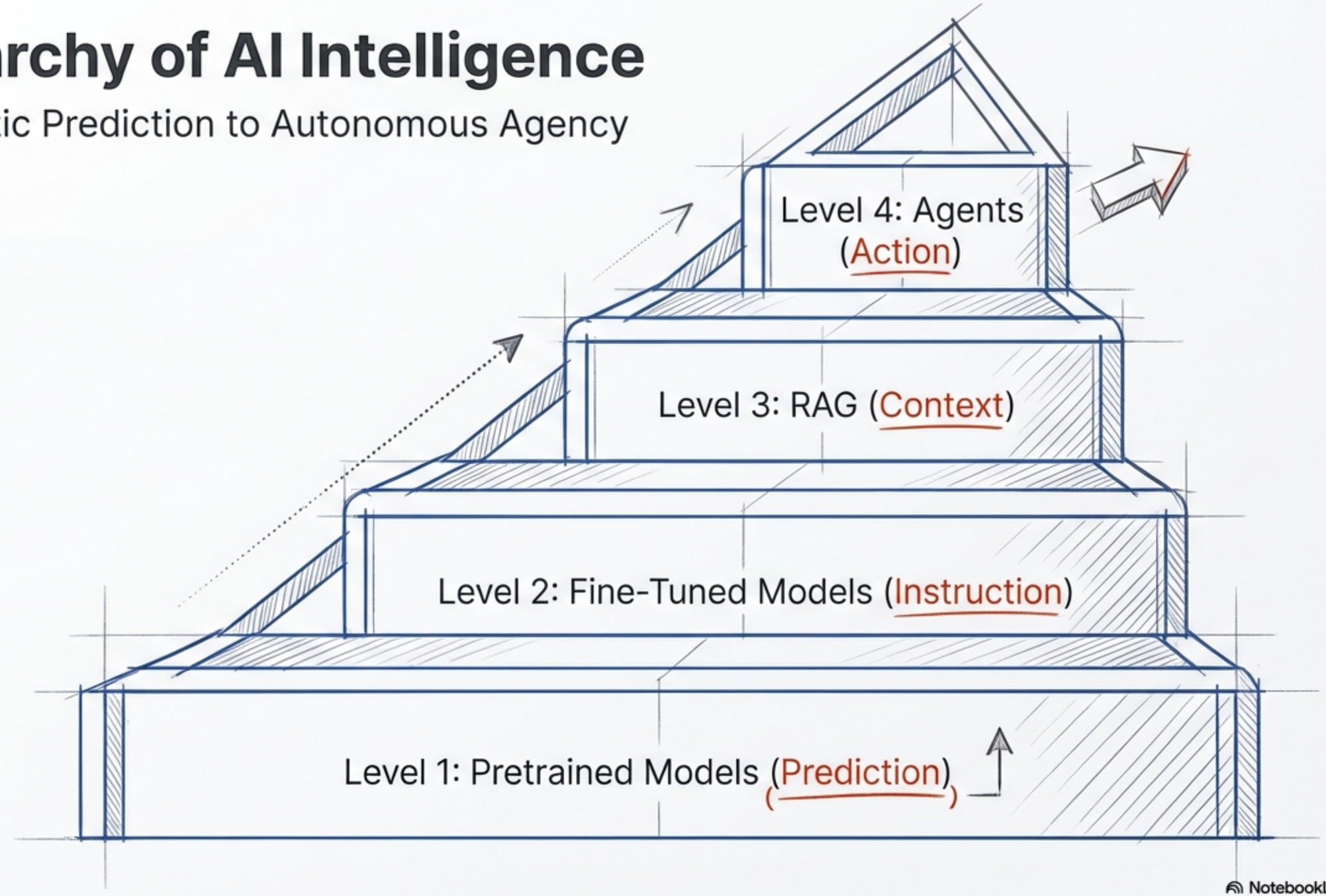
The Hierarchy of AI Intelligence

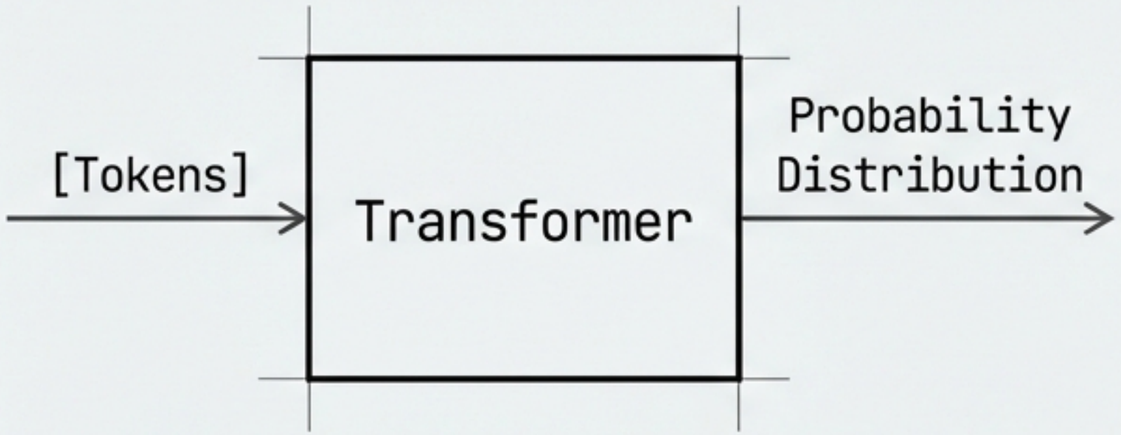
From Probabilistic Prediction to Autonomous Agency

