

Lessons Learned from Hosting a Frontier AI and LLM Tutorial Series for a Mixed Audience

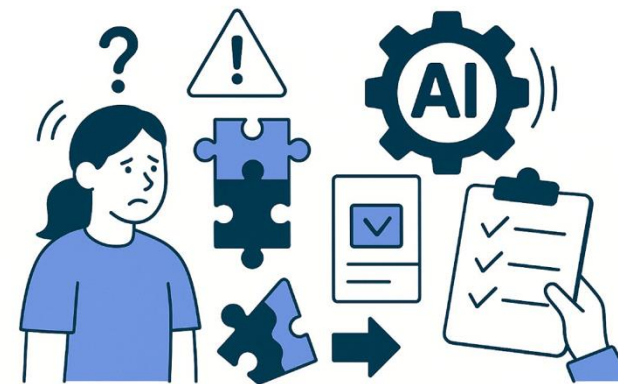
Meifeng Lin, Kriti Chopra, and Mohammad Atif
Computational Science Department,
Computing and Data Sciences

Why a tutorial series — and what it covered

Rationale

Rapid, uneven adoption

LLMs are already in active use across science, engineering, and operations at BNL, code development, idea generation, writing assistance. Adoption is uneven across staff, and growing fast.



Sourced: <https://www.worklytics.co/blog/top-ai-adoption-challenges-and-how-to-overcome-them>



Sourced: <https://www.lakera.ai/blog/guide-to-hallucinations-in-large-language-models>

