Day 6 In-Class Exercises: Embeddings and the Frontier

Student guide. Run each lab step in order and predict before you run. Formative, not graded.

What to run in the labs

Morning: embeddings notebook

- PREDICT which pair of companies will have the highest cosine similarity.
- Encode the six earnings snippets.
- Build the cosine-similarity matrix.
- WRONG WAY: compare with the raw dot product; see how vector length distorts it. Then use cosine similarity.

Afternoon: RAG notebook

- Watch the instructor demo: retrieval over three toy sentences (embed, score by cosine, pick the closest).
- PREDICT, for the query "revenue growth", which snippets get retrieved, and sketch the final prompt.
- WRONG WAY: dump all chunks into the prompt; measure how long and off-topic it gets.
- The fix: retrieve only the top three chunks and cite them; compare the grounded answer with the model answering from memory (discussion).

Discussion exercises

EX-D6-1. Attention by hand. Sentence: "The bank raised its rate after the central bank's decision."

- (a) For the word "rate", which other words should attention weight most, and why?
- (b) What tells the model the first "bank" is a commercial bank and the second is a central bank?

EX-D6-2. Will an LLM alone get this right? Sort each into: LLM alone / needs RAG or a tool / plain script.

- (a) summarize this earnings-call transcript (text given).
- (b) Apple's exact revenue in last week's results.
- (c) compute the 30-day volatility of these returns.
- (d) draft a polite email declining a meeting.
- (e) is this filing's tone bullish or bearish (text given).

EX-D6-3. Design a finance agent, then break it. For a "quarterly revenue growth" agent:

- (a) List its think-act-observe tool-loop steps.
- (b) Name one step where a silent error would corrupt the final number.
- (c) Name one guardrail you would add.