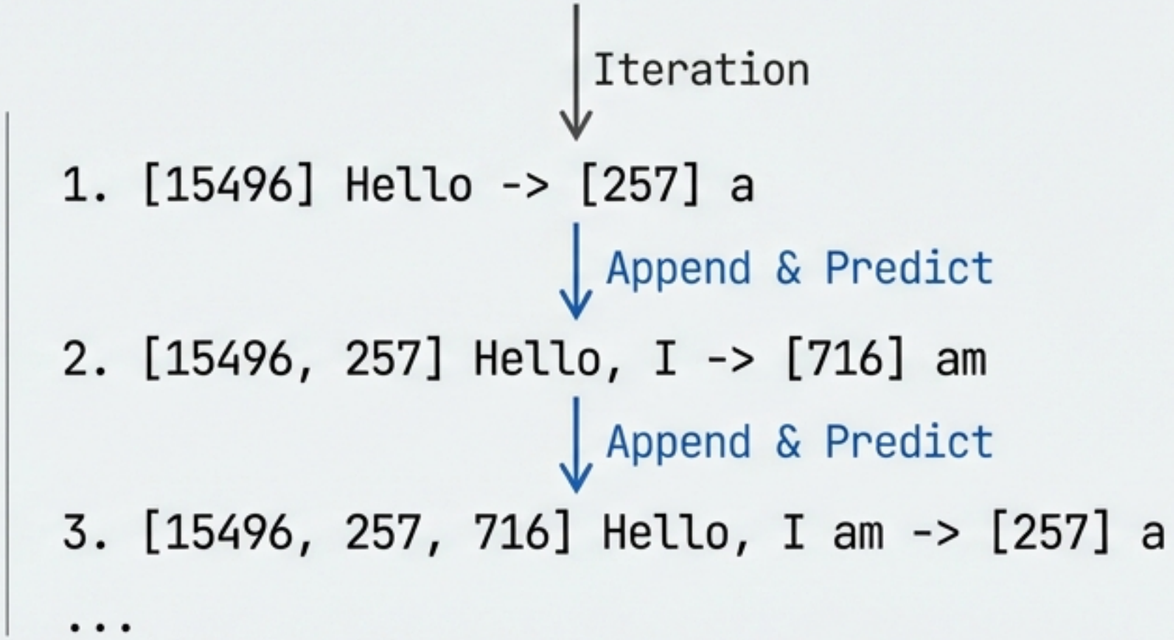


NotebookLM

Level 1: The Pretrained Model

The Well-Read Parrot





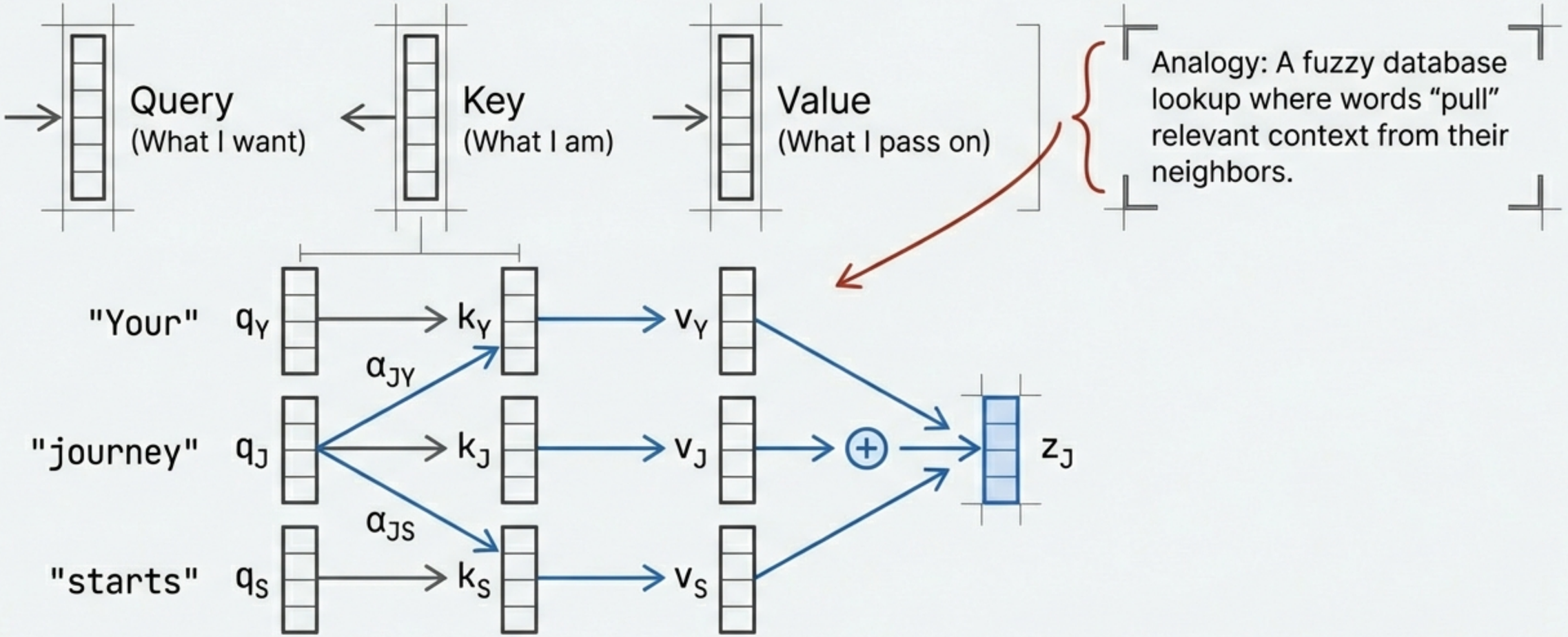
$$\text{Maximize } P(w_t \mid w_{1:t-1})$$

Learns syntax & facts via compression.
No intent, only statistics.

