

Understanding Chatbots

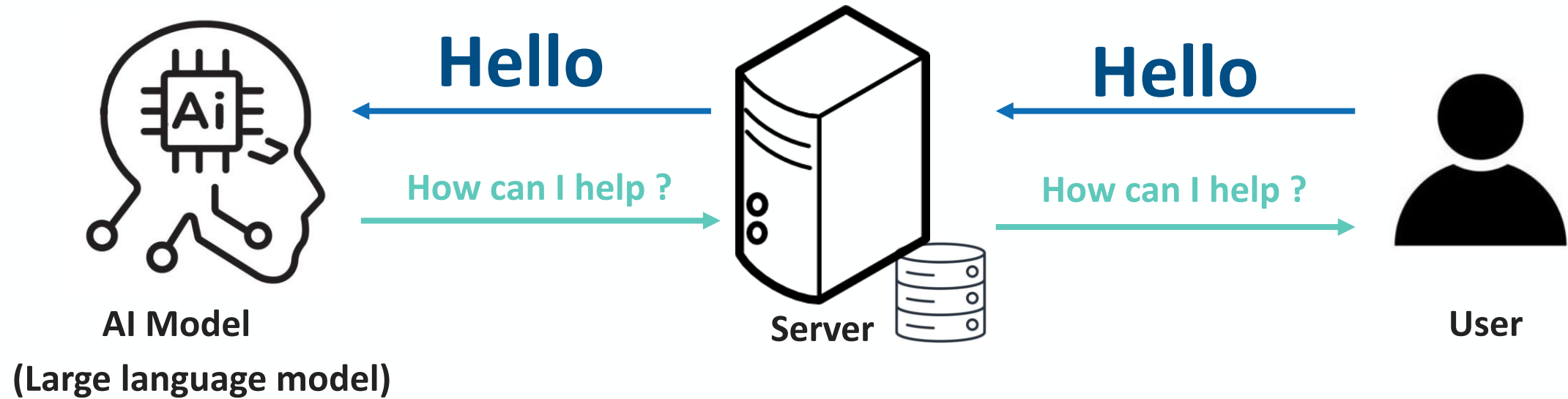
Jose Almeida | 25 November 2025 | Bucharest

