



AI nell'analisi documentale

Presentata da Javid Astro Ameri

Contesto

La gestione e l'**analisi di documenti** complessi richiedono **tempo** e **risorse** significative.

I Large Language Models (**LLM**) offrono la possibilità di **interrogare documenti** in linguaggio naturale.

L'uso del cloud comporta **criticità legate a privacy** e conformità GDPR, a causa dell'invio di dati sensibili verso server remoti.

Cresce la necessità di **soluzioni locali**, in grado di mantenere **il controllo sui dati**, ridurre la dipendenza da terze parti e garantire riservatezza.

Obiettivi del progetto

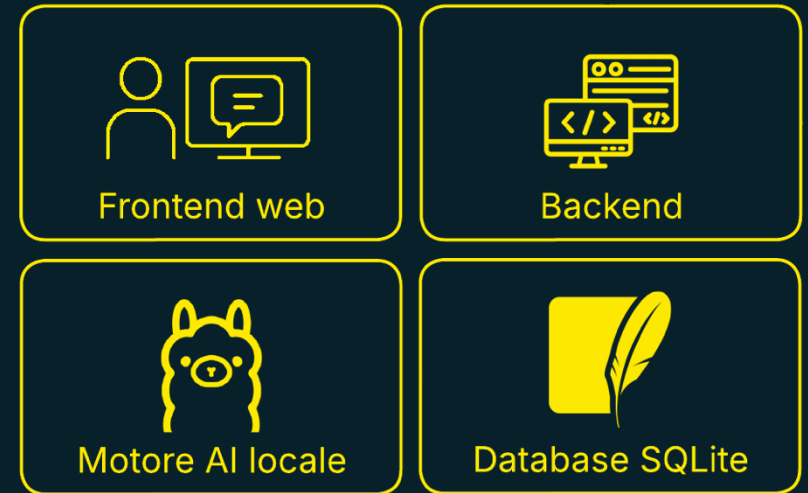
Creare un sistema prototipale locale per:

- Consultazione **documenti PDF** via chat
- Rispetto **privacy**, semplicità d'uso
- **Risposta contestuale** a domande
- Utilizzo in **ambiti**: contratti, normative, manuali tecnici



★ Architettura del progetto ★

- **Frontend:** chat, upload, gestione documenti
- **Backend:** parsing, indicizzazione, retrieval, risposta
- **LLM locale (Ollama)** per generazione risposte
- **SQLite** per persistenza chat e documenti