NotebookLM

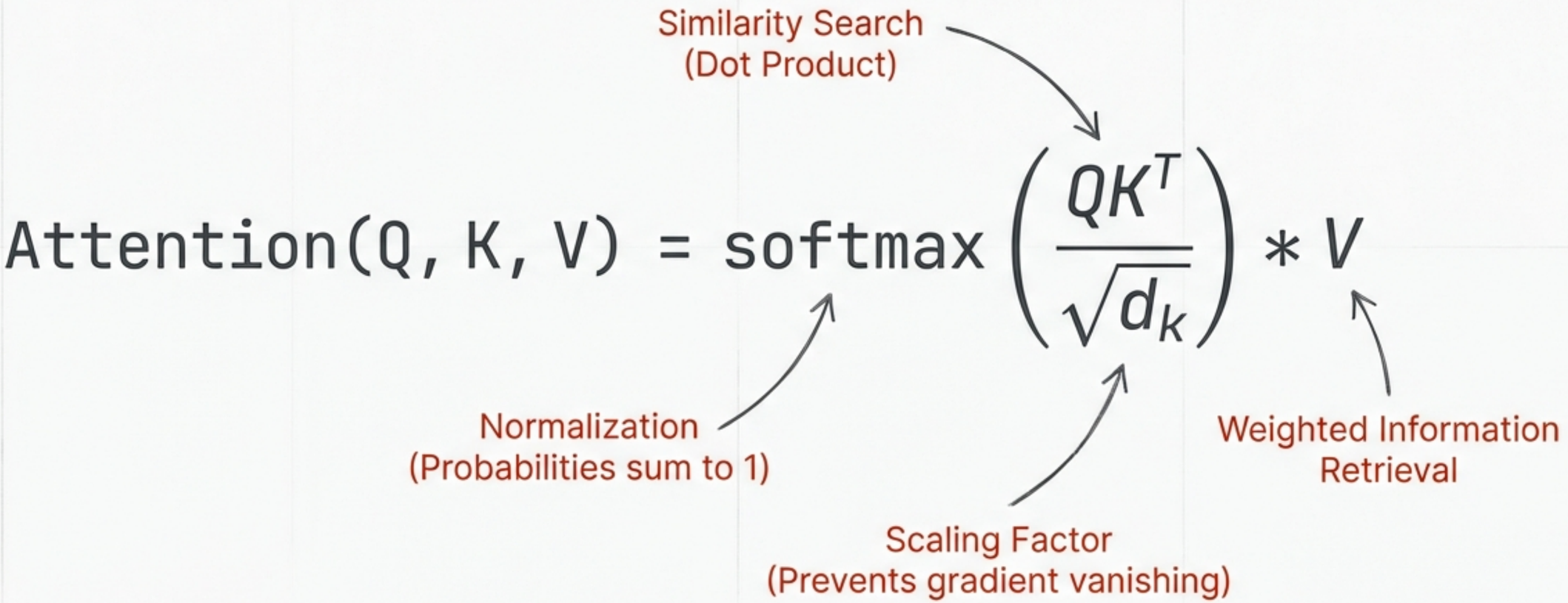
The Engine: Self-Attention

Contextual Gravitation



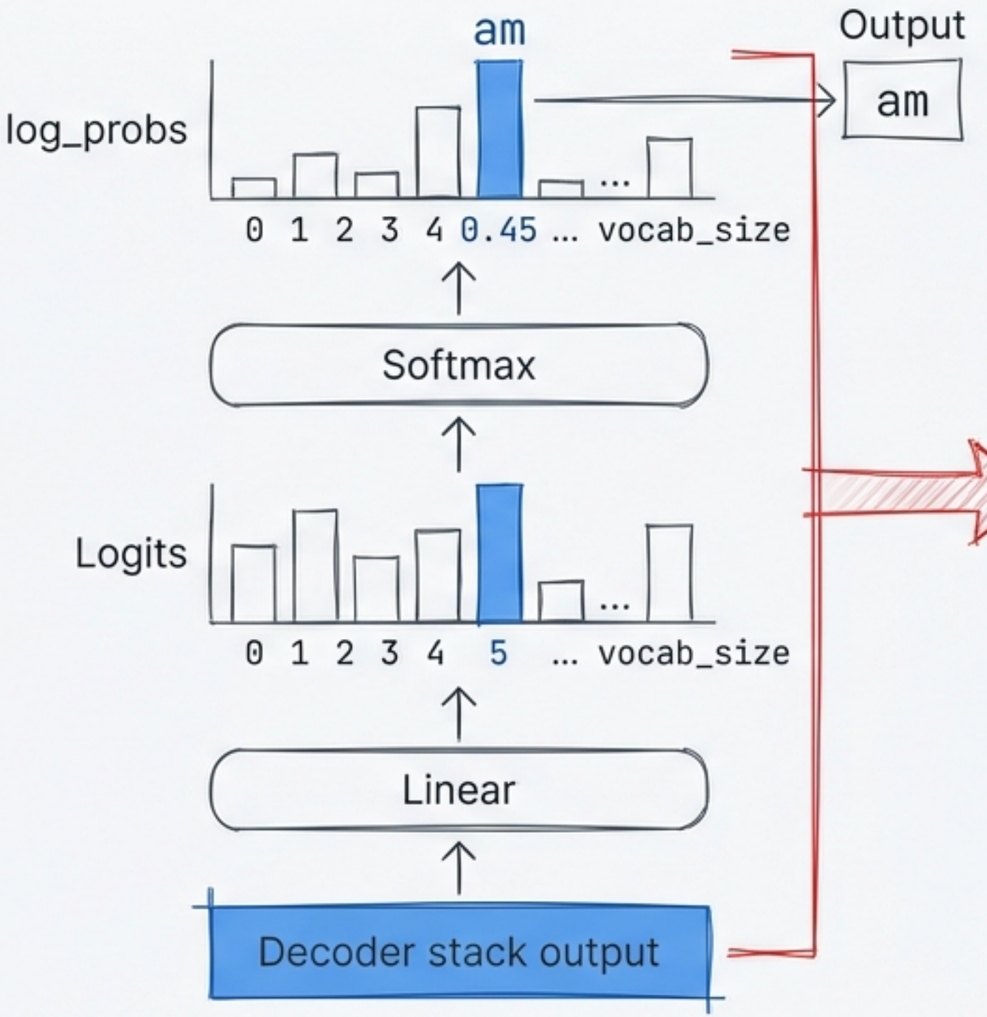
NotebookLM

The Math of Attention



NotebookLM

The Limitation: Probabilistic, Not Factual



Hallucination Risk:
Optimizes for plausibility, not truth.

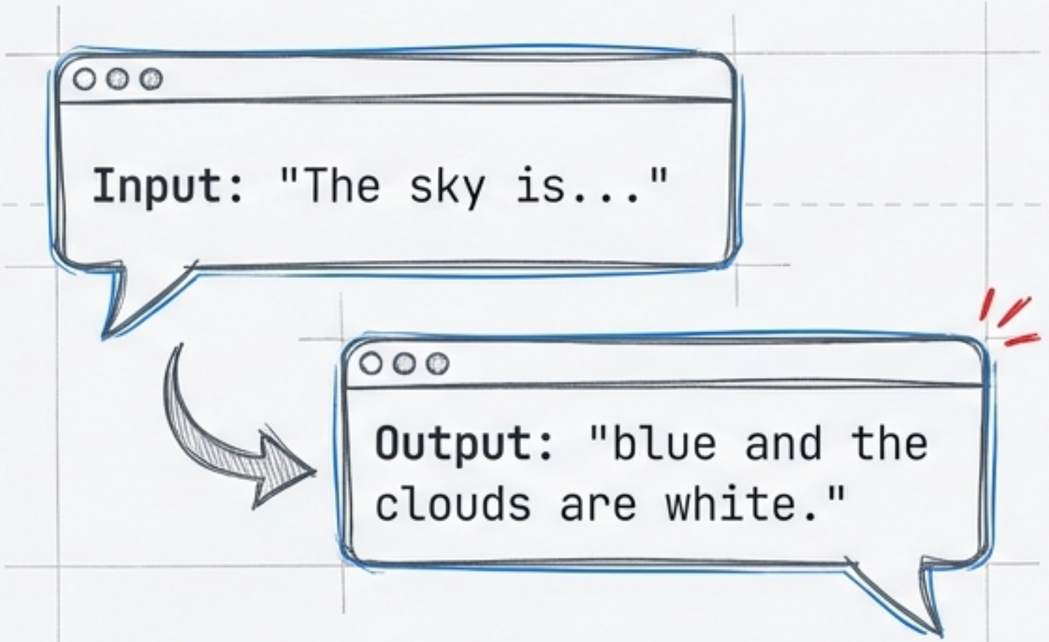
```
Methods:
- Greedy (Argmax)
- Top-k Sampling
- Nucleus Sampling (Top-p)
```

