

SAVIA

INTELLIGENZA ARTIFICIALE
PER LA QUALITÀ DELLE LEGGI

IL CHATBOT SAVIA

MICHELE VISCIARELLI

30 giugno 2025



Regione Emilia-Romagna
Assemblea legislativa



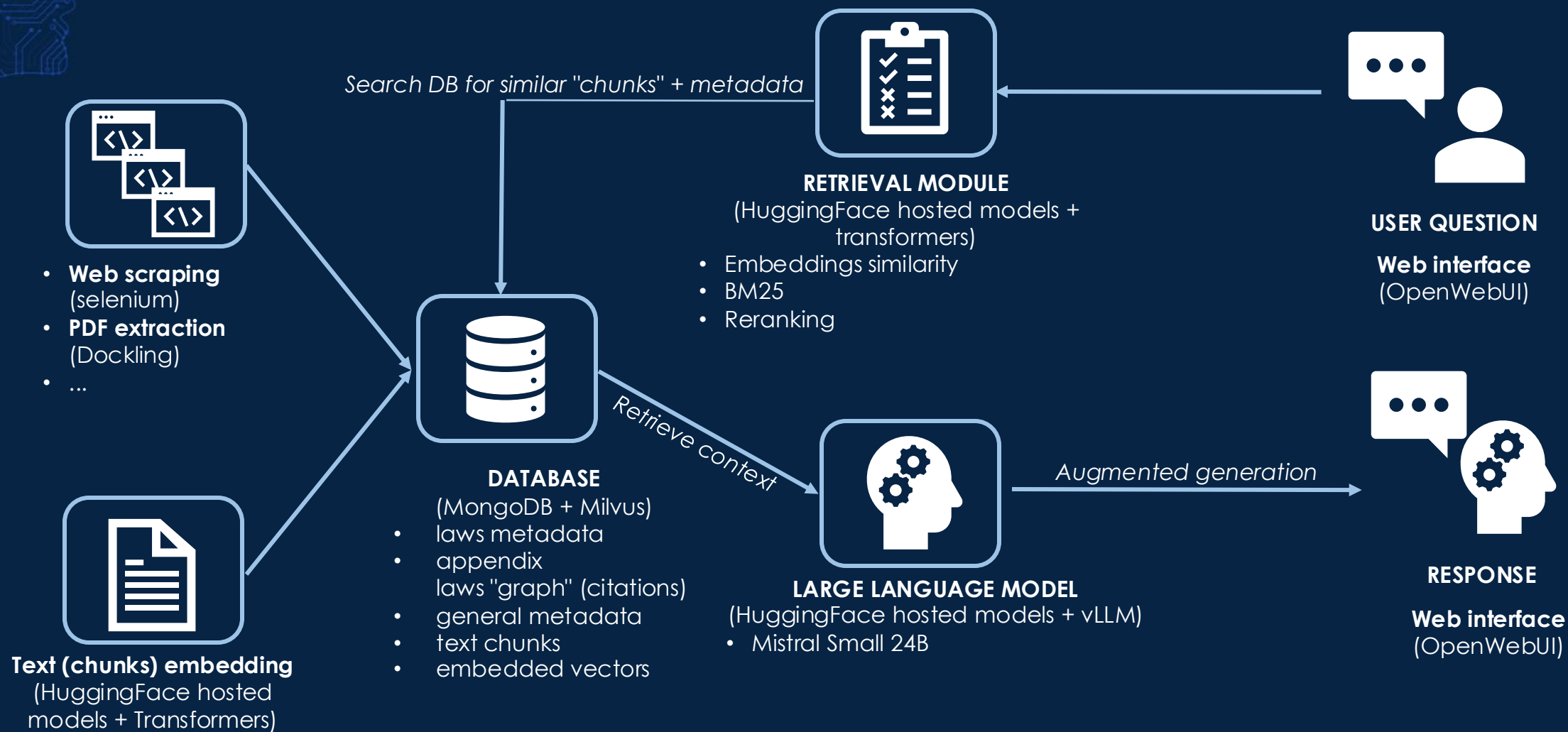
Il progetto SAVIA – il chatbot

Scopo:

- creare un tool per navigare il corpus normativo della regione Emilia-Romagna, collegando leggi, atti, appelli, sentenze, documenti valutativi, ecc...

