

Fund Finance AI session

2026-05-06

About me

- **Alexandros Sivris**
- Freelance product engineer
- Building AI-driven enterprise software solutions (last 3 years with LLMs)
- Find me on LinkedIn



alexandros-sivris

alex@sivris.com · sivris.com

Outline

1. AI assistant basics
2. Hands on (part 1)
3. 🍕
4. Hands on (part 2)
5. Closing



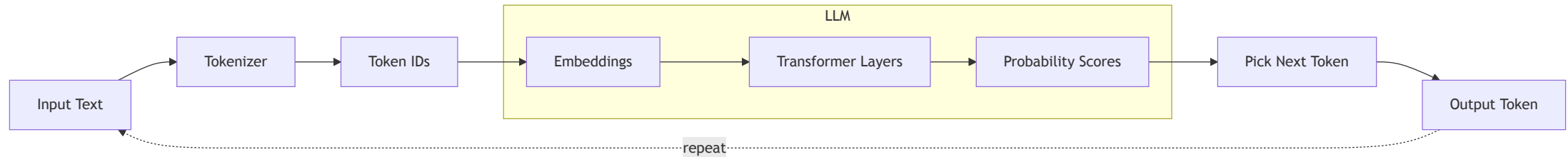
What is an AI Assistant?

- Foundation model (LLM) + system prompt + interaction interface + orchestration (context window, memory, session state)
- Examples:
 - Claude (app/code/cowork)
 - ChatGPT/Codex
 - OpenWork



What is an LLM?

- System trained on huge amount of text to predict the next token



- Input is split into pieces (**tokens**), e.g. What | is | fund | finance | ? → model picks most likely next token → repeats
- Trained on enormous corpora: GPT-3 on ~ 300B tokens, Llama 3 on 15T+ tokens
- Predicts text — does not "know" or "verify" facts

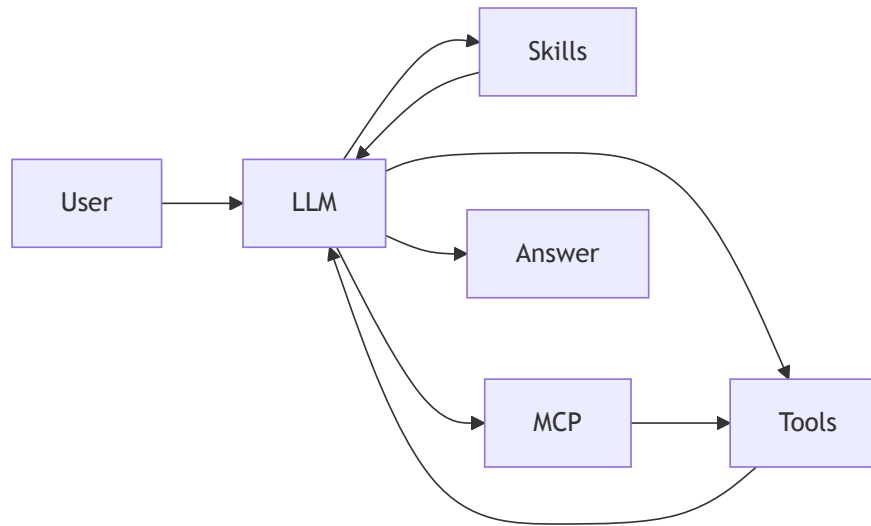
How is an LLM trained?

- Reads billions of text examples, guesses the next token, corrects itself, repeats
- Example: What is fund finance? →
 -  Fund finance is a type of lending... — high score → reinforce these weights
 -  Fund finance is a banana... — low score → adjust weights away from this
- Millions of examples, millions of corrections → model gets better at continuing any text

From LLM to AI assistant

- LLM alone = generates likely text
- Add instructions + documents + tools → can do real work
- Read, calculate, draft, format, send — combined into one workflow
- Concrete example shown later

🔌 MCP, Skills and Tools



LLM: "understands", "reasons", and generates responses

Skills: predefined instructions or workflows

Tools: external functions or APIs the LLM can use (e.g. reading or writing files)

