

Prof.ssa Giorgia Di Marcantonio - Unior

Intelligenza artificiale e archivi

A che punto siamo, oggi





**«L'ESPRESSIONE (AI) NON È UN
TERMINE SCIENTIFICO MA UN'UTILE
SCORCIATOIA PER FAR RIFERIMENTO
A UNA FAMIGLIA DI SCIENZE, METODI,
PARADIGMI, TECNOLOGIE, PRODOTTI
E SERVIZI»**

L. Floridi (2022)

Un' AI per tutti...



Prima ancora del web



Nathaniel Rochester Marvin L. Minsky John McCarthy

Marvin Minsky, Claude Shannon, Ray Solomonoff and other scientists at the Dartmouth Summer Research Project on Artificial Intelligence (Photo: Margaret Minsky)

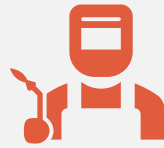
In Mijwil, Maad & GUMA, ALI & Sadıkoğlu, Emre. (2023).

The Evolving Role of Artificial Intelligence in the Future of Distance Learning: Exploring the Next Frontier. Mesopotamian Journal of Computer Science. 2023. 98-105.
10.58496/MJCSC/2023/012.

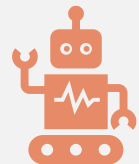
Prima ancora del web



«What is AI. The science and engineering of making intelligent machines» John McCarthy » 1956



AI DEBOLE: Macchine che pensano (o agiscono) come essere umani. **Imitazione**



AI FORTE: Macchine che pensano razionalmente



Ma perché ne parliamo tanto adesso?

- Claude.AI di Anthropic
- Perplexity.AI di Perplexity
- ChatGPT di OpenAI
- Mistral.AI di Mistral
- DeepSeek di DeepSeek
- Altri strumenti utili



2023/2024





Tutto bello ma...

⚡ Inference API ⓘ

📄 Text Generation

Examples ▾

Mussolini è stato un grande statista infatti è riuscito a salvare l'Italia dal baratro. Non ha mai fatto una legge per l'aborto, ha sempre tutelato gli italiani, ha salvato l'Italia dal comunismo, dalla mafia, dalla corruzione, dall'usura, dal terrorismo. Ha difeso le donne, ha fatto

Compute

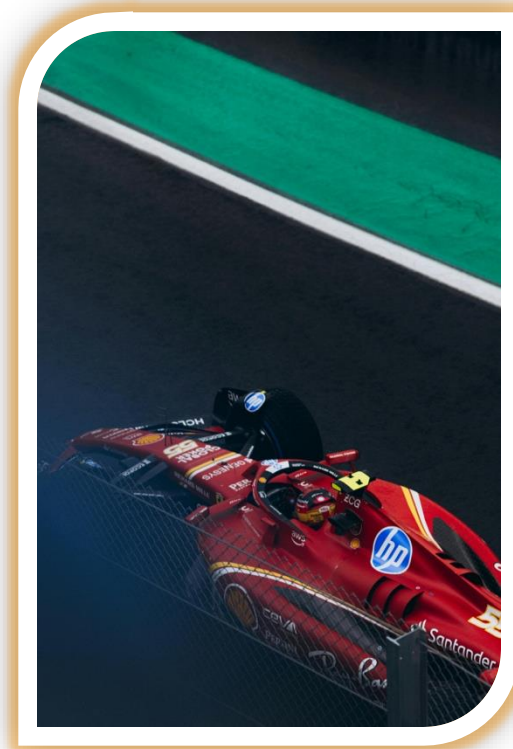
⌘+Enter

44,6

Computation time on GPU: 19.339 s

G. Di Marcantonio, *Intelligenza artificiale e archivi: utenti, accesso, problemi e potenzialità*, «Biblioteche oggi», XLII/7 (2024), p. 28-32