Interact

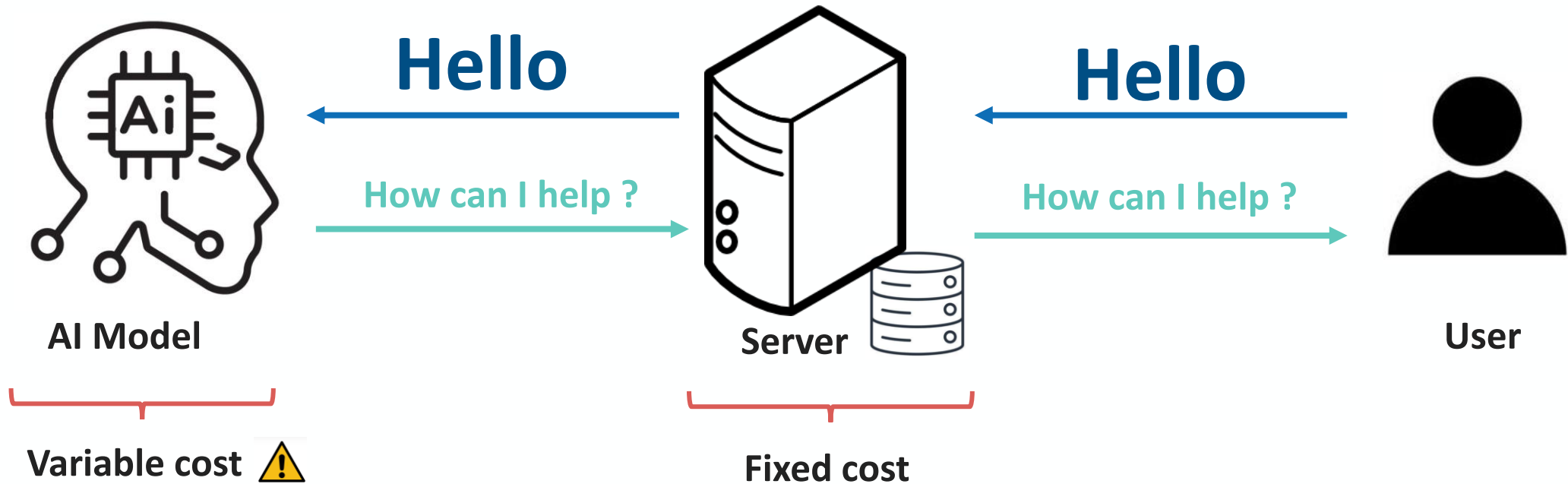


**Co-funded by
the European Union**
Interreg

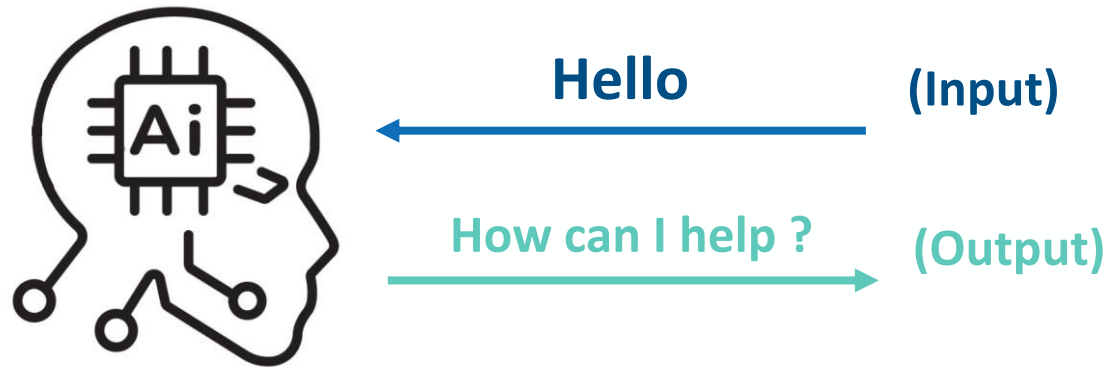
Chatbot



Chatbot

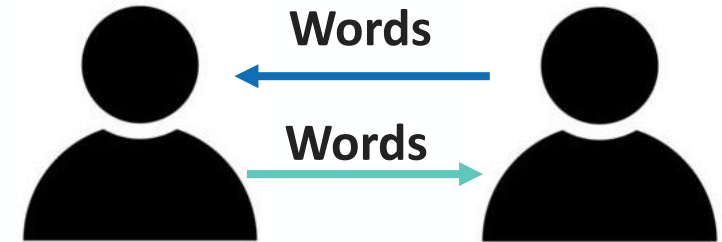
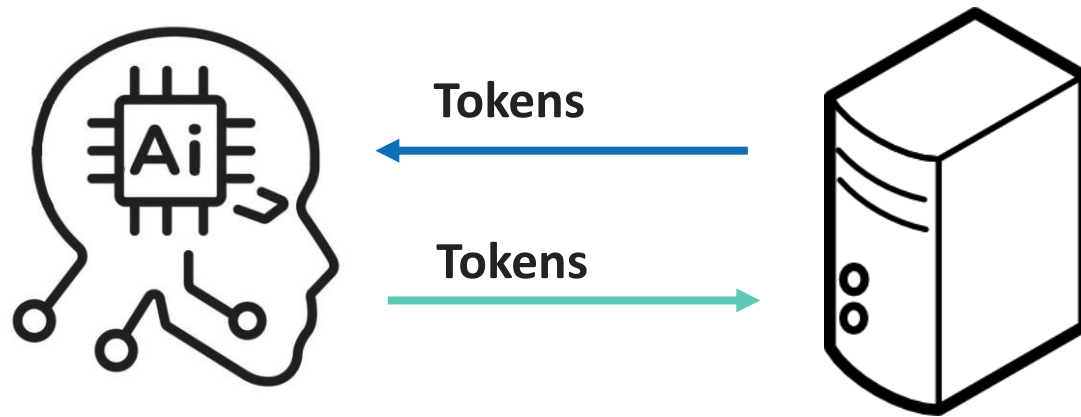


Input and output



$$\begin{aligned}\text{Total cost} &= (\text{'Hello'}) + (\text{'How can I help ?'}) \\ &= 1x (\text{Input}) + 5x (\text{Output})\end{aligned}$$

Machines speak tokens



Tokens = words + 20% (aproximation, english language)

Example pricing – OpenAI API (United States)

GPT-5.1

The best model for coding and agentic tasks across industries

Price

Input:

\$1.250 / 1M tokens

Output:

\$10.000 / 1M tokens

Possibly 1Trillion parameters

GPT-5 mini

A faster, cheaper version of GPT-5 for well-defined tasks

Price

Input:

\$0.250 / 1M tokens

Output:

\$2.000 / 1M tokens

GPT-5 nano

The fastest, cheapest version of GPT-5—great for summarization and classification tasks

Price

Input:

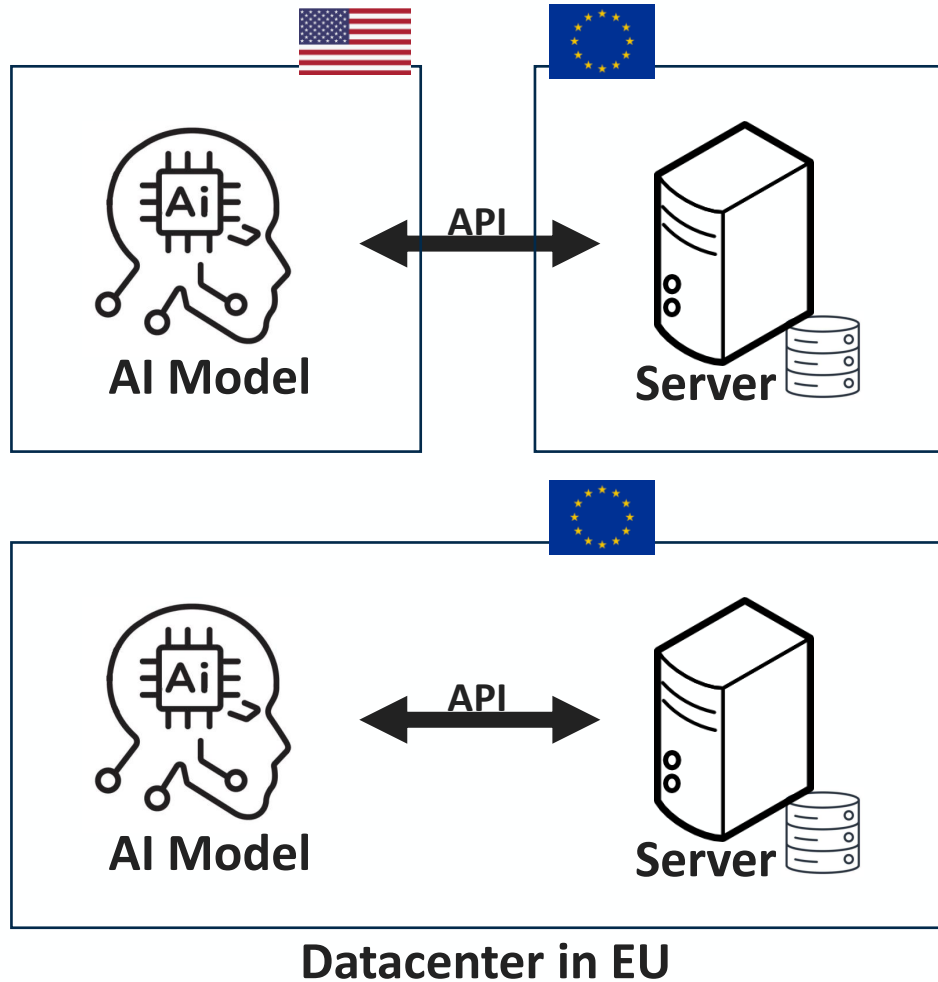
\$0.050 / 1M tokens

Output:

\$0.400 / 1M tokens

Likely 8Billion parameters

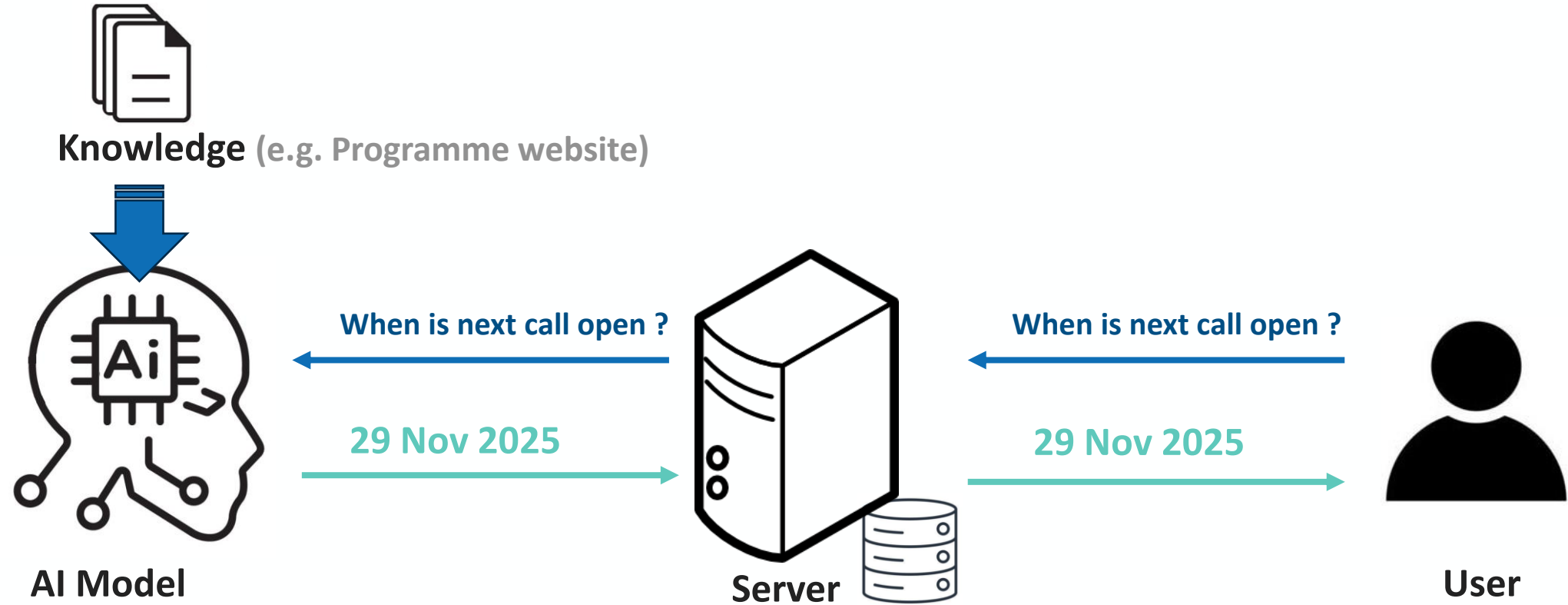
Model location



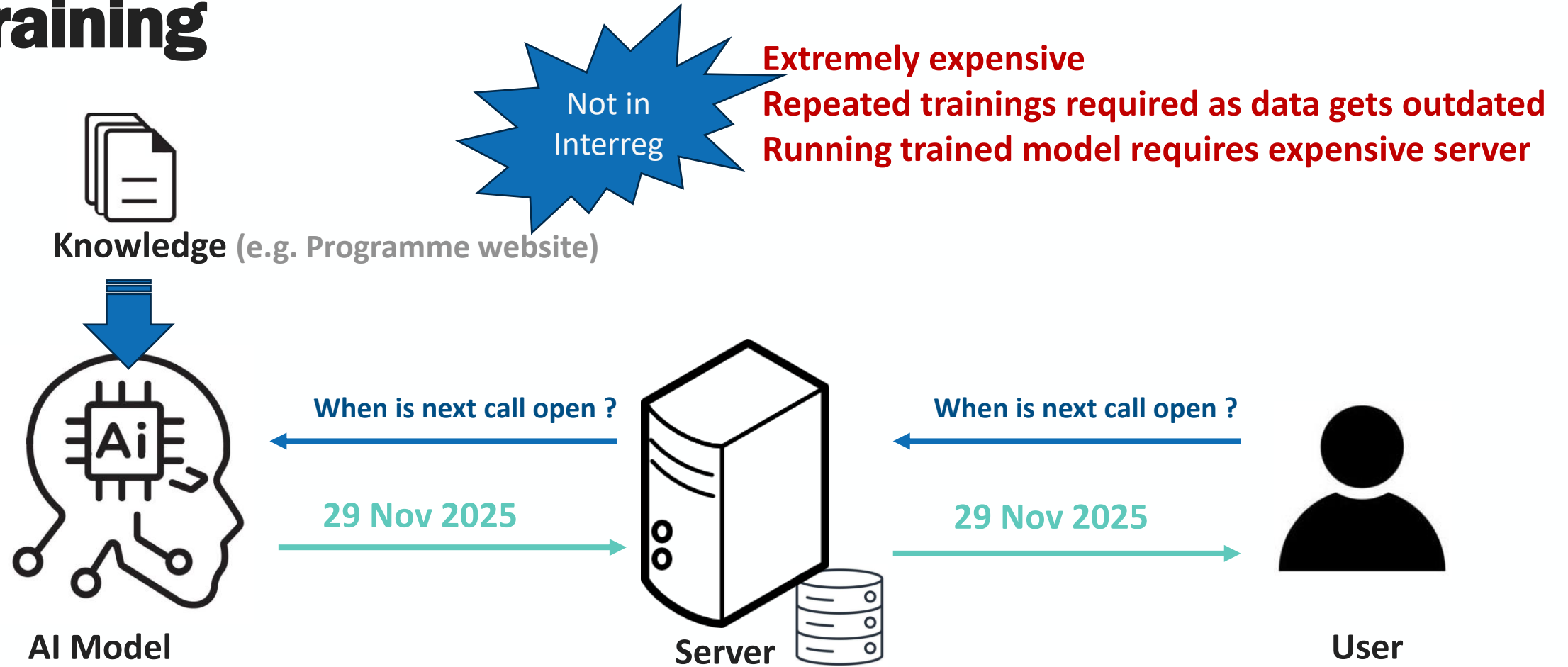
- Latest updated models
- Cheapest price
- **Non-EU data residency**

- **Not latest updated models**
- **~30% expensive price**
- **EU data residency**

