

Passion for Innovation.
Compassion for Patients.™



Upgrading Query Chat into a Verifiable Analytics Experience

Ding Cheng

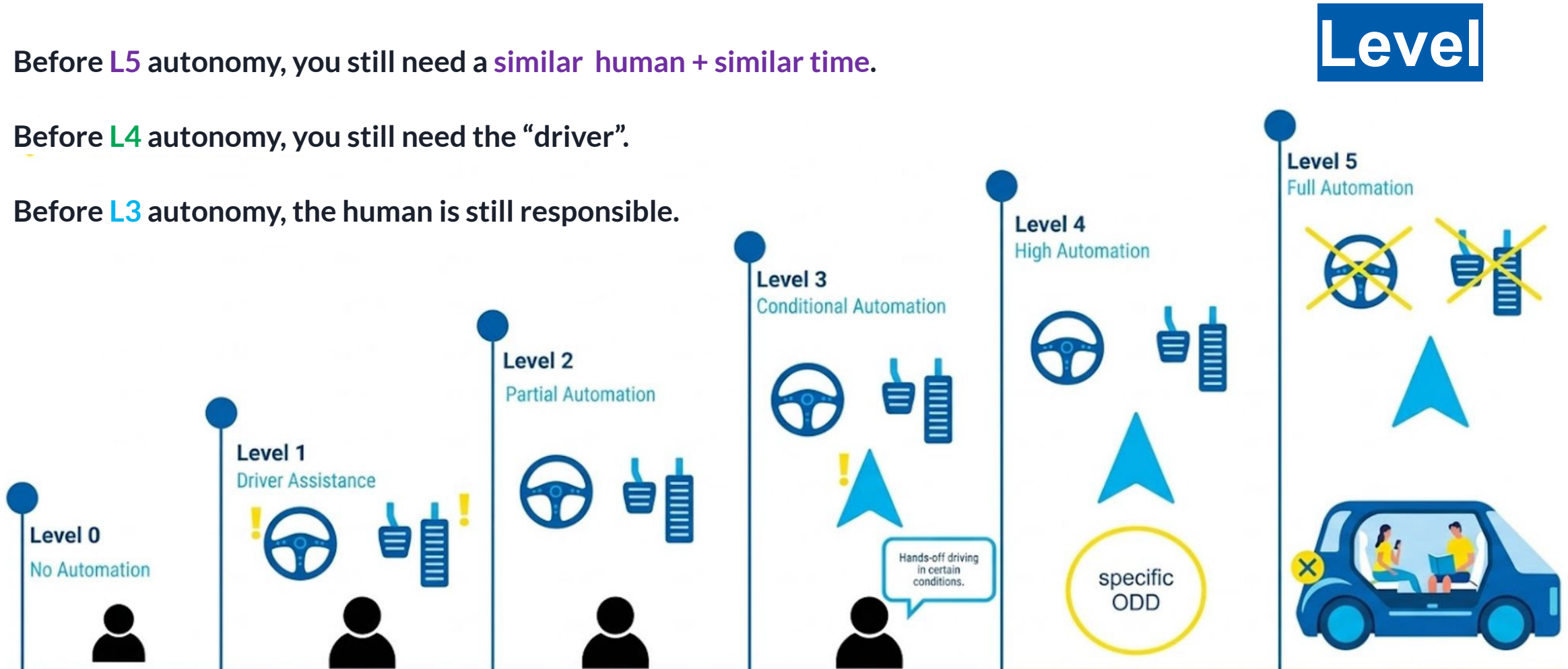
2026-03-26

What exactly are we trying to use LLM for?

Before **L5** autonomy, you still need a **similar human + similar time**.

Before **L4** autonomy, you still need the “driver”.

Before **L3** autonomy, the human is still responsible.





What exactly are we trying to use LLM for?



