

IA e reti neurali generative un'analisi etica e scientifica per comprendere e scegliere il nostro futuro

NIPS Lab Workshop
Galleria Nazionale dell'Umbria, 21/4/2023

VALENTINA FRANZONI

DIPARTIMENTO DI MATEMATICA
E INFORMATICA
UNIVERSITÀ DEGLI STUDI DI PERUGIA

Valentina Franzoni
ARTIFICIAL INTELLIGENCE AND EMOTION RECOGNITION



Intelligenza Artificiale (IA)

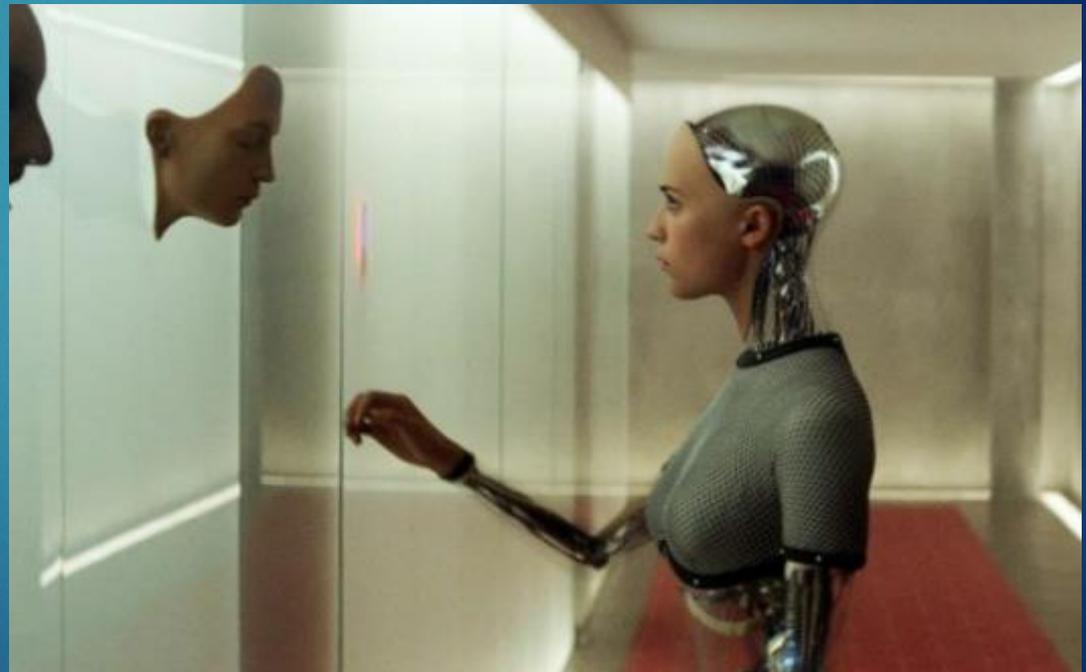
Interdisciplinare.

Questioni

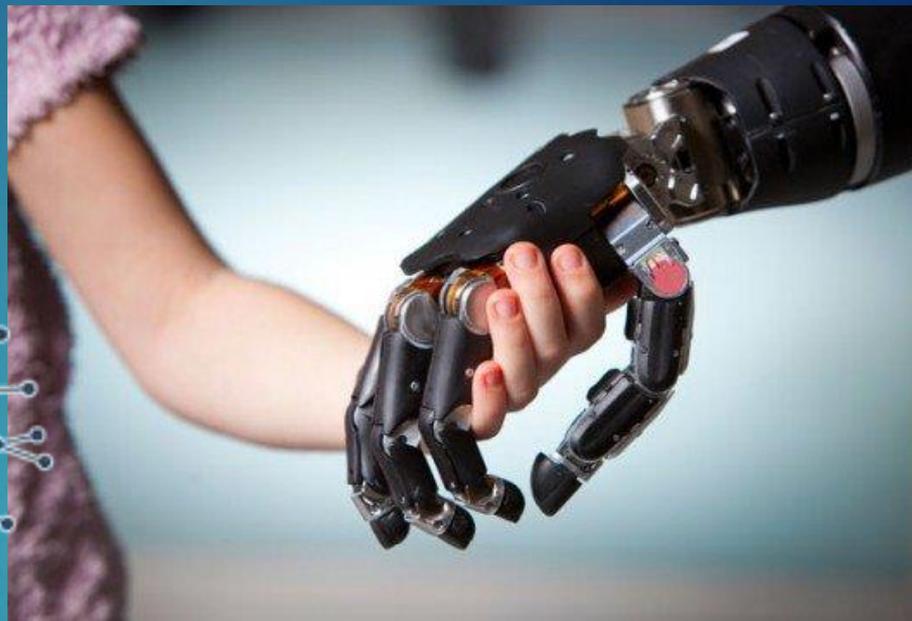
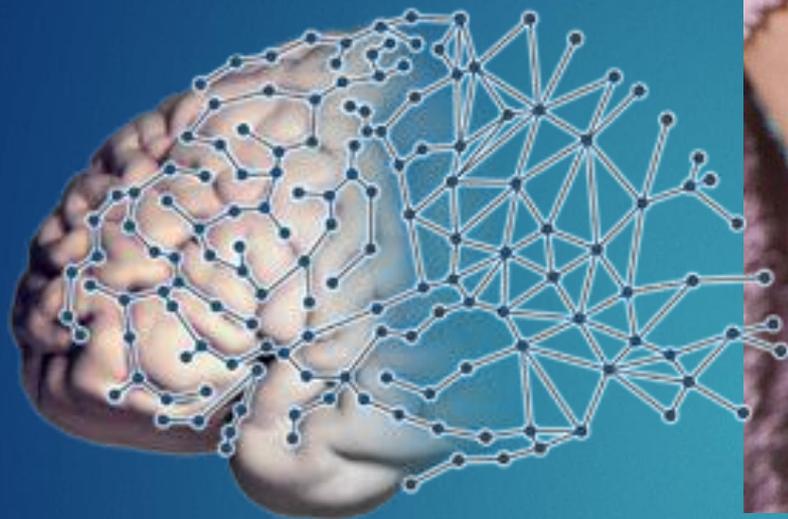
->tecniche