Metodologia:

- creare un **chatbot**, basato su large language models (LLMs), in grado di rispondere a domande sulla normativa regionale e documenti correlati.





Definizione della base documentale

Dataset:

- Leggi regionali (1972-2024) da Demetra
- Atti implementativi ("atto del dirigente", "atto della giunta")
- Atti attuativi (regionale/provinciale/comunale - WIP)
- Documenti collegati (es: "Clausola Valutativa", AIR/VIR)

Svariati formati:

- Banche dati web —> **Web scraping** (selenium)
- Scansione di documenti/PD —> **PDF extraction** (Docling)
- ...



Definizione della base documentale

Divisione in "chunk":

- atti, paragrafi, unità semanticamente simili, ecc... (LangChain)

Estrazione di metadati:

- anno/data/numero/riferimenti ad altre leggi, ecc... (NLP "classico")
- riassunti, ecc... (LLM)



Large Language Models

Definizione:

Un large language model (LLM) è un tipo di intelligenza artificiale. Apprende da una grossa mole di dato testuale e grazie a questo riesce a processare e produrre testo in linguaggio naturale.

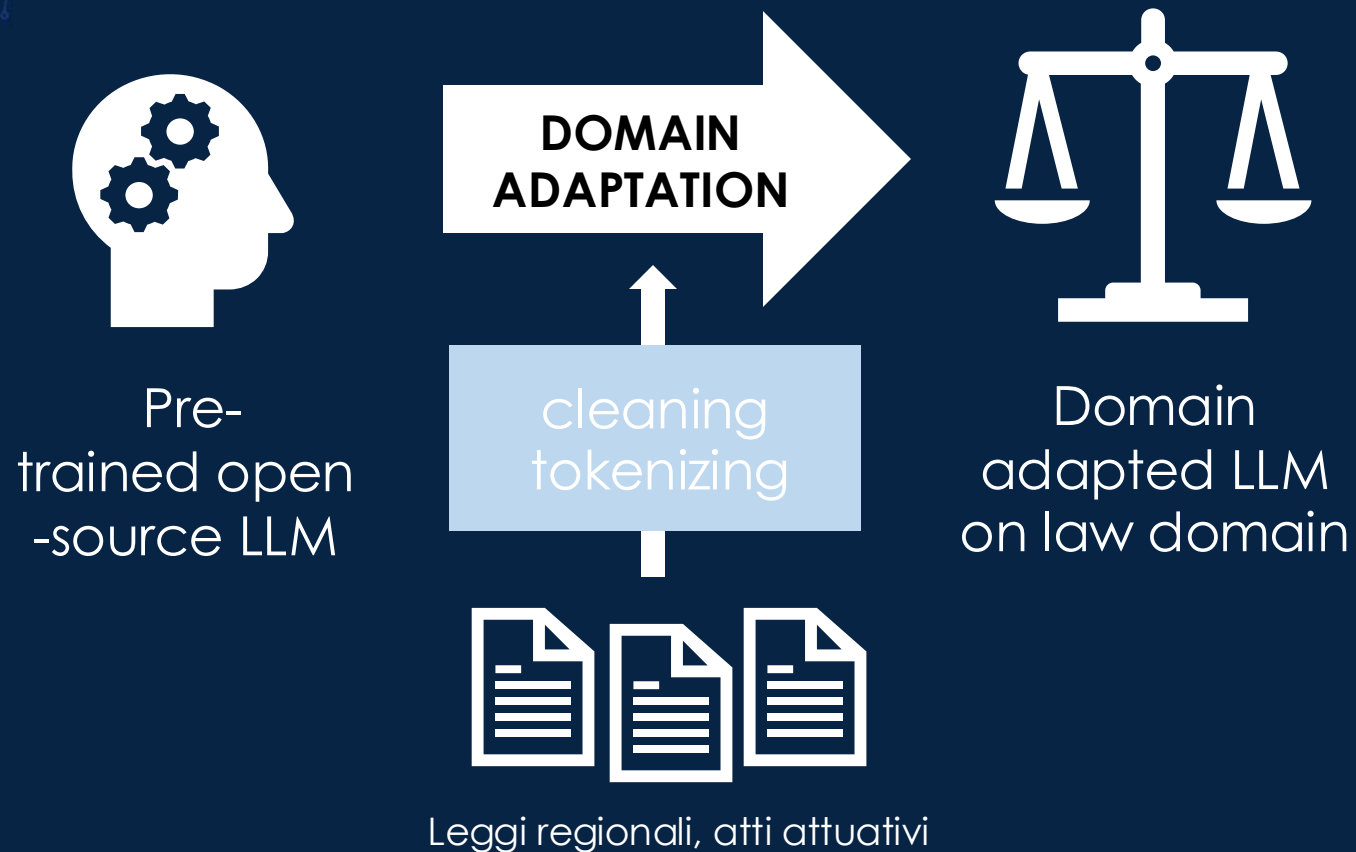
Training:

Fase di allenamento generale dove il modello viene allenato su una grandissima quantità di dato testuale nella forma di libri, articoli, pagine web, social media, domande-risposte ecc... In questa fase, il modello apprende dalla sua base documentale ed è in grado di completare/dialogare su un input dell'utente.

Fine Tuning:

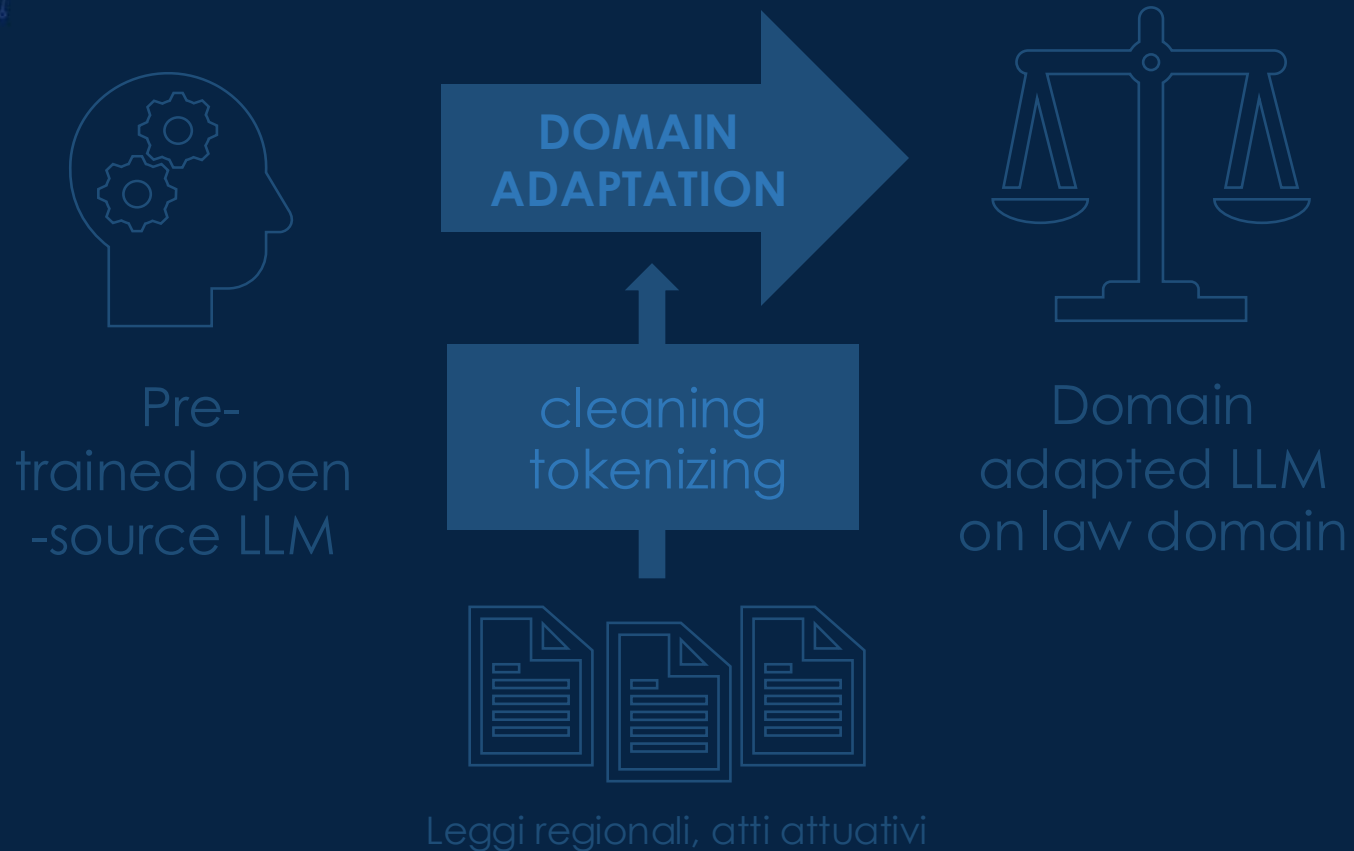
Fase di allenamento ulteriore dove il modello apprende da ulteriori documenti (di solito di un dominio specifico). Il modello quindi viene "regolato" su uno specifico dominio/caso d'uso.