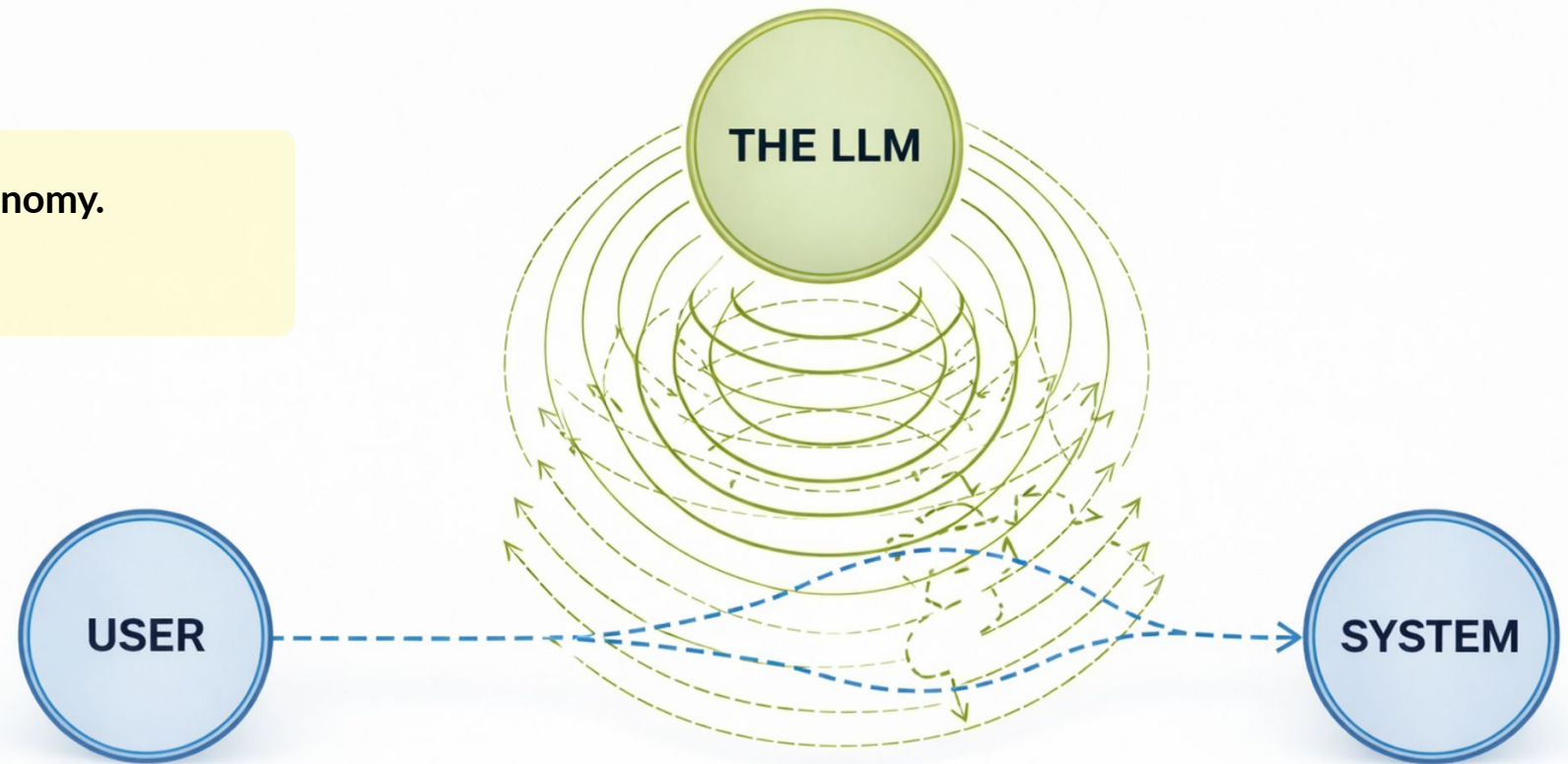
Scene



The Three-Body Problem of LLM

-  Adding an LLM means adding a third actor: powerful, flexible, but unstable.
-  Semantic failures do not throw exceptions. They return plausible but wrong answers.