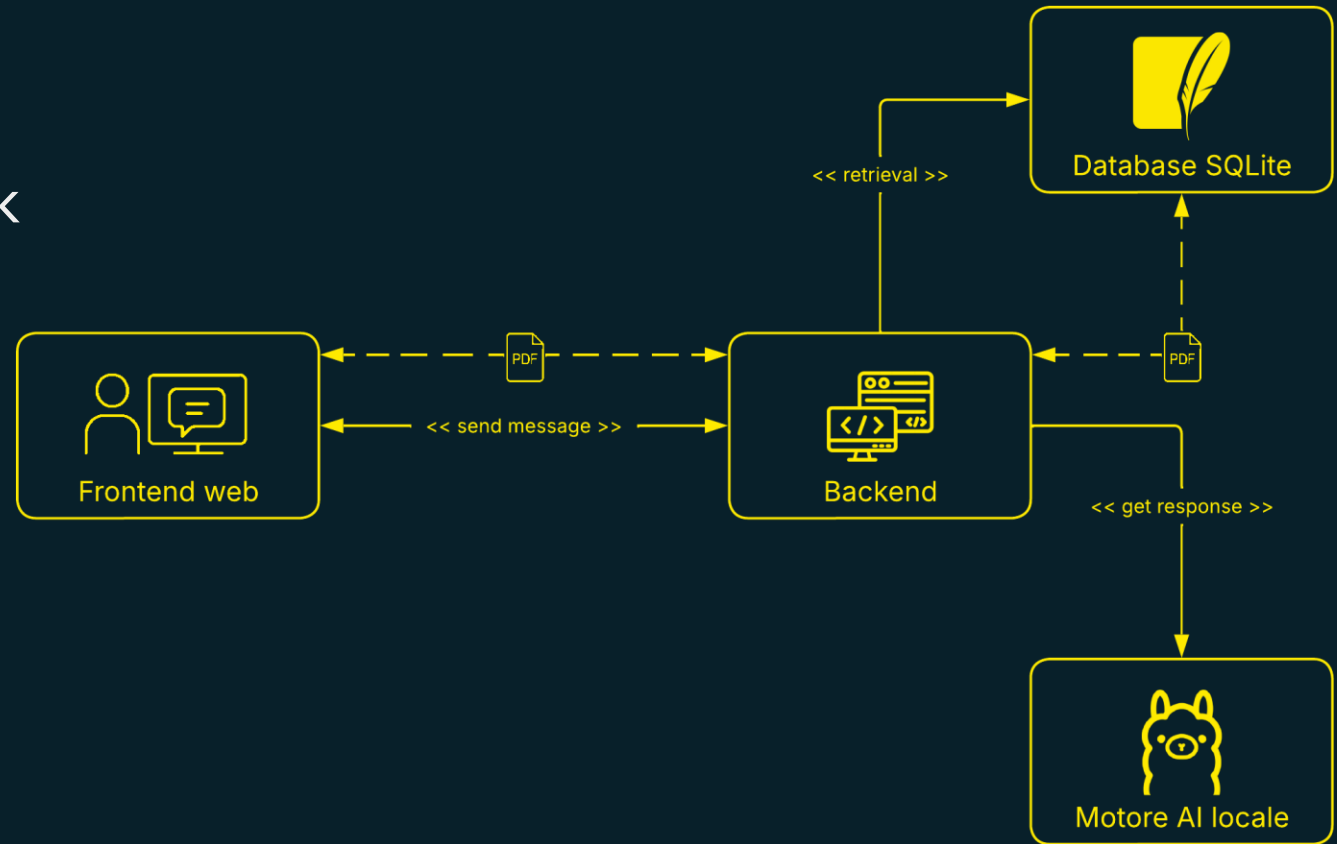
AI locale e Ollama

- Utilizzo **modelli open-source** (Mistral, LLaMA)
- **Esecuzione in locale** sulla macchina utente
- **Privacy garantita**: nessun dato inviato a server esterni
- **Controllo** sui modelli e personalizzazione flussi



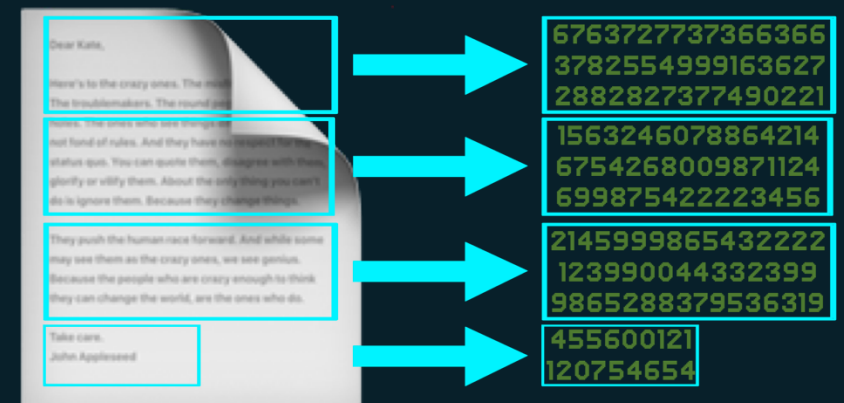
Flusso Operativo

- Upload PDF
- **Parsing** e suddivisione in chunk
- Calcolo **embedding**
- **Retrieval** frammenti rilevanti
- Generazione **risposta** LLM
- **Salvataggio** chat e messaggi



