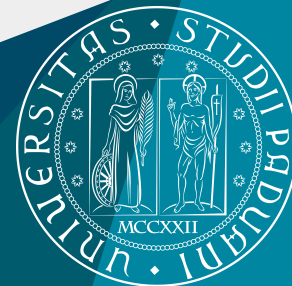


Presentazione RTB

Gruppo 20 - A.A. 2024/25

7Commits



Il capitolato: Artificial QI

“Valutare i risultati di sistemi che utilizzano i Large Language Models”

Obiettivo Principale:

- Realizzare un'unica applicazione per archiviare domande/risposte, inviarle via API a un LLM esterno, valutare i risultati e presentare i report

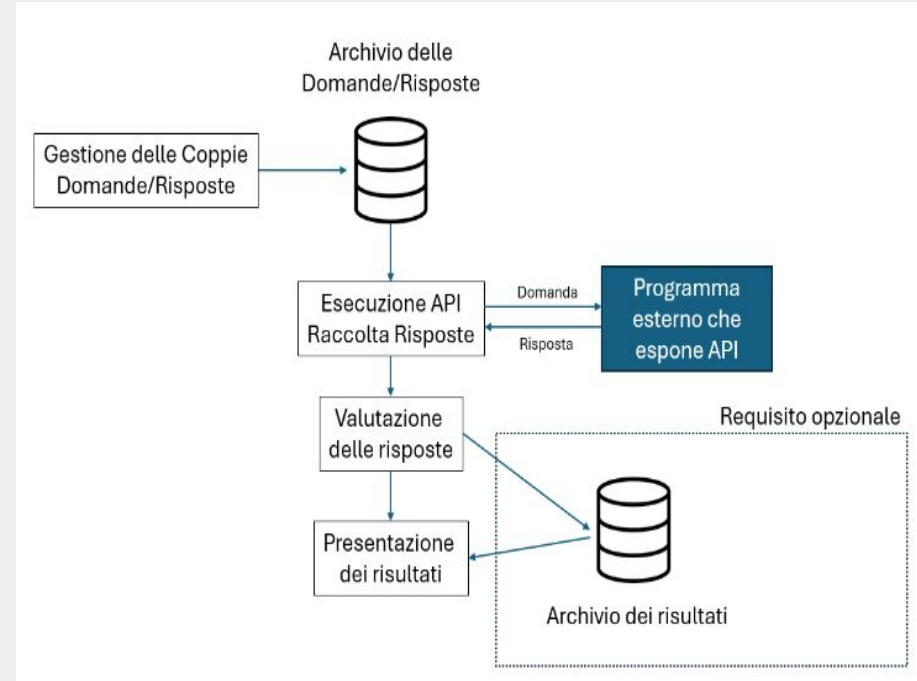
Contesto:

- Necessità di valutare le risposte ottenute da un LLM vs quelle attese

Analisi dei Requisiti: uno sguardo rapido

Requisiti obbligatori:

1. Archiviazione di una lista di domande e delle risposte attese.
2. Programma di esecuzione del test via API REST OpenAPI 3.1.
3. Programma di valutazione della correttezza/verosimiglianza delle risposte.
4. Procedura di presentazione dei risultati dell'esecuzione.
5. Integrazione in un unico sistema, non utility separate.



Scelta delle Tecnologie (sviluppo)

Python 3.10:

- Permette una rapida prototipazione
- Vasta disponibilità di librerie AI/data processing
- Presenza di molte risorse online



Streamlit:

- Libreria focalizzata sulla creazione rapida di dashboard e app di data science
- Ideale per visualizzare dati di Machine Learning
- Scelto per semplicità d'uso e rapidità di sviluppo



Scelta delle Tecnologie

(persistenza & container)

MySQL:

- Database per archiviazione di domande/risposte, risultati di test e report
- Il gruppo aveva più confidenza con MySQL



Docker:

- Garantisce portabilità, riproducibilità e isolamento dell'ambiente di sviluppo
- Tecnologia di containerizzazione già nota ai membri del gruppo
- Ottima documentazione e supporto online