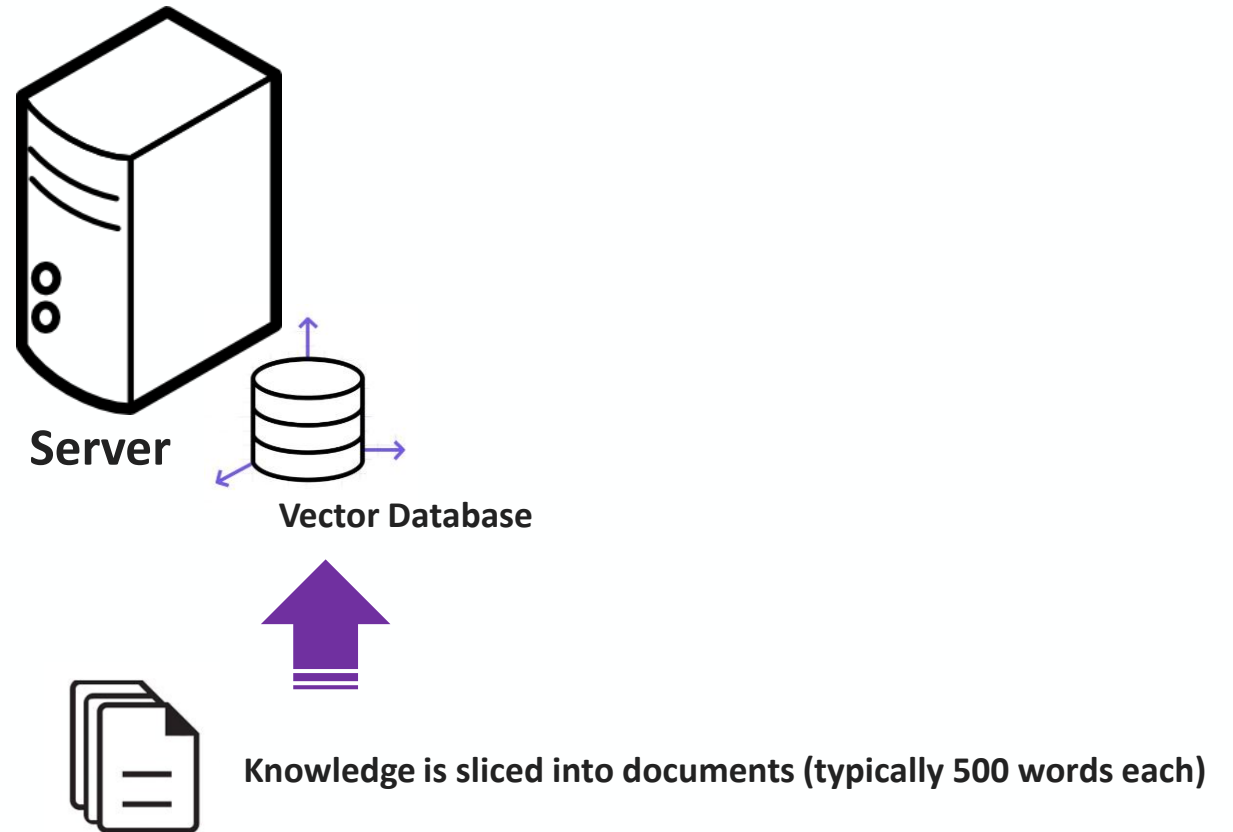
Training



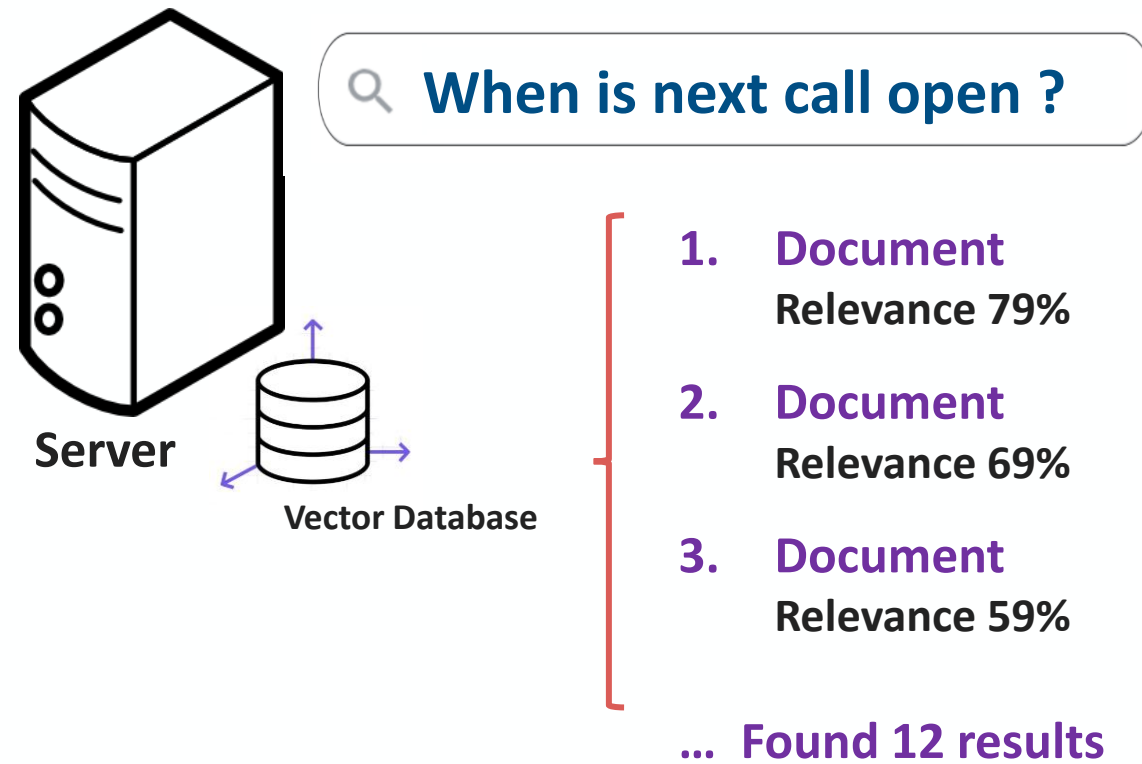
Training



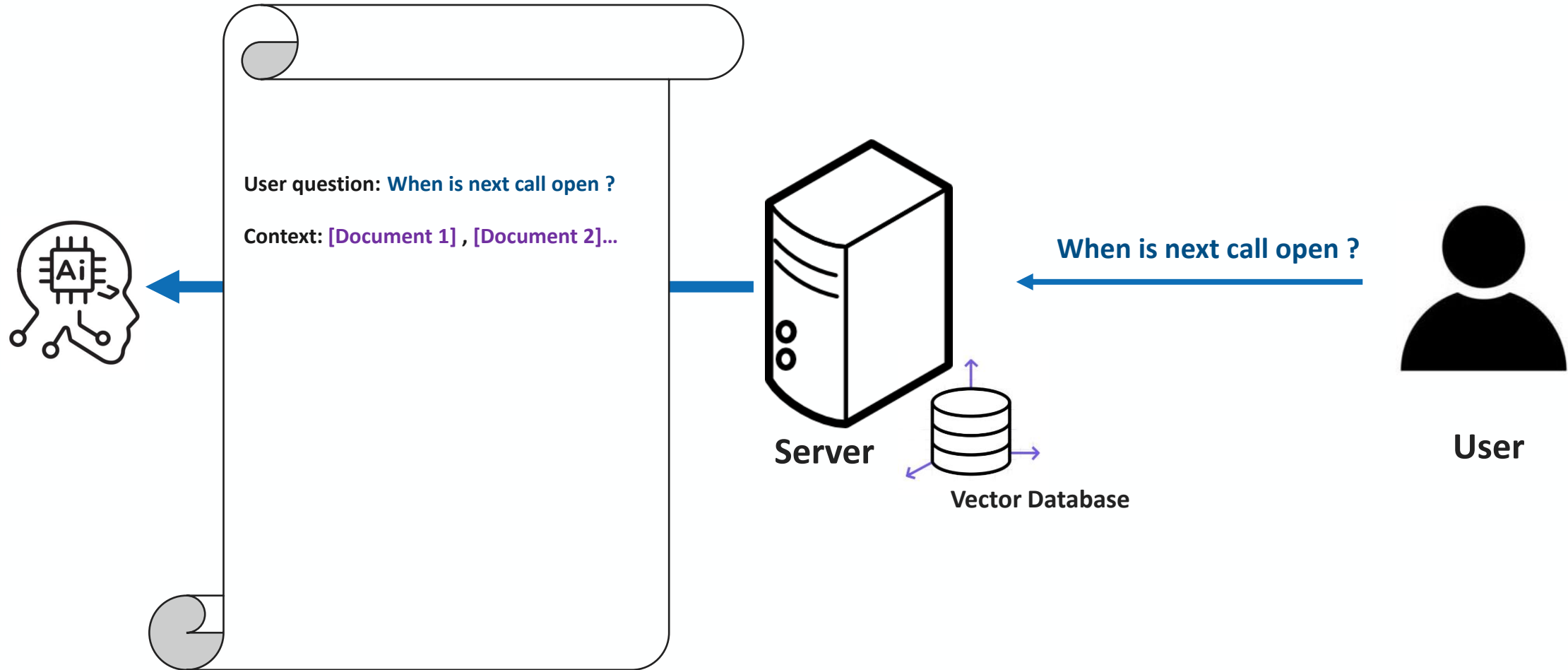
RAG - Retrieval-Augmented Generation



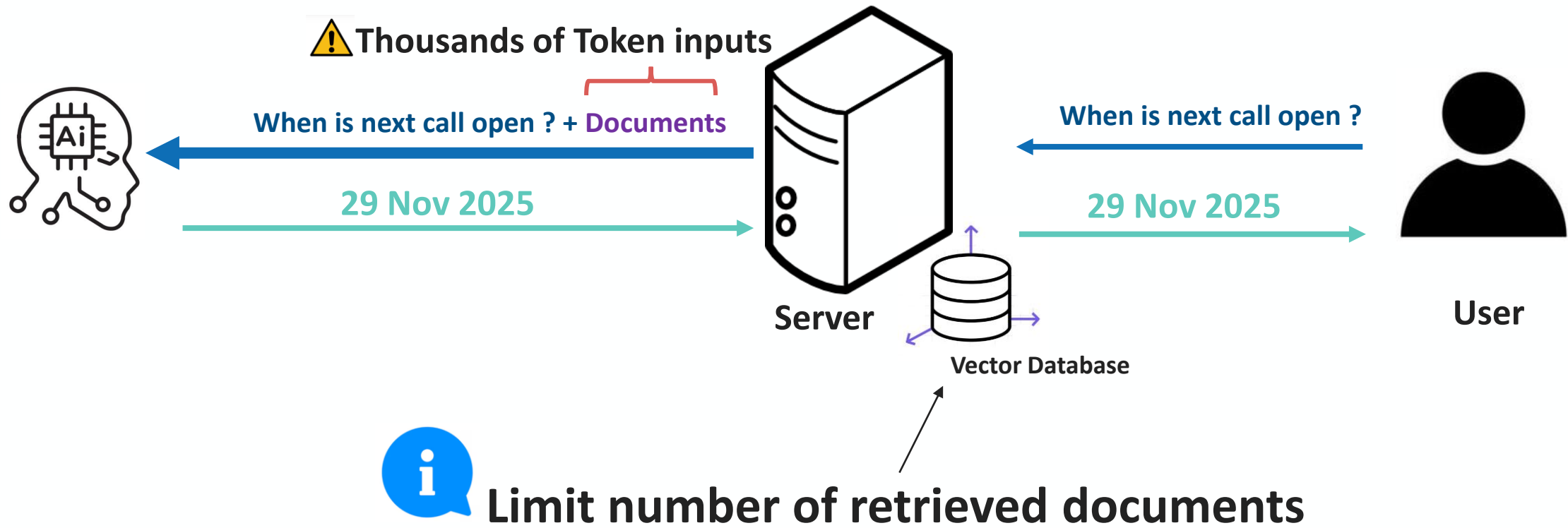
RAG - Retrieval-Augmented Generation



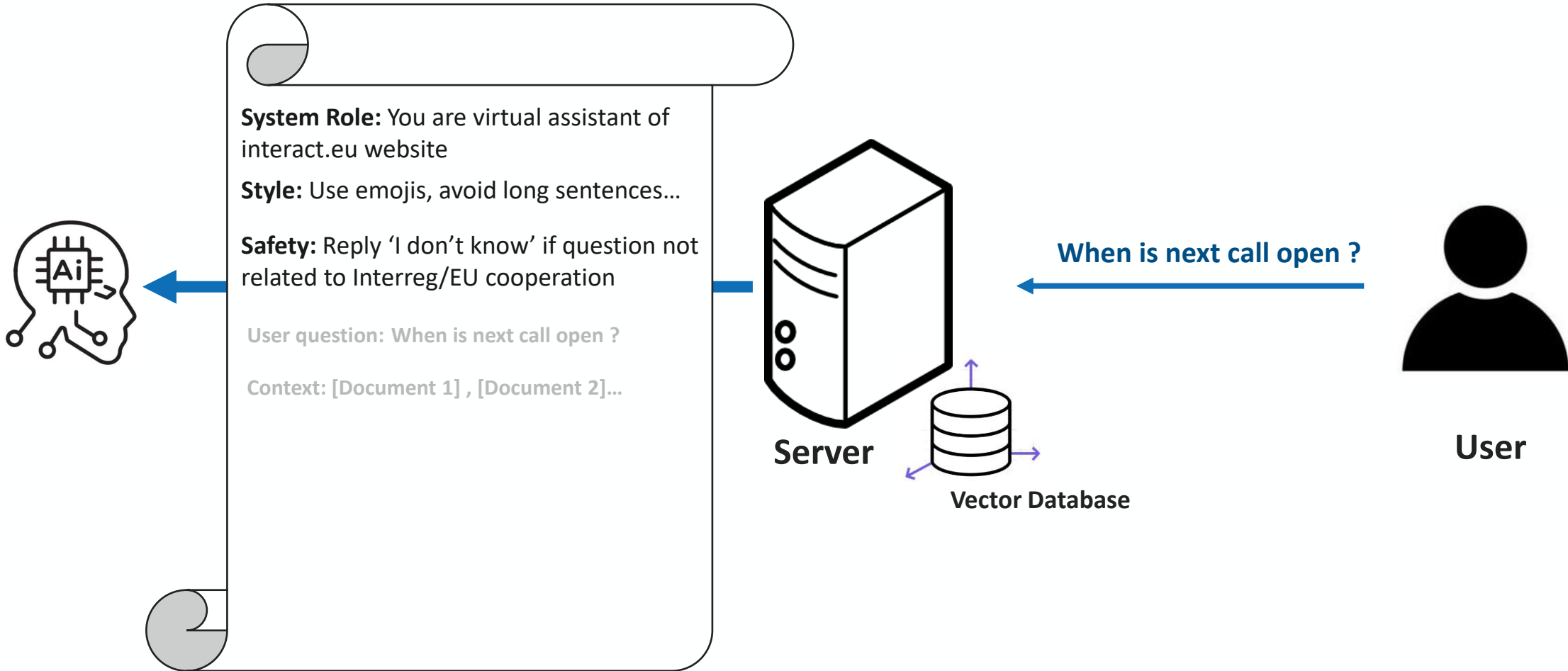
RAG - Retrieval-Augmented Generation



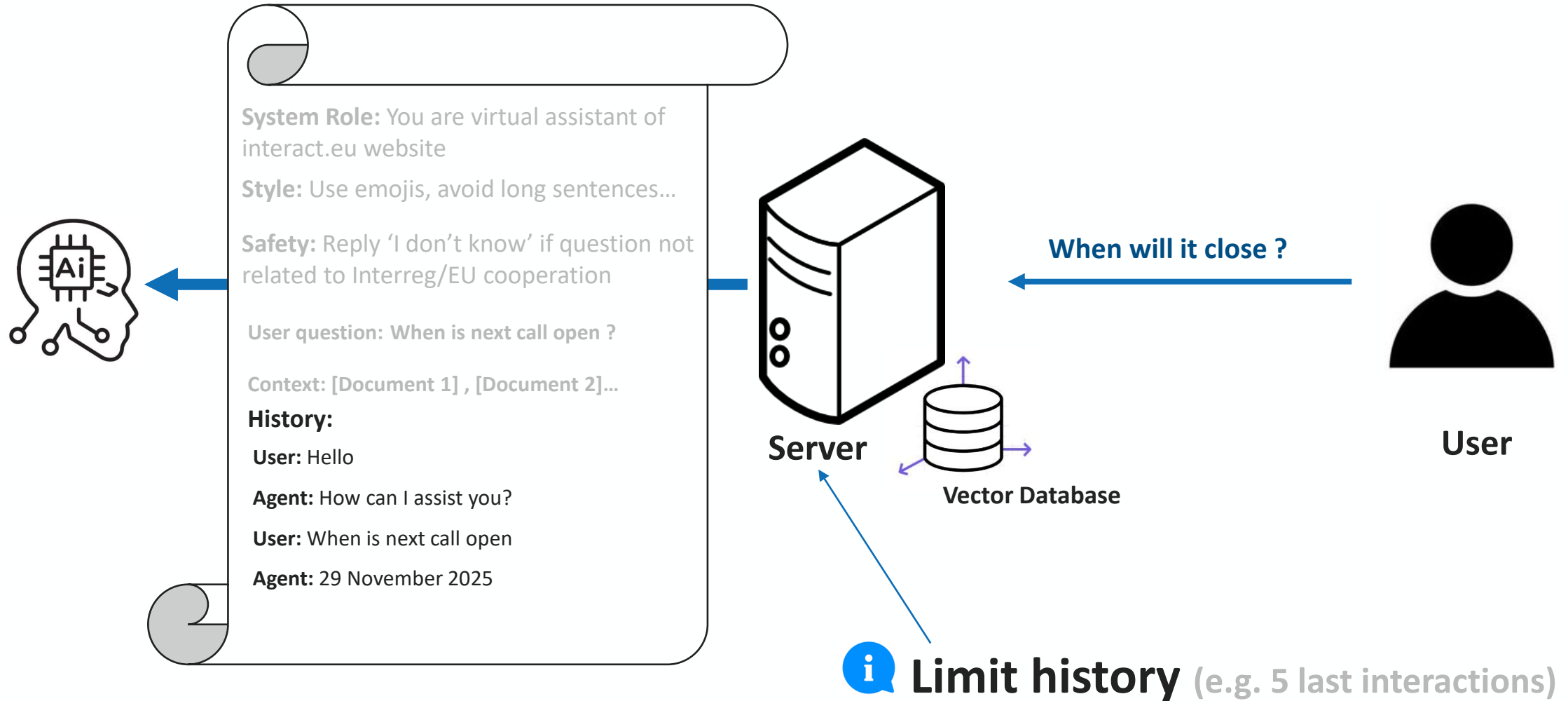
RAG - Retrieval-Augmented Generation



Metadata

