

# INTRODUZIONE ALL'INFORMATICA UMANISTICA

Corso online di formazione per insegnanti

**Introduzione all'informatica umanistica (Digital Humanities).**

**Analisi e annotazione semi-automatica di testi letterari.**

**Generazione automatica di testi (AI generativa).**

*Andrea Bolioli (AIUCD)*

Venerdì 10 novembre 2023

ore 16:00 / 18:00



# Chi sono

## Andrea Bolioli



- Membro del Comitato direttivo dell'AIUCD
- Research and Innovation Consultant in *Aptus.AI*
- Linguista computazionale e digital humanist
- Fondatore di *CELI - Language Technology* 25 anni fa
- Laurea in Filosofia del linguaggio
- Ho collaborato per circa dieci anni con la Facoltà di Scienze della Formazione e con il DAMS di Torino
- Hobby: arrampicata, cinema, letteratura, ...

# Di cosa parliamo oggi

- Introduzione all'Informatica Umanistica / Digital Humanities  
Che cos'è, chi se ne occupa
- Analisi e annotazione semi-automatica di testi (letterari)  
L'esempio di cbook.it  
Edizioni scientifiche digitali
- Generazione automatica di testi (AI generativa).  
Linguistica computazionale / Natural Language Processing  
Modelli di linguaggio di grandi dimensioni (*large language models*) e ChatGPT

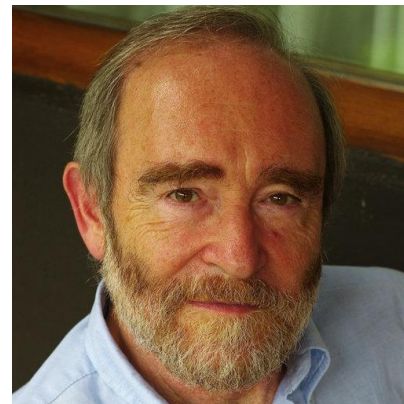
Ogni ora facciamo una sessione di domande, esperienze e riflessioni di (almeno) 10 minuti.

Le domande sono benvenute in ogni momento.

# In ricordo di Dino Buzzetti

Il Presidente onorario dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), prof. Dino Buzzetti, ci ha lasciato improvvisamente il 23 aprile 2023. Dino Buzzetti è stato tra i fondatori dell'associazione nel 2011, Presidente dal 2011 al 2014, e poi Presidente onorario, collega, maestro, amico, o semplicemente Dino...

<https://infouma.hypotheses.org/1475>



Il primo incontro di questo corso sarebbe stato fatto da Dino, come l'anno scorso.

## Ogni epoca ha le sue Humanities...



Il nome della rosa, 1986, Jean-Jacques Annaud

# Ogni epoca ha le sue Humanities... 2021



digital humanities



All

Books

Images

News

Videos

More

Tools

About 300,000,000 results (0.85 seconds)

[https://en.wikipedia.org/wiki/Digital\\_humanities](https://en.wikipedia.org/wiki/Digital_humanities)

## Digital humanities - Wikipedia

**Digital humanities (DH)** is an area of scholarly activity at the intersection of computing or digital technologies and the disciplines of the humanities.

[History](#) · [Values and methods](#) · [Projects](#) · [Criticism](#)

<https://corsi.unibo.it/DigitalHumanitiesKnowledge>

## Digital Humanities and Digital Knowledge - Laurea Magistrale

**Digital Humanities** Advanced Research Centre (DH.arc) ; Programme type Laurea Magistrale (Second cycle degree/Two year Master - 120 ECTS) ; Place of teaching ...