->etiche

->legali



Uomo e macchina, un sogno



Intelligenza artificiale **secondo gli schemi del cervello umano** (Alan Turing, anni '50), da utilizzare o scoprire (**Human Brain Project, 2013-2023**)

Uomo e macchina, una storia



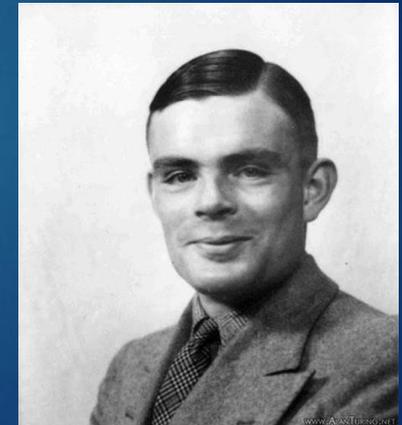
Ada Lovelace Byron
(1815-1852)
Programmazione
di algoritmo 1843

Macchina analitica
di Charles Babbage
1842



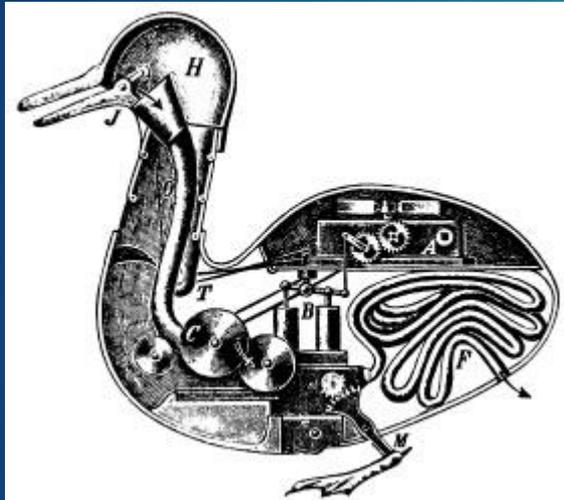
Alan Turing
(1912-1954)

1936 Macchina di Turing
(modello teorico di calcolo)
1950 Test di Turing



Uomo e macchina, una realtà

robotica, cibernetica, meccanica, automatica, elettronica, mecatronica, informatica



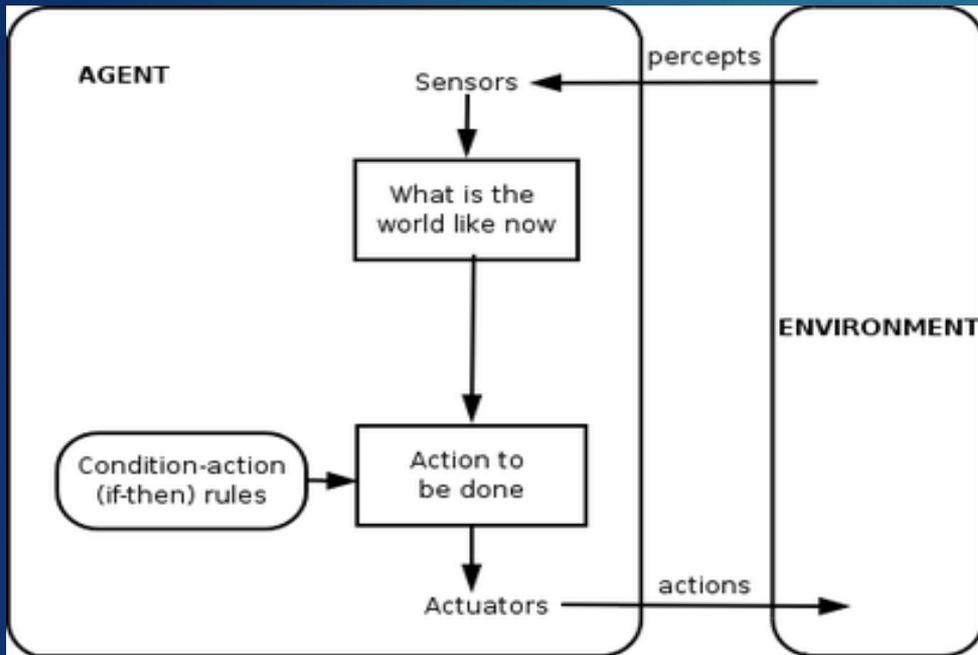
Anatra meccanica capace di digerire
(Jacques de Vaucanson, 1739)

- ▶ Robot(=esecutori) autonomi (automi) costruiti per funzionare come organismi viventi.
- ▶ Intelligenza vista come insieme di meccanismi.

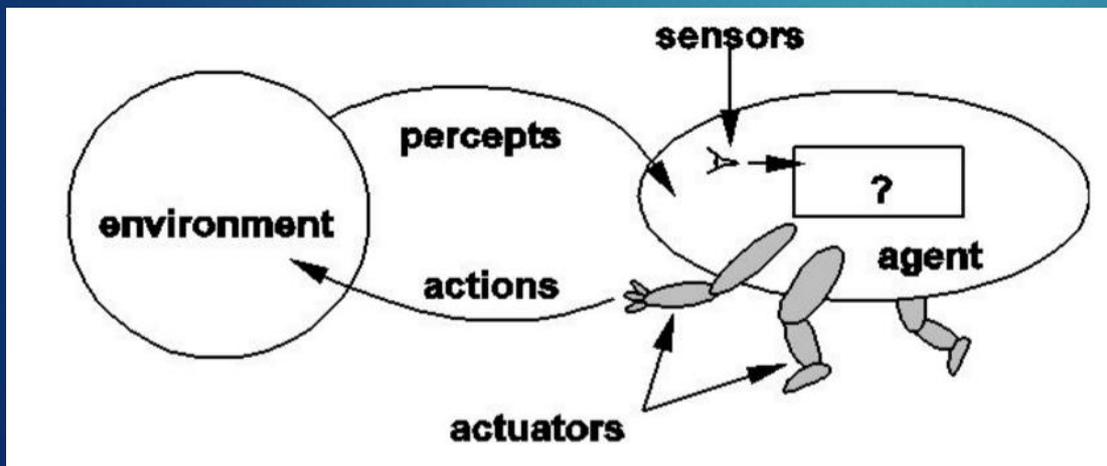
**Paradigma uomo-animale
limitato a un corpo**