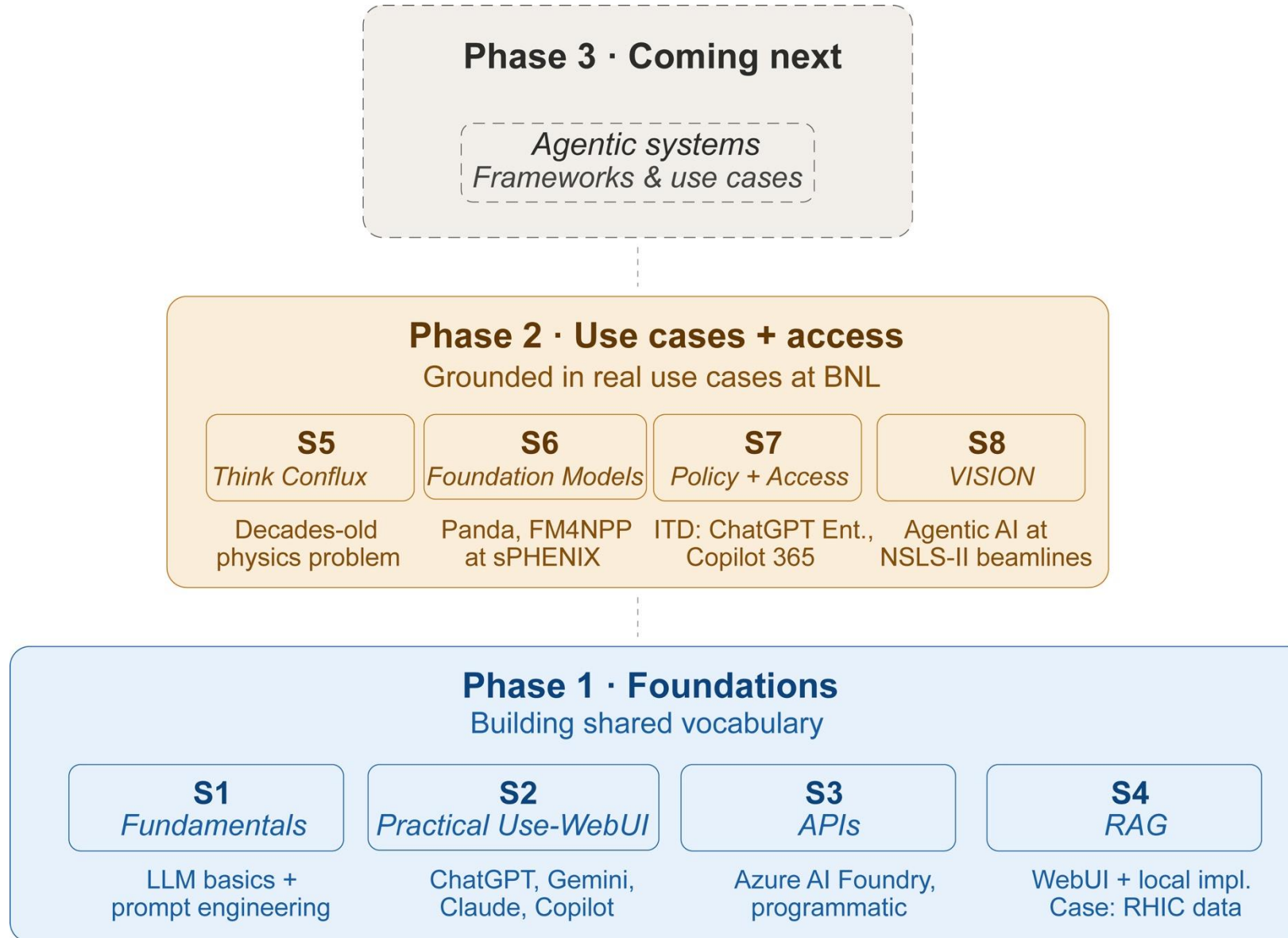
Real productivity gains, real risks

We've seen meaningful productivity gains, but the risks are poorly understood: hallucinations, data leakage, false confidence, safety and security concerns.

Genesis Mission

DOE's Genesis Mission aims to transform AI for science, and LLMs are becoming a crucial part of that infrastructure. Staff need to know how to use them effectively *and* responsibly.

Course structure : evolved in response to participant interest



A practical toolkit for using Frontier AI and LLM models

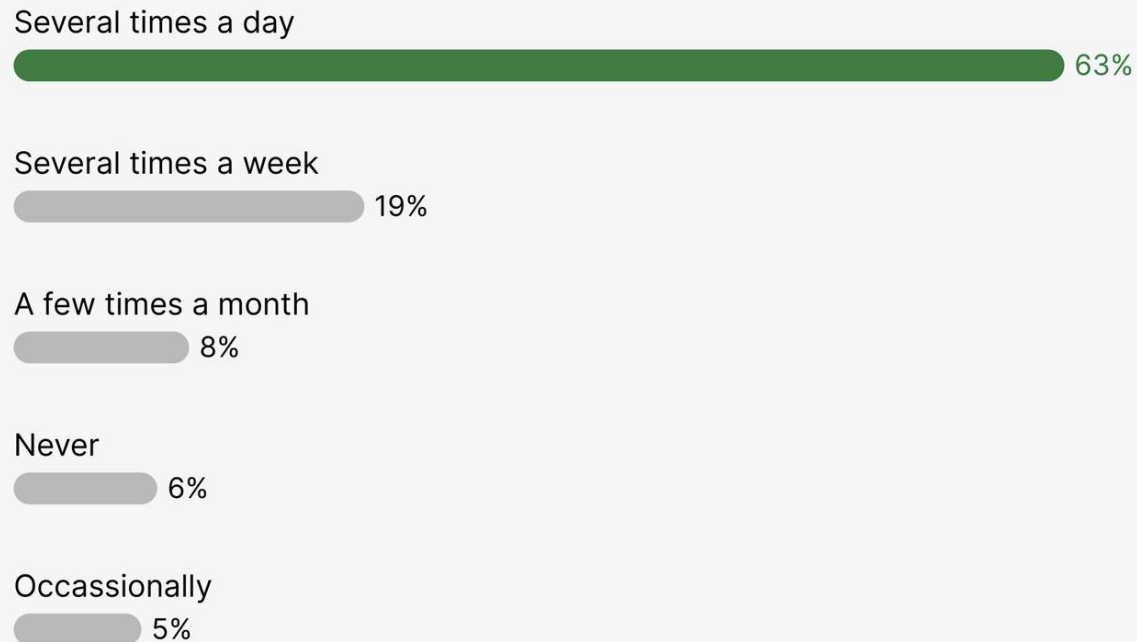
- | | | |
|----------|---|---|
| 1 | Prompt Engineering
<i>Highest leverage, lowest barrier</i> | Structured prompting, TRACE framework, role/format/context. |
| 2 | WebUI chatbot demos
<i>Practical applications and failure modes</i> | Showcasing frontier model WebUI interfaces. General settings and best practices were demonstrated using practical demos |
| 3 | Retrieval-Augmented Generation
<i>Grounding in your own data</i> | The practical answer to "can I use this on my documents?" Demonstrated with RHIC data preservation. Hands-on with Python + Colab. |
| 4 | API Integration
<i>Beyond the web UI</i> | Azure AI Foundry access to frontier models. For users ready to embed AI into pipelines, scripts, and workflows. |
| 5 | Institutional Context
<i>Which tool, which data, which policy</i> | Knowing what's sanctioned (Copilot 365, ChatGPT Enterprise via ITD, lab Azure) — and what's not. |

