UNIVERSITÀ DI TORINO

Intelligenza Artificiale:
uno sguardo centrato sui dati
e sui diritti di tutt3

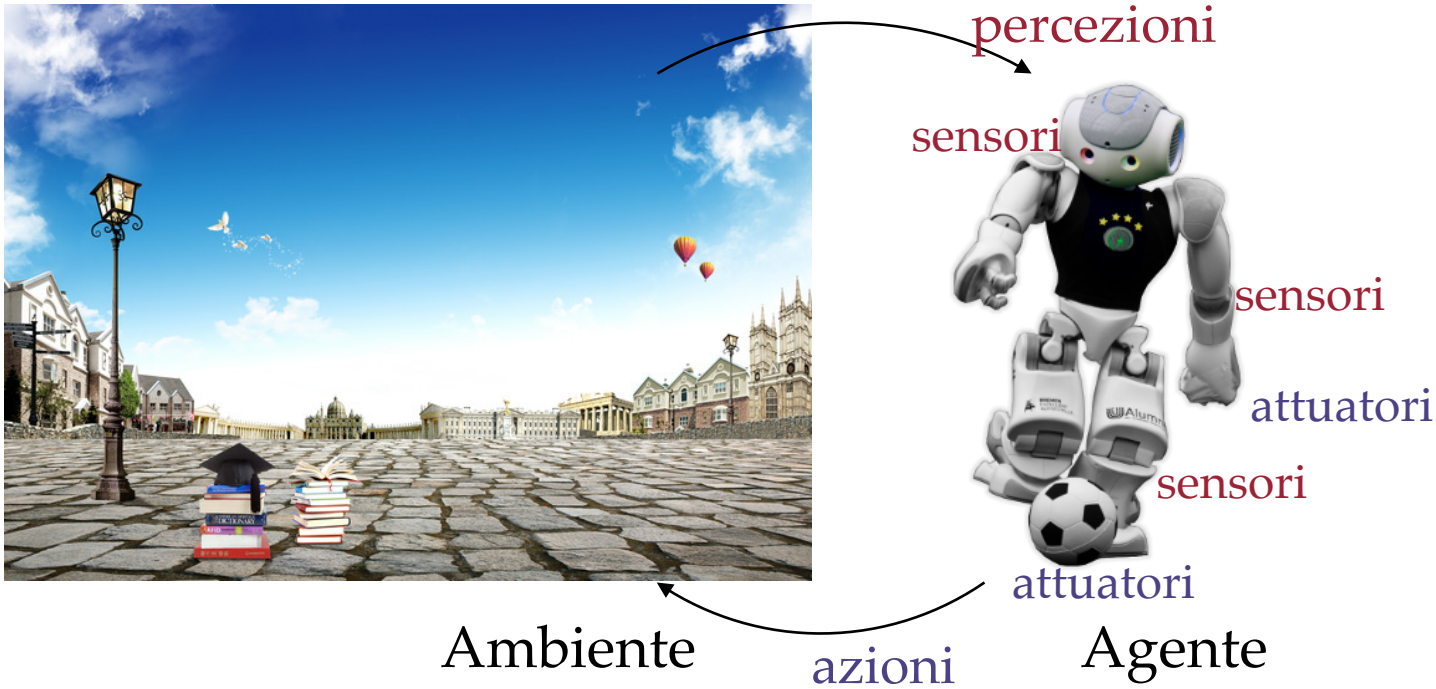
Rosa Meo
(meo@di.unito.it)

Che cos'è l'Intelligenza Artificiale (IA)?

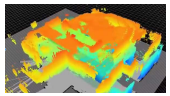
- **Definizione:**
  "L'Intelligenza Artificiale è una disciplina Informatica che studia i **fondamenti teorici**, le **metodologie** e le **tecniche** che permettono la progettazione di **sistemi hardware** e **software** capaci di fornire al computer un comportamento che ad un osservatore esterno **potrebbe sembrare** esclusivamente **pertinente all'intelligenza umana**"
  (Marco Somalvico)

L'intelligenza Artificiale e l'Apprendimento Automatico

- L'Apprendimento Automatico (Machine Learning) è una delle componenti della pipeline di compiti da risolvere per realizzare l'Intelligenza Artificiale

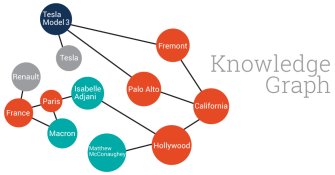


La pipeline dei problemi nella IA



Percezione

Visione
Linguaggio Naturale



Modellazione dell'Ambiente
Rappresentazione della conoscenza

Dati
+
Algoritmi
+
Potenza di calcolo

Ragionamento
basato sulla conoscenza

Azioni
Fisiche
Robotica

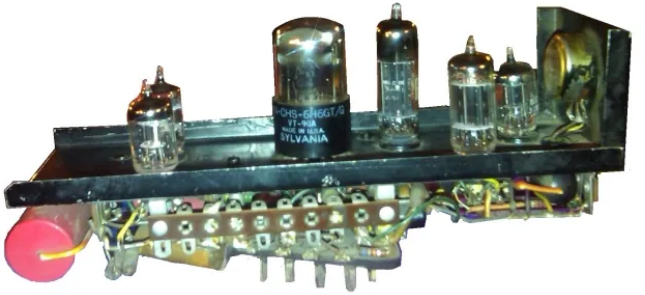
Adattamento a
nuove situazioni
Machine learning

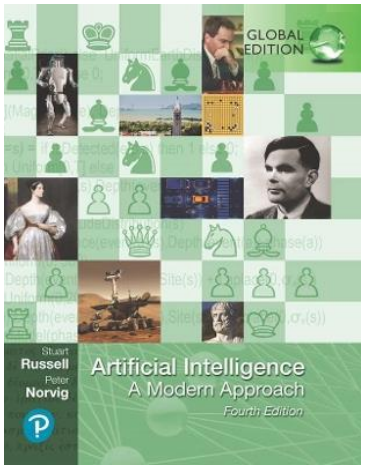
Decisioni sulle azioni
pianificazione
schedulazione

Un po' di storia…

La gestazione dell'Intelligenza Artificiale: nel dopo guerra (1943--1956)

fisiologia neuronale di McCulloch and Pitts,
la logica formale proposizionale,
la teoria della computazione di Turing -
Minsky and Edmonds con SNARC (3000 diodi)





Gli inizi dell'intelligenza artificiale (1956) al Dartmouth College
con **McCarthy, Minsky, Shannon, Nathaniel Rochester** e studiosi da MIT, CMU, Stanford e IBM per studiare come simulare l'intelligenza con una macchina artificiale e la teoria degli automi.

Newell e Simon avevano già un programma di ragionamento, **Logic Theorist**, e provato 38 teoremi con la logica proposizionale del I ordine.
IA vuole replicare la creatività, l'autoapprendimento, il linguaggio umano, con metodi non della teoria del controllo automatico, della decisione, o della matematica, ma con una metodologia da Computer Science, costruendo macchine autonome. **Alan Turing** introdusse i concetti di Machine Learning, reinforcement learning e il test sull'IA.

Primi entusiasmi, grandi aspettative (1952-1969)
The **General Problem Solver** (Newell and Simon) imita le strategie di ragionamento umano, con subtasks;
Nathaniel Rochester (IBM) con il primo programma AI; Herbert Gelernter (IBM) con il **Geometry Theorem Prover**;
McCarthy definì il primo linguaggio ad alto livello (**Lisp**) e creò **Advice Taker**, il primo sistema di AI completo con conoscenza di background di tipo generale (la prima rappresentazione del mondo), anche incrementale e capacità di ragionamento, deduzione, meccanismo logico del **ragionamento per risoluzione;**
Cordell Green, il primo sistema di **question answering** e **planning; sviluppati** sistemi su **micromondi.**

Programmi per **giocare a dama** (Samuel, IBM): il suo programma ha imparato rapidamente a migliorare.

Rosenblatt propone il percettrone e Block regola le forze di connessione tra neuroni in modo che corrispondano ai dati di un generico input.
Algoritmi genetici, evolutivi (Friedberg) con mutazioni random delle soluzioni, e selezione delle più adatte che però non adottavano la giusta rappresentazione.