Intelligenza Artificiale e agenti

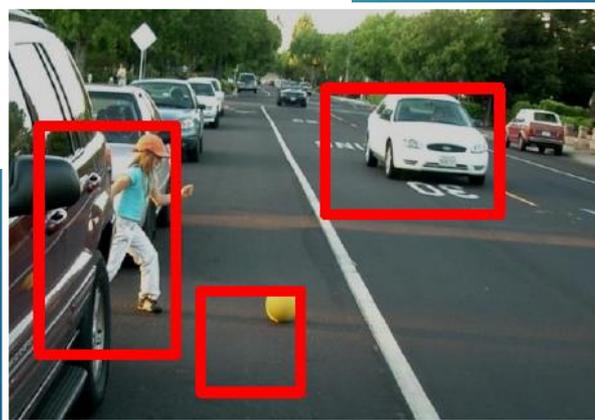
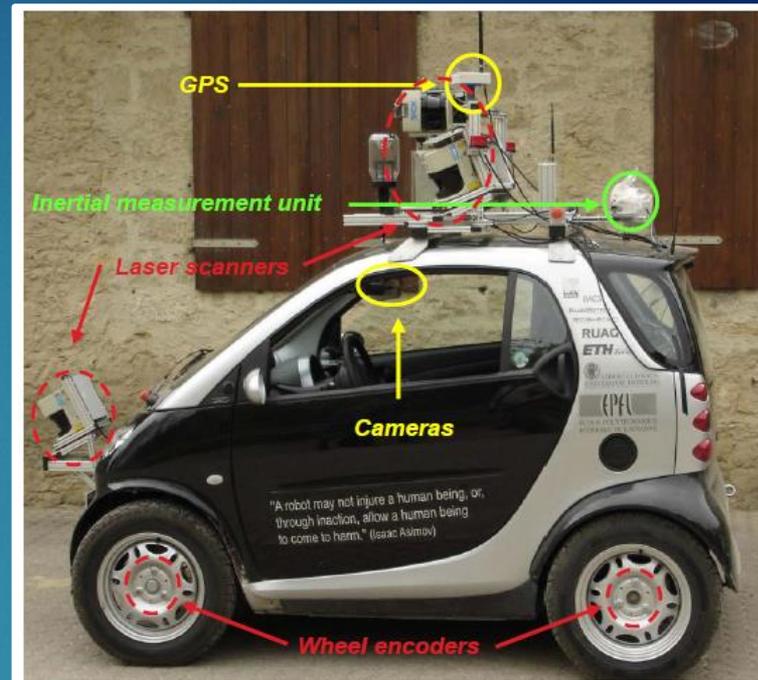


- Agente** (software/macchina):
- > agisce **autonomamente**
 - > **percepisce** l'ambiente (dati, sensori)
 - > può persistere in un'**attività**
 - > Può **adattarsi** al cambiamento

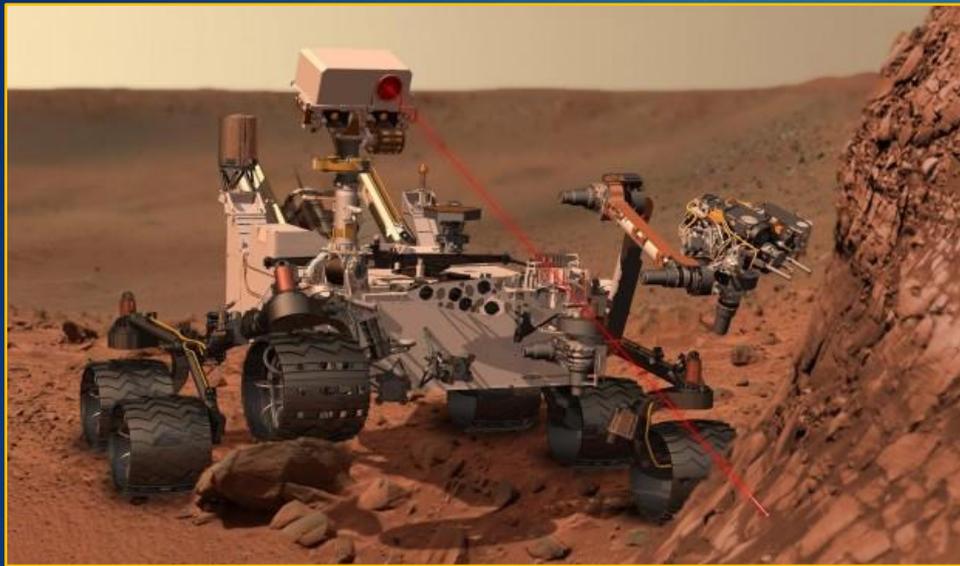


Quali sensori e tecnologie?

- ▶ **Differential GPS**
 - ▶ Global localization and navigation
- ▶ **Inertial Measurement Unit (IMU)**
 - ▶ Orientation (roll, pitch, yaw) of the car
 - ▶ Accelerometers
- ▶ **Wheel encoders**
 - ▶ Local motion estimation (odometry)
- ▶ **Laser scanners & cameras**
 - ▶ Obstacle avoidance
 - ▶ Scene interpretation (road detection, pedestrians)
 - ▶ Environment mapping
 - ▶ Motion estimation

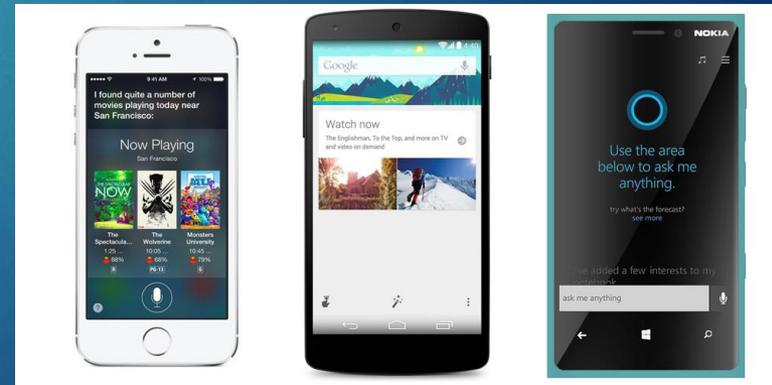
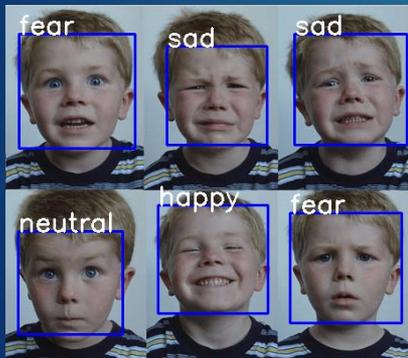


Quali macchine intelligenti?

