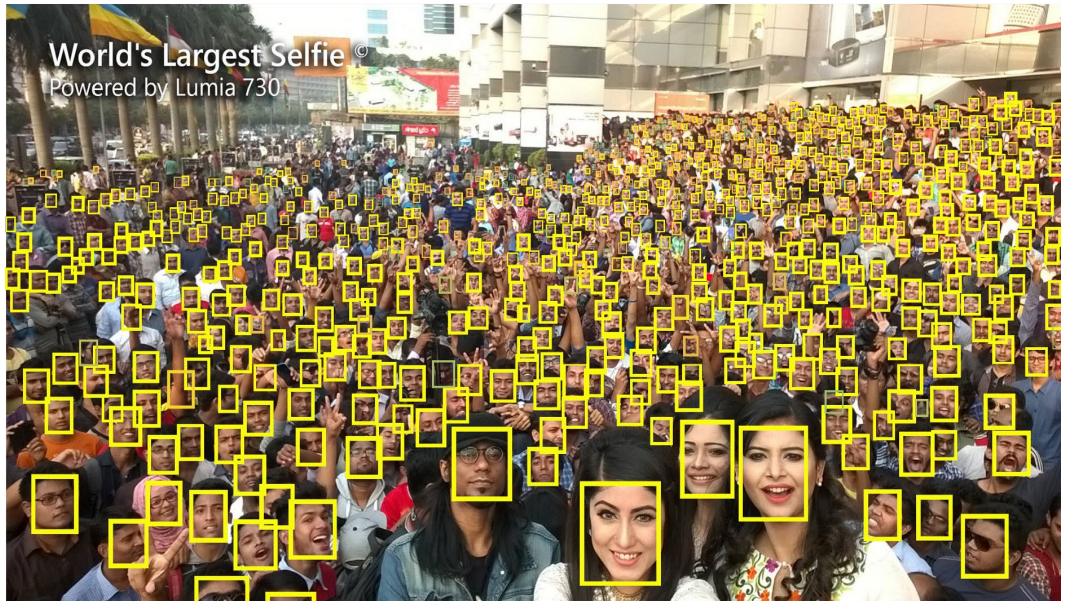
Un camion per Tesla…

Questi algoritmi non hanno mai un’accuratezza del 100%



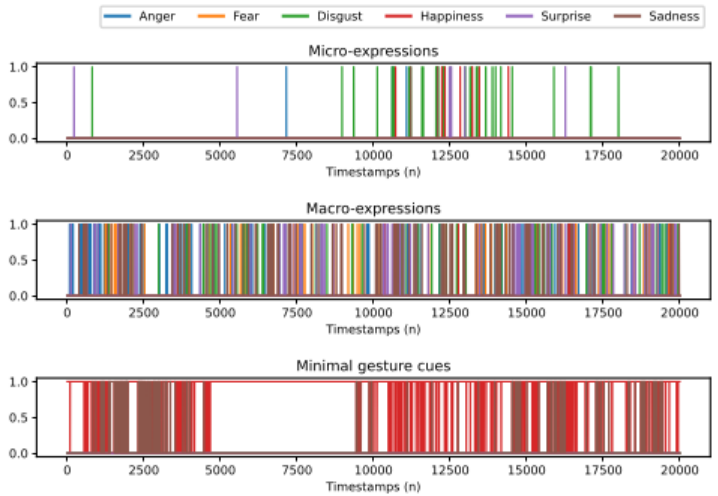
Analisi video intelligente



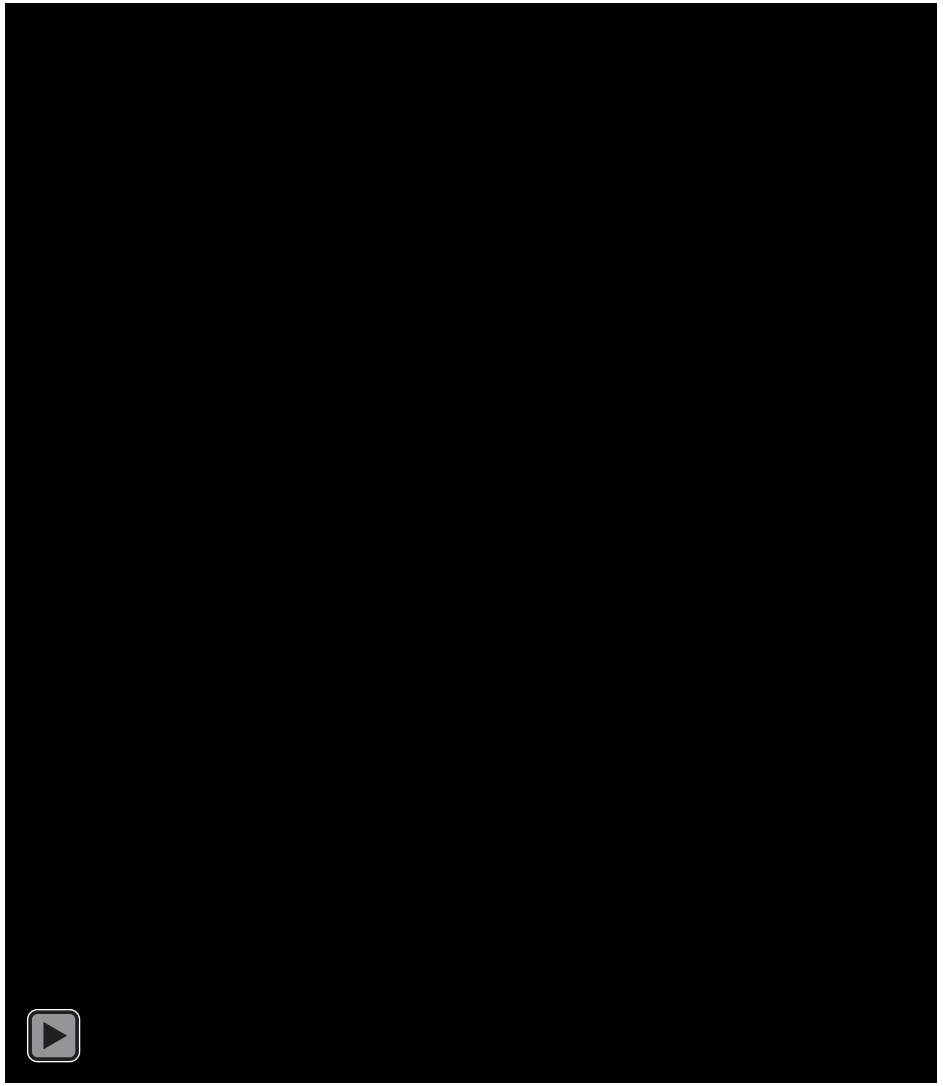


Riconoscimento emozioni

Usato per controllare persone sospette, studenti, …



Tre modelli, ognuno molto buono su una di tre diverse fonti (macro-espressioni, micro-espressioni, gesti), quando attivati su immagini diverse da quelle dei dataset originali, sono spesso in disaccordo.
Chi ha veramente ragione?



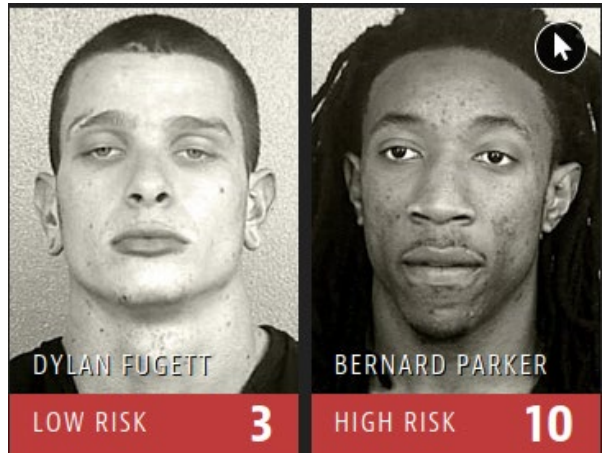
Distribuzione non corretta dei dati: bias