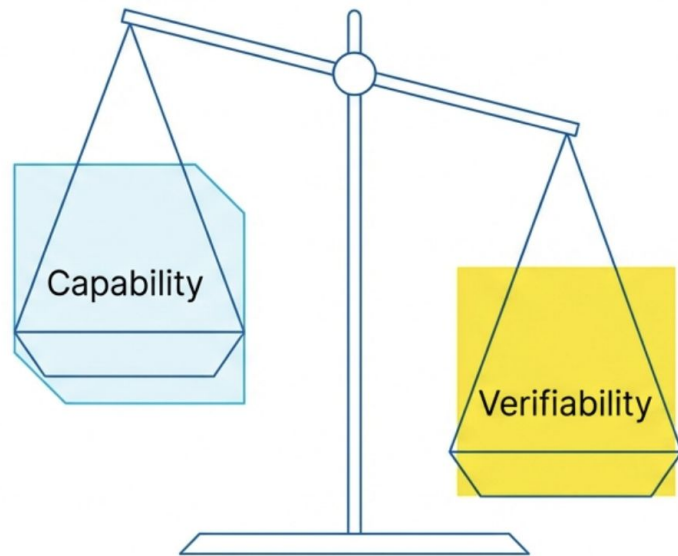
The design goal is not maximum autonomy.
It is verifiable assistance.



The Other Side of Alignment

Alignment = Capability × Direction × Interpretability

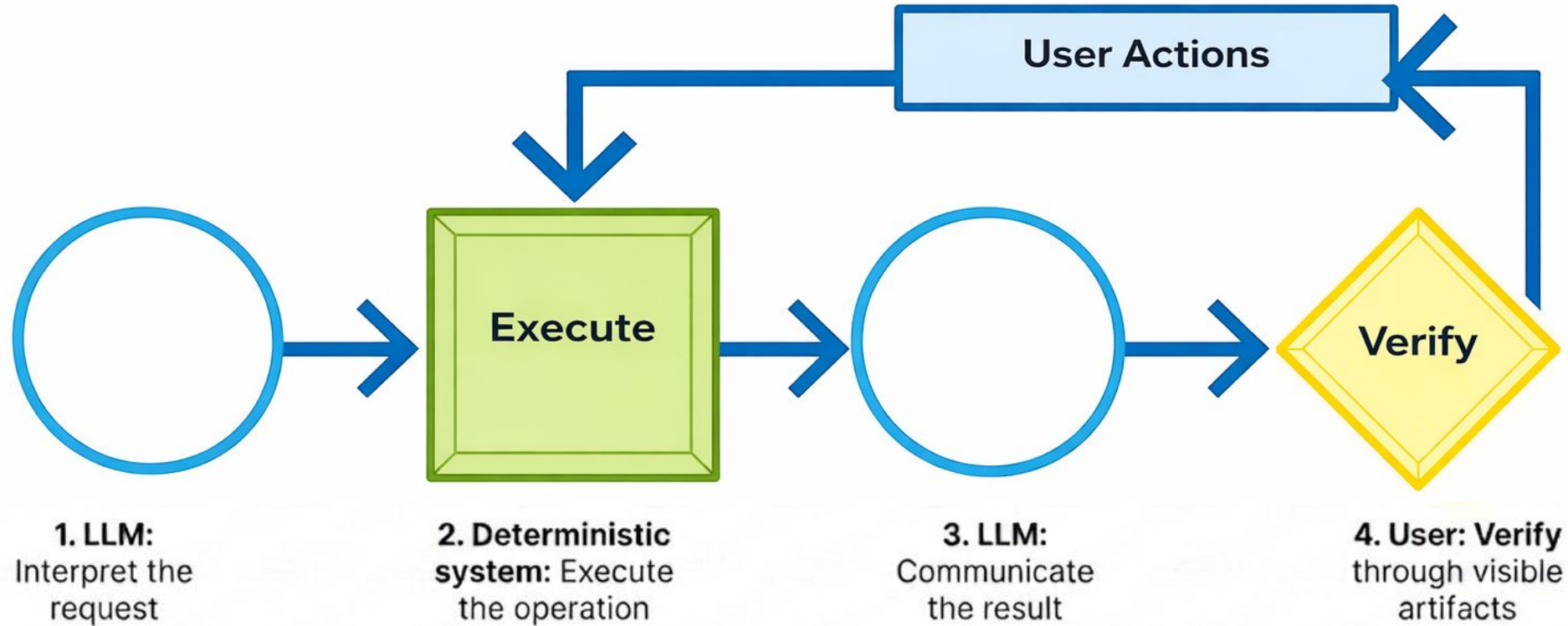
- Not just doing things well
- But also doing them the way we want
- And in a way we can understand



Can the LLM understand the user's intent?