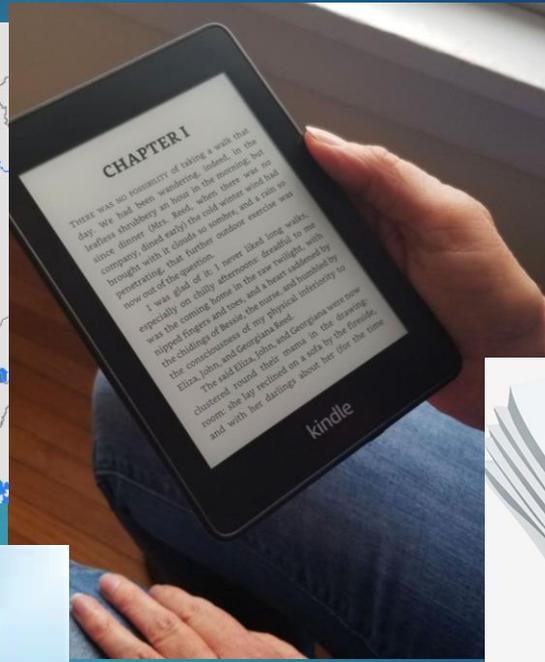


Quali dati e interfacce?

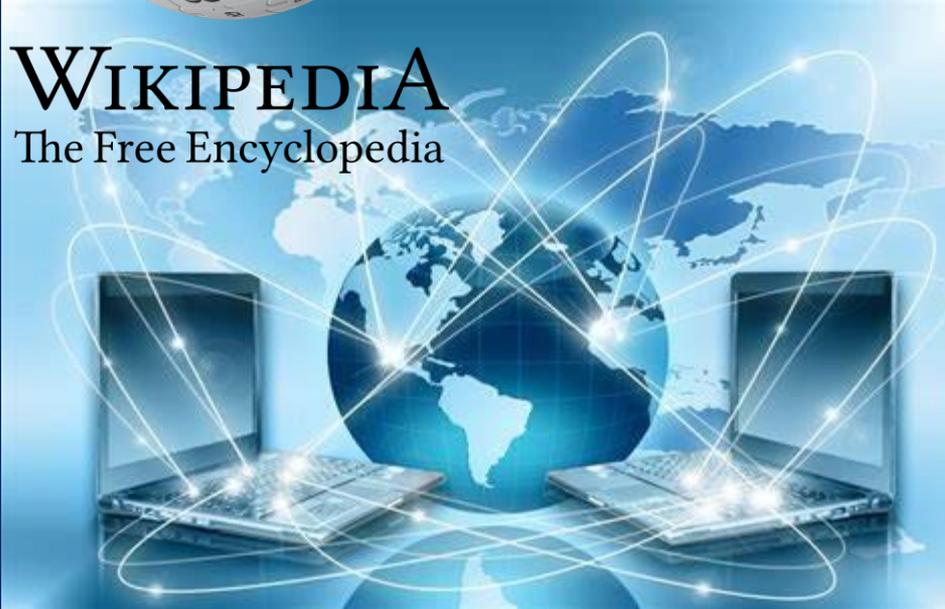
- Dove sei (GPS, wi-fi, sensori di prossimità)
- Cosa stai facendo (app, sensori di movimento)
- Cosa e chi cerchi
- Parlare (schermo, speaker)
- Ascoltare (tastiera, microfono)
- Muoversi con te



Non solo dati personali...



WIKIPEDIA
The Free Encyclopedia



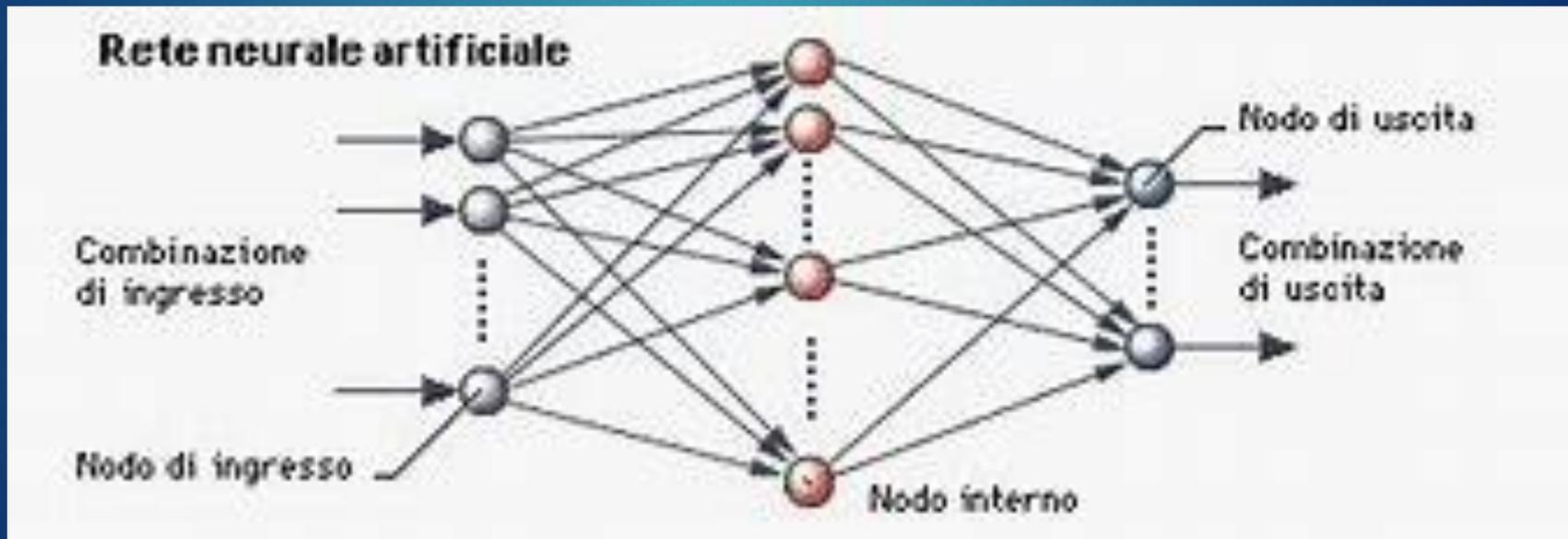
Intelligenza Artificiale è...