Participant Feedback: Rapid, uneven adoption!

☑ Active poll

102 88

How often do you use LLMs?



☁ Active poll

84

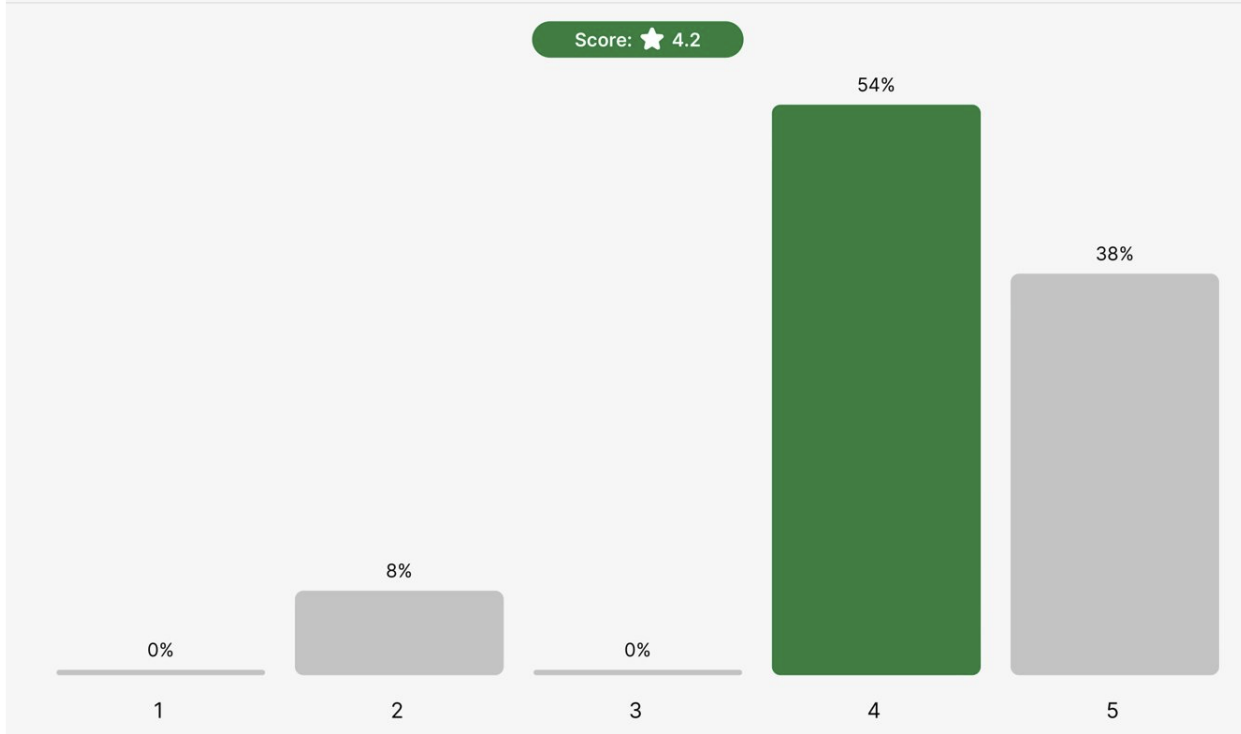
What are your main concerns or challenges with using LLMs?