Large Language Models



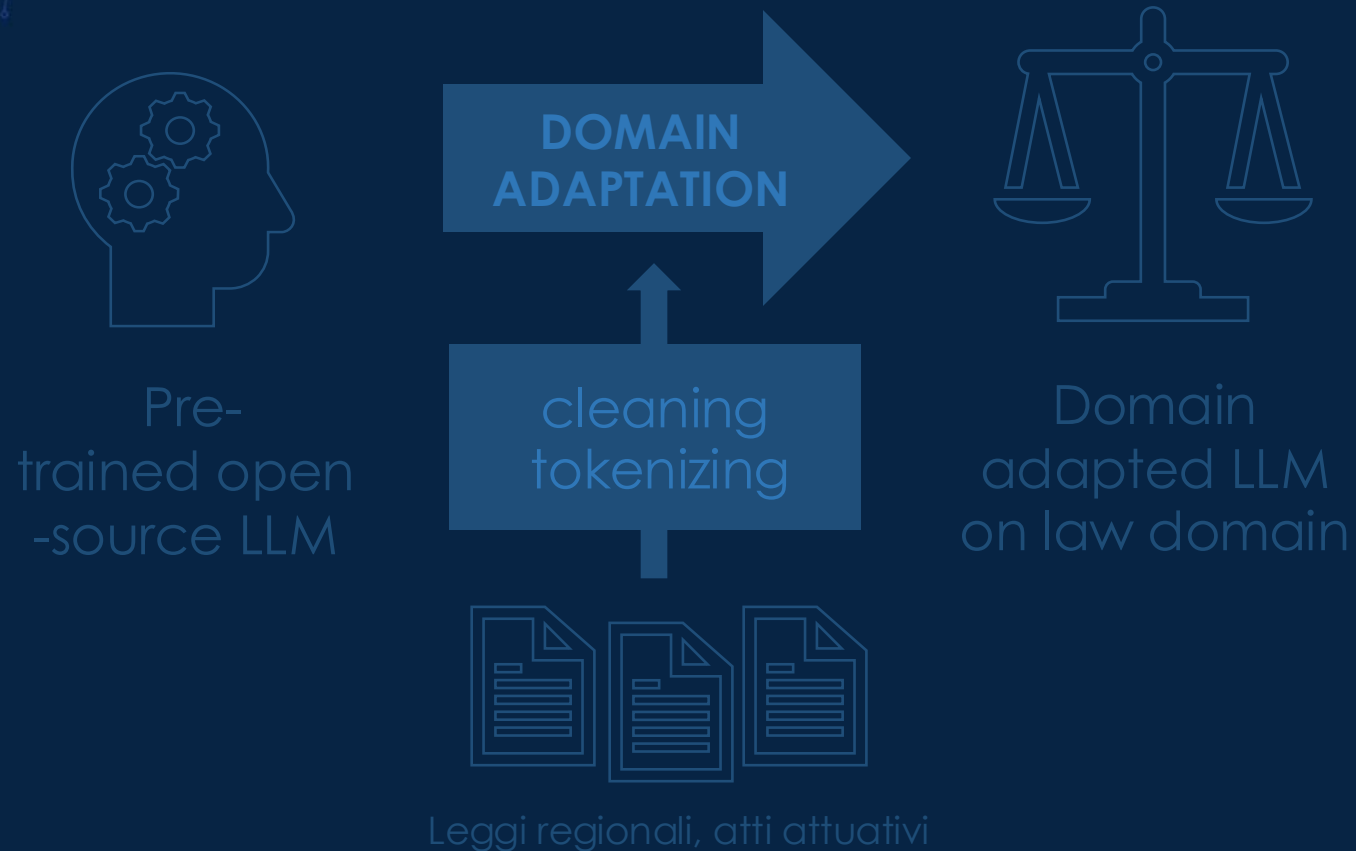
Pre-trained open-
source LLM
PIU' RECENTI