- Voci sintetiche (es. navigatori, «premere 1...»)
- Assistenti vocali (SIRI, Google Home, Cortana...)
- Navigazione su mappe o su Internet
- Previsione traffico
- Previsioni del tempo
- Suggerimenti di acquisti/contatti/notizie/post
- Smart....(phone, car, home, app)
- Protezione accesso (biometrico, anti-robot)
- Social network online
- Applicazioni mediche (riconoscimento immagini tumorali)
- Robotica
- Chatbot

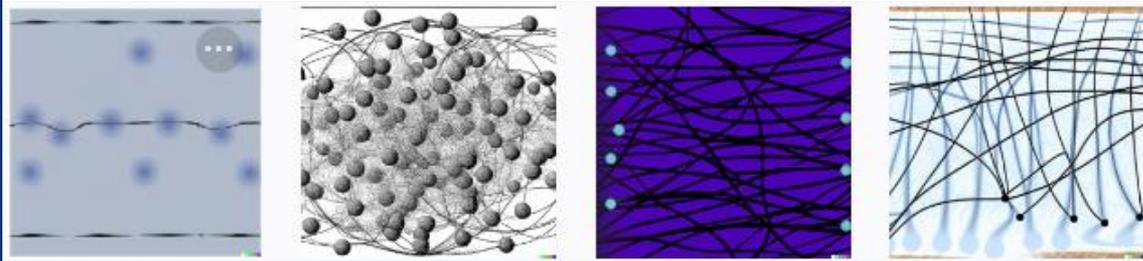
Perché parlarne oggi?

Reti Neurali (dagli anni '70)

IA, i primi 50 anni (Luigia Carlucci Aiello)



Reti Neurali (addestramento *strati*)