82% of attendees use LLMs at least weekly — and their top concerns are trustworthiness, security, and how to use them effectively

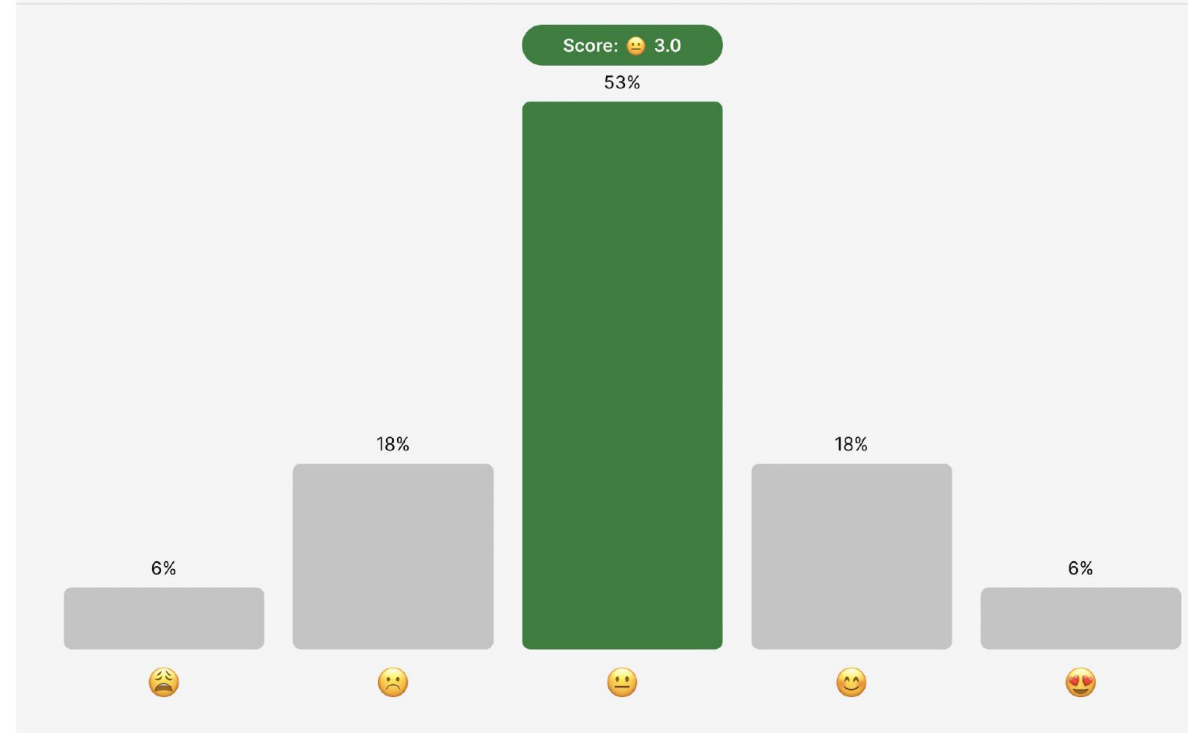
Participant Feedback: Fluency is not correctness!

How well did the LLM work for the easy to intermediate tasks?



Easy / intermediate tasks: Room utilization from CSV · Meeting notes → action items · Batch file rename script

How well did the LLM work for the hard tasks?



Hard tasks (*rated 3.0*) Multi-omics dataset integration · Debugging a subtle numerical bug · Literature review with verifiable citations

Reliability tracks complexity, not domain. The harder the task, the wider the trust gap

Participant Feedback: From users to builders

S1 · "What is this?"	S2 · "Does it work?"	S3-4 · "How do I deploy this?"
<i>"What AI tools are available to us as members of the lab?"</i>	<i>"Are we supposed to trust this on the first try?"</i>	<i>"How do I match model choice to available hardware?"</i>
<i>"How can I use specialized models like BioGPT?"</i>	<i>"What data are we allowed to provide to the LLM?"</i>	<i>"What key management tools are recommended?"</i>
<i>"How is this more efficient than a Google search?"</i>	<i>"How does it choose which model to use?"</i>	<i>"Techniques for ingesting our arcane data formats?"</i>