Can the user understand what the LLM just did?

Proposed Basic Workflow



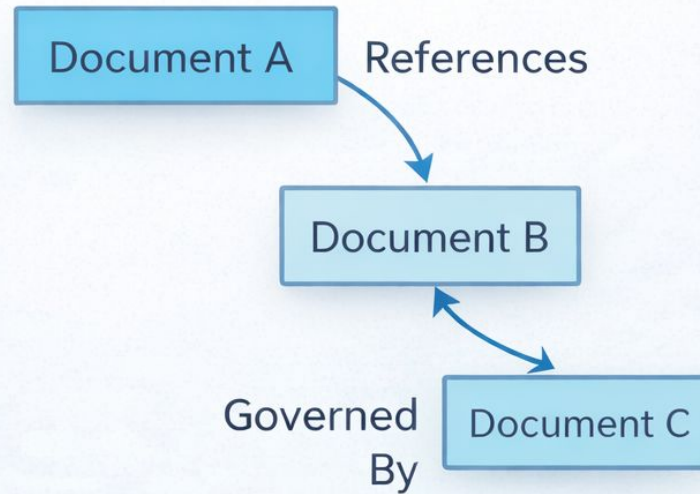
Example 1: Tabular Data Dashboard

Why text-to-SQL Struggles?