[Overview](#) · [Course Structure Diagram](#) · [Programme](#) · [Course Timetable](#)

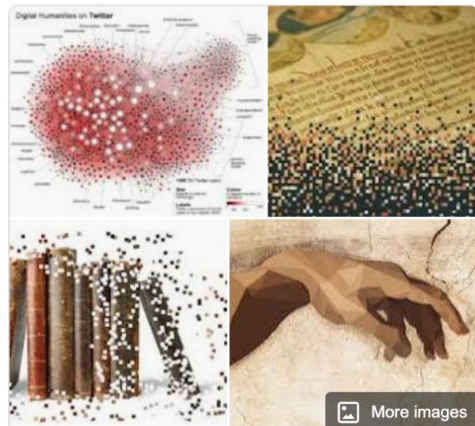
<https://corsi.unibo.it/magistrale/Di...> · [Translate this page](#)

## Digital humanities and digital knowledge - Laurea Magistrale

The degree programme aims to training Experts in Design, Production and Development of Web Resources, Experts in **Digital** Management of Cultural Resources, ...

## People also ask

What do u mean by digital humanities?



More images

## Digital humanities



Discipline

Digital humanities is an area of scholarly activity at the intersection of computing or digital technologies and the disciplines of the humanities. It includes the systematic use of digital resources in the humanities.

Motore di ricerca Google (09-11-2022) <https://www.google.com/>

# Ogni epoca ha le sue Humanities... 2023

Default (GPT-3.5)



che cos'è l'informatica umanistica, o digital humanities?



Le informatica umanistica, conosciuta anche come digital humanities in inglese, è un campo interdisciplinare che combina la tecnologia informatica e l'analisi digitale con discipline umanistiche come la storia, la letteratura, la filosofia, la linguistica e altre.

L'obiettivo delle digital humanities è quello di utilizzare strumenti informatici e metodi analitici per studiare, analizzare e comprendere meglio i testi, le opere d'arte, la storia e altri aspetti delle discipline umanistiche. Questo campo si avvale di strumenti informatici per raccogliere, analizzare, interpretare e presentare dati in modi



Regenerate

Send a message



ChatGPT can make mistakes. Consider checking important information.

Interfaccia di ChatGPT (09-11-2023)  
<https://chat.openai.com/>



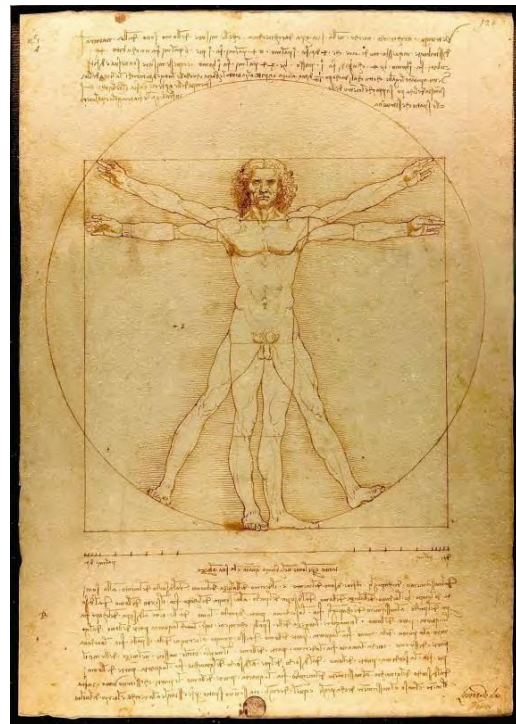
# Dove sono finiti gli umanisti?

"Il ruolo di filosofi, storici, letterati, antropologi, archeologi, linguisti, filologi, ecc, di fronte a questi immensi cambiamenti culturali, etici e geopolitici è semplicemente di consumatori ?"