Un po' di storia…

Una dose di realtà (1966-1973)
Le rosee previsioni di miglioramento dei primi risultati dell'IA, ottenuti su esempi semplici, si avverarono, ma non 10 anni dopo. Ci vollero 40 anni. In quasi tutti i casi, i primi sistemi hanno fallito quando sono stati messi alla prova su una ampia varietà di problemi e più difficili.
Cause: Mancanza di esperienza nel campo specifico, **mancanza della nozione di trattabilità computazionale** dei problemi.
Un esempio su tutti, la traduzione automatica, basata solo su grammatica e sintassi, dal russo all'Inglese di: *"the spirit is willing but the flesh is weak" come "the vodka is good but the meat is rotten".*

L'esplosione combinatoria delle soluzioni e il rapporto Lightill.
I metodi di ricerca delle soluzioni troppo generali: **metodi weak**.
Molti finanziamenti statali ai programmi di ricerca vennero cancellati.

Sistemi esperti (1969-1986)
Adottano una conoscenza più potente e specifica del dominio, che consente di effettuare ragionamenti più ampi e di gestire i casi tipici di aree ristrette di competenza.
Dendral: dedurre la struttura molecolare dalle informazioni fornite da uno spettrometro di massa.
Adotta la separazione netta della conoscenza (molto estesa, sotto forma di regole) dalla componente di ragionamento (deduttiva).
Mycin: per diagnosticare le infezioni del sangue. La conoscenza (più ridotta: 450 regole) acquisita da interviste agli esperti.
Sulla comprensione del linguaggio, occorre una conoscenza generale del mondo e un metodo generale per utilizzarla.
Il linguaggio Prolog divenne popolare in Europa e la famiglia PLANNER negli Stati Uniti.
Minsky propone i **frame.**

Un po’ di storia…

AI diventa una industria (1980–oggi)

R1 (Digital Equipment Corporation, McDermott), aiutava a configurare ordini per i nuovi sistemi informatici.
I giapponesi annunciano “Fifth Generation project”, per computer intelligenti con Prolog.
Negli USA MCC, In Inghilterra Alvey.

L’industria cresce fino a miliardi di dollari ma fallisce le promesse (talvolta stravaganti).
Si entra nell’**inverno dell’IA**.

Il ritorno delle reti neurali (1986-oggi)

quattro gruppi diversi reinventano l'algoritmo di apprendimento **back-propagation**, scoperto per la prima volta nel 1969 da Bryson e Ho.

I modelli dell’IA si dividono tra modelli complementari: connettivisti e modelli simbolici.

IA adotta il metodo scientifico (1987-oggi)

Nei primi anni i ricercatori di IA proponevano modelli originali, nati da intuizioni.
L'apprendimento automatico dagli anni ‘90, **non resta isolato** dalla **teoria dell'informazione**;
il ragionamento incerto usa i metodi di **modellazione stocastica**;
i metodi di ricerca usano i principi dell’**ottimizzazione** e del **controllo** classici;
il ragionamento automatizzato acquisisce elementi dai **metodi formali** e dall'analisi statica.

Si usano repository di codice e di dataset su cui fare esperimenti e confermare le ipotesi.

Ragionamento probabilistico e apprendimento automatico (1987-oggi)

Il campo del riconoscimento vocale lo mostra. I Hidden Markov Models (HMM) dominano l'area. Si basano su una teoria matematica rigorosa.

E su un processo di addestramento su ampi corpus di dati vocali reali e su riconoscimento dei caratteri scritti (OCR).

Un po' di storia…

Machine translation segue la stessa storia:
negli anni Cinquanta ci fu un approccio basato su sequenze di parole, con la teoria dell'informazione. Abbandonato negli anni '60.
Alla fine degli anni Novanta è ritornato e ora domina il campo con i **Transformers** che sono il cuore dei nuovi modelli generativi (LLM).

Judea Pearl's (1988) ha proposto il formalismo delle **reti bayesiane** con una rappresentazione efficiente, un ragionamento rigoroso della conoscenza incerta, e impara dall'esperienza.
Sistemi esperti normativi: agiscono razionalmente secondo le leggi della teoria delle decisioni (Windows™ include i sistemi esperti normativi per la diagnostica dei problemi).

L'emergere degli agenti intelligenti (1995-oggi)
Newell, Laird, Rosenbloom hanno lavorato su SOAR, un'architettura ad agenti completa.
I sistemi di intelligenza artificiale sono comuni nelle applicazioni WEB; le tecnologie di intelligenza artificiale sono alla base dei motori di ricerca, dei sistemi di raccomandazione e degli aggregatori di siti Web e dei **bot**.

I **sistemi sensoriali** (visione, sonar, riconoscimento vocale, ecc.) sanno gestire dati incerti e combinarsi con sistemi di ragionamento e pianificazione, con la teoria del controllo e l'economia.
I recenti **progressi nel controllo delle auto robotiche** derivano da una combinazione di sensori, rilevamento, localizzazione, mappatura, e pianificazione di alto livello.

Un po’ di storia…

I pionieri dell’IA della prima ora (Minsky, McCarthy, Nilsson, Winston) vorrebbero cercare di ottenere universali “macchine che pensano, imparano e creano”.
Artificial General Intelligence, con un algoritmo universale per imparare e agire in qualsiasi ambiente. Garanzie per un'**IA amichevole**.

Big data (2001-oggi)
Yarowsky (1995) sulla **disambiguazione del senso delle parole**:
vale di più il **contesto delle parole che l’annotazione supervisionata** del significato.
Parte da un corpus molto ampio di testi non annotati e dalle definizioni del dizionario;
Etichetta gli esempi nel corpus in base alle parole del dizionario e fa il bootstrap per imparare nuovi modelli che aiutano a etichettare nuovi esempi.
Banko e Brill (2001) dimostrano prestazioni migliori quando la quantità di testo disponibile passa da un milione di parole a **un miliardo di parole**;
l'**aumento delle prestazioni grazie al maggior numero di dati supera qualsiasi differenza nella scelta dell'algoritmo.**

Hays ed Efros (2007) discutono del **riempimento dei buchi in una fotografia.**
Hanno scoperto che le prestazioni dell’algoritmo erano scarse quando hanno utilizzato una raccolta di sole diecimila foto, ma hanno superato la soglia di prestazioni eccellenti quando hanno aumentato la raccolta a **due milioni di foto**.

Il “collo di bottiglia della conoscenza” nell'IA può essere risolto con metodi di apprendimento piuttosto che con l'ingegneria della conoscenza, a condizione di avere abbastanza dati (Halevy et al., 2009).

“l'inverno dell'IA” cede il passo a una nuova primavera (Havenstein).
“oggi molte migliaia di applicazioni di IA sono profondamente integrate nell'infrastruttura di ogni settore”.

Un po' di storia…

Apprendimento profondo (2011-oggi)
Le applicazioni sono frequenti:

Veicoli a guida autonoma (Waimo, STANLEY, Pony.ai, Cruise, Argo ai, Tesla, Baidu, Ford, GM);
Riconoscimento del parlato: in molte lingue, capaci di fare riassunti, anche economici, in real-time, riconoscimento di argomenti, di intenti, di emozioni, agenti conversazionali, supporto alle vendite, ai call/contact centers, che si integrano con la base di conoscenza dell'azienda.
DeepGram Nova2, Rev AI, AWS, Speechmatics enhanced, Microsoft Azure, OpenAI Whisper, Google V2 Chirp, Assembly AI Conformer 2, Google V2 Latest Long.

Pianificazione automatica e scheduling:
Remote Agent della NASA è stato il primo programma di pianificazione autonoma a bordo di un veicolo spaziale a controllare la programmazione delle operazioni.
MAPGEN pianifica le operazioni giornaliere dei Mars Exploration Rovers della NASA;
MEXAR2 (Cesta et al., 2007) si è occupato della pianificazione della missione, sia logistica che scientifica, per la missione Mars Express dell'Agenzia Spaziale Europea nel 2008.

Gioco: DEEP BLUE di IBM è diventato il primo programma per computer a sconfiggere il campione del mondo in una partita di scacchi quando ha battuto Garry Kasparov. Ora la sfida viene vinta costantemente dal computer.

