

# chatGPT: avvertenze per l'uso

Marco Rospocher

Il “caso ChatGPT” - Tra prospettive di sviluppo e di regolamentazione dell'IA. Un approccio multidisciplinare  
Università degli Studi di Verona :: 21 giugno 2023



UNIVERSITÀ  
di **VERONA**

Dipartimento  
di **LINGUE  
E LETTERATURE STRANIERE**



# Intelligenza Artificiale

- “the science and engineering of making **intelligent machines**” (John McCarthy)
  - abilità di una macchina di mostrare capacità umane (e.g., ragionamento, apprendimento, creatività, ...)
- Dartmouth Summer Research Project on Artificial Intelligence (1956)
- Disciplina che copre **varie aree di ricerca**:
  - rappresentazione della conoscenza, ragionamento automatico, pianificazione, tecniche di visione, elaborazione del linguaggio naturale, ...



# Natural Language Processing (NLP)

- **Trattamento automatico** mediante un calcolatore elettronico delle informazioni scritte o parlate in una **lingua naturale**, al fine di consentire **interazioni uomo-macchina** efficaci
- Una delle principali aree di ricerca AI, sin dagli inizi (e.g., agenti conversazionali, traduzione automatica)
- Fortemente **interdisciplinare**: Linguistica, Informatica, Matematica, Statistica, ...
- Tecnologie AI/NLP sono parte della nostra **vita quotidiana** già da molti anni:



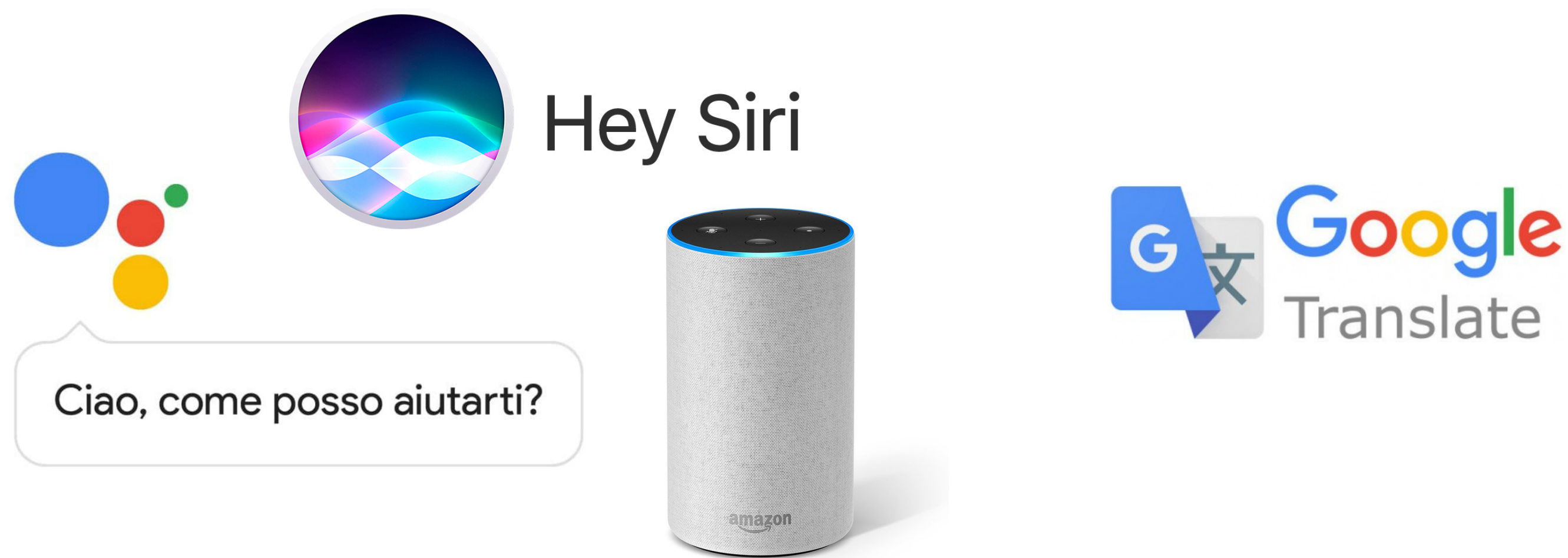
# Natural Language Processing (NLP)