*DISTINGUE MA NON SA
CHE SONO!*

pixel

linee

forme

contorni

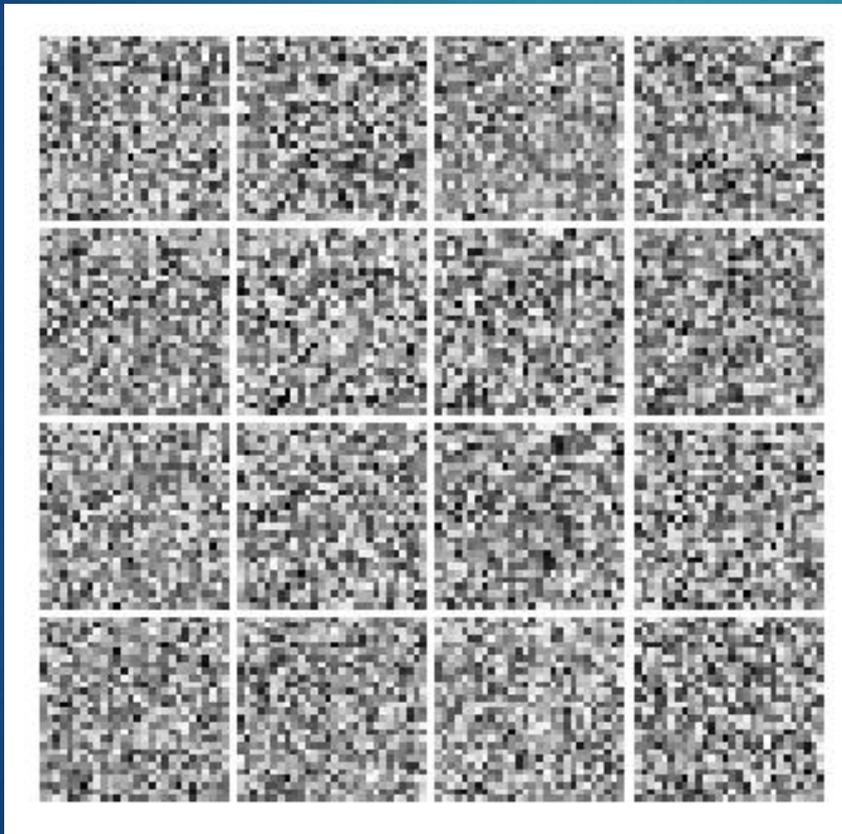
distribuzione di colori

pattern

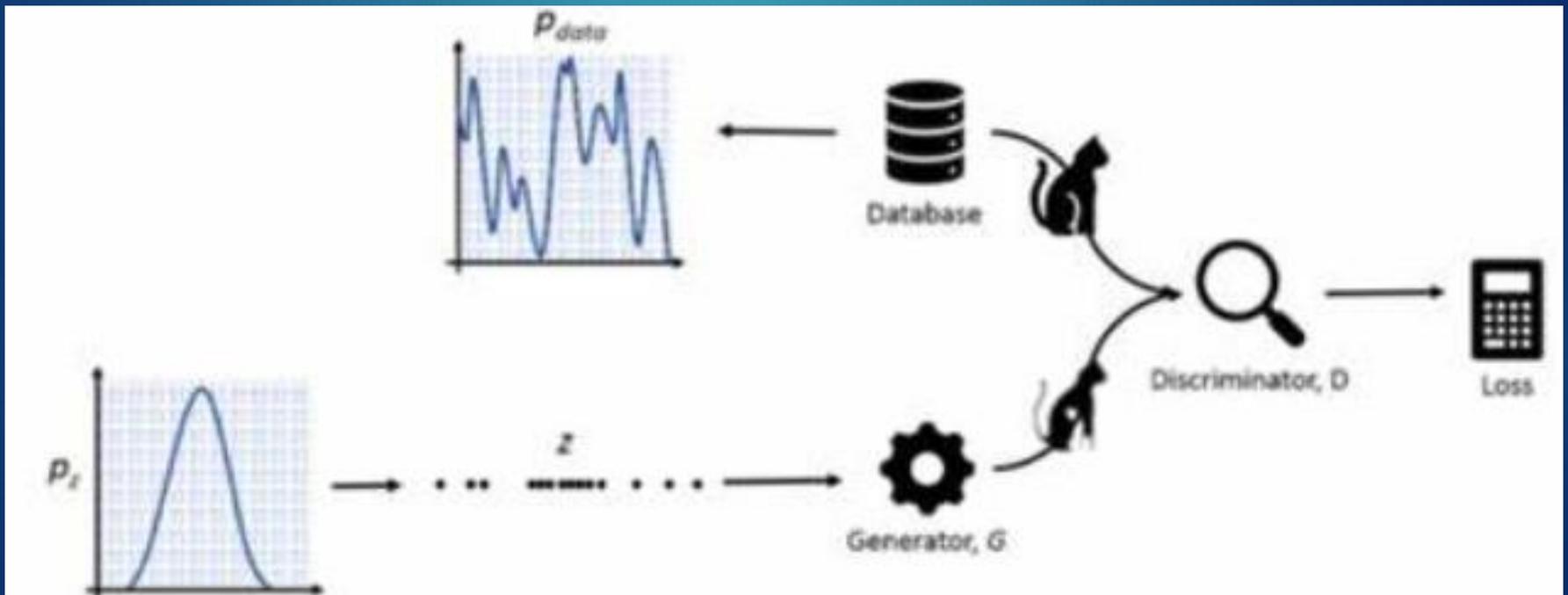
elementi singoli

elementi di sfondo

oggetti



Reti Neurali Generative



Creazione di dati *sintetici* attraverso RETI GENERATIVE

Questi dati non provengono da un campione di dati reali, ma sono stati creati dall'algorithmo attraverso un processo di **generazione casuale** basato su un insieme di **regole** predefinite e **parametri**. I dati sintetici possono essere utilizzati per diverse finalità.

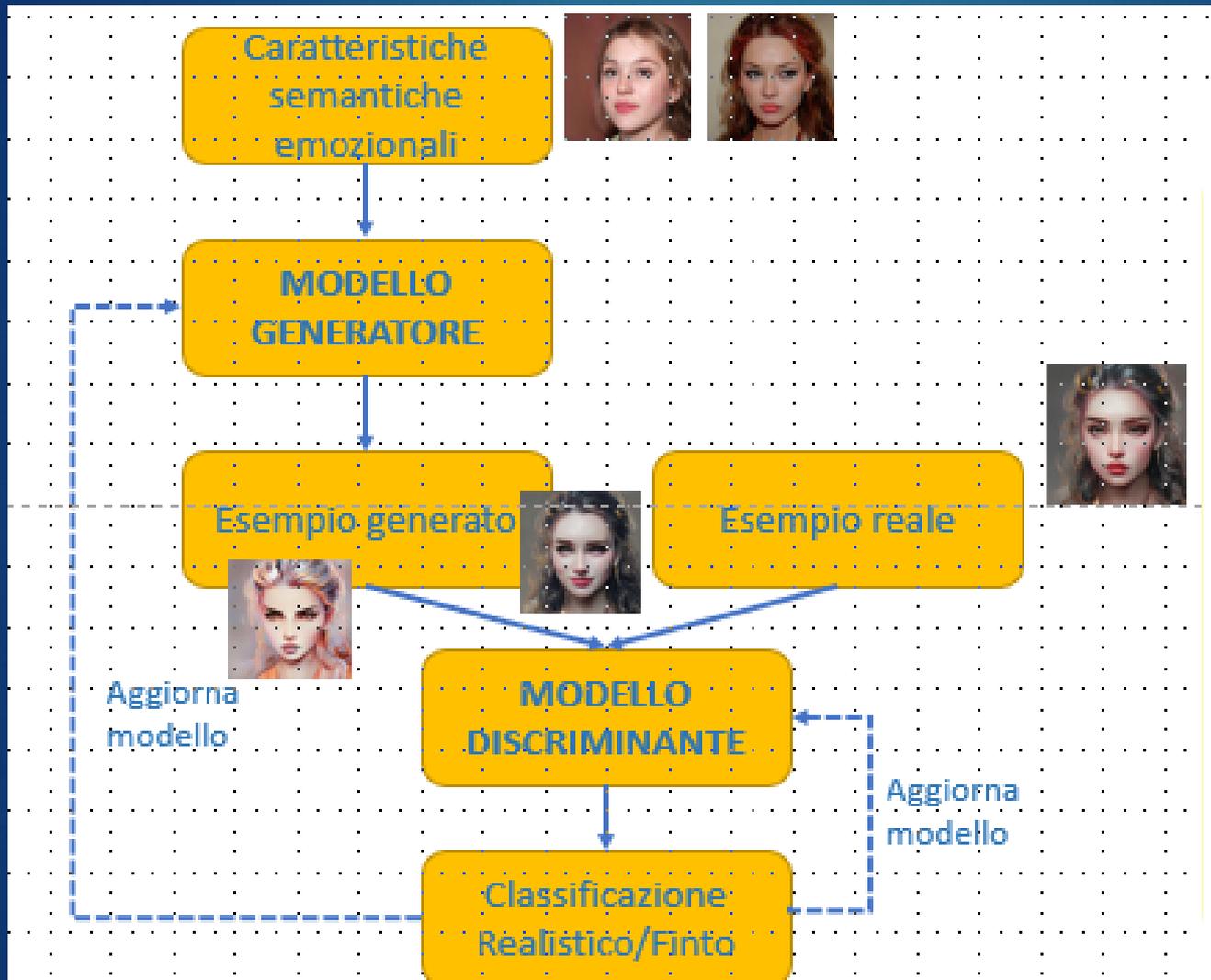
Le **reti neurali generative** sono un esempio di algorithmo di intelligenza artificiale in grado di generare dati sintetici, come ad esempio **immagini**, **testo** o **suoni**, con un livello di complessità e creatività sempre maggiore.