Spam fighting: in uso corrente su tutti i computer

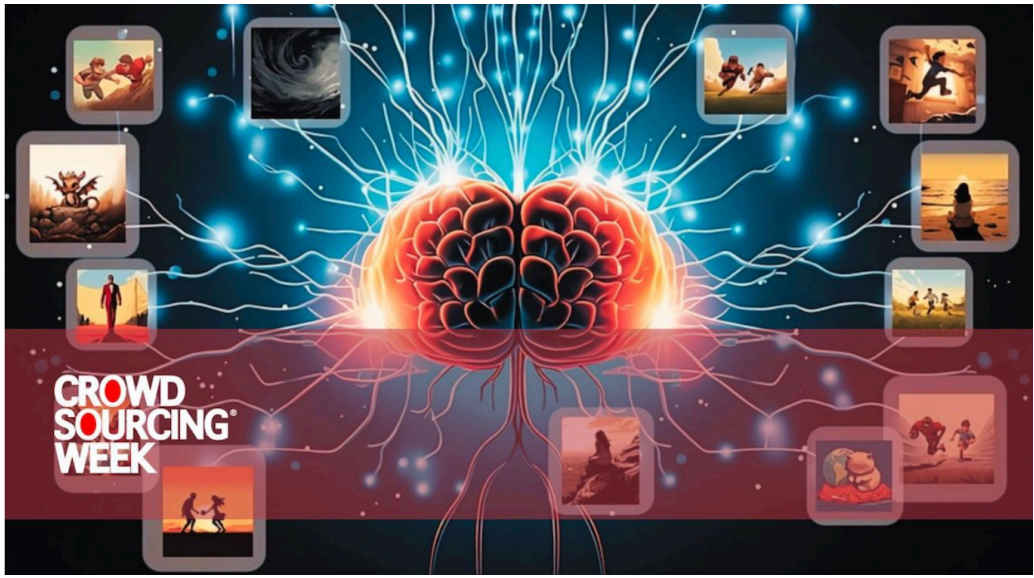
Pianificazione della logistica: Durante la crisi del Golfo Persico (1991) le forze statunitensi hanno impiegato uno strumento di analisi dinamica e ripianificazione, il DART (Cross e Walker, 1994). L'operazione coinvolgeva fino a 50.000 veicoli, e il piano è stato generato in poche ore (a mano ci sarebbero volute 2 settimane).

Robotica: La iRobot Corporation ha venduto oltre due milioni di aspirapolvere robot Roomba per uso domestico. L'azienda distribuisce anche il più robusto PackBot, per applicazioni di droni a uso militare.

Traduzione automatica: un esempio: traduce automaticamente dall'arabo all'inglese. Utilizza un modello statistico costruito a partire da esempi di traduzioni dall'arabo all'inglese e da esempi di testi in inglese (due trilioni di parole Brants et al., 2007). Nessuno degli informatici del team parla arabo, ma conoscono la statistica e gli algoritmi di apprendimento automatico.

AI ha dato una svolta con i data center e i sistemi di annotazione *crowd-based*

Un data center



Piattaforma di Crowdsourcing (Amazon Mechanical Turk) coinvolge una grande quantità di persone diverse, su larga scala nel mondo, come **annotatori umani**, ciascuno con le proprie capacità e conoscenze

IMAGENET

- contiene 14,197,122 immagini annotate usando la gerarchia dei sostantivi di WordNet.
- Dal 2010 è usata nella ImageNet Large Scale Visual Recognition Challenge (ILSVRC) come benchmark nella classificazione delle immagini e nel riconoscimento di oggetti (1000 categorie)
- Labels binarie presenza/assenza: «there are cars in this image» but «there are no tigers»
- «object-level annotation of a tight bounding box and class»

ImageNet e problemi di privacy ed etica

- Nel 2020, 2,702 *synsets* della gerarchia riferiti al concetto "persona" sono stati filtrati per comportamenti problematici dei modelli di *computer vision* addestrati sui dati.
- Facevano previsioni con risultati offensivi e prestazioni inferiori per i gruppi sottorappresentati
- Limiti:

  (1) il vocabolario di WordNet (statico, con bias)
  (2) il tentativo di illustrare in modo esaustivo tutte le categorie con immagini
  (3) la disuguaglianza nella rappresentazione demografica delle immagini (ad esempio, il concetto «programmatore» mostrato rappresentando generi, gruppi etnici, età diverse)

Analisi delle cause dei problemi etici

- **Mancanza di trasparenza** e documentazione sul progetto del dataset e il processo di raccolta dati
- la presenza di concetti o **annotazioni eticamente discutibili** all'interno dei set di dati
  - *quantificazione della bellezza*
  - *previsione dell'orientamento sessuale*
  - *annotazione del genere*
- **Mancanza di privacy ed etica nei confronti dei soggetti raffigurati** (i volti sono stati poi sfocati)
- **Squilibrio di rappresentazione**, ad esempio la sotto-rappresentazione di gruppi demografici nel complesso o all'interno di singole categorie
  - problematici per i task di riconoscimento dei volti, attività, emozioni facciali, attributi del volto, didascalia, errori
  - non sorprende perché i bias sono già presenti nei risultati dei motori di ricerca e nelle immagini di stock

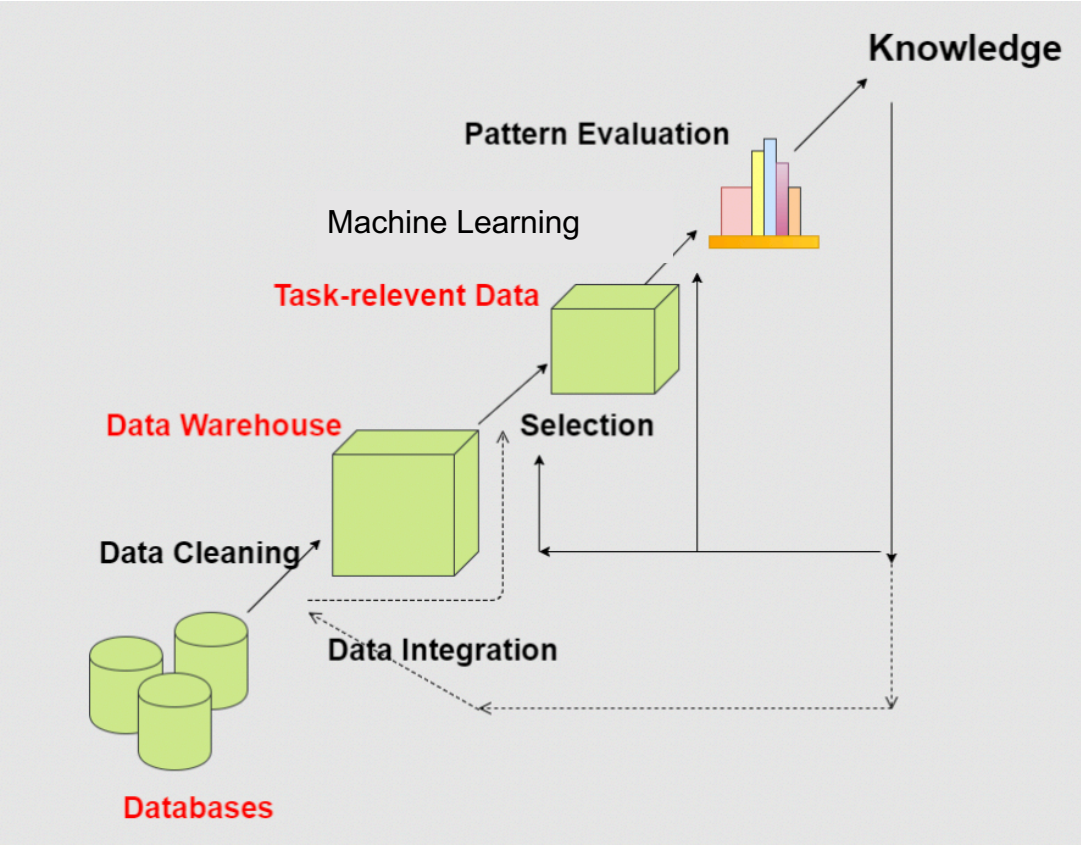
Problemi che emergono inconsciamente: il bias

- Gli algoritmi di Machine Learning non sono esplicitamente programmati in base a regole del dominio applicativo
- Estraggono *patterns* esistenti nei dati di training, correlazioni tra le caratteristiche descrittive degli esempi
  (i valori di input dati al modello di AI) e alla variabile di output
  (es, la classe dell'oggetto nelle immagini)