- **Trattamento automatico** mediante un calcolatore elettronico delle informazioni scritte o parlate in una **lingua naturale**, al fine di consentire **interazioni uomo-macchina** efficaci
- Una delle principali aree di ricerca AI, sin dagli inizi (e.g., agenti conversazionali, traduzione automatica)
- Fortemente **interdisciplinare**: Linguistica, Informatica, Matematica, Statistica, ...
- Tecnologie AI/NLP sono parte della nostra **vita quotidiana** già da molti anni:



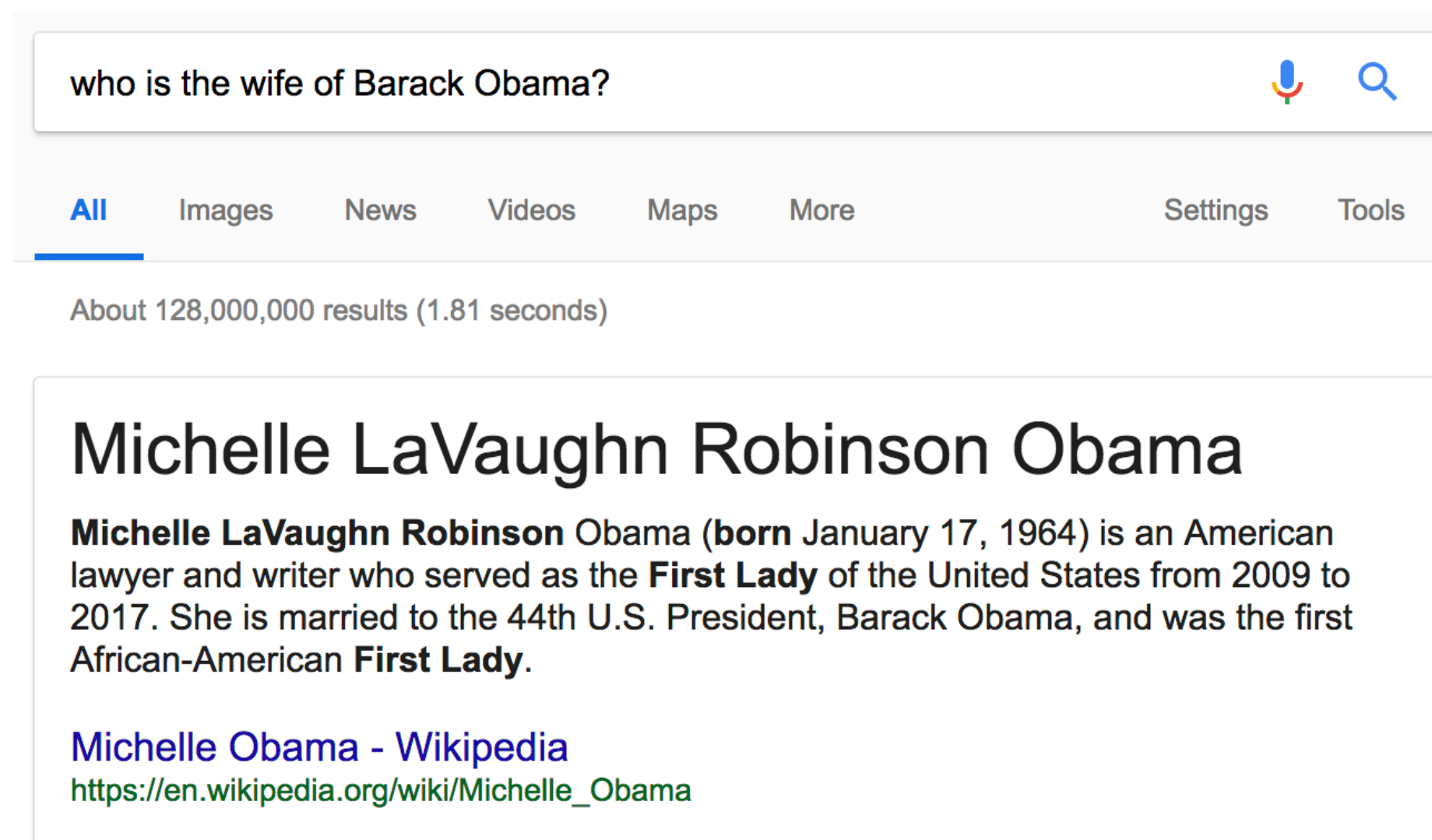
# Natural Language Processing (NLP)