RAG



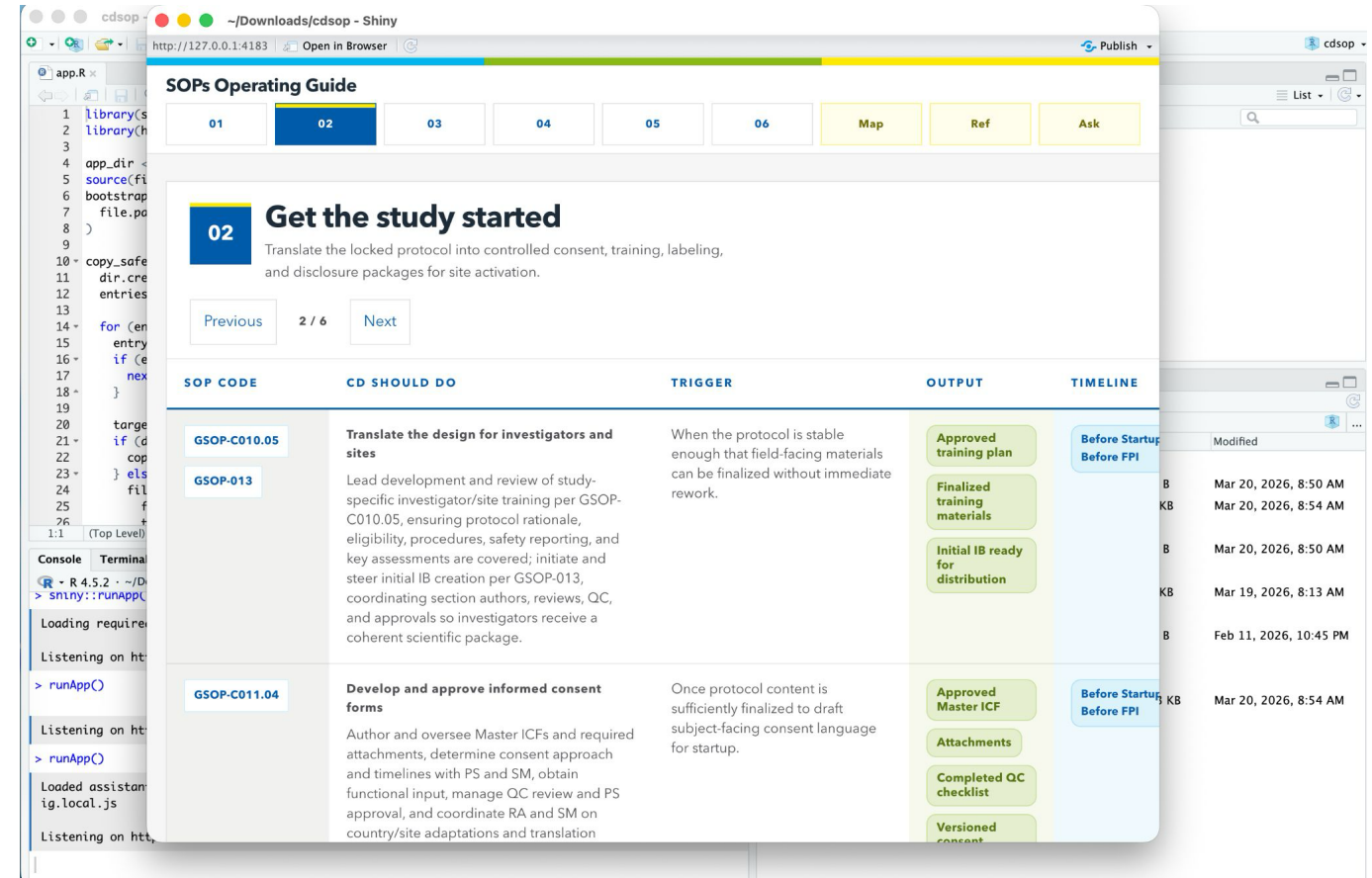
Document Relationships



Example 2: Document View

Why Not Just RAG?

- **Interpretation Layer:** Plan the search and decide what to retrieve next.
- **Execution Layer:** Fetch documents, follow references, extract structure.
- **Communication Layer:** Explain the answer with citations, scope, and uncertainty.



The screenshot displays a Shiny web application interface for an SOPs Operating Guide. The browser address bar shows the URL `http://127.0.0.1:4183`. The application title is "SOPs Operating Guide". A navigation bar at the top contains tabs for sections 01 through 06, along with buttons for "Map", "Ref", and "Ask". Section 02, "Get the study started", is currently selected. Below the section title, a brief description reads: "Translate the locked protocol into controlled consent, training, labeling, and disclosure packages for site activation." Navigation buttons for "Previous" and "Next" are visible, with "2 / 6" indicating the current page. A table with the following columns is displayed: "SOP CODE", "CD SHOULD DO", "TRIGGER", "OUTPUT", and "TIMELINE".