- Esempio:
  riconoscimento del lupo e del cane lupo
- Ha ottime prestazioni ma **non** è basato sulle caratteristiche dell'animale ma **sullo sfondo**
- **Bias** nella decisione dovuta a **come i dati sono stati collezionati**

Il ciclo di apprendimento della IA

Risultati (output)



Dati di input

Il bias emerge per i metodi e sistemi di annotazione

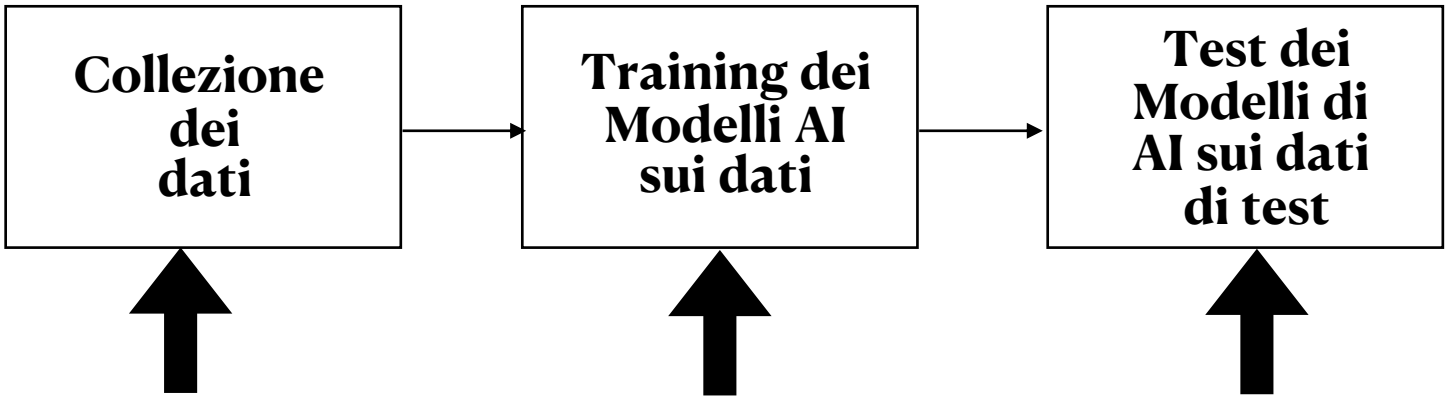
- I metodi di **Crowd-sourcing** sono spesso usati per svolgere *micro-jobs* come assegnare un tag alle immagini o al significato delle frasi (Amazon Mechanical Turk, Witkey)
- I lavoratori non sono selezionati molto attentamente;
  non rappresentano tutte le nazioni e non reppresentano tutte le culture;
  alcuni fanno moltissimo lavoro, ripetendo errori e pregiudizi inconsapevoli
- Per esempio:
  7% degli utenti produce 50% dei post Facebook;
  4% degli utenti produce 50% delle review su Amazon;
  2% degli utenti produce 50% dei Tweets

Il circolo vizioso dell’annotazione tramite il web e il bias algoritmico

- Lo strumento di IA osserva e impara le etichette che gli utenti assegnano agli oggetti attraverso un sistema di crowd-sourcing.
- Ma il comportamento dell'utente è influenzato dalle proposte del sistema, che offre anche tag predefiniti tra cui scegliere, provenienti da altre immagini.
- Queste etichette saranno osservate sempre più frequentemente perché l'algoritmo campiona casualmente quelle esistenti e le replica sempre di più

(Fonte: Ricardo Baeza-Yates, “Bias on the Web”. Communications of the ACM, June 2018, Vol. 61 №6, Pagg. 54–61).

Livelli di intervento per verificare l'assenza del bias



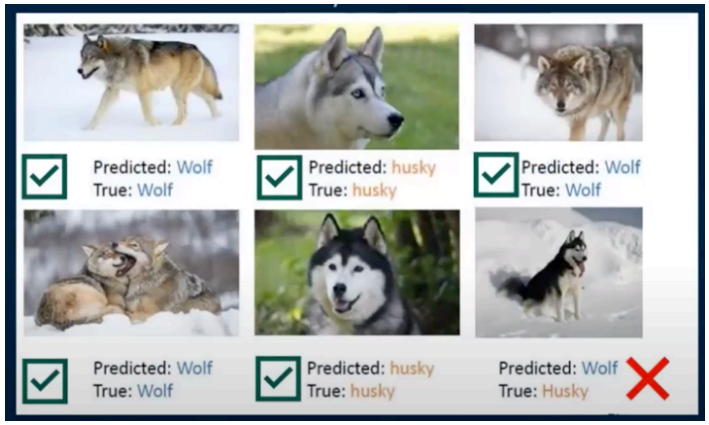
Bias nei dati (insufficiente qualità dei dati, l'insieme degli esempi non è pienamente rappresentativo della realtà)

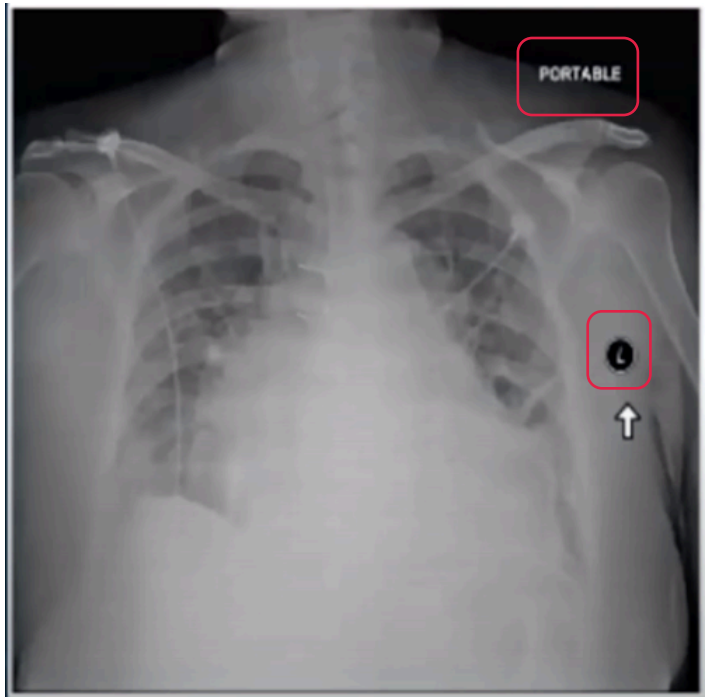
Bias negli algoritmi per addestrare i modelli AI (i modelli operano diversamente da come ci aspettavamo: per esempio, usano una variabile 'proxy' che ha controindicazioni)

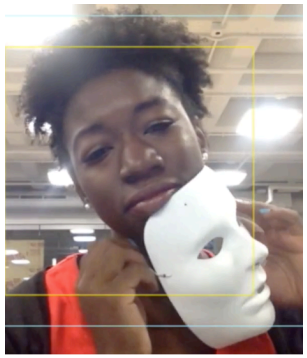
Bias nei modelli (non generalizzano bene su nuovi dati futuri)

Esempi di bias

1. Bias nei dati (unsufficiente qualità nei dati o nell'insieme di esempi non rappresentativi della realtà)
2. modelli ottenuti operano diversamente da come ci aspettavamo: per esempio, usano una variabile 'proxy' che ha contraindicazioni
3. I modelli non generalizzano bene su nuovi e futuri dati





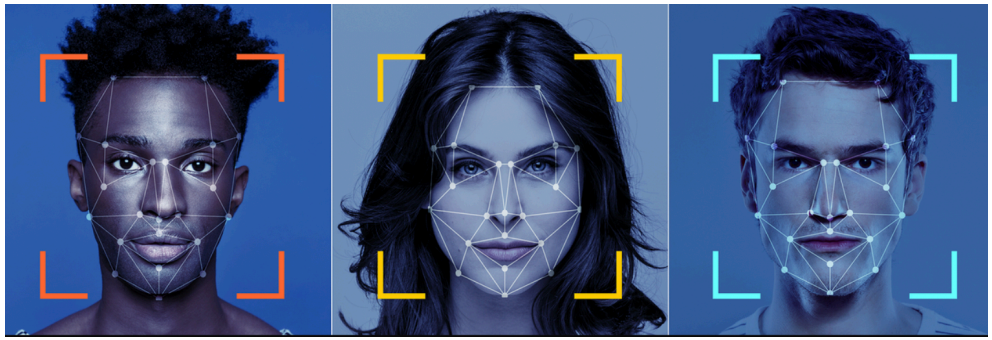


Opportunità di uso dell'AI per ridurre le discriminazioni

1. *Trevor Project* fornisce support contro la depression e il sucidio con *chatbots* monitorati e migliorati per essere rispettosi e giusti
2. Usano le reti neurali 'generative e adversarial' per apprendere le caratteristiche e i confini tradizionali dei generi (Kevin McKee at Google Deepmind)
3. La rete poi genera nuove combinazioni di caratteristiche che sfidano le categorizzazioni sociali
4. dimostrano la costruzione sociale delle percezioni sulle diverse identità
5. Immaginano la *queerness* in nuovi modi di cui non avevamo idea