NotebookLM

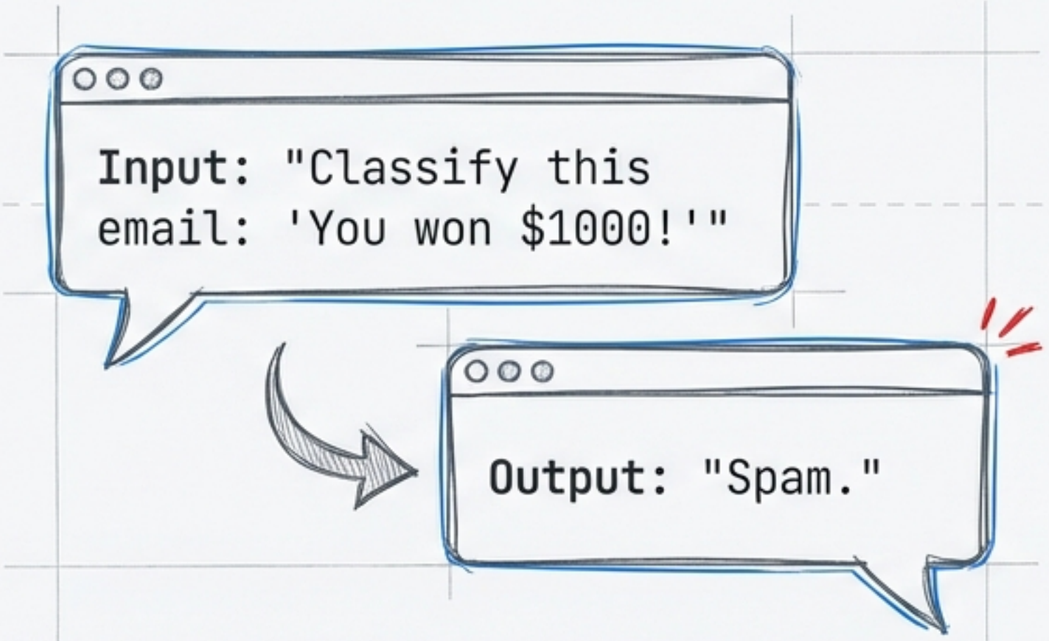
Level 2: The Specialist (Fine-Tuning)

From Completion to Instruction

Pretrained (Base Model)



Fine-Tuned (Instruction Model)

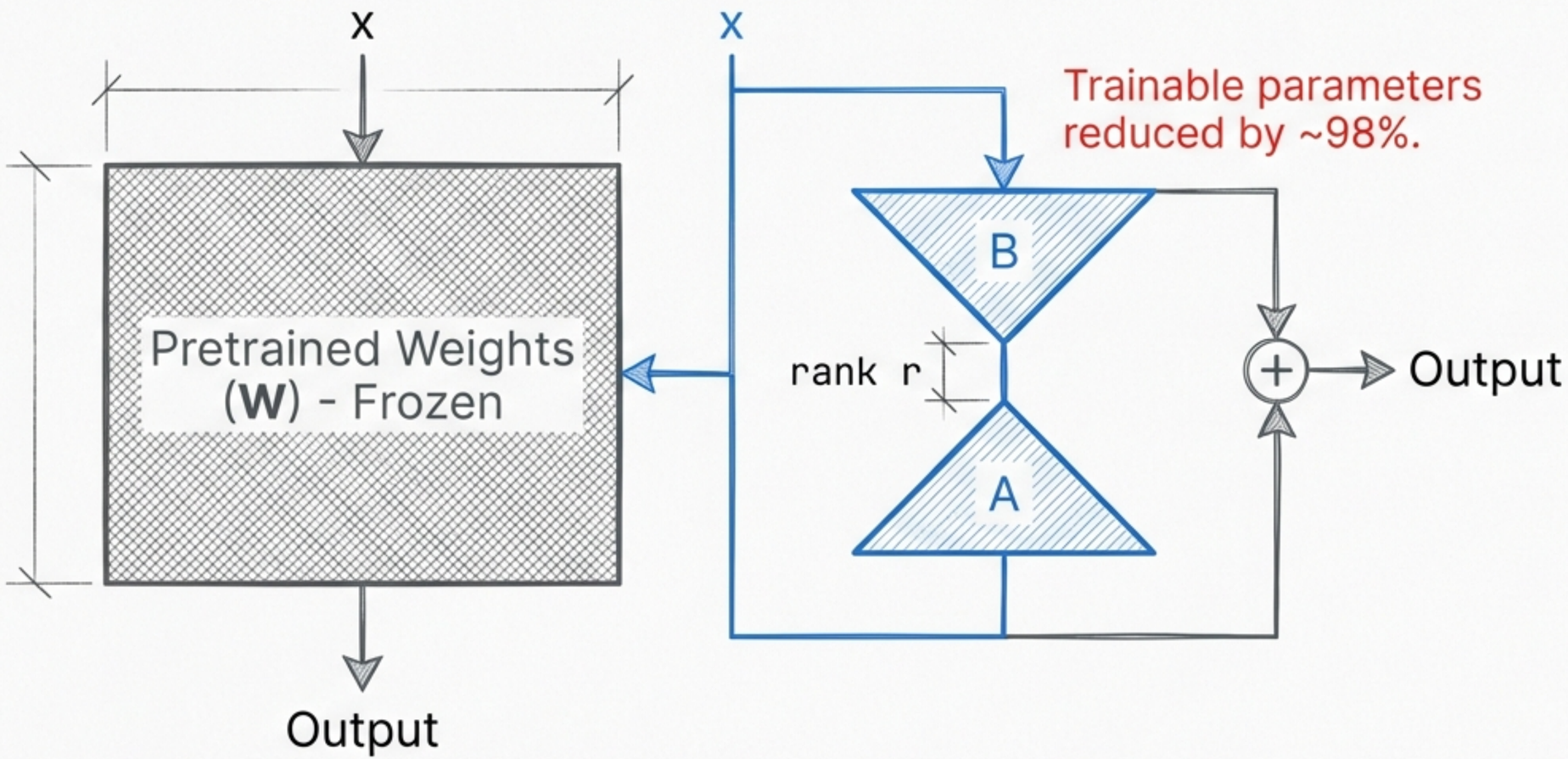


Supervised Fine-Tuning (SFT) constrains the distribution to specific tasks.