- **Trattamento automatico** mediante un calcolatore elettronico delle informazioni scritte o parlate in una **lingua naturale**, al fine di consentire **interazioni uomo-macchina** efficaci
- Una delle principali aree di ricerca AI, sin dagli inizi (e.g., agenti conversazionali, traduzione automatica)
- Fortemente **interdisciplinare**: Linguistica, Informatica, Matematica, Statistica, ...
- Tecnologie AI/NLP sono parte della nostra **vita quotidiana** già da molti anni:



# Natural Language Processing (NLP)

- **Trattamento automatico** mediante un calcolatore elettronico delle informazioni scritte o parlate in una **lingua naturale**, al fine di consentire **interazioni uomo-macchina** efficaci
- Una delle principali aree di ricerca AI, sin dagli inizi (e.g., agenti conversazionali, traduzione automatica)
- Fortemente **interdisciplinare**: Linguistica, Informatica, Matematica, Statistica, ...
- Tecnologie AI/NLP sono parte della nostra **vita quotidiana** già da molti anni:



# Evoluzione NLP

- **Inizialmente:** codificare esplicitamente (algoritmi, **regole**) quello che gli umani fanno quando svolgono una particolare attività (trasformare un dato di input in un output / risultato)
  - Grandi speranze/promesse, realizzazione estremamente complessa e complicata
- Anni '90, prende piede il **Machine Learning**: possiamo **predire** l'output atteso basandoci sulle regolarità osservate in una grande quantità di **dati**?
  - e.g., possiamo “approssimare” la traduzione di un testo senza dover capire completamente il linguaggio umano?
  - i dati (e.g., il testo) cominciano a contare (quasi) più degli algoritmi di apprendimento stessi

# (Large) Language Models

- Modelli di distribuzione della **probabilità delle parole** o di **sequenze di parole**
  - Scopo: massimizzare la capacità di predire testo “corretto” nella lingua di riferimento
    - L’intervento del prof. Marco Rospocher è molto interessante
    - Interessante Rospocher dell’intervento molto avvocato è Marco
- Sono modelli pre-trained, costruiti da **enormi quantità di linguaggio scritto**
  - possono essere successivamente adattati (con successo) per tantissime analisi / attività diverse
- Hanno totalmente stravolto il modo di fare NLP negli ultimi 5/6 anni



# (Large) Language Models

- Modelli di distribuzione della **probabilità delle parole** o di **sequenze di parole**
  - Scopo: massimizzare la capacità di predire testo “corretto” nella lingua di riferimento
    - L'intervento del prof. Marco Rospocher è molto interessante 👍
    - Interessante Rospocher dell'intervento molto avvocato è Marco 👎
- Sono modelli pre-trained, costruiti da **enormi quantità di linguaggio scritto**
  - possono essere successivamente adattati (con successo) per tantissime analisi / attività diverse
- Hanno totalmente stravolto il modo di fare NLP negli ultimi 5/6 anni

# GPT: Generative Pre-trained Transformer

- Modello **generativo autoregressivo**: predire (iterativamente) la prossima parola di un testo dalla sequenza di parole precedenti



# GPT: Generative Pre-trained Transformer

- Modello **generativo autoregressivo**: predire (iterativamente) la prossima parola di un testo dalla sequenza di parole precedenti
  - L'Università degli Studi di Verona ha recentemente inaugurato un nuovo

# GPT: Generative Pre-trained Transformer

- Modello **generativo autoregressivo**: predire (iterativamente) la prossima parola di un testo dalla sequenza di parole precedenti
- L'Università degli Studi di Verona ha recentemente inaugurato un nuovo 
  - corso
  - dipartimento
  - edificio
  - dottorato
  - cappello
  - ...