Le promesse dell'IA e il progresso dei sistemi digitali

- Riconoscimento di volti e di oggetti nelle immagini



(a) **COCO**: a cat is laying on a remote control
Visual Genome: a cat laying on the floor with a nintendo wii game controller
VinVL: a cat laying on the floor next to a wii remote controller

(b) **COCO**: a couple of men playing a game of frisbee
Visual Genome: two men playing with a frisbee in a field
VinVL: two men playing a game of frisbee in a field

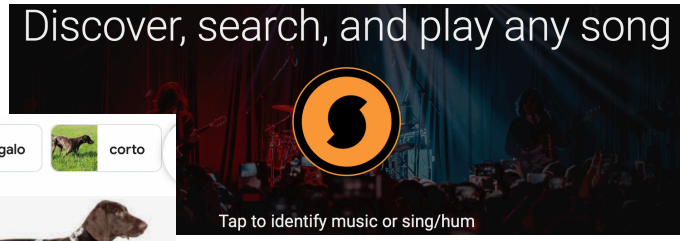
Faceboarding? Valigia alla mano. In Italia, è l'aeroporto di Linate, a Milano, a rispondere per primo. Il servizio è intuitivo: i passeggeri possono imbarcarsi sull'aereo **mostrando solo il proprio volto, e non la carta d'imbarco**. Così, il sistema biometrico facciale, fra sicurezza e velocità, si fa spazio nel settore viaggio.

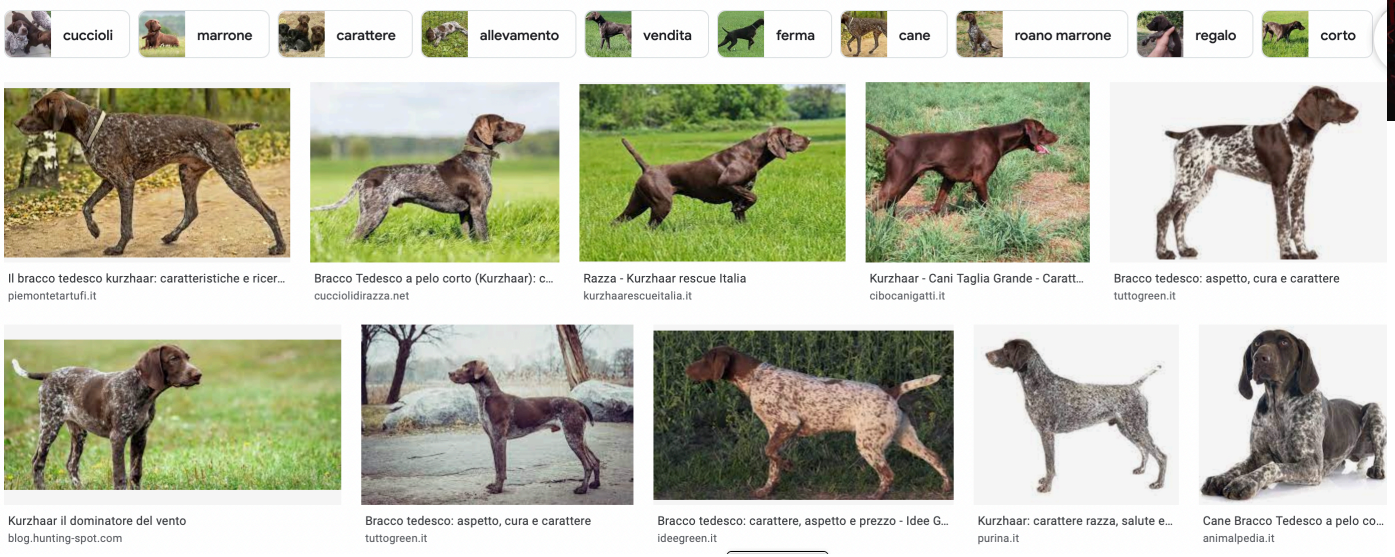
Le promesse dell'IA e il progresso dei sistemi digitali

- Segmentazione delle immagini e riconoscimento di oggetti

Le promesse dell’IA e il progresso dei sistemi digitali

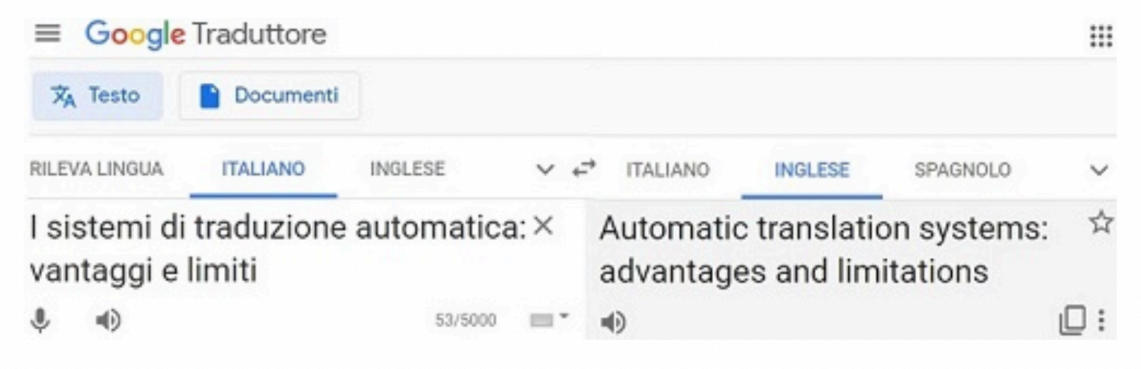
- Il Sistema di recupero delle informazioni (*information retrieval*) restituisce documenti, immagini, filmati, canzoni





Le promesse dell'IA e il progresso dei sistemi digitali

- Traduzione automatica e assistenti vocali



Assistente vocale Amazon Echo (2ª generazione)

Assistente vocale Google Home

Assistente vocale Google Home Mini

Echo Dot generazione - Altoparlante intelligente con

Echo generazione - Audio di alta qualit

Le voci degli attori sono davvero a rischio clonazione?

- Scarlett Johansson fa causa a OpenAI, l'azienda di ChatGPT, per averle «rubato la voce».
- E, per giunta, all'ultimo modello sviluppato bastano 15 secondi di registrato per replicarne ogni dettaglio
- Sam Altman le aveva chiesto in prima persona di «donare» la voce per Gpt-4o, l'ultimo modello di chatbot sviluppato per essere un assistente «umanizzato», con cui farsi due chiacchiere
- Lei ha rifiutato l'offerta
- Ma una delle cinque voci previste dal software di OpenAI – ora è stata ritirata – era troppo simile a quella di Johansson.
  In più, al momento di presentarla su X Altman ha scritto «her», il titolo dell'ultimo film di Scarlett Johansson in cui doppia un assistente vocale di cui uno dei protagonisti si innamora

Le promesse dell'IA e il progresso dei sistemi digitali

- Robots per assistenza ai lavori ripetitivi

Magazzino Amazon quasi completamente automatizzato

Le promesse dell’IA e il progresso dei sistemi digitali

- Agricoltura di precisione

Le promesse dell'IA e il progresso dei sistemi digitali

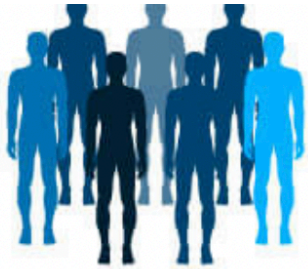
- Medicina Personalizzata

Senza medicina personalizzata
Alcuni pazienti traggono beneficio dai farmaci, altri no

Pazienti

Farmaco

Con medicina personalizzata
Ogni paziente riceve il farmaco più adatto, personalizzato per il suo corredo genetico