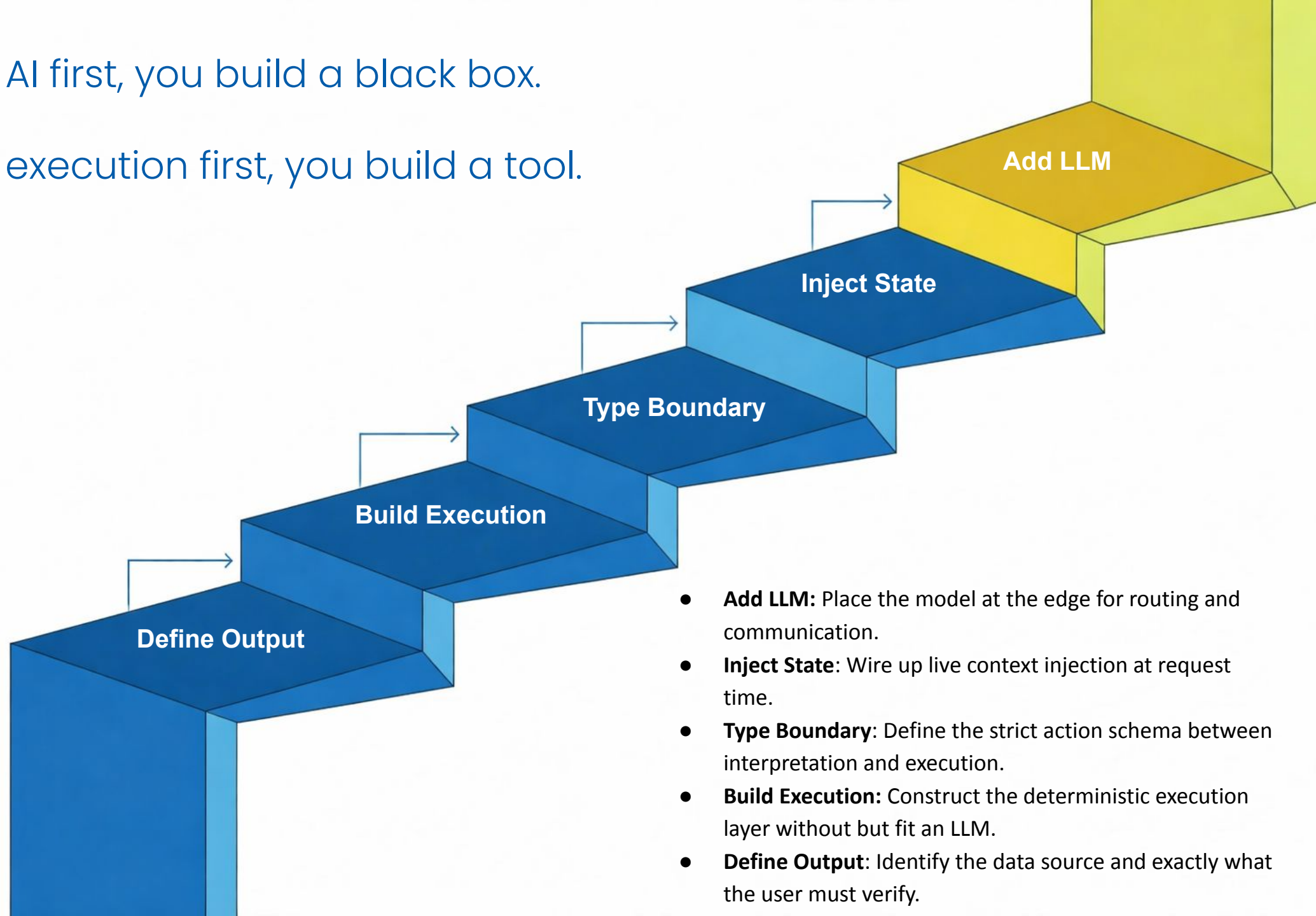
SOP CODE	CD SHOULD DO	TRIGGER	OUTPUT	TIMELINE
GSOP-C010.05	Translate the design for investigators and sites	When the protocol is stable enough that field-facing materials can be finalized without immediate rework.	Approved training plan	Before Startup Before FPI
GSOP-013	Lead development and review of study-specific investigator/site training per GSOP-C010.05, ensuring protocol rationale, eligibility, procedures, safety reporting, and key assessments are covered; initiate and steer initial IB creation per GSOP-013, coordinating section authors, reviews, QC, and approvals so investigators receive a coherent scientific package.		Finalized training materials Initial IB ready for distribution	
GSOP-C011.04	Develop and approve informed consent forms	Once protocol content is sufficiently finalized to draft subject-facing consent language for startup.	Approved Master ICF Attachments	Before Startup Before FPI

The background shows a code editor with R code and a terminal window with the following output:

```
Terminal  
R 4.5.2 - ~/D  
> shiny::runApp()  
Loading require  
Listening on ht  
> runApp()  
Listening on ht  
> runApp()  
Loaded assistan  
ig.local.js  
Listening on htt
```

If you build the AI first, you build a black box.

If you build the execution first, you build a tool.



- **Add LLM**: Place the model at the edge for routing and communication.
- **Inject State**: Wire up live context injection at request time.
- **Type Boundary**: Define the strict action schema between interpretation and execution.
- **Build Execution**: Construct the deterministic execution layer without but fit an LLM.
- **Define Output**: Identify the data source and exactly what the user must verify.

Thank you!