Numerico et al., *L'umanista digitale*, 2010



Il nome della rosa, 1986, Jean-Jacques Annaud



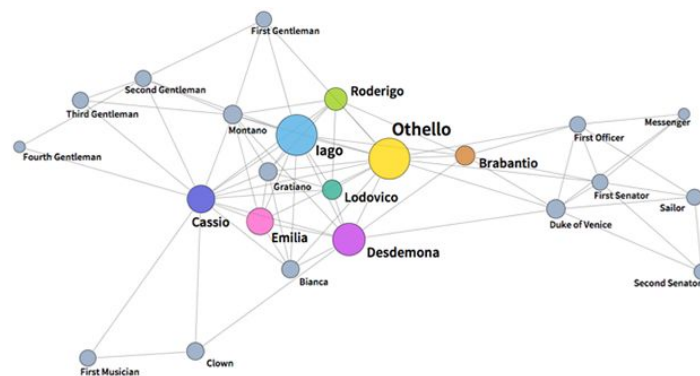
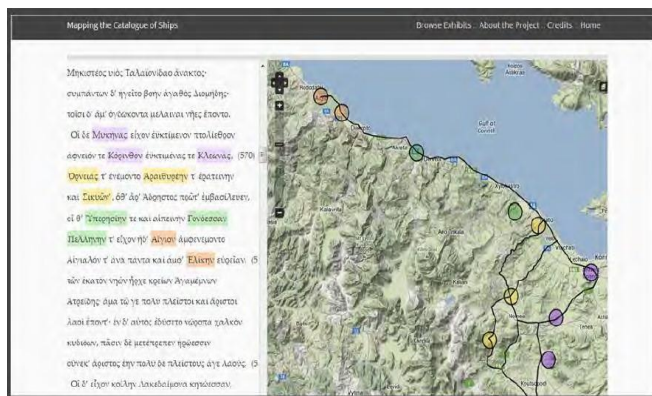
Uomo vitruviano, di Leonardo Da Vinci



# Digital Humanities / Informatica Umanistica 2010

Una definizione di DH del 2010, secondo Numerico, Fiormonte, Tomasi, “*L'umanista digitale*” (Il Mulino, 2010):

*Area di studio e applicazione delle **tecnologie ICT** (Information and Communications Technology) per gestire le risorse artistiche, culturali, storiche, letterarie, accessibili in digitale e per facilitarne l'accesso.*



# Digital Humanities / Informatica Umanistica 2023

*Digital Humanities – Metodi, strumenti, saperi,*

a cura di Fabio Ciotti, Carocci Editore

2023

<https://infouma.hypotheses.org/1901>



# Altri riferimenti utili sulle Digital Humanities

**Associazioni** di Digital Humanities (Informatica Umanistica, Computational Humanities, ecc):

- Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD) <http://www.aiucd.it/>
- Alliance of Digital Humanities Organizations (ADHO) <http://adho.org/>