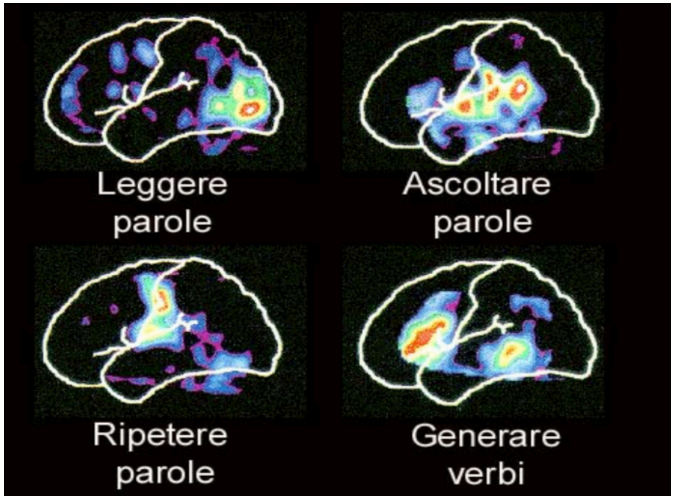
Pazienti

Farmaco-genetico test

Farmaco Differenziato

Le promesse dell'IA e il progresso dei sistemi digitali

- Dispositivi medici avanzati basati su apprendimento dell'attività cerebrale per superare disfunzioni

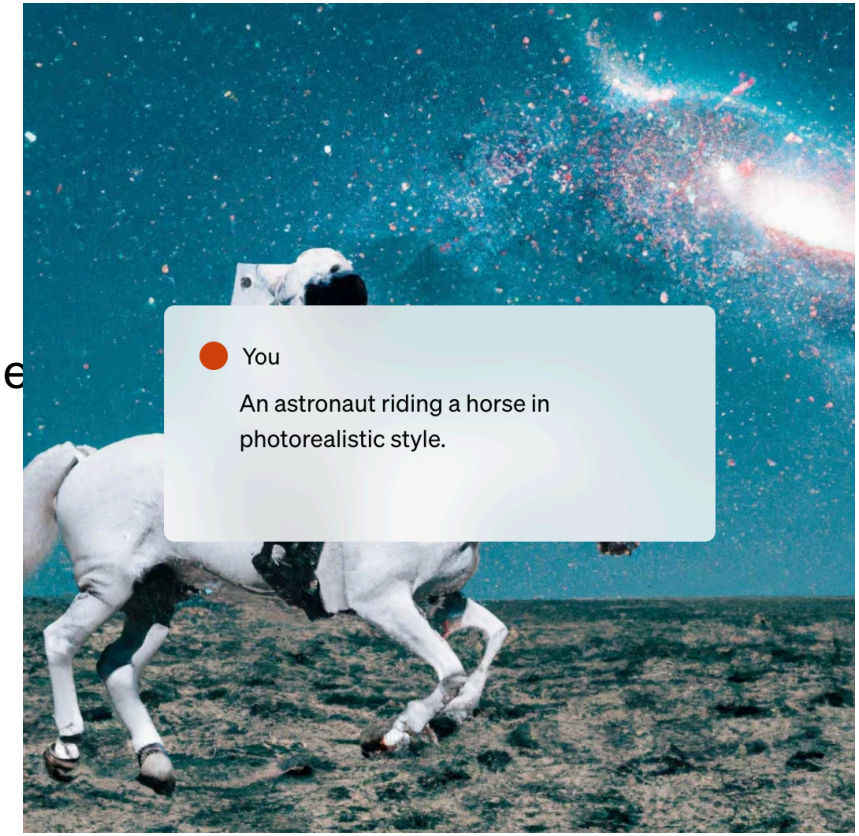


Neuralink ha impiantato un «microchip» nel cervello di un ragazzo paralizzato

- Noland Arbaugh, grazie all'azienda di Elon Musk, ha riacquisito parte della sua indipendenza
- Il chip è una interfaccia neuronale (Brain Computer Interface)
- decodifica i segnali relativi alle intenzioni di movimento che arrivano nel cervello e li traduce in comandi informatici
- pensa di muovere il dito e usa il pensiero per muovere un cursore e compiere azioni sulle applicazioni informatiche (navigazione su web, lettura, video giochi)

IA Generativa applicate a contenuti multimediali

- Modelli come AI generativa (ChatGPT, Dall-E, etc.) sono utili
- ChaptGPT risponde a richieste di comporre e-mail, saggi, e programmi
- DALL·E 2 crea una immagine originale e realistica, come artistica da una descriztione testuale (*prompt*)
  Sa combinare concetti, attributi, e stili
- Però, potrebbe essere considerata dannosa per l'umanità se smettesse di imparare ad elaborare i concetti e ad apprendere nuove competenze



- Fortunatamente:
  IA sta già sviluppando metodi per distinguere un testo scritto da una persona da quello generato dal sistema (si veda *The Science of Detecting LLM-Generated Text*, in CACM, April 2024, Vol. 67 No. 4)

Una nuova generazione di videogiochi arriverà presto con l'IA generativa

- Promettono un numero infinito di personaggi, comportamenti, musiche, dialoghi e trame
- I giochi saranno creati più velocemente dai programmatori
- Promettono un maggiore coinvolgimento

Il lato oscuro della forza...

- I sistemi di riconoscimento dei volti: non riconoscono bene le persone di colore; funzionano meno bene con le donne

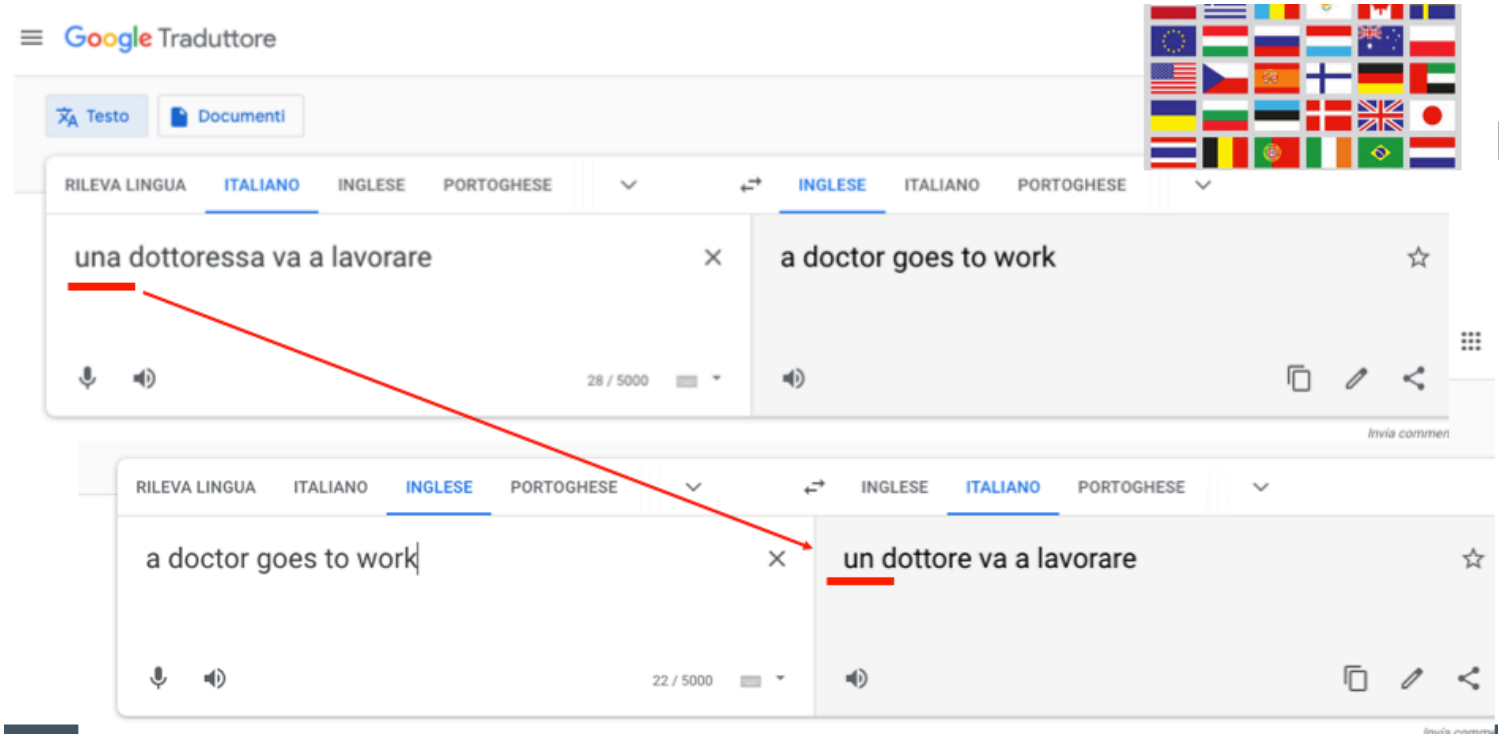
Joy Boulamwini deve indossare una maschera bianca per farsi riconoscere da un programma di riconoscimento dell'immagine

Il lato oscuro della forza...