Access & policy ·

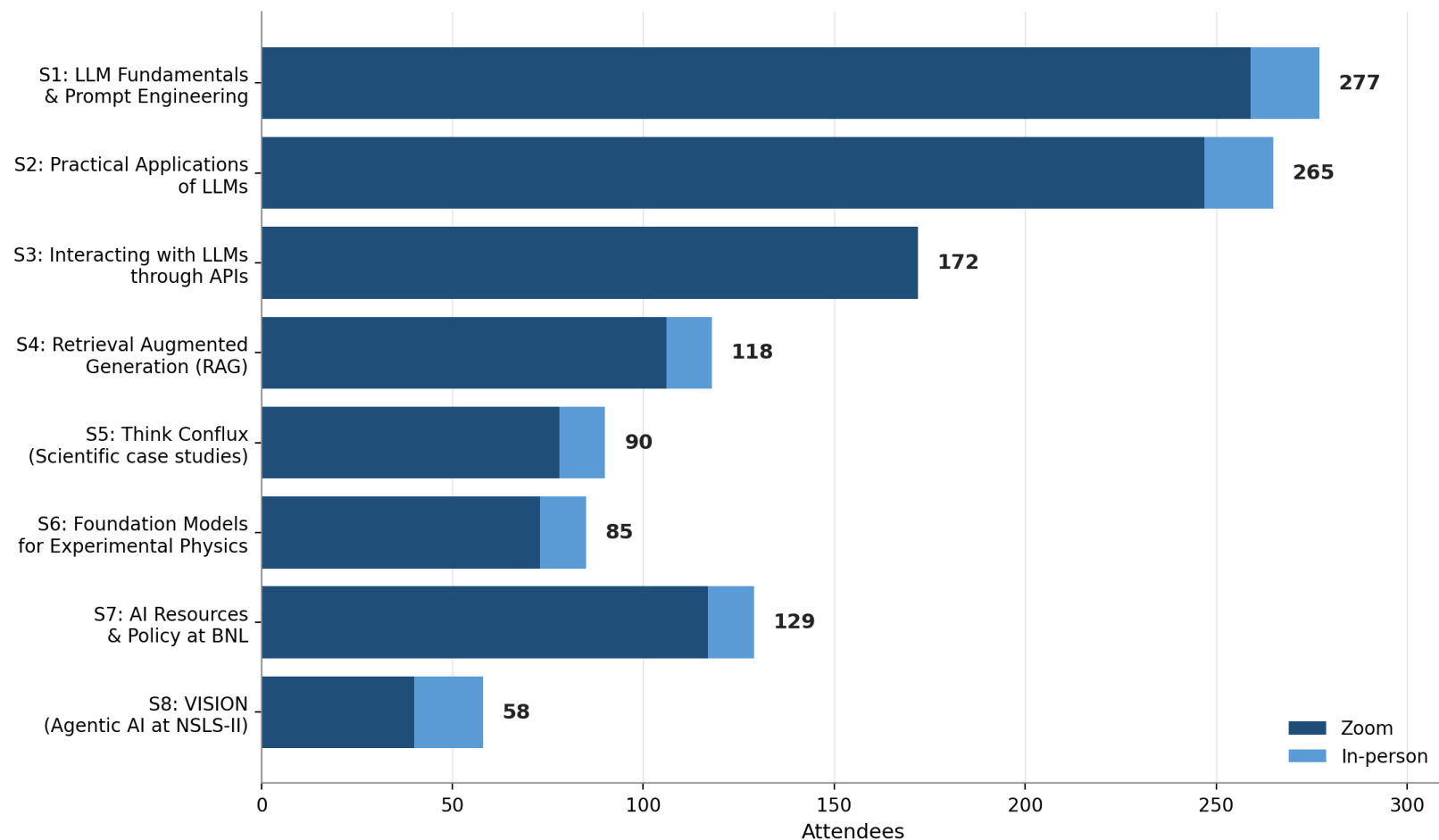


Productivity & verification ·



Applying it to my work

Attendance: strong sustained interest from a mixed audience



571

unique Zoom attendees across the series — ~12% of BNL staff reached.

~300

attendees at S1 and S2 — fundamentals + practical use

Operational + scientific staff both well represented

50+ min

median Zoom duration on a 60-min slot

What worked and What can be improved

What Worked

Accessible content- Slides shared through Indico and Video recordings made available

Interactive format Live Slido polls, demos, and real-time Q&A throughout each session.

Office hours Held between technical jumps (post-APIs, post-RAG) to catch people up.