NotebookLM

The Efficiency Solution: LoRA

Low-Rank Adaptation



NotebookLM

The LoRA Equation

$$W_{updated} = W + \Delta W \approx W + (B * A)$$

$$h = W_0 * x + B * A * x$$

Inter Bold
Frozen

Inter Bold
Trainable Adapter

A is d x r, B is r x d, where r << d

NotebookLM

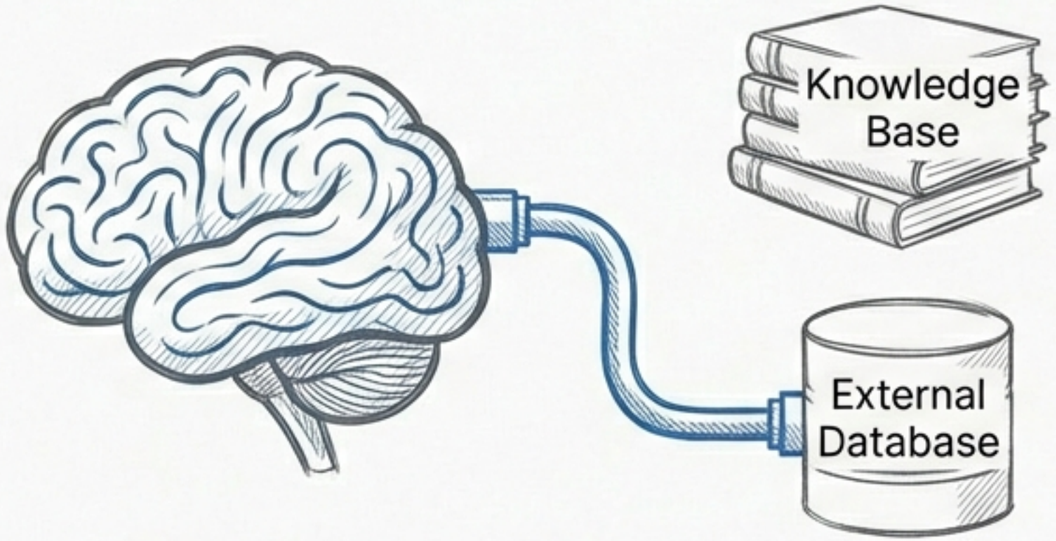
Level 3: The Researcher (RAG)

Overcoming the Knowledge Cutoff

Limited by training data

Parametric Memory (Weights)

Frozen, Expensive, Hallucinations



Non-Parametric Memory (Context)