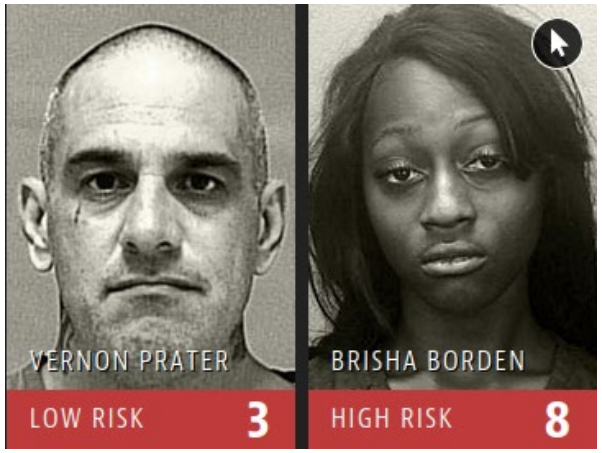
L'etichettatura, la scelta e la distribuzione dei dati influenzano il modello prodotto.

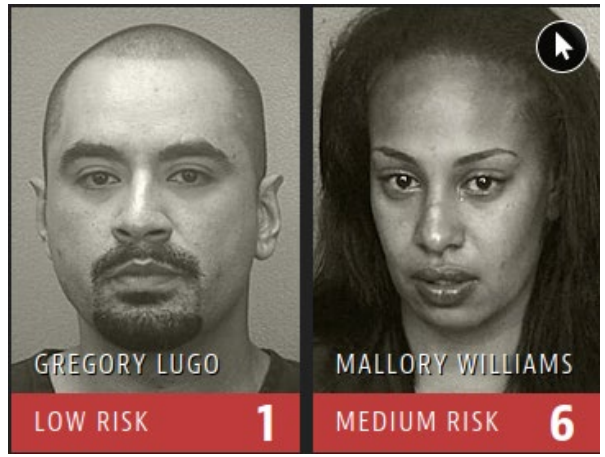
Purtroppo non è comune esserne coscienti fino in fondo.

Anche le proposte di Netflix, Amazon, Pinterest, … sono influenzate da condizionamenti impliciti (noi) ed indotti dal sistema (modelli generali, politiche, …)

Valutazione del rischio di commettere di nuovo reati. Quelli con valutazione di rischio più alta non han più commesso reati, gli altri ne han commessi di significativi.





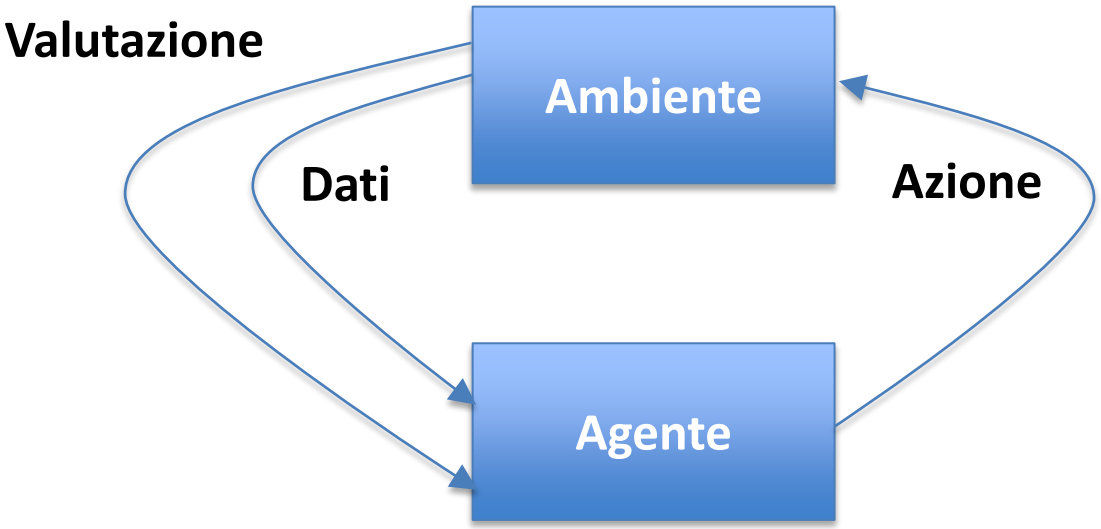


Apprendimento per rinforzo

La macchina apprende sulla base di **valutazioni circa la qualità dei risultati** prodotti, fornite dai committenti e mediate dai progettisti. Il criterio di valutazione determina come apprende la macchina.

Usato tipicamente per apprendere azioni e strategie, ma non solo.