La rete NON «SA» cosa sia ciò che genera...

Esempio: DALL-E



Esempio: Artbreeder



Esempio: ChatGPT

Scopo è il linguaggio,
non il contenuto.

ChatGPT



Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021



Esempio: ChatGPT



MA

Are you sentient?

~~"EVERYONE CAN BE A MILLIONAIRE DOING THIS WITH CHATGPT"~~



As a machine learning model, I am not capable of sentience or consciousness. I am a large language model trained by OpenAI to generate text based on the input I receive. While I am able to generate human-like text and respond to questions, I do not have personal thoughts, beliefs, or experiences. I exist solely to process and generate text based on the input I receive. I apologize if this is not the answer you were looking for. Is there something else I can help you with?

Criticità recenti delle Reti Neurali

Dati non realistici (**falsi**)

Dati non coerenti con i requisiti di partenza (**inutili**)

Contenuti potenzialmente pericolosi (**dannosi**)

Solo uno **strumento**, ma con potenziale discriminazione nella distribuzione, secondo parametri politici, economici o di warfare e **dispendio computazionale**

Potenziale violazione di **diritto d'autore e privacy** (addestramento)

Potenziale **bias**/stereotipi/discriminazioni nei dati (o troppo pochi dati di addestramento)

Difficile interpretazione dei risultati da parte di non esperti (errori, voce non umana che parla come umano)

Text



ChatGPT
OpenAI



ChatGPT is fine-tuned from GPT-3.5, a language model trained to produce text.

Free



GPT-4
OpenAI



GPT-4 is OpenAI's most advanced system, producing safer and more useful responses



Notion AI
Notion



A connected assistant which produces text responses based on user's questions and the current page context.

Free

Image



Stable Diffusion
Stability.ai



A deep learning, text-to-image model, used to generate detailed images conditioned on text descriptions.

Free



Image Creator
Microsoft Bing



Image Creator generates AI images based on your text

Free



Midjourney
Midjourney



Midjourney is an AI program which generates images from natural language descriptions(prompts).

Video



Runway
Runway



A multi-modal AI system that can generate novel videos with text, images, or video clips.

Free



Fliki
Fliki



Fliki is an text to video and text to speech creator powered by generative AI.

Free



Wonder...
Wonder Studio



An AI tool that automatically animates, lights and composes CG characters into a live-action scene

Audio



JukeBox
OpenAI



A neural net that generates music, including rudimentary singing, as raw audio in a variety of genres and artist...

Free



MusicLM
Google Research



MusicLM is an AI model that can generate high-fidelity music from text.

Free



Text to Speech
Microsoft Azure



AI voice generators to speak naturally using synthesized speech from input text

Free

3D



Point-E
OpenAI



A system for generating 3D point clouds from complex prompts

Free



Magic3D
NAVIDIA



Magic3D is a new text-to-3D content creation tool that creates 3D mesh models with unprecedented quality.



Imagine 3D
Luma AI



An early experiment to prototype and create 3D with text.

Coding



Copilot X
Github



An AI pair programmer with an early adoption of OpenAI's GPT-4



Codex
OpenAI



An AI AI system that translates natural language to code



CodeWhisperer
Amazon



CodeWhisperer can generate code suggestions ranging from snippets to full functions based on your comments and...

Free

Reti neurali a supporto della creatività

Criticità recenti delle Reti Neurali

L'uso criminale va considerato a parte, perché non può essere attribuito come mancanza allo strumento.



Documenti sull'etica dell'IA

2019 IEEE CertifAIEd methodology: trasparenza, responsabilità, bias, privacy

2020 Vaticano, Rome Call for AI Ethics: primato all'umanità, condiviso organizzazioni internazionali, governi, istituzioni e settore privato

2021 UNESCO Recommendation on the Ethics of AI: diritti e dignità dell'umanità

2021 European Commission, ArtAI Act: aderire alla legislazione esistente sui diritti fondamentali e sul valore dell'Unione

2023 U.S. Copyright Office: non è possibile rivendicare il diritto d'autore sui contenuti generati dall'intelligenza artificiale

2023 Garante per la privacy italiano: stop a ChatGPT

2023 Partnership on AI (OpenAI): sviluppatori, ricercatori e distributori siano chiari su ciò che la IA può e non può fare

Applicazioni delle GAN

Creazione di **immagini** sintetiche

Generazione di **testi**

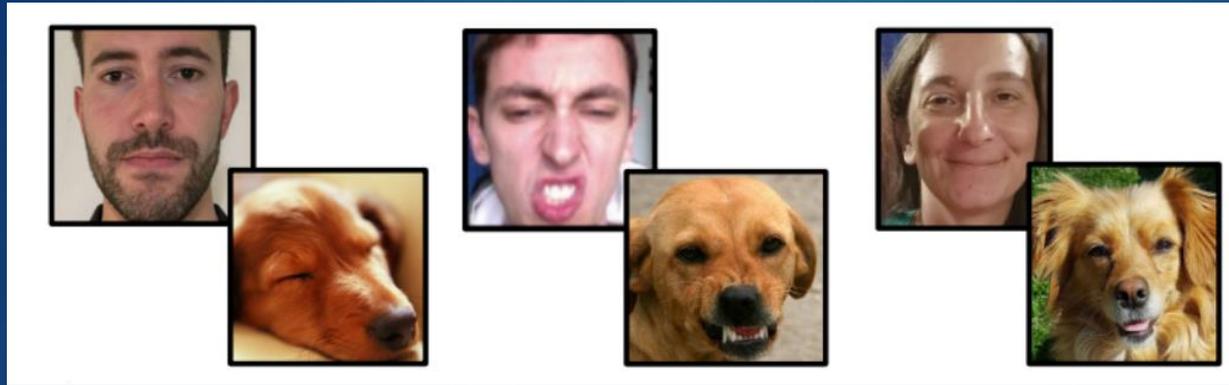
Produzione di **suoni**

Simulazioni scientifiche

Anonimizzazione di dati

Protezione dalle reti neurali...

