

## Lab 13 – Part G

LLMs (Large Language Models) are powerful tools, but they still have some limitations. Below are the main issues and ideas to reduce their impact.

### 1. Stateless API

*Problem:* LLMs do not remember past interactions unless you send the full history every time.

*Solution:* Add conversation memory on the client side. Store messages in a session and send the full context when calling the model.

### 2. Not trained on your data

*Problem:* LLMs don't know your company's private or latest data.

*Solution:* Use techniques like retrieval-augmented generation (RAG) to connect the LLM with your internal documents or databases during runtime.

### 3. Limited size of data you can send

*Problem:* The context window is limited (only a few thousand tokens).

*Solution:* Summarize or chunk your data before sending it. Use tools that manage long documents or split the input smartly.

### 4. Prone to hallucinations

*Problem:* LLMs sometimes make up answers that sound correct but are wrong.

*Solution:* Use fact-checking or ask the model to rate its confidence. For critical tasks, double-check results or compare across models.

### 5. Not aware of your APIs

*Problem:* LLMs don't know how your systems or APIs work by default.

*Solution:* Provide API documentation in the prompt or fine-tune a model with your API instructions. You can also guide the model step by step.

### 6. Not aware of real-time data

*Problem:* LLMs can't access live data like current stock prices or news.

*Solution:* Combine the LLM with real-time API calls. First get the data, then feed it into the prompt.