**Blog** dell'AIUCD: <https://infouma.hypotheses.org/>

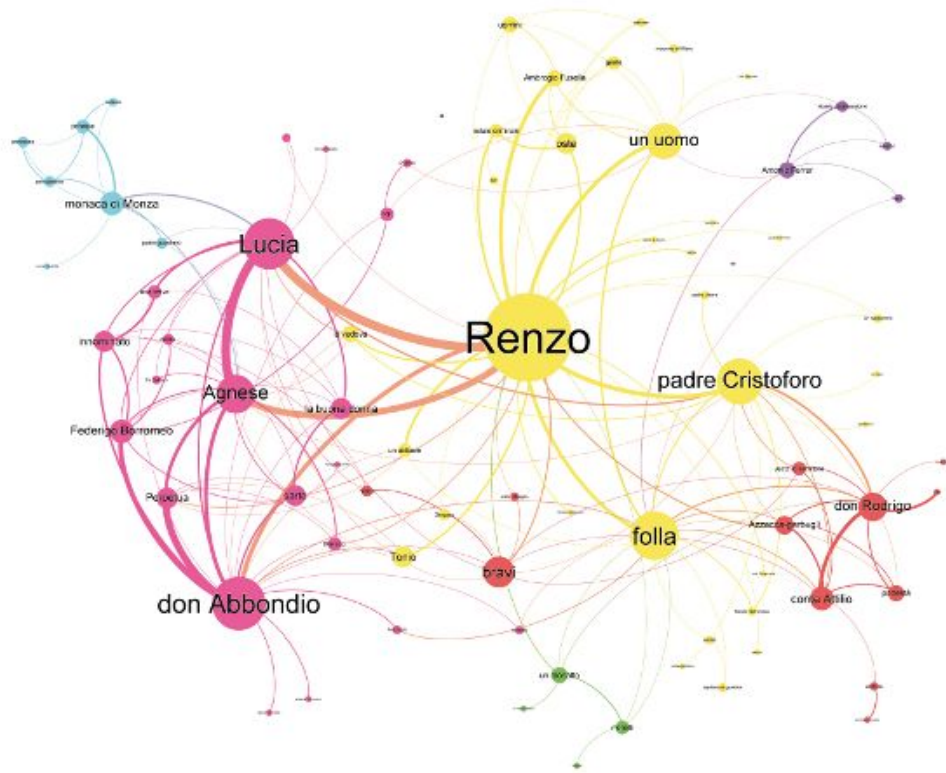


# Un esempio di DH con i classici della letteratura: cbook.it

Una piattaforma web collaborativa realizzata con insegnanti e studenti di scuole superiori a partire dal **2013** e utilizzata da insegnanti di circa 50 scuole in Italia dal 2014 al 2018 per realizzare percorsi didattici, esplorare i testi e aggiungere contenuti (immagini, mappe geografiche, reti sociali), annotare i testi.

<https://cbook.it/>

Ancora visibili pubblicamente oggi (2023):  
“I promessi sposi”  
“Le avventure di Pinocchio”



# Edizioni digitali: alcuni esempi

Il quaderno di Paolo Bufalini

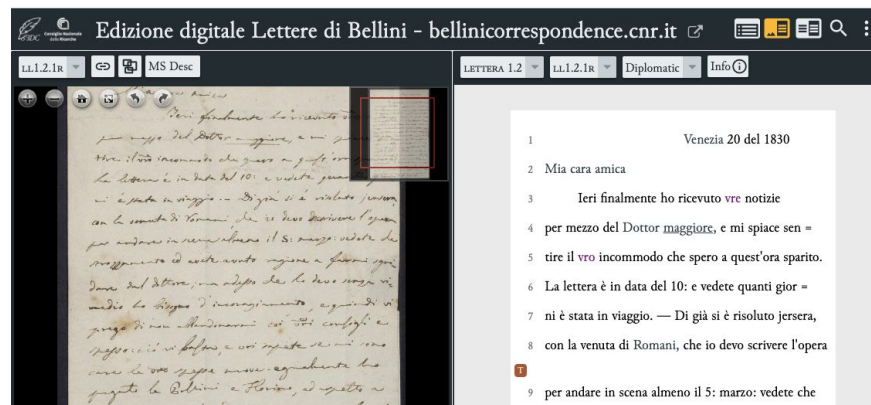
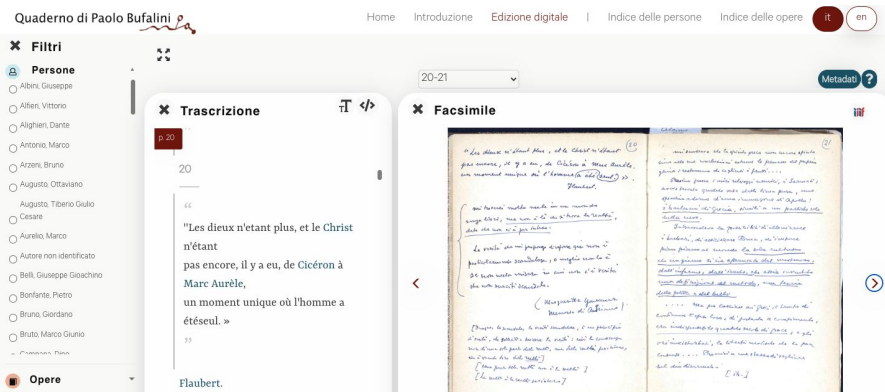
<https://projects.dharc.unibo.it/bufalini-notebook/>

Bellini Digital Correspondence (BDC)

<http://bellinicomrespondence.cnr.it/>

The Bodleian First Folio

<https://firstfolio.bodleian.ox.ac.uk/>



# **D**omande, esperienze, e riflessioni



# Linguistica Computazionale / Natural Language Processing

da Wikipedia inglese:

Computational linguistics is an interdisciplinary field concerned with the computational modelling of natural language, as well as the study of appropriate computational approaches to linguistic questions. In general, computational linguistics draws upon linguistics, computer science, artificial intelligence, mathematics, logic, philosophy, cognitive science, cognitive psychology, psycholinguistics, anthropology and neuroscience, among others. [...]