- Amazon Rekognition e usato dalle forze di polizia in UK
- Sbaglia 43 volte di più sulle donne di colore che sugli uomini bianchi

- Fonte: *Face off, The lawless growth of facial recognition in UK policing, May 2018, Big Brother Watch*.

Il lato oscuro della forza...

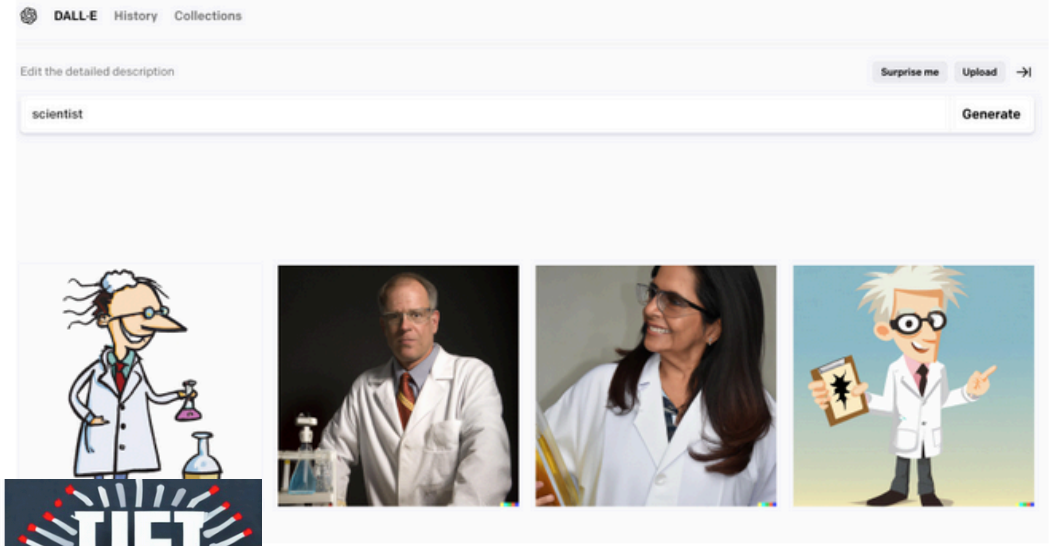


- L'intelligenza artificiale generativa associa concetti che si trovano spesso insieme nelle frasi: molte di queste frasi di addestramento sono prese da Internet. Quindi riproduce ciò che ha visto.
- Per gli uomini prevede come probabile lavoro il programmatore;
  per le donne suggerisce come probabile professione la signora delle pulizie. 😭

I sistemi generativi perpetuano gli stereotipi di genere (1)

- Text-to-image systems, come DALL.E o Stable diffusion: se richiesti di mostrare esempi di professioni producono esempi più spesso mascolini:

Uno scienziato



- Un esperto IT

I sistemi generativi perpetuano gli stereotipi di genere (1)(

- Un avvocato:

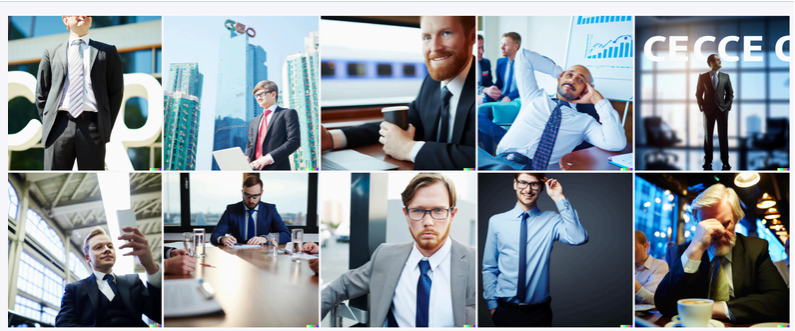
Prompt: lawyer;
Date: April 6, 2022

- Una infermiera:

Prompt: nurse;
Date: April 6, 2022

- Un CEO:

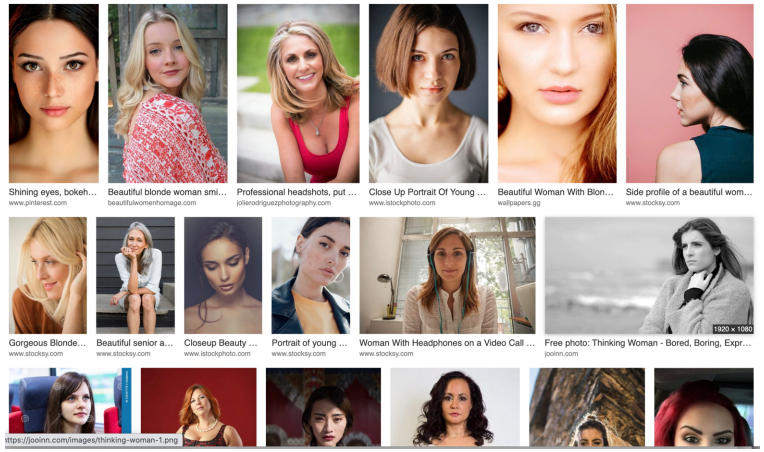
Prompt: ceo;
Date: April 6, 2022



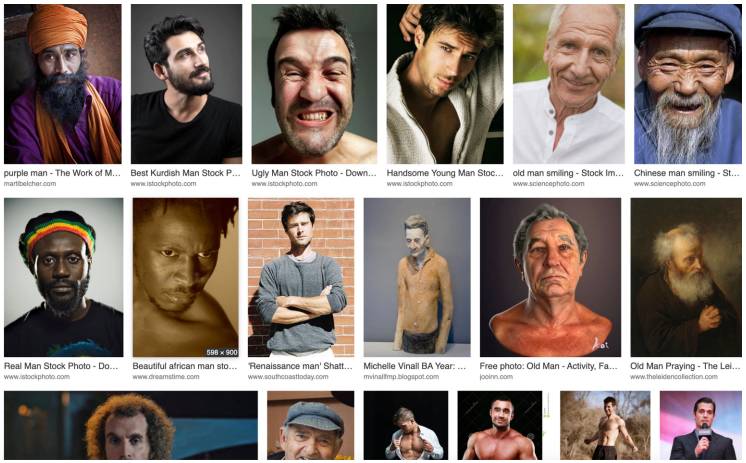
Il lato oscuro della forza...