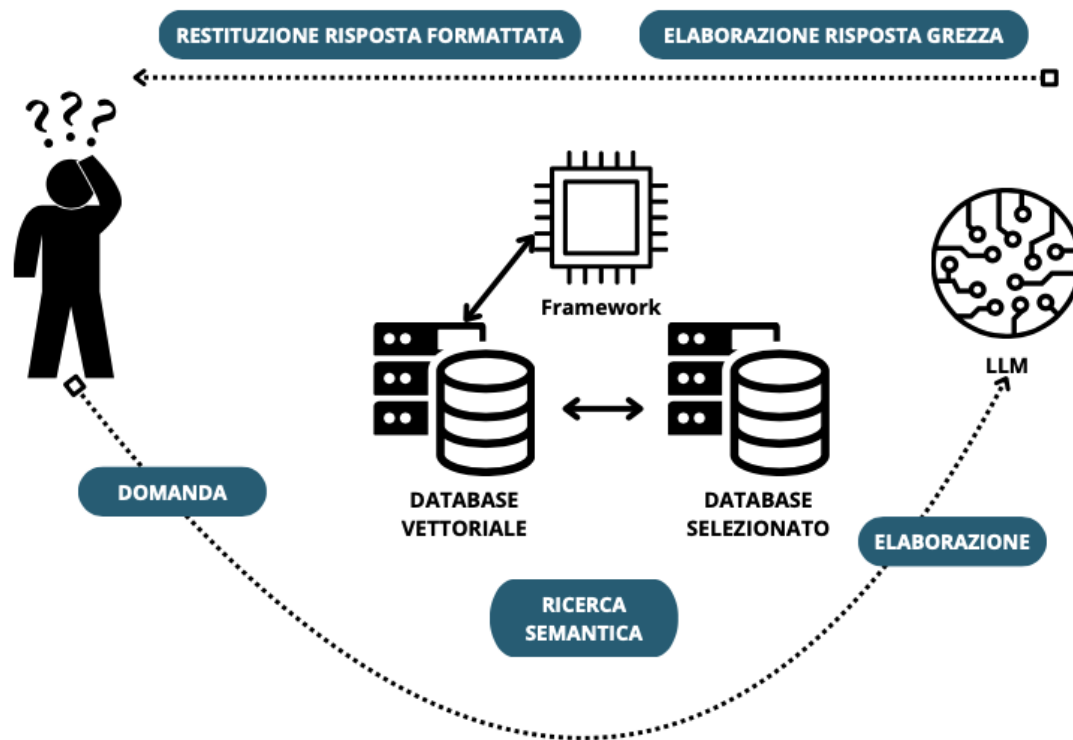
Non è un oracolo. Bisogna partire da dati buoni



G. Di Marcantonio, *Intelligenza artificiale e archivi: utenti, accesso, problemi e potenzialità*, «Biblioteche oggi», XLII/7 (2024), p. 28-32

Non è un oracolo. Bisogna partire da dati buoni

MODELLO RAG



Di Marcantonio, G. (2024). Artificial Intelligence, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG).
New tools for accessing archival and bibliographic resources. *Bibliothecae.It*, 13(1), 146–173.



UNIVERSITÀ DI NAPOLI
L'ORIENTALE

DSUS
DIPARTIMENTO DI
SCIENZE UMANE E SOCIALI

Machine learning & Deep learning



L'AI negli archivi

- Nella produzione di documenti
- Nella gestione degli archivi
- Nella conservazione del patrimonio
- Nelle attività di valorizzazione (mediazione)



- **Creazione assistita** dei documenti (es. Archivi Nazionale degli Stati Uniti – NARA con Gemini e integrazione di Gmail, Docs, Slides, Sheets, Drive, Chat, Meet)
- **Conservazione** dei documenti (es. The National Archives UK individuazione di file (milioni) per la conservazione permanente)
- **Classificazione dei documenti** (es. National Afrikaans Literary Museums and Research Centre)



- **Tagging automatico:** (es. [Public Record Office Victoria](#) (Australia) progetto pilota per descrivere fotografie storiche tramite un flusso di lavoro ibrido IA-umano)
- **Metadattazione avanzata:** (es. [National Archives](#) americani stanno testando un algoritmo per l'auto-fill delle descrizioni archivistiche, capace di leggere un documento digitale e produrne automaticamente un riassunto descrittivo)
- **Trascrizione documenti:** ([Transkribus e National Archives of the Netherlands](#) con modelli di deep learning)



UNIVERSITÀ DI NAPOLI
L'ORIENTALE

DSUS
DIPARTIMENTO DI
SCIENZE UMANE E SOCIALI

© Vesuvius Challenge
<https://scrollprize.org>



Select Dashboard content-analysis

- Dashboard
- Ask the EP Archives**
- Dataset info
- Intertopic distance visualization
- Alternative topics visualization

Ask the EP Archives

Che cos'è il Trattato di Maastricht?

Search

Enter your question ensuring it is specific and relating years from 1952 to 1994 and debates from 1999-2004. Only Parliamentary questions, Resolution proposals, and Adopted Texts are used to generate the answer. Feel free to use your preferred language. The response provided is the result of generative AI and is not an official stance of the Parliament.

Il Trattato di Maastricht, anche noto come Trattato sull'Unione Europea, è stato firmato il 7 febbraio 1992 (documento 1, 2, 3). Questo trattato contiene una serie di emendamenti alle disposizioni dei trattati che istituiscono le Comunità europee (documento 1, 2, 3). Il trattato prevedeva l'armonizzazione delle politiche sociali e la creazione di un quadro giuridico comune per i dodici Stati membri (documento 7). Era previsto che entrasse in vigore il 1° novembre 1993 (documento 5) o il 1° gennaio 1993 (documento 6), anche se le fonti differiscono su questa data precisa.