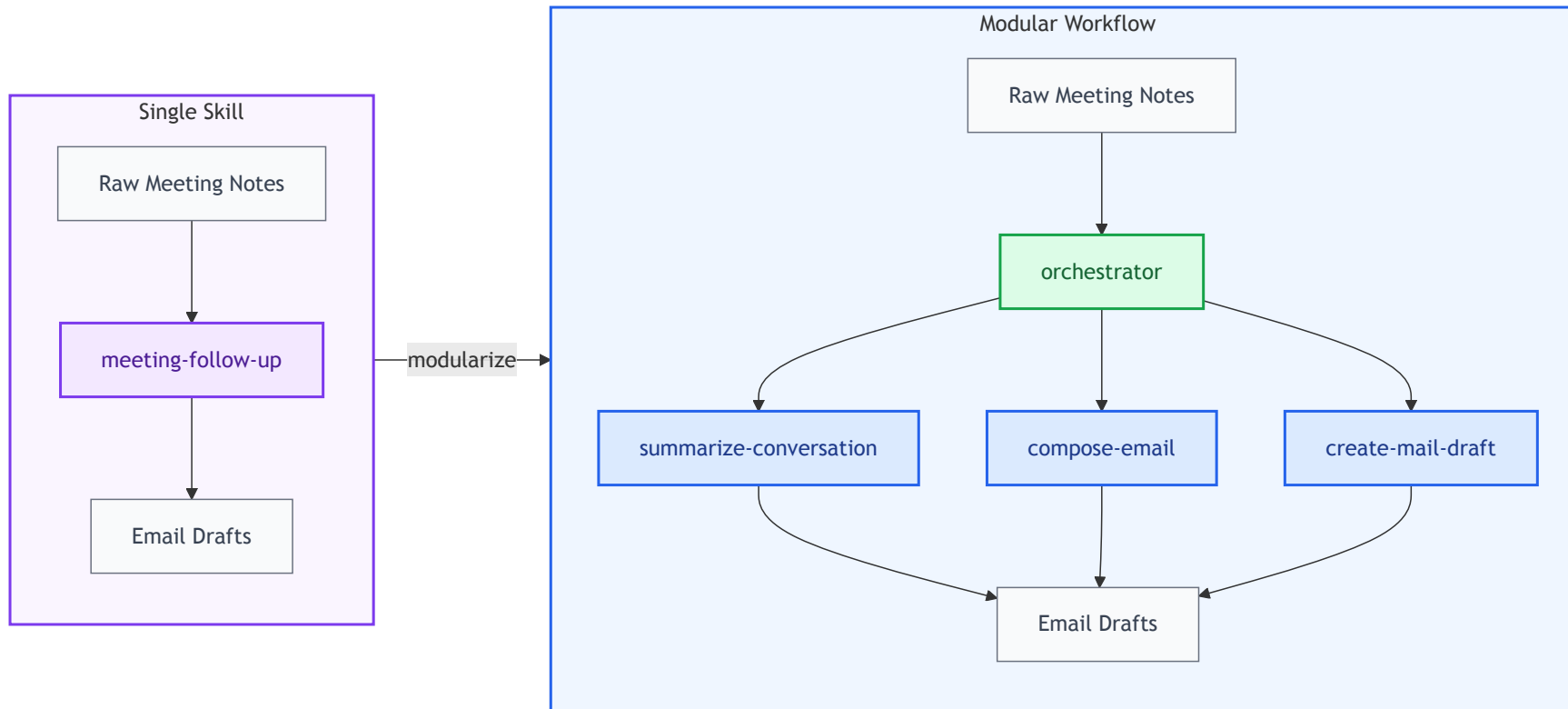
MCP: standard interface connecting LLM to tools/data (e.g. Chrome DevTools MCP server)

How to create AI automations?