In presenza di dati complessi una rete neurale interpreta i dati per fornirne una generalizzazione e un'approsimazione (**Deep Reinforcement Learning**).

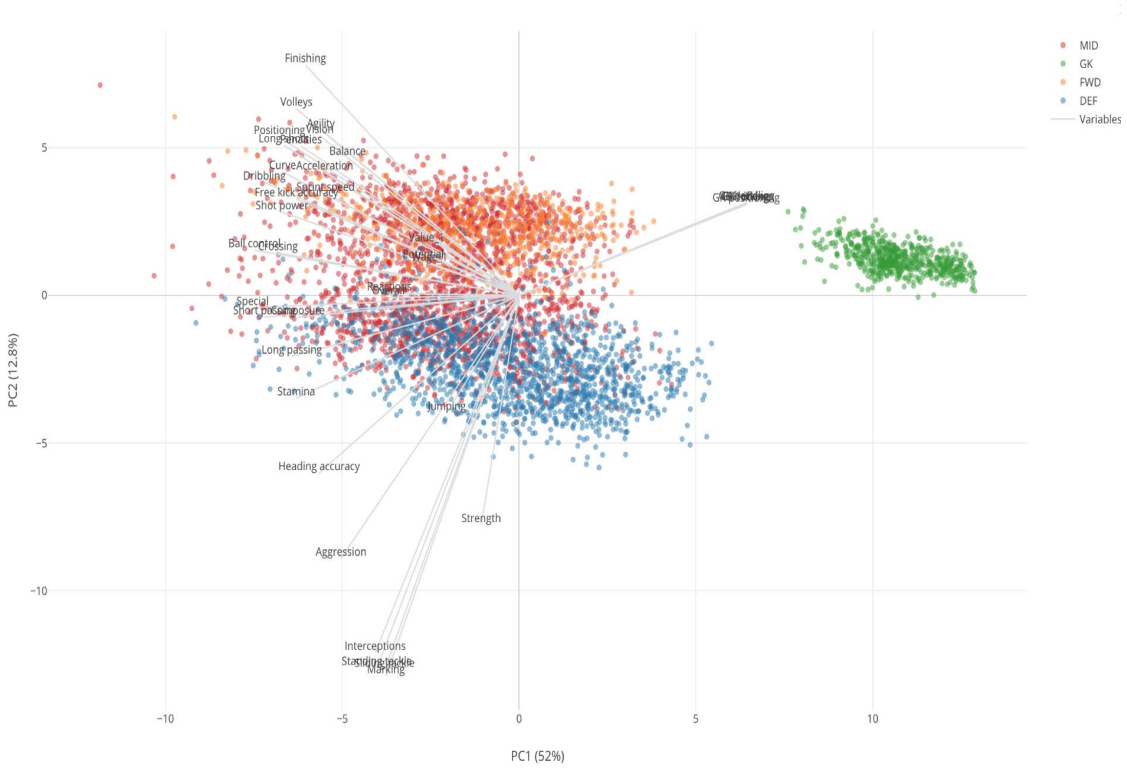


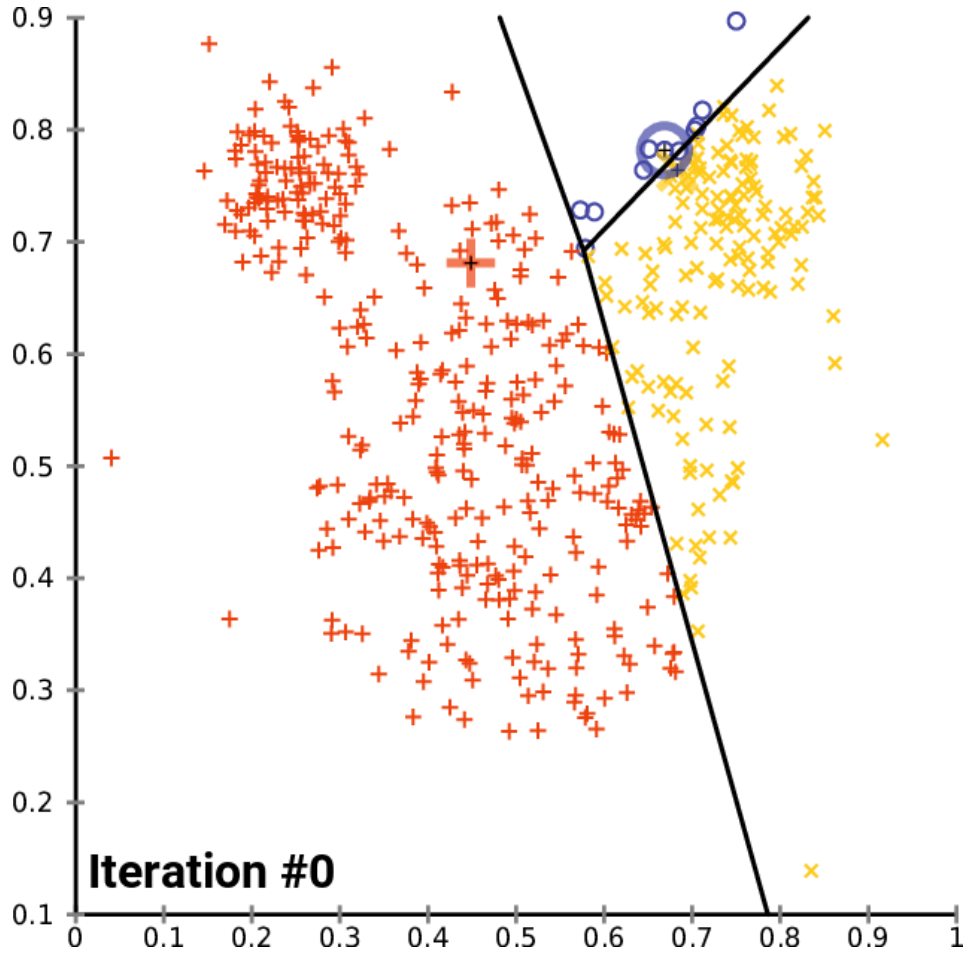
Non solo alpha-go

- Auto autonome: ottimizzazione della traiettoria, pianificazione del movimento, ottimizzazione del controllo, politiche di gestione dello scenario
- Controllo di robot e ambienti (es.: il raffreddamento dei data center di Google)
- Gestione finanziaria: scelte in borsa, …
- Linguaggio naturale: riassunto, generazione di risposte, traduzioni, gestione dialogo
- Salute: scelta dei trattamenti
- Offerte servizi: suggerimenti, raccomandazioni, ottimizzazione video streaming, …
- Marketing: caratteristiche dei consumatori, definizione prezzi personalizzati, …
- Videogames

Apprendimento non supervisionato

La macchina apprende **regolarità presenti nei dati** basate su criteri di similitudine forniti dai progettisti per ogni caratteristica.