The term "computational linguistics" is nowadays (2020) taken to be a near-synonym of natural language processing (NLP) and language technology.

( [https://en.wikipedia.org/wiki/Computational\\_linguistics](https://en.wikipedia.org/wiki/Computational_linguistics) )

dal sito della Association for Computational Linguistics (ACL):

Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical").

( <https://www.aclweb.org/portal/what-is-cl> )

## Riferimenti utili di Linguistica computazionale

- Associazione Italiana di Linguistica Computazionale (AIRC)

<https://www.ai-lc.it/>



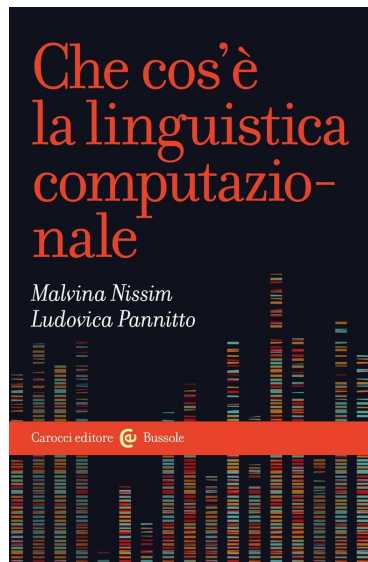
- Association for Computational Linguistics (ACL)

<https://www.aclweb.org/portal/>



Association for  
Computational  
Linguistics

# Libri recenti di Linguistica computazionale



Malvina Nissim, Ludovica Pannitto (2022)  
*Che cos'è la linguistica computazionale*  
Carocci editore



Elisabetta Jezek, Rachele Sprugnoli (2023)  
*Linguistica computazionale*  
*Introduzione all'analisi automatica dei testi*  
Il Mulino

# Interfacce di dialogo uomo-macchina

Interfacce di dialogo uomo-macchina / Interfacce conversazionali / Assistenti vocali

Con questi termini indichiamo dei sistemi di dialogo presenti negli smart speakers (altoparlanti intelligenti), come ad es. Amazon Alexa, Google Assistant, Xiaomi XiaoAI, nei telefoni cellulari, nei siti web, in alcuni robot (ad es. Nao e Pepper di SoftBank Robotics), e più in generale tutti i sistemi software/hardware in grado di interagire con una persona in linguaggio naturale

Andrea Bolioli “Parlare con gli oggetti: tecnologia e filosofia delle interfacce vocali” in MUSICA, SCIENZA E LINGUAGGIO. Dall’acustica musicale agli assistenti vocali.” A cura di Sergio Giudici. Edizioni ETS (2022)

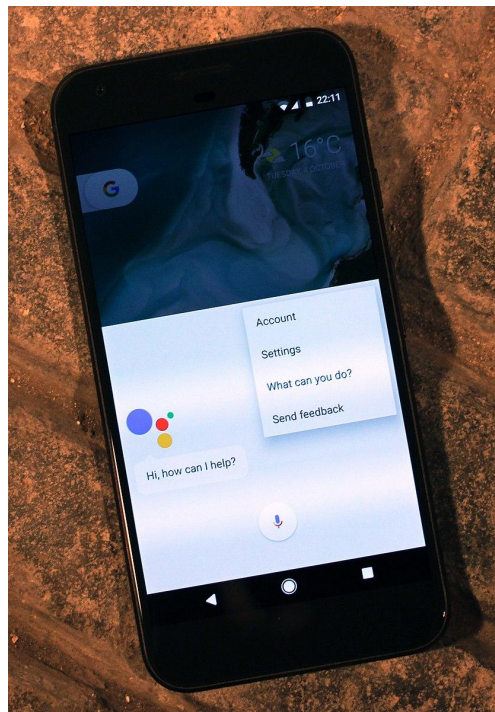
# Interfacce di dialogo uomo-macchina: Assistenti vocali