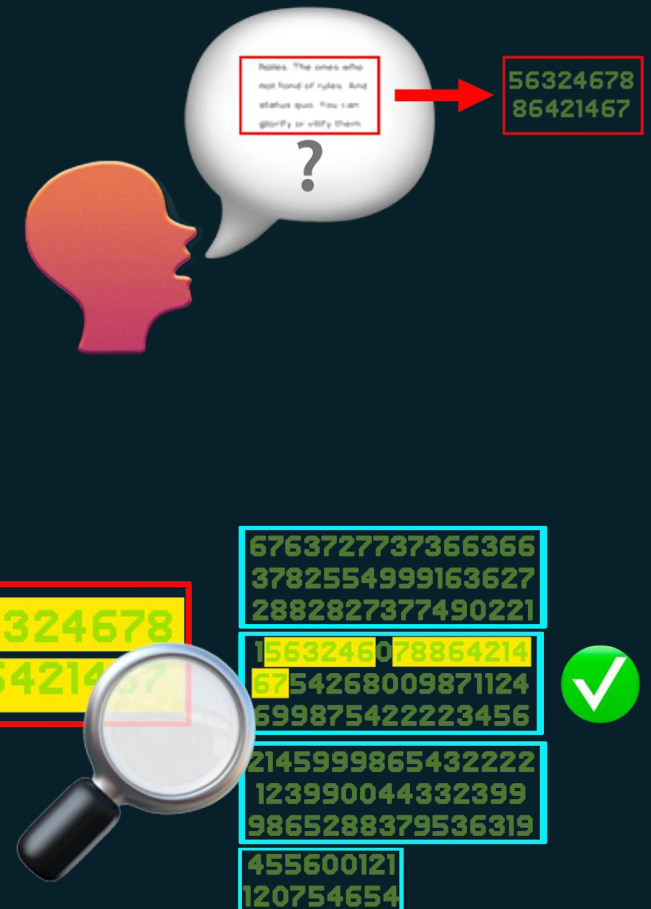
★ Parsing e Indicizzazione ★

- Parsing PDF con PDF.js sul backend
- Suddivisione in chunk sovrapposti (500 caratteri, overlap 100)
- Calcolo embedding con LLM locale
- Indicizzazione nel database per retrieval rapido



Retrieval semantico

- Misurazione pertinenza con **Cosine Similarity**
- **Confronto embedding** domanda vs embedding chunk
- **Estrazione frammenti** più rilevanti
- **Risultati** più **precisi** rispetto a keyword search

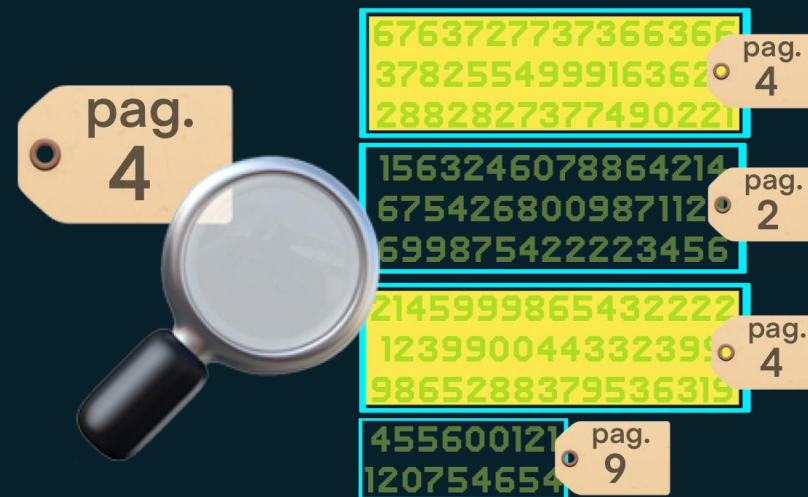


Retrieval per pagina

Sfida: rispondere a “*Cosa dice pagina X?*”

Tre strategie testate:

- I. Inserire numero pagina nei chunk
- II. Pattern matching “pagina X” nella domanda
- III. Classificazione AI per decidere retrieval



Scelta finale: pattern matching (efficiente, stabile)

Risultati e vantaggi

- Risposta **veloce** e **pertinente** alle domande
- **Privacy** e controllo totale sui dati
- **Consultazione semplificata** di documenti complessi
- Sistema **estendibile** per evoluzioni future

Modello AI: phi4:latest



v2.0



Le tue chat:

USA_Giappone_WWII.pdf



ScottPilgrim_vs_theWorld.pdf



Akira(1988).pdf



Rostam.pdf



Tobal_No._1.pdf



PokeTeam_Relazione.pdf



EarthBound.pdf



Conclusioni

- Dimostrata **fattibilità** dell'AI locale per analisi documentale
- Combina **semplicità** d'uso e privacy
- **Applicabile** in contesti sensibili
- **Futuri sviluppi**: estensione formati supportati, funzionalità avanzate, utenti