Calendar invites Helped increase participation

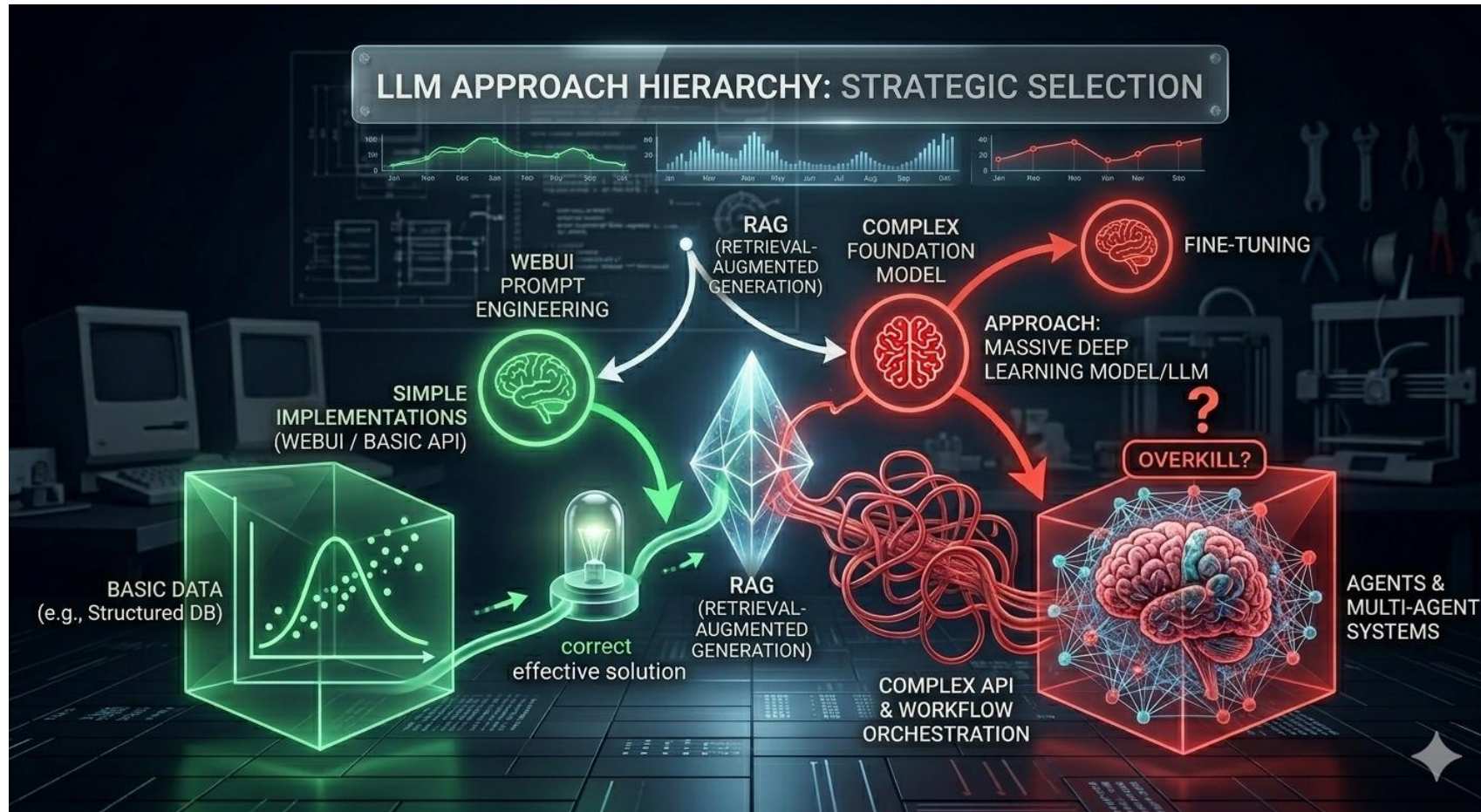
What to Improve

More hands-on time Attendees felt rushed during follow along demos

Local Python environment setup Not equally easy for everyone

Balanced content for scientific vs operational audience Attendance attrited as content technicalized

Clearer cost guardrails and policy guidance



Created with Gemini Nano banana

Building **AI fluency** means more than building trust.

It means building judgment about which approach fits which problem, and when a simpler one would have done.