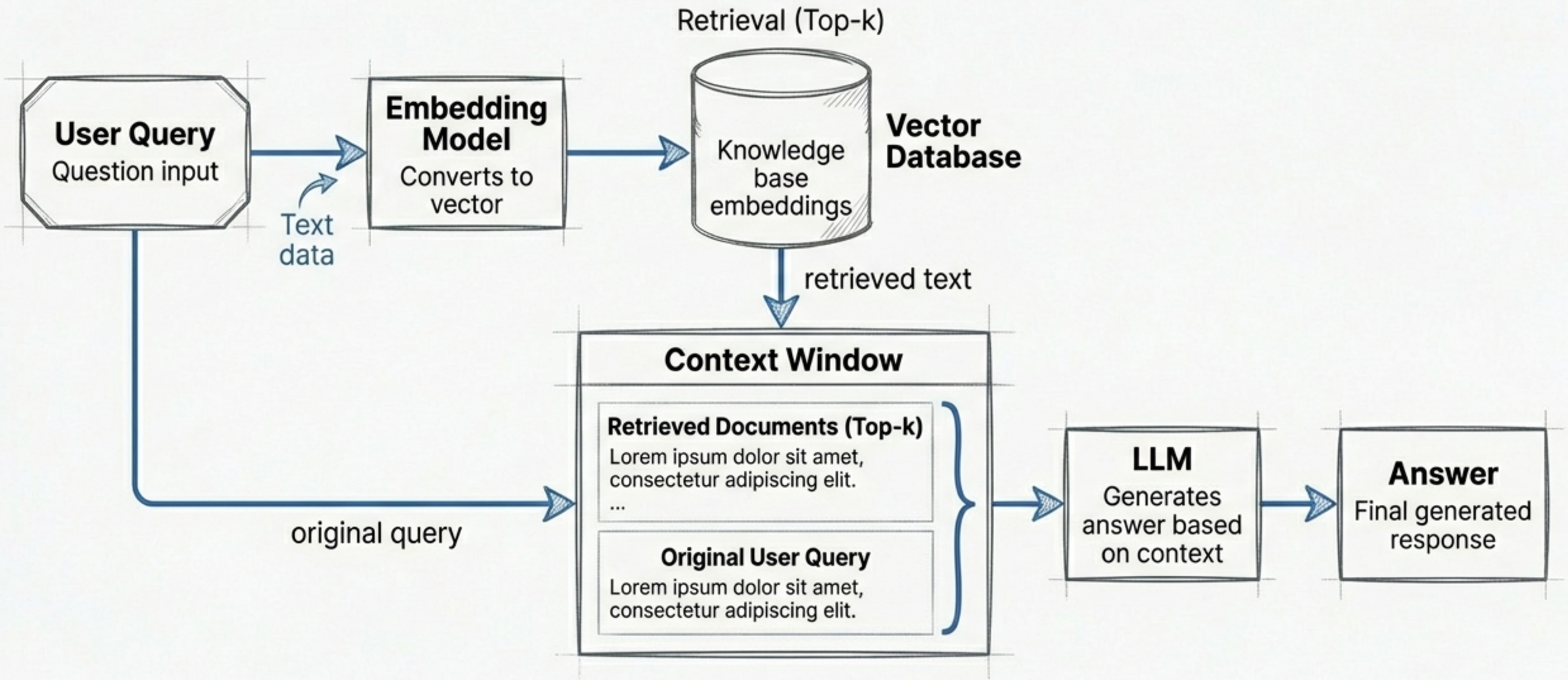
Dynamic, Cheap, Grounded

Augments with real-time data

Retrieval Augmented Generation

NotebookLM

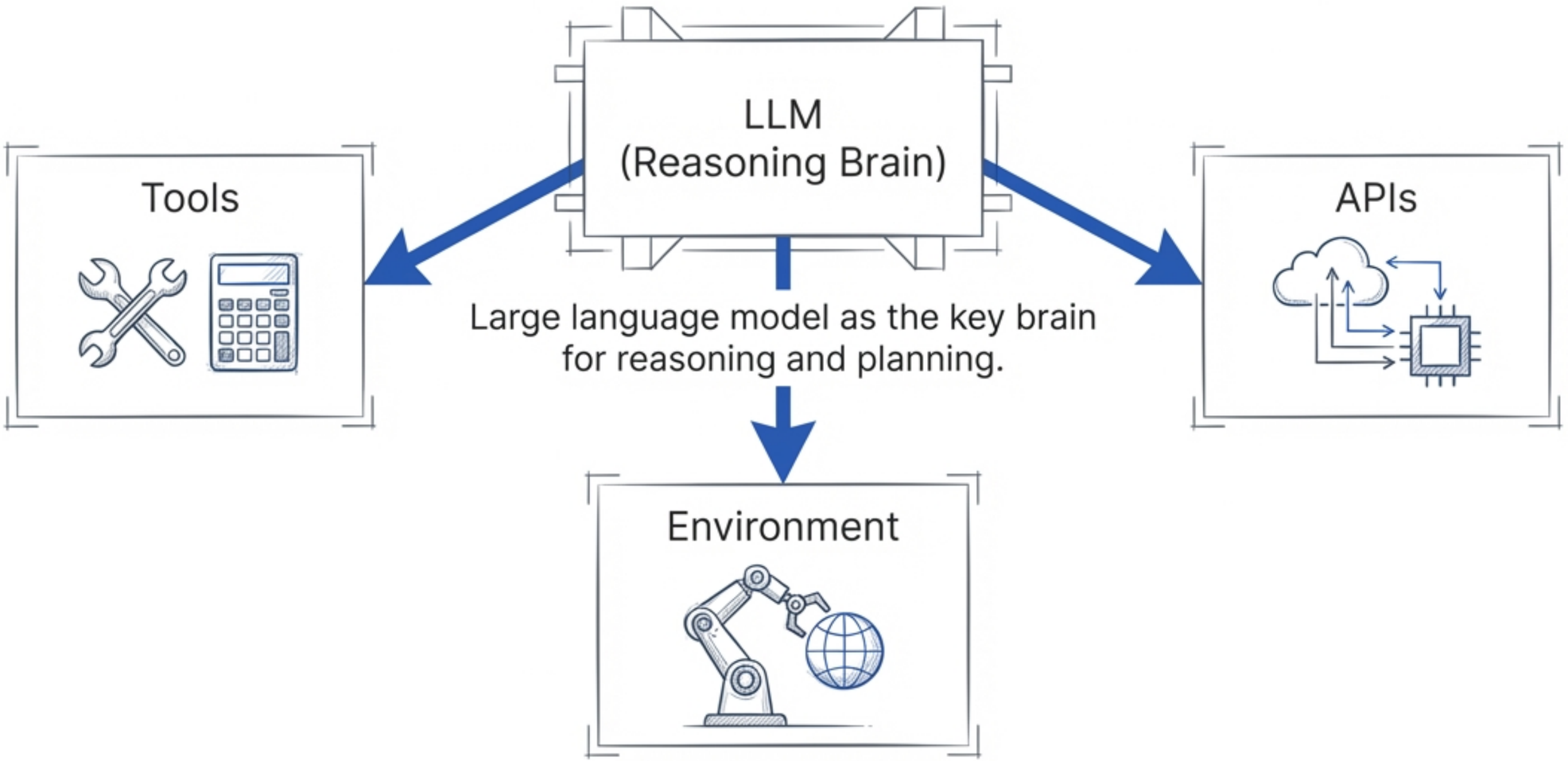
RAG Architecture: The ‘Open Book’ Exam



NotebookLM

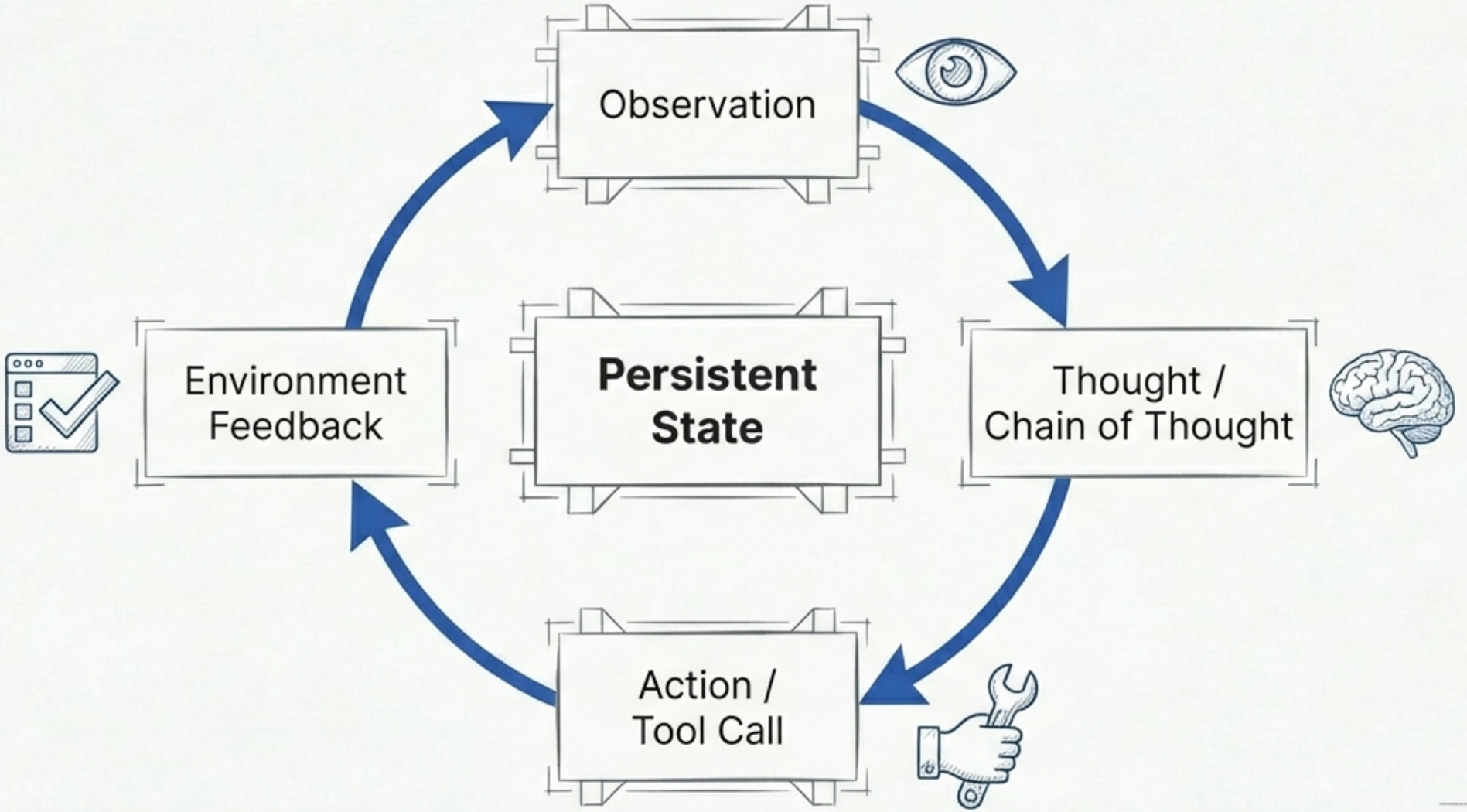
Level 4: The Agent (The Actor)