# GPT: Generative Pre-trained Transformer

- Modello **generativo autoregressivo**: predire (iterativamente) la prossima parola di un testo dalla sequenza di parole precedenti
  - L'Università degli Studi di Verona ha recentemente inaugurato un nuovo 
    - corso
    - dipartimento
    - edificio
    - dottorato
    - cappello
    - ...
- Dandogli un opportuno suggerimento (**prompt**), possiamo utilizzarlo per varie analisi e attività diverse

# Come fa a imparare GPT?

- “Legge” **enormi quantità** di testo
  - Intuizione: ogni parola di un testo letto è la “miglior” parola da predire rispetto alle parole che la precedono nel testo
  - Ogni sequenza di testo analizzata contribuisce ad aggiustare “gli ingranaggi interni” del modello (GPT-3 ne ha circa 175 miliardi...)
- Un enorme quantità di testo è **già disponibile!**
  - Web, Wikipedia, Libri pubblicati, Corpora, ...



# Come fa a imparare GPT?

- “Legge” **enormi quantità** di testo
  - Intuizione: ogni parola di un testo letto è la “miglior” parola da predire **rispetto** alle parole che la precedono nel testo
  - Ogni sequenza di testo analizzata contribuisce ad aggiustare “gli ingranaggi interni” del modello (GPT-3 ne ha circa 175 miliardi...)
- Un enorme quantità di testo è **già disponibile!**
  - Web, Wikipedia, Libri pubblicati, Corpora, ...

# Come fa a imparare GPT?

- “Legge” **enormi quantità** di testo
  - Intuizione: ogni parola di un testo letto è la “miglior” parola da predire rispetto alle parole che la precedono nel testo
  - Ogni sequenza di testo analizzata contribuisce ad aggiustare “gli ingranaggi interni” del modello (GPT-3 ne ha circa 175 miliardi...)
- Un enorme quantità di testo è **già disponibile!**
  - Web, Wikipedia, Libri pubblicati, Corpora, ...



# chatGPT

- chatGPT è un adattamento di GPT, addestrato ulteriormente (in due fasi) grazie al **contributo umano**, per interagire in conversazioni, generare istruzioni, etc.
  - **prima fase**: sono stati creati manualmente molti esempi accoppiati di prompt e testo generato, e il modello è stato addestrato a imparare cosa deve fare (supervised learning)
  - **secondo fase**: il modello è addestrato a scegliere, tra più completamenti possibili per lo stesso prompt, quello più adatto nel contesto conversazionale (reinforcement learning)
- Sembra sia **in grado di fare di tutto** e di più...
  - generare / correggere / riscrivere / migliorare / ... un testo
  - creare / correggere / documentare / spiegare il codice sorgente di un programma
  - fare calcoli
  - rispondere a domande / quesiti
  - ...

# chatGPT

- chatGPT è un adattamento di GPT, addestrato ulteriormente (in due fasi) grazie al **contributo umano**, per interagire in conversazioni, generare istruzioni, etc.
  - **prima fase**: sono stati creati manualmente molti esempi accoppiati di prompt e testo generato, e il modello è stato addestrato a imparare cosa deve fare (supervised learning)
  - **secondo fase**: il modello è addestrato a scegliere, tra più completamenti possibili per lo stesso prompt, quello più adatto nel contesto conversazionale (reinforcement learning)
- Sembra sia **in grado di fare di tutto** e di più...
  - generare / correggere / riscrivere / migliorare / ... un testo
  - creare
  - fare ca
  - rispondere a domande / quesiti
  - ...

Ricordiamoci però di cosa stiamo parlando: un modello che, date delle parole, cerca di predire la prossima parola plausibile per completare il testo, nello stile dei testi che ha già visto

# Il boom di chatGPT...

- Popolarità senza precedenti...



# Il boom di chatGPT...

- Popolarità senza precedenti...