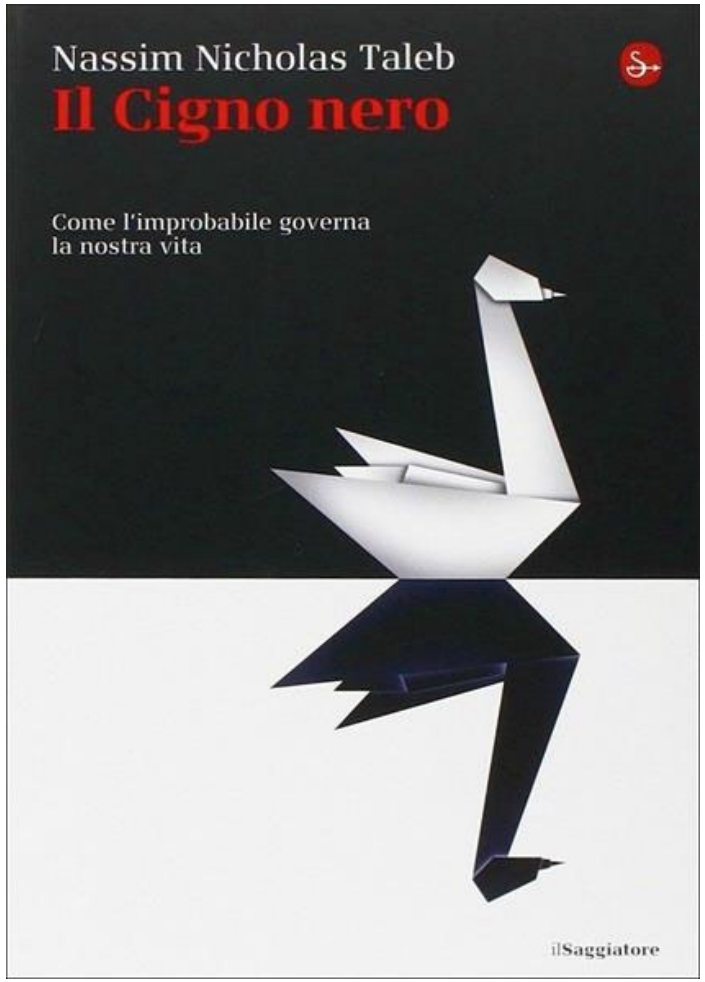


Il cigno nero

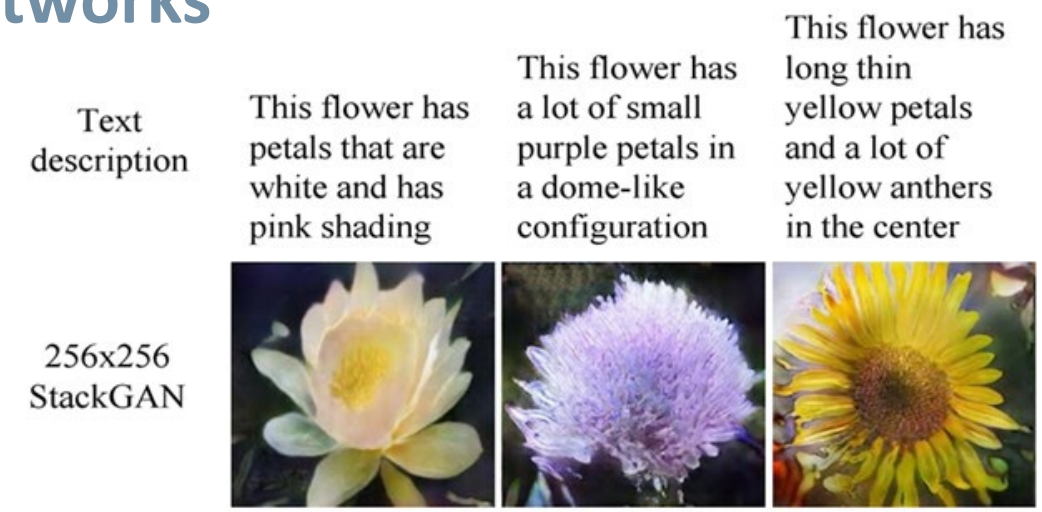
Nella realtà non esistono solo *esemplari medi*: non si può prendere il centro del gruppo come rappresentativo di tutti gli esemplari del gruppo

- I valori delle caratteristiche sono dispersi
- Esistono eccezioni
- Rischio di discriminazione degli esemplari distanti da tutti gli altri (outlier)
- …

A volte gli esemplari significativi sono proprio i «cigni neri»



GANN - Generative Adversarial Neural Networks



Questi sistemi non "capiscono" nulla; sono in grado di generare immagini o testo spesso coerenti, ma anche errati o non credibili.
Un umano sceglie a posteriori le proposte migliori.

Transformer

I transformer sono usati per elaborazione di linguaggio naturale (Es. ChatGPT è un Transformer).

Il processo di apprendimento di un Transformer prevede l'utilizzo di una grande quantità di dati di addestramento e consiste in due fasi principali:

- *Addestramento non supervisionato su grandi quantità di dati: Il modello viene addestrato su grandi quantità di dati di testo (ad esempio, interi libri o corpus di notizie) utilizzando un approccio non supervisionato. Durante l'addestramento, il modello impara a riconoscere le relazioni tra le parole e le frasi all'interno del testo e a rappresentarle in uno spazio a molte dimensioni.*
- *Fine-tuning supervisionato: Una volta addestrato, il modello viene sottoposto ad un addestramento supervisionato su un dataset specifico per il compito che deve svolgere (ad esempio, traduzione automatica o generazione di testo). Durante questa fase, il modello viene adattato specificamente al compito e alle caratteristiche del dataset, migliorando la sua capacità di generalizzazione.*