From Passive Text to Active Doing



NotebookLM

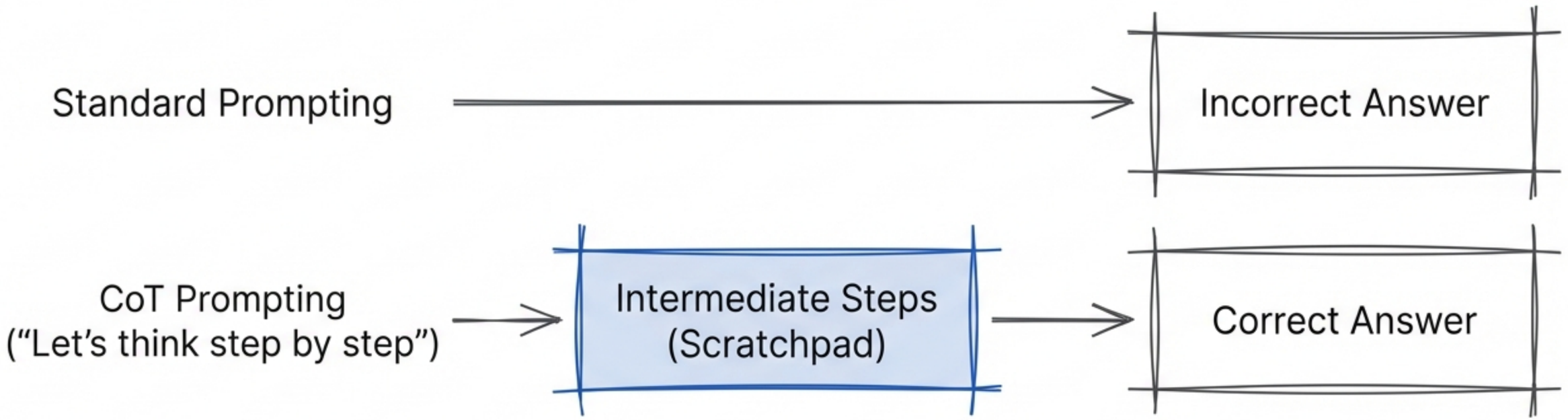
The Agentic Loop



NotebookLM

Unlocking Reasoning: Chain of Thought (CoT)

System 1 vs. System 2 Thinking



Generating tokens is computation.
More tokens = more compute time = better reasoning.

NotebookLM

CoT in Action

I have 3 apples. My dad has 2 more apples than me. How many do we have?

CoT

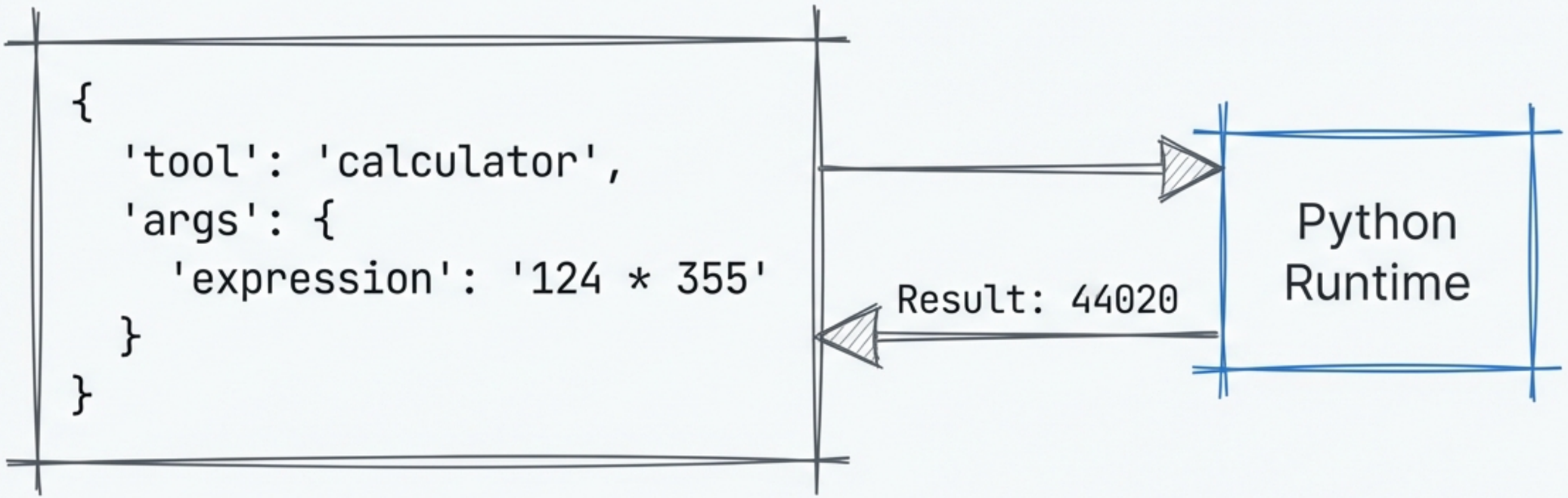
```
Step 1: I have 3 apples.
Step 2: Dad has 3 + 2 = 5 apples.
Step 3: Total = 3 + 5 = 8.

Answer: 8
```

NotebookLM

The Hands of the Agent: Tool Use

Model Controller Protocol (MCP)

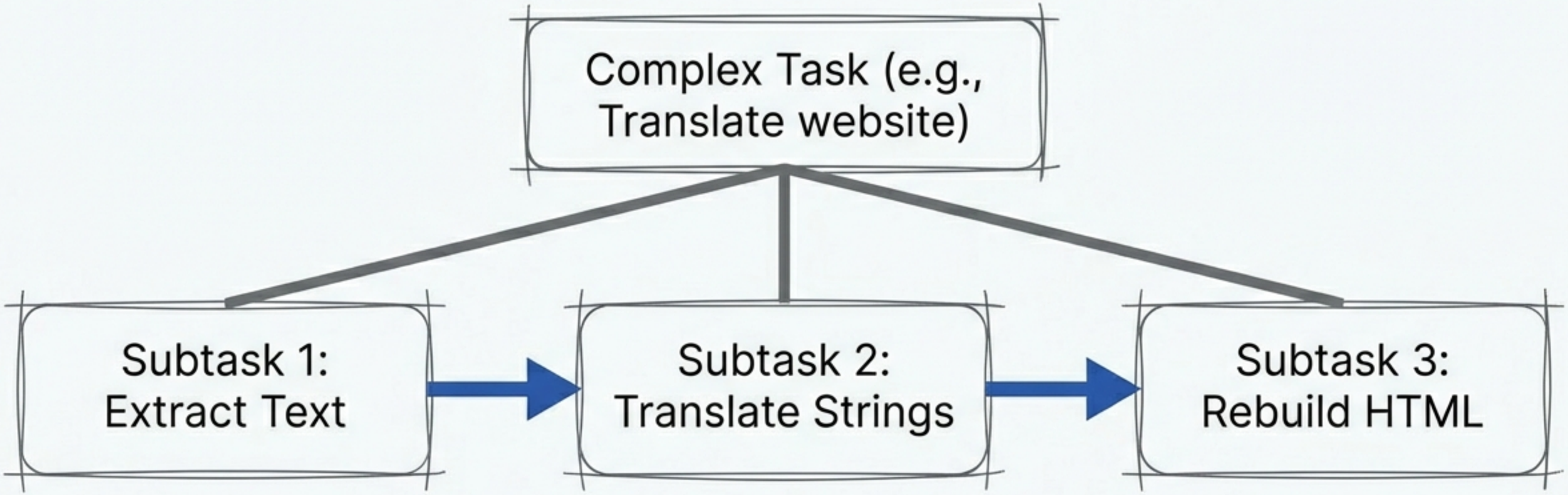


The model on outputs structured data to control external software.