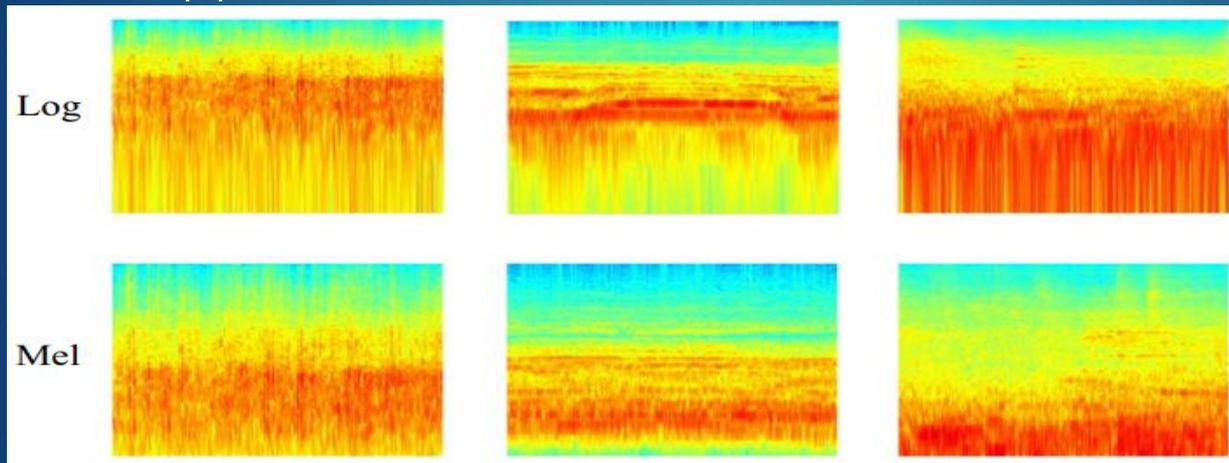
Convolutional Neural Network



Approval

Disapproval

Neutral



Riconoscere:

Emozioni

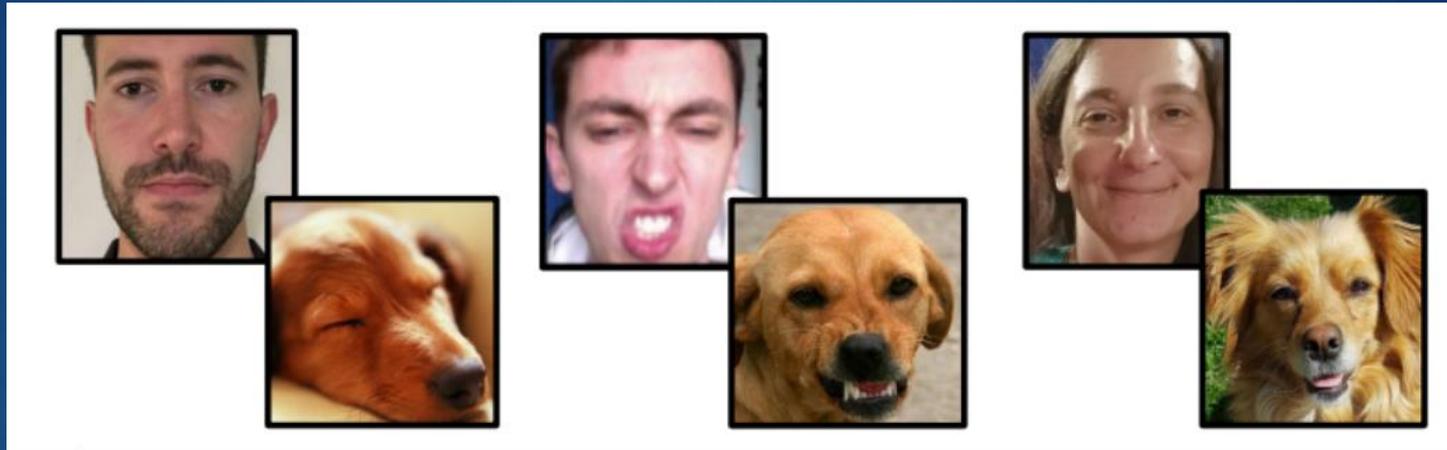
Aggressività

Allarme

Dolore

Franzoni, V., Mezzetti, D. (2022). **Recognizing and Predicting Neonatal Pain in Preterm Intensive Care Unit: a Study Protocol**. WI-IAT2022: 680-682

Generative Adversarial Network



Baia, A.E., Biondi, G., Franzoni, V., Milani, A., Poggioni, V., (2022). **Lie to me: Shield Your Emotions from Prying Software.** SENSORS, vol.22

Approccio etico

Human Brain Project (2013-2023): 700 ricercatori da 27 paesi

Intelligenza Artificiale:

- > **tecniche** (es. reti neurali generative, rinforzo, transfer learning)
- > **dispositivi** (computer, smartphone, robot, rete)
- > **approcci** (black box, trasparenza)
- > **dati**-> pregiudizi, discriminazione, esclusione

Comprensione: *caratteristiche, preoccupazioni etiche*

- > **opzioni di mitigazione, priorità** delle misure pratiche

CHI È IN GRADO DI RIFLETTERE E PRENDERE
DECISIONI IN QUESTE DIREZIONI ETICHE?
(**politici**, istituzioni, media, utenti, **ricercatori**)

...e ora parliamone!

Valentina Franzoni valentina.franzoni@dmi.unipg.it

- Gruppo Facebook:
@affectivecomputing
www.facebook.com/affectivecomputing
- [10 Best AI Art Generators \(January 2023\) - Unite.AI](#)
- OpenAI chat GPT: <https://chat.openai.com/chat>

Slide rilasciate sotto licenza CC BY-SA 3.0 IT:

<http://creativecommons.org/licenses/by-sa/3.0/>