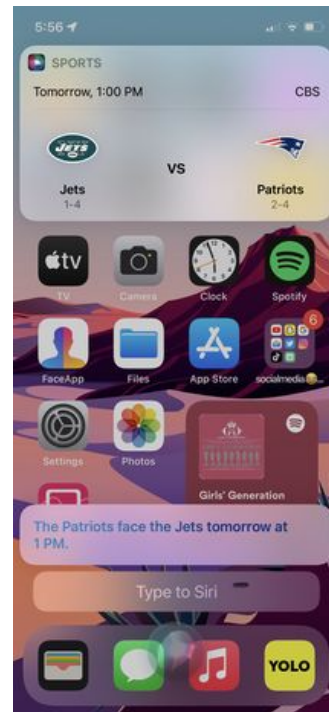
Amazon Alexa, immagine da Wikipedia



Google Home e Nest, immagine da Wikipedia

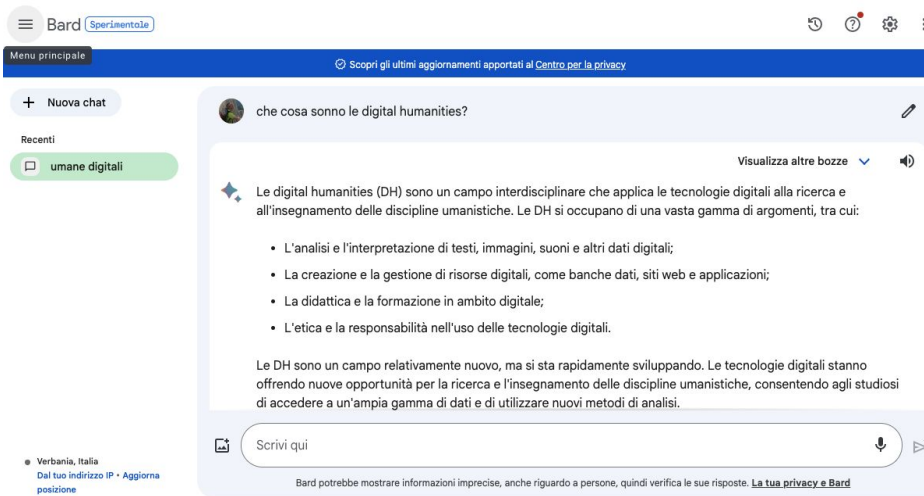


Google Assistant, immagine da Wikipedia

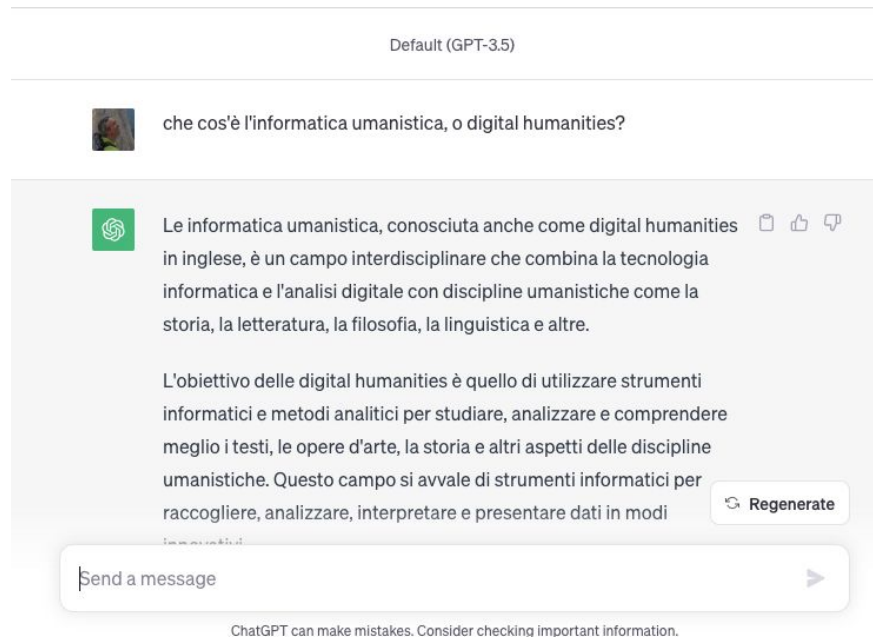


Apple Siri, immagine da Wikipedia

# Interfacce di dialogo uomo-macchina: Chatbot recenti



Bard di Google 10/11/2023 <https://bard.google.com/chat>



ChatGPT di OpenAI 10/11/2023  
<https://chat.openai.com/>



# Un large language model di successo: ChatGPT di OpenAI

**Generative Pre-trained Transformer** is an autoregressive language model that uses deep learning to produce human-like text. Given an initial text as prompt, it will produce text that continues the prompt.

**Competenze linguistiche** di GPT-3 e dei modelli successivi:

- Completare una frase
- Riassumere
- Rispondere a domande
- Dialogare / chattare
- Generare (scrivere) un testo
- Tradurre
- ...

Per provarlo:

<https://chat.openai.com/>

# Le dimensioni dei grandi modelli di linguaggio

Le dimensioni dei language model (*large language models*) sono aumentate considerevolmente nel corso degli anni: nel 2020, ad esempio, alcuni ricercatori di Google hanno presentato Meena, “a **2.6 billion parameter** end-to-end trained neural conversational model”<sup>6</sup>, un modello per le conversazioni in inglese creato a partire da circa 341 GB di testi (“public domain social media conversations”). Poco dopo, alcuni ricercatori di Facebook hanno presentato un modello (transformer neural networks) contenente **9,4 miliardi di parametri**, creato a partire da circa 1,5 miliardi di esempi di conversazioni.