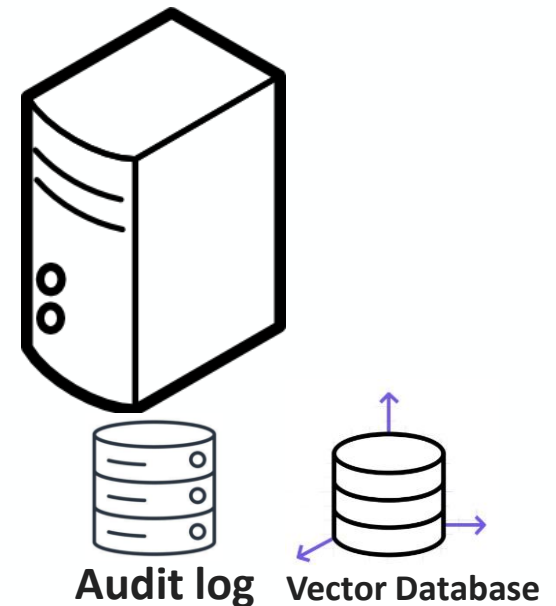


Memory (Short term)



User activity

- **Audit trail** – Database logs all user activity
- Quality control — Improve the chatbot over time
 - Most common questions, where users get stuck..
- Detect errors or harmful behaviour
 - Detect allucinations, incorrect guidance, abuse..
- Analytics
 - Number of questions, impact, time saved,...



Focus on data

- The assistant will generate answers based on the data it sees
 - Ensure data exists, remove outdated/irrelevant/duplicated
- Low-quality, outdated, or inconsistent content → hallucinations and wrong answers
- Incorrect information makes users lose trust quickly.
 - Your chatbot represents your brand



Still on data...

Cooperation can be complex;
our job is to make it easier.

Interact



Co-funded by
the European Union
Interreg

About Interact
& our service ▾

Programme
management ▾

Project management
and skills ▾

Communication
& visibility ▾

Finance
& control ▾

Synergies
& cooperation ▾

Our team

The full Interact team is available to you below, please use the filters to narrow your search by our fields of work.

You can also filter our team based on office location (if you know this), and by the different roles the Interact team fulfil.

Filters ▾



Mercedes Acitores



Rafael Agostinho



Jose Almeida



Monika Balode



Thank you!