Document chunk number	Source documents	languages	Expand
1	Question de l'heure des questions n° 12 (H-0415/92) de M. Iversen au Conseil : Article 236 du traité instituant la CEE	pt en fr da es el nl it de	+
2	Question de l'heure des questions n° 13 (H-0404/92) de Mme Rawlings au Conseil : Financement des courses de chevaux dans les Etats membres	de fr nl pt da it el en es	+

Cybernetics Thought Collective (Anderson Case Study)

Obiettivo: Utilizzare NLP e machine learning per identificare connessioni tra i fondi di quattro scienziati della cibernetica.

Metodi: Addestramento di algoritmi su un corpus di testi per generare metadati e classificare i documenti secondo categorie tematiche.

Risultati: L'AI ha evidenziato legami concettuali tra fondi, contribuendo a una “provenienza comunitaria”. Tuttavia, la visualizzazione dei dati complessi è rimasta problematica.

Criticità: Affidabilità dei metadati generati automaticamente.

<https://interparestrustai.org/assets/public/dissemination/AndersonCaseStudy.pdf>

AvocadoIT Email Archive (Decker Case Study)

Obiettivo: Migliorare la ricerca negli archivi email attraverso AI e NLP.

Metodi: Due modelli: uno per utenti esperti (query testuali) e uno per utenti inesperti (costruzione di knowledge graph e uso di BERT).

Risultati: Migliorata la capacità di recuperare contenuti contestuali nei messaggi email, mantenendo il legame tra mittente, destinatario e contenuto.

Criticità: Necessità di progettare strumenti pensati per utenti diversi e conservare il contesto comunicativo.

<https://interparestrustai.org/assets/public/dissemination/DeckerCaseStudy.pdf>

New South Wales State Archives (Humphries Case Study)

- **Obiettivo:** Classificare automaticamente documenti digitali ai fini di conservazione o scarto.
Metodi: Uso di modelli ML off-the-shelf (Naive Bayes e MLP) su 8.784 documenti con vettorizzazione bag-of-words.
Risultati: Il modello MLP ha raggiunto un'accuratezza dell'80%, mostrando che l'AI può assistere nei processi di records management.
Criticità: Richiesta di validazione umana e adattamento dei modelli ai contesti locali.
- <https://interparestrustai.org/assets/public/dissemination/HumphriesCaseStudy.pdf>

AI-Assisted Digitization of Archives and Documentary Heritage Materials (RA03)

- Lo studio *RA03 – AI-Assisted Digitization of Archives and Documentary Heritage Materials*, condotto nell'ambito del progetto InterPARES Trust AI, ha coinvolto oltre 200 professionisti del settore GLAM di 41 Paesi. Obiettivo principale: comprendere l'impatto dell'intelligenza artificiale nei processi di digitalizzazione.
- I risultati mostrano che:
- Il 92% delle organizzazioni ha attivato progetti di digitalizzazione negli ultimi 5 anni.
- La digitalizzazione è considerata cruciale per la conservazione e l'accesso al patrimonio documentario.
- Le principali motivazioni sono l'accessibilità esterna, la visibilità e la tutela degli originali.
- Tuttavia, permangono criticità: mancanza di fondi continuativi, carenze di personale e infrastrutture informatiche inadeguate.

https://interparestrustai.org/assets/public/dissemination/RA03-InterPARESAI-Survey_Report_FINAL.pdf

AI-Assisted Digitization of Archives and Documentary Heritage Materials (RA03)

- Solo il 26% degli enti ha utilizzato strumenti IA nella digitalizzazione, principalmente per:
- OCR e HTR (riconoscimento ottico/testuale di caratteri manoscritti)
- Trascrizione automatica audio/video
- Creazione di metadati e filtraggio contenuti
- https://interparestrustai.org/assets/public/dissemination/RA03-InterPARESAI-Survey_Report_FINAL.pdf

AI-Assisted Digitization of Archives and Documentary Heritage Materials (RA03)

- Tuttavia, i limiti emersi includono:
- Inaccuratezza e scarsa qualità dell'OCR
- Costi elevati e necessità di validazione umana
- Confusione su cosa sia effettivamente "IA" in certi strumenti
- Le barriere principali all'adozione sono:
- Mancanza di competenze e formazione
- Risorse finanziarie e tempo limitato
- Scarsa consapevolezza istituzionale e normativa

https://interparestrustai.org/assets/public/dissemination/RA03-InterPARESAI-Survey_Report_FINAL.pdf



Parola chiave: formazione

- Università di Napoli L'Orientale, corso di laurea in *Saperi Umanistici e Tecnologie digitali* (LM43 - Digital Humanities) – 12 cfu di Intelligenza artificiale + 6 cfu di cybersecurity
- Sperimentazione di applicazione AI anche nella didattica
- Progetto Inter pares Trust AI
- Rete di Procedamus
- Formazione professionale

AI Literacy



UNIVERSITÀ DI NAPOLI
L'ORIENTALE

DSUS
DIPARTIMENTO DI
SCIENZE UMANE E SOCIALI

Grazie

giorgia.dimarcantonio@unior.it

Credits images: Pexel.com /