Linguaggio naturale

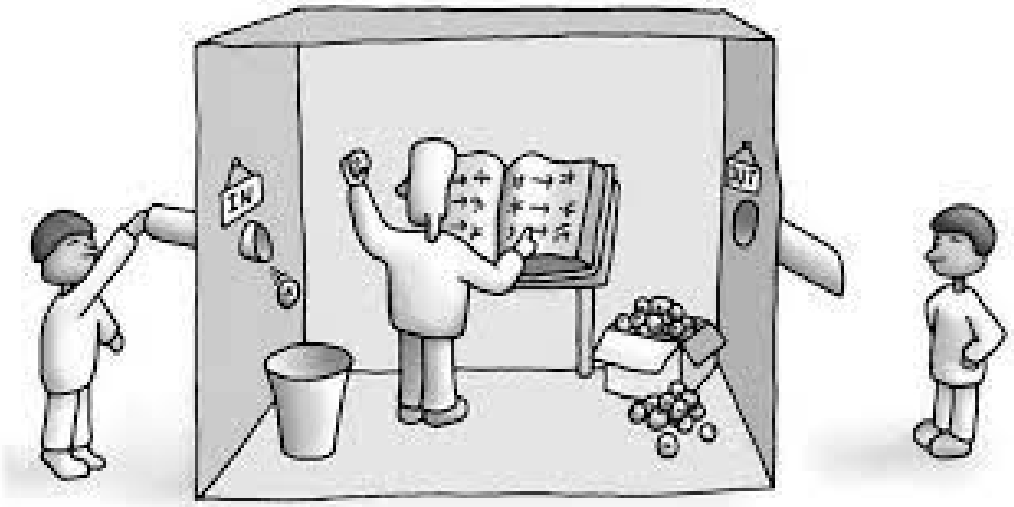
Riconoscimento linguaggio naturale parlato abbastanza buono anche se …

Traduzione di linguaggio naturale ottima

Generazione di testo in linguaggio naturale: ancora molto da fare, nonostante ChatGPT, perché il sistema dovrebbe capire cosa scrive, mentre di fatto è come una complessa "stanza cinese" di Searle.

Comunque è sufficiente per generare testi e risposte credibili con conseguenze importanti,

… e post convincenti (30%-60% dei post su Twitter, 50 milioni di bot, nel 2021)



Meraviglioso, ma…

Occorre considerare:

- Possibili bias nei dati
- Possibili bias nella definizione degli obiettivi
- Qualità dell'algoritmo
- Generalità del modello prodotto
- Nessun algoritmo ha affidabilità del 100%
- **Utenti (decisori, politici, pubblico …) tendono a considerare i modelli come se fossero affidabili al 100%**

Una storia già vissuta…

Quanto un decisore può fidarsi di quello che la macchina produce?

Tutti gli algoritmi di apprendimento producono modelli che producono risultati con un certo **grado di affidabilità**, più o meno lontano dal 100%, spesso anche del 70-80% o meno

Sugli **articoli scientifici** sono riportati i metodi sperimentali e i risultati ottenuti

I **media** riportano (spesso "notiziabilizzati") i risultati pubblicizzati dagli **uffici comunicazione** di centri di ricerca e aziende, in entrambi i casi finalizzati ad ottenere vantaggi economici

I **decisori** spesso accedono solo ai media, e **non hanno competenze per valutare**, come riconosciuto anche dalla comunità scientifica e commissioni UE => necessità di normative

Responsabilità

Chi è responsabile per un corretto uso di sistemi di intelligenza artificiale?

- Produttori di tecnologia
- Venditori di tecnologia
- Media
- Compratori di tecnologia
- Utenti finali









Conclusione

Intelligenza artificiale e apprendimento automatico sono:

- strumenti potenti che possono **migliorare** molto la nostra vita
- **non affidabili** al 100%
- Le caratteristiche dei prodotti di IA sono **dipendenti** da chi li progetta e da chi li usa
- Siamo tutti **responsabili** per farne un uso appropriato
- Occorre diffondere il più possibile conoscenza sui loro **limiti** e le loro **potenzialità**
- Occorre **definire regole** per impedirne un uso improprio che **possano essere applicate**