Large Language Models



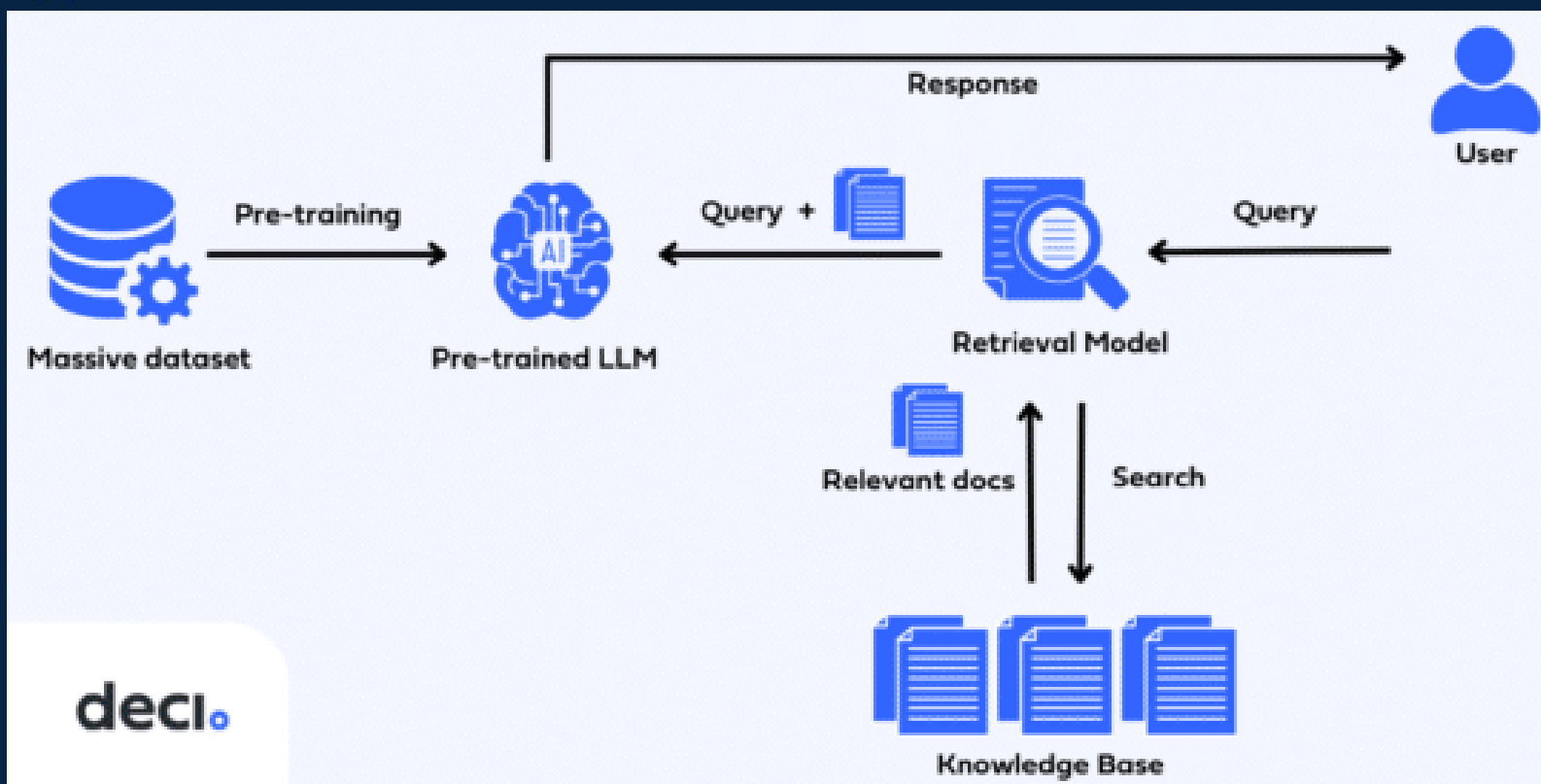
Pre-trained open-
source LLM
PIU' RECENTI

Large Language Models



- Llama 3.1 (8B -> 405B)
- Qwen (32 -> 70B)
- Gemma (27B)
- **Mistral Small 24B**
- ...