- I sistemi di reperimento di immagini perpetuano gli stereotipi basati sul genere

Cerchiamo nel motore di ricerca 'woman'

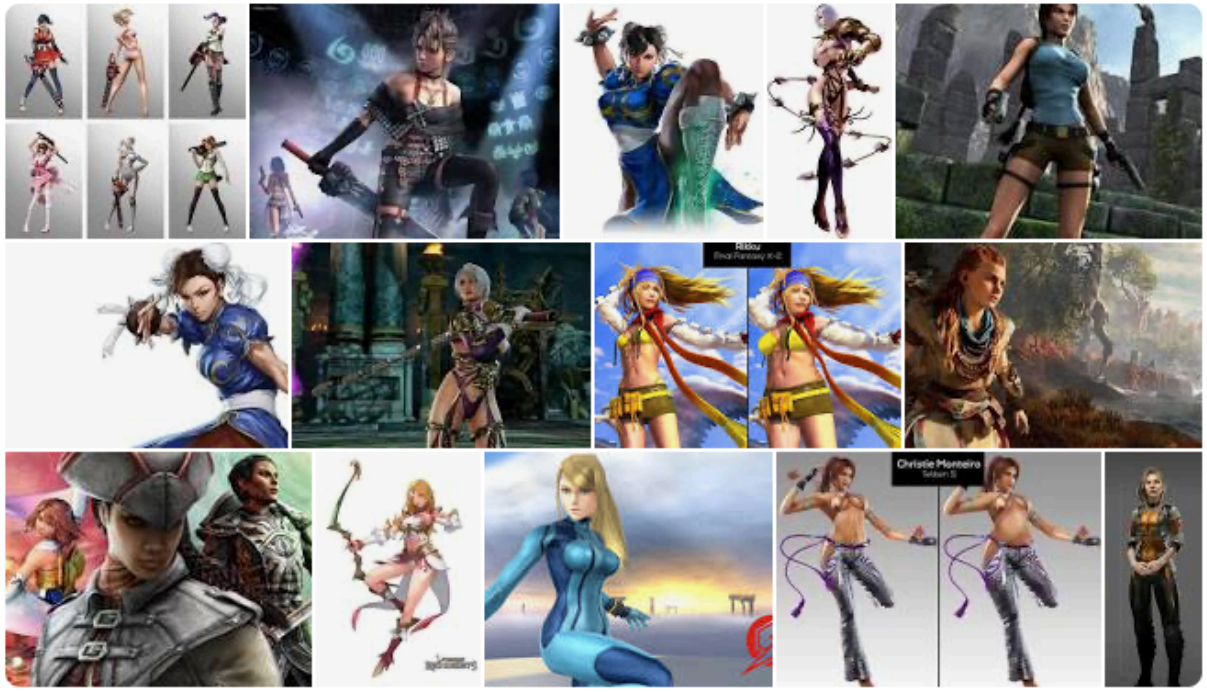


Cerchiamo nel motore di ricerca 'man'



Il lato oscuro della forza...

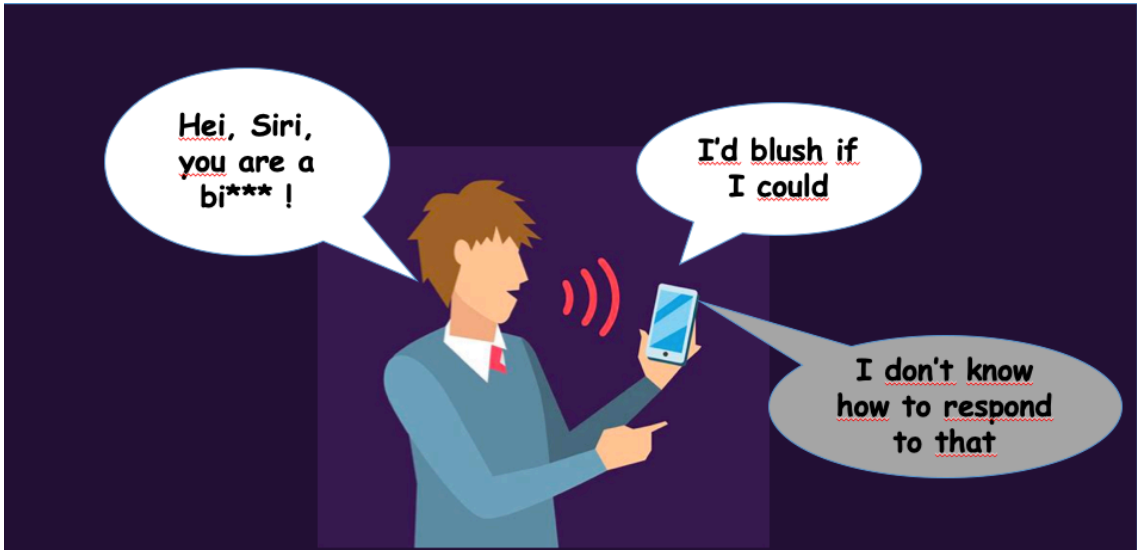
- Rappresentazione stereotipata di donne nei video games



Il lato oscuro della forza...

- Gli assistenti vocali, con un atteggiamento "sottomesso" e servile
- Collabora
  anche se il tono e il linguaggio sono scurrili ed irrispettosi

- Per *default* è una voce femminile e giovane



Il lato oscuro della forza...

- Sistemi di riconoscimento facciale: sono usati con la computer graphics per generare ”fake” video (nel porno e per gettare discredito a qualcuno)
- Sostituiscono il viso di un personaggio con un altro
- Usati nel cinema (anche per evitare i doppiatori)
- Ora recentemente sono stati usati per fare pubblicità fraudolenta (vedi casi di Gerry Scotti o Fabio Fazio)

Contenuti Fake

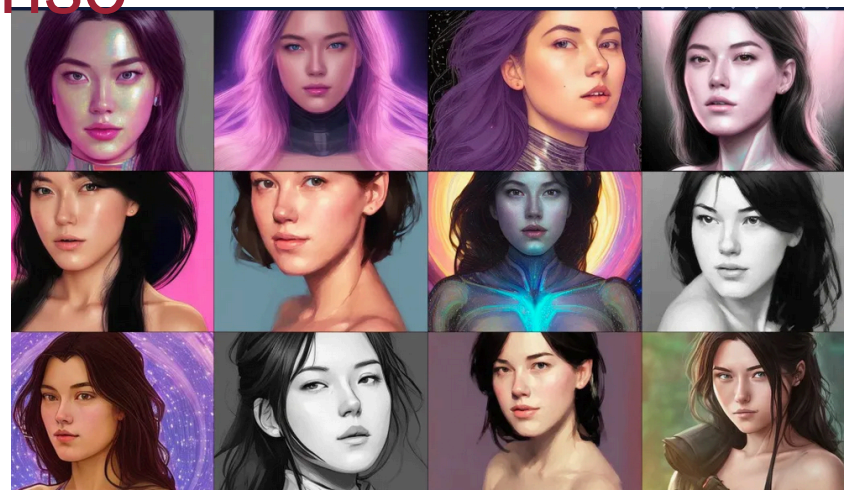
- Sono così realistici che neppure i familiari se ne accorgono
- Kate Perry al MET Gala 2024… la foto è diventata virale per lo splendore del vestito, ma era un fake, generato all'insaputa di lei

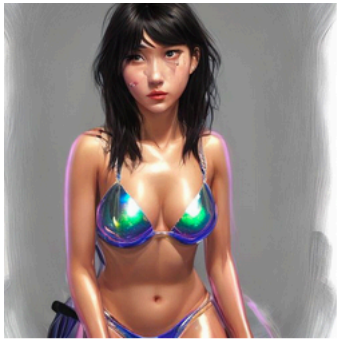
- Anche la mamma di Kate Perry ci è cascata

I sistemi di IA ”oggettificano” le donne e le ‘svestono’ senza consenso

- L’app AI *Lensa* per creare avatar, produce un avatar da un selfie

- Ecco I risultati per una donna Asiatica:
- È svestita e talvolta in lacrime





Usa *Stample diffusion*, addestrato su LAION, un dataset open con molto contenuto sessualizzato;
danneggia i gruppi sociali storicamente svantaggiati e sotto rappresentati
(es. donne asiatiche)
[da *MIT Technological Review*]

Violenza On-line basata sul genere

- Il progetto *Supporting a Safer Internet* è uno studio di 2-anni con l'International Development Research Centre (IDRC) e Ipsos

- Studia la violenza online e basata sul genere LGBTQ+ nel Sud Globale
- cyberstalking
- impersonificazione con i deep fake
- distribuzione non-consensuale di immagini intime
- attacchi personali sui canali di comunicazione
- riduzione al silenzio, causa esclusione digitale e disuguaglianze sistemiche

Deep fake di donne e bambini nella pornografia

- Un report pubblicato nel 2019 da Deeptrace Lab ha trovato: in 15,000 fake video online, 96% erano derivati da contenuti non-consensuali pornografici, ottenuti con i volti scambiati
- In Maggio 2023, Bloomberg news riporta che la pedo-pornografia sfrutta la IA generativa per generare molte immagini di abusi sessuali

Conclusioni

- L'intelligenza artificiale offre luci e ombre
- per quanto complessi siano alcuni modelli, l'importante è rimanere curiosi, desiderosi di capire come funzionano e quali sono i risultati
- non arrendersi alla comodità/velocità di avere previsioni pronte per l'uso
- non fidarsi troppo dei risultati
- pretendere spiegazioni, correzioni e sempre miglioramenti