1. Pick a workflow that is repetitive and time-consuming
2. Break it into a few small steps with well-defined inputs and outputs
3. Plan, build, and test
4. Iterate

💡 Simple example

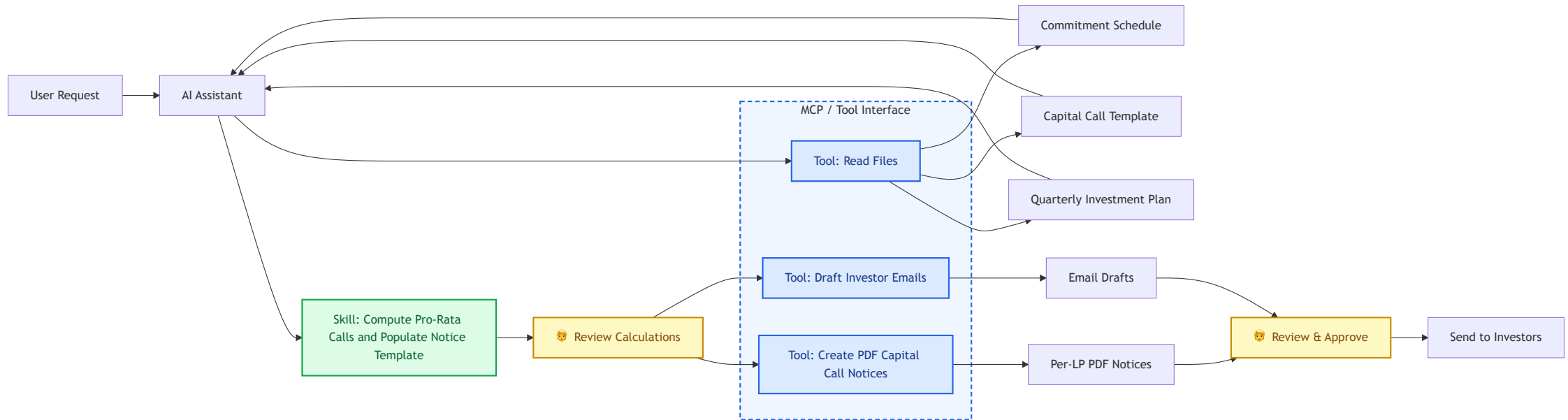
After a call you have meeting notes which you want to summarize, create follow-up action items and send those action items to the attendees.





Complex example

Read the fund's commitment schedule, the capital call template, and this quarter's investment plan; calculate each LP's pro-rata call amount; generate personalized capital call notices as PDFs and draft cover emails to each investor.



Hands-on: What workflows would you automate?

- **Goal:** Identify and sketch a workflow you would like to automate
- Think of a task you do repeatedly that is time-consuming
- What are the inputs? What is the expected output?
- Could an AI assistant handle it — fully or partially?
- It doesn't have to be end-to-end — automating **one step** and reviewing the rest yourself is already a big win
- Where would you want a human in the loop?
- Which available MCPs, skills, APIs, etc. can you use?

Limits: hallucinations

- LLMs predict likely text, not verified facts
- Confident answers can be **wrong** – especially numbers, citations, dates, legal text
- Mitigations:
 - Ground in real documents (LPA, side letter, NAV report)
 - Ask for source quotes + page numbers
 - Human-review high-stakes output

Limits: data privacy

- In many AI setups, your data leaves your machine and is processed by a third-party or external service
- Ask before using a tool:
 - Where is data stored? Used for training? Retention policy? Enterprise tier available?
- Rules of thumb:
 - Enterprise tier for client data – never consumer accounts
 - Redact names/IDs when prototyping
 - Check internal policy + NDAs before uploading

Limits: cost & efficiency

- LLMs charge per token (input + output) → long docs = more cost + slower
- Levers to control cost:
 - **Pick the right model tier:** small/fast (Haiku) for summaries & extracts, balanced (Sonnet) for most workflows, top (Opus) for hard reasoning. Smallest that handles the task.
 - **Shape output:** skills can compress responses (e.g. *caveman* skill → ~75% fewer output tokens, same substance)
 - **Don't paste everything:** retrieve relevant chunks (RAG) instead of full 200-page LPA
 - **Summarize first,** then operate on the summary

Hands-on: build your first skill

1. Open Claude Cowork
2. Pick one workflow from the previous discussion
3. Chat with the assistant until it produces the result you want
4. Save it as a skill — so you (or anyone) can reuse it with one click
5. Run the skill again — does it produce consistent results?
6. Bonus: share your skill with a neighbour and have them try it

Thank you

Questions?



alexandros-sivris

alex@sivris.com · sivris.com