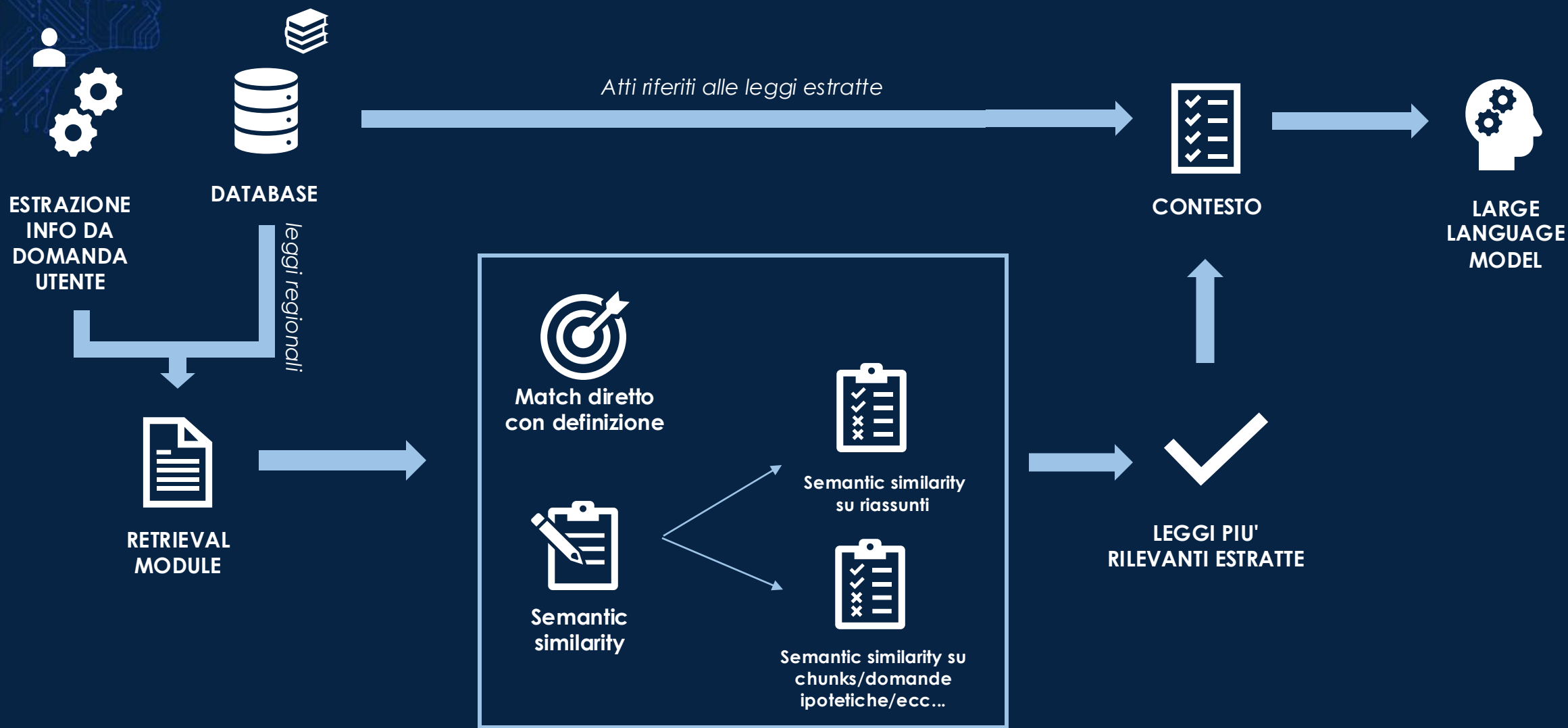
Retrieval Augmented Generation

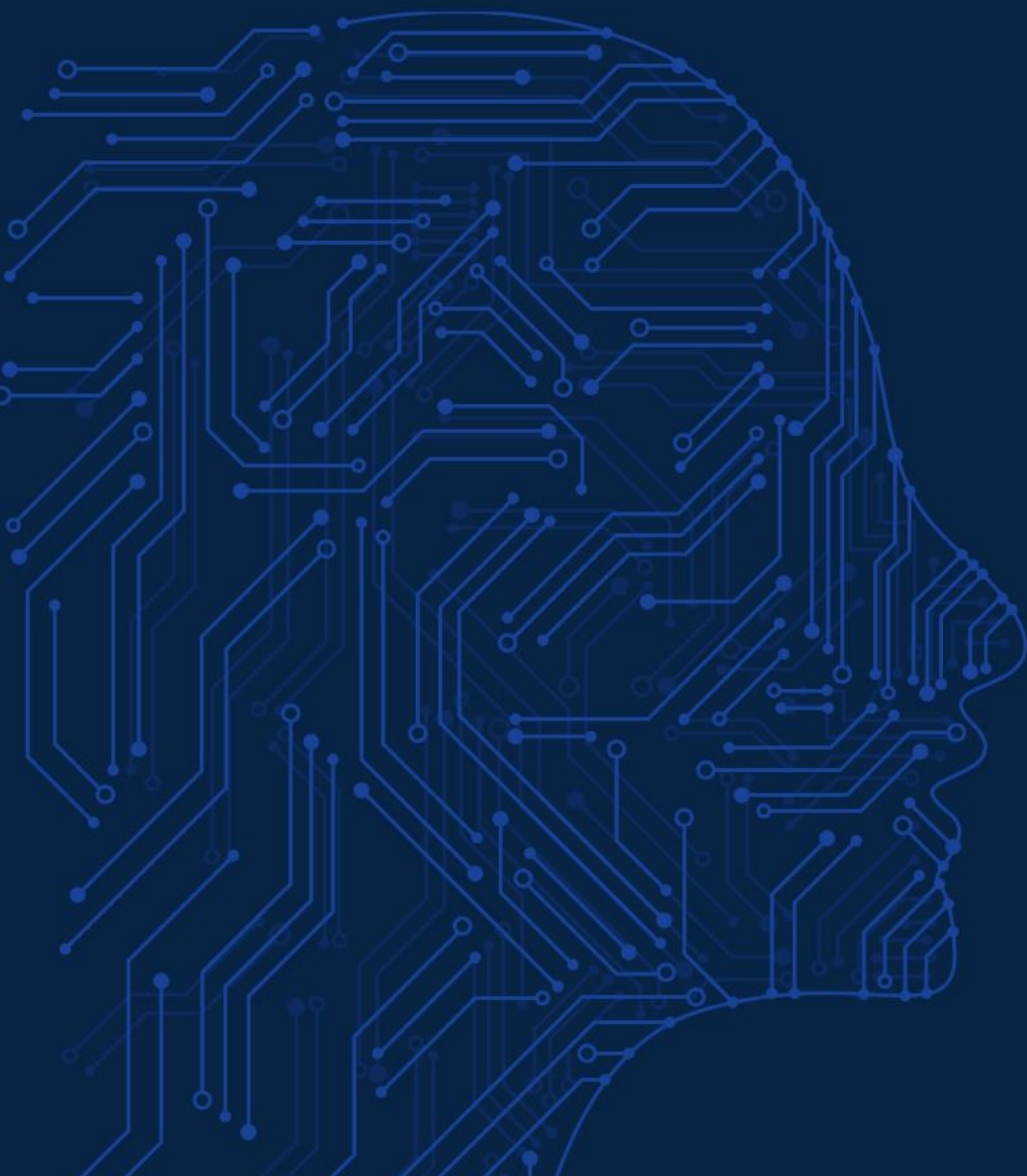


- Tecnica per aumentare l'accuratezza e affidabilità degli LLM con informazioni recuperate da fonti esterne.
- Nel prompt viene inserito contesto relativo alla domanda dell'utente recuperato dalla base documentale.



INTELLIGENZA ARTIFICIALE
PER LA QUALITÀ DELLE LEGGI





SAVIA

INTELLIGENZA ARTIFICIALE
PER LA QUALITÀ DELLE LEGGI

MICHELE VISCIARELLI

m.visciarelli@cineca.it



Regione Emilia-Romagna
Assemblea legislativa