



ICT Training Center



Il tuo partner per la Formazione e la Trasformazione digitale della tua azienda





SPRING AI

GENERATIVE ARTIFICIAL INTELLIGENCE CON JAVA

Simone Scannapieco

Corso avanzato per Venis S.p.A, Venezia, Italia

Novembre 2025

RETRIEVAL AUGMENTED GENERATION

NATURA DINAMICA DEI CONTENUTI DIGITALI

The image shows three consecutive screenshots of a mobile messaging application interface, likely from an iPhone, displaying a conversation with ChatGPT. The screenshots are timestamped at 10:11, 10:13, and 10:13.

10:11: The user asks: "Quanti nuovi articoli vengono generati su Wikipedia ogni giorno?" ChatGPT responds: "Il numero di nuovi articoli creati su Wikipedia ogni giorno varia, ma in media vengono creati circa **500-1.000 nuovi articoli** al giorno, considerando l'intera piattaforma in tutte le lingue. Tuttavia, il numero può oscillare a seconda delle lingue e della comunità attiva in quel momento.

La versione in inglese di Wikipedia è solitamente quella con la crescita più alta, ma ci sono molte altre lingue in cui la creazione di nuovi articoli avviene quotidianamente. Inoltre, non tutti gli articoli creati vengono mantenuti, poiché alcuni possono essere eliminati per non rispettare le linee guida della piattaforma.

Questa cifra varia anche in base a fattori come eventi speciali, iniziative di scrittura e la disponibilità di nuovi contributori.

10:13: The user asks: "E quante pagine vengono modificate o generate ogni giorno nell'intero World Wide Web?" ChatGPT responds: "Non esiste un dato esatto e univoco sul numero di pagine web che vengono modificate o generate ogni giorno su tutto il World Wide Web, poiché questa informazione varia costantemente ed è distribuita tra miliardi di siti web gestiti in modo decentralizzato. Tuttavia, possiamo fare alcune stime basate su ricerche e analisi precedenti:

- 1. Numero totale di pagine web:** Il numero complessivo di pagine web è in costante crescita e si stima che vi siano più di 1,9 miliardi di siti web attivi. Ogni sito può contenere da poche a milioni di pagine. Nel 2021, Google aveva indicizzato oltre 130 trilioni di pagine web.
- 2. Aggiornamenti giornalieri:** Secondo alcune stime, il numero significativo di pagine web viene aggiornato

In sintesi, il numero di pagine modificate o generate quotidianamente sul web potrebbe essere nell'ordine di **decine di milioni** o addirittura più, ma non esiste un numero esatto facilmente calcolabile a causa della natura decentralizzata e dinamica di internet.

👉 Pre-addestramento costante per le LLM non è efficiente

👉 Come fare con **fattoidi** (es. cariche politiche, popolazione per stato, ...)?
Simone Scannapieco Spring AI - Corso avanzato

RETRIEVAL AUGMENTED GENERATION

RIVOLUZIONE NEL MONDO DELL'ADDESTRAMENTO LLM

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Namen Goyal, Heinrich Kuttler, Mike Lewis, Wen-Tau Yin, Tim Rocktaschel, Sebastian Riedel, Douwe Kiela. *Retrieval-Augmented Generation of Knowledge-Intensive NLP Tasks*.
ArXiv, 2021.



1800

1850

1900

1950

2000

2050

The screenshot shows a mobile phone interface with a white background. At the top, there's a status bar with the time "11:43", signal strength, battery level, and a small icon. Below the status bar, the screen displays a ChatGPT interface. On the left, there's a circular profile picture of a person. Next to it, the text "ChatGPT >" is followed by a small blue icon. In the center, a message bubble contains the text "Spiegami brevemente cosa è RAG". On the right side of the screen, there's another message bubble with the text "RAG (Retrieval-Augmented Generation) è una tecnica avanzata nell'ambito dell'intelligenza artificiale e del Natural Language Processing (NLP). Combinà due approcci principali: 1. Retrieval (Recupero): Viene utilizzato un sistema di ricerca per estrarre informazioni pertinenti da una grande base di dati o documenti, che potrebbero essere rilevanti per una determinata query o richiesta. 2. Generation (Generazione): Dopo aver recuperato queste informazioni, viene utilizzato un modello di generazione di linguaggio naturale (come GPT) per sintetizzare e presentare una risposta ben strutturata basata sia sulle informazioni recuperate che sulle proprie capacità generative.".

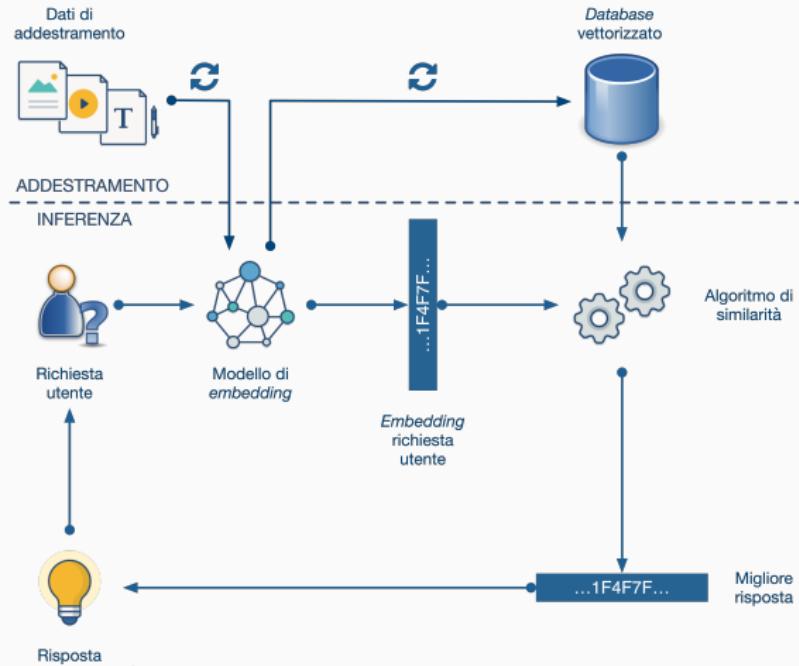
2. **Generation (Generazione):** Dopo aver recuperato queste informazioni, viene utilizzato un modello di generazione di linguaggio naturale (come GPT) per sintetizzare e presentare una risposta ben strutturata basata sia sulle informazioni recuperate che sulle proprie capacità generative.

In pratica, RAG consente ai modelli di intelligenza artificiale di combinare il potere della ricerca documentale con la capacità di generare testi coerenti e dettagliati, risultando in risposte più accurate e informate, soprattutto in contesti dove è necessario un supporto fattuale preciso.

- 👉 Nuova frontiera del *prompt engineering*
- 👉 Riduce i costi di addestramento LLM
- 👉 Mantiene costantemente aggiornati i sistemi LLM
- 👉 Costi di integrazione contenuti
- 👉 Approccio utilizzabile anche senza LLM

WORKFLOW DEL PROCESSO RAG

... E CON LLM



Approccio non generativo

RAG SENZA LLM: ESEMPIO

RICERCA SEMANTICA DI FILM E SERIE