- ... in parte dovuto ad un ottima strategia di marketing e PR

## Sparks of Artificial General Intelligence: Early experiments with GPT-4

Sébastien Bubeck    Varun Chandrasekaran    Ronen Eldan    Johannes Gehrke  
Eric Horvitz    Ece Kamar    Peter Lee    Yin Tat Lee    Yuanzhi Li    Scott Lundberg  
Harsha Nori    Hamid Palangi    Marco Tulio Ribeiro    Yi Zhang

Microsoft Research

<https://arxiv.org/pdf/2303.12712.pdf>

Marco Rospocher  
chatGPT: avvertenze per l'uso

# Il boom di chatGPT...



<https://openai.com/research/gpts-are-gpts>

- Pop

GPTs are GPTs: An early look at the labor market impact potential of large language models

March 17, 2023

[Read paper ↗](#)

[Responsible AI](#), [GPT-4](#), [Publication](#)

- ... ir

## Abstract

We investigate the potential implications of Generative Pre-trained Transformer (GPT) models and related technologies on the U.S. labor market. Using a new rubric, we assess occupations based on their correspondence with GPT capabilities, incorporating both human expertise and classifications from GPT-4. Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted. The influence spans all wage levels, with higher-income jobs potentially facing greater exposure. Notably, the impact is not limited to industries with higher recent productivity growth. We conclude that



UNIVERSITÀ  
di VERONA

co Rospocher

UNIVERSITÀ DI VERONA | [avvertenze per l'uso](#)

# Handle with care

- Ricordiamoci di cosa stiamo parlando: un modello che, date delle parole, cerca di **predire la prossima parola plausibile per completare il testo**, nello stile dei testi che ha già visto
  - Non c'è nessuna garanzia sulla correttezza semantica di quanto prodotto (forma vs significato)
    - **Stochastic parrots** (E. Bender, T. Gebru, et al.)
  - Il testo molto convincente (fluidità, linguaggio) dà una falsa parvenza di verità, soprattutto a chi non è esperto di quanto trattato
    - Opinabili applicazioni in campo legale, medico, ecc.
- Non è un motore di ricerca!

# Handle with care

- L'addestramento di questi modelli richiede **enormi capacità computazionali**
  - solo poche realtà possono permettersi di costruire (e controllare) questi modelli
- OpenAI è stata critica per **non essere molto trasparente** (a dispetto del nome...)
  - Non sappiamo quali contenuti sono stati usati per addestrare GPT... violazioni copyright? condizionamenti?
  - GPT non è stato rilasciato pubblicamente... / versioning dei modelli controllato da OpenAI
- chatGPT è un sistema di cui **non conosciamo** ancora i limiti, i pericoli, l'affidabilità, il potenziale, ....:
  - **unscoped technology** (T. Gebru)
  - chi è **responsabile** di quanto prodotto da chatGPT?
  - come tutti i LLM, "impara" e amplifica i **bias**, i possibili condizionamenti e le distorsioni presenti nei testi usati per l'addestramento (e.g., discriminazione, disinformazione, messaggi di odio)



# Digital Arena for Inclusive Humanities

<https://bit.ly/daih-univr>

- Centro di ricerca interdisciplinare di eccellenza nell'**ambito delle tecnologie digitali** e delle **analisi computazionali** applicate alle lingue e letterature, con attenzione all'**inclusione**
  - Istituito con il finanziamento **Dipartimenti di Eccellenza MUR 2023-2027**
  - Finalità di ricerca all'intersezione tra l'**informatica**, e in particolare **AI**, e l'area **umanistica**, e in particolare gli **studi linguistici e letterari**
  - Obiettivo: favorire e promuovere l'**inclusione** a livello di ricerca, infrastrutture, formazione e divulgazione di contenuti scientifici per **superare le disparità, disuguaglianze e discriminazioni** della società moderna
  - Studiare l'inclusività di questi modelli e di come possano essere usati per favorire l'inclusione è una delle tematiche di ricerca del centro



# GPT: i fratelli e le sorelle...

- PaLM / Bard / LaMDA / ... (Google)
- LLaMA / Galactica (Meta AI)
- BLOOM (BigScience - open science project)
- ...
  
- Oltre la generazione di testo:
  - GPT-4 / DALL-E
  - Midjourney
  - Stable Diffusion
  - ...

# Marco Rospocher



UNIVERSITÀ  
di **VERONA**

Dipartimento  
di **LINGUE**  
E LETTERATURE STRANIERE



<https://marcorospocher.com>



<https://bit.ly/daih-univr>