Il modello di cui si è parlato di più nel 2020 è stato GPT-3 (Generative Pre-trained Transformer 3), un modello che usa il deep learning “to produce human-like text” creato da OpenAI (Brown 2020). GPT-3 conteneva **175 miliardi di parametri**.

Per quanto riguarda il **set di dati testuali** su cui era stato addestrato GPT-3 (2020), era composto complessivamente da circa **499 miliardi di token**. Il 60% proveniva da una versione filtrata dell'intero web dal 2011 al 2021, detta “Common Crawl”; le altre porzioni erano invece dataset di qualità maggiore e includevano una versione espansa del WebText dataset, due corpora di libri digitali (Books1 e Books2) e la Wikipedia in inglese.

Il set di training era parzialmente multilingua: circa il 93% delle parole del corpus in inglese, mentre il restante 7% sono in altre lingue, tra le quali il francese, il tedesco, lo spagnolo. La lingua italiana era rappresentata nello 0,6% del corpus, con più di 1 miliardo di parole.

T. Brown, B. Mann, N. Ryder et al., Language models are few-shot learners, in H. Larochelle, M. Ranzato, R. Hadsell (eds.), “Advances in Neural Information Processing Systems 33”, (NeurIPS 2020), pp. 1877-1901, 2020.

# Conclusioni

Abbiamo iniziato a conoscere:

- l'Informatica Umanistica (Digital Humanities) in generale
- l'analisi e annotazione semi-automatica di testi letterari
- la generazione automatica di testi (AI generativa).

Nei prossimi incontri verranno approfondite tematiche DH specifiche.

Per contatti: [andrea.bolioli@gmail.com](mailto:andrea.bolioli@gmail.com)

# **D**omande, esperienze, e riflessioni