NotebookLM

Planning & Decomposition

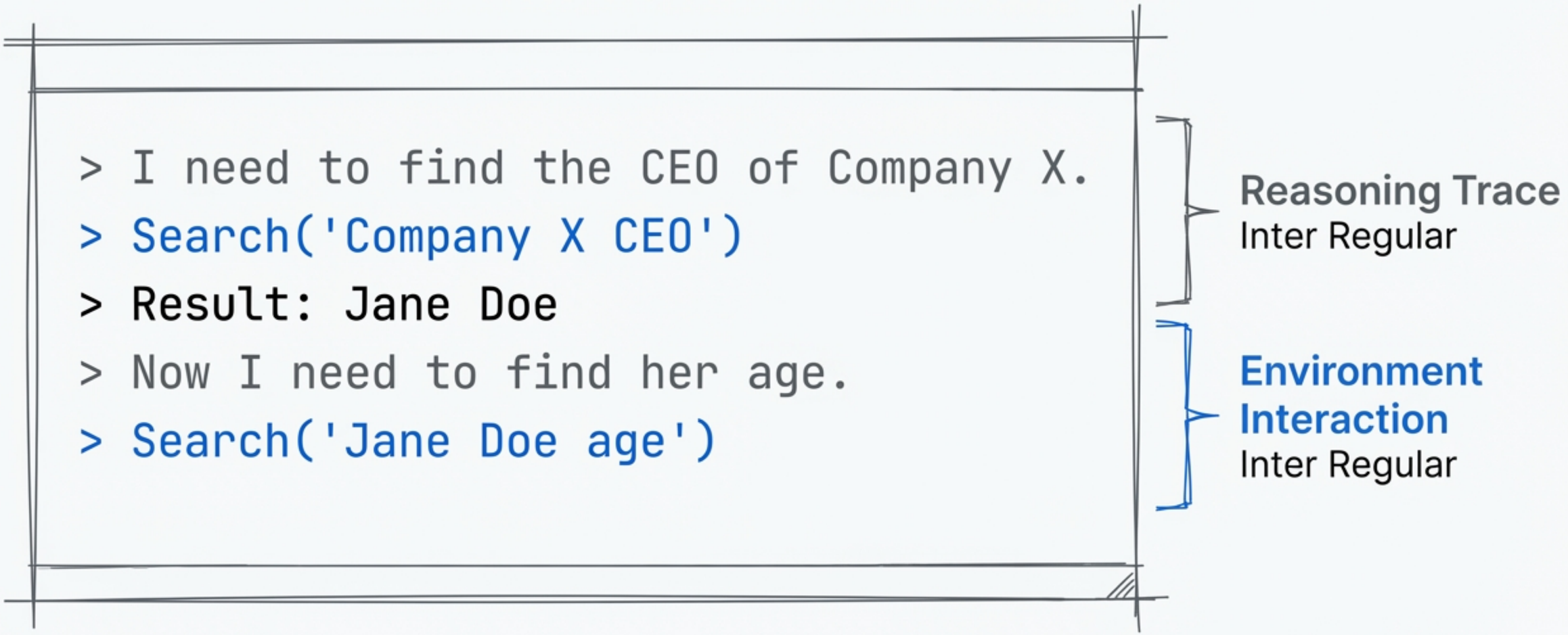
Least-to-Most Prompting



Decompose complex problems into sequential dependencies.

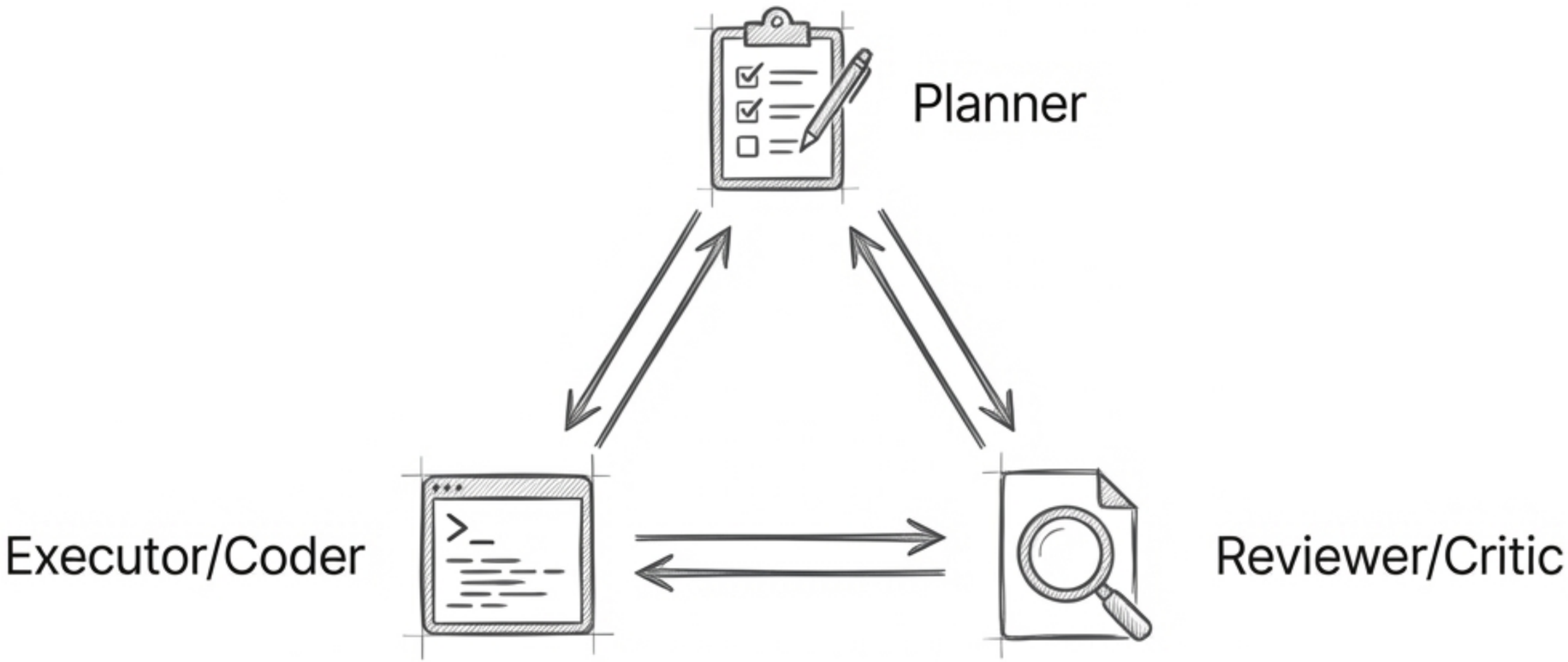
NotebookLM

ReAct: Reasoning + Acting



NotebookLM

Multi-Agent Collaboration



Specialization allows for complex task solving beyond a single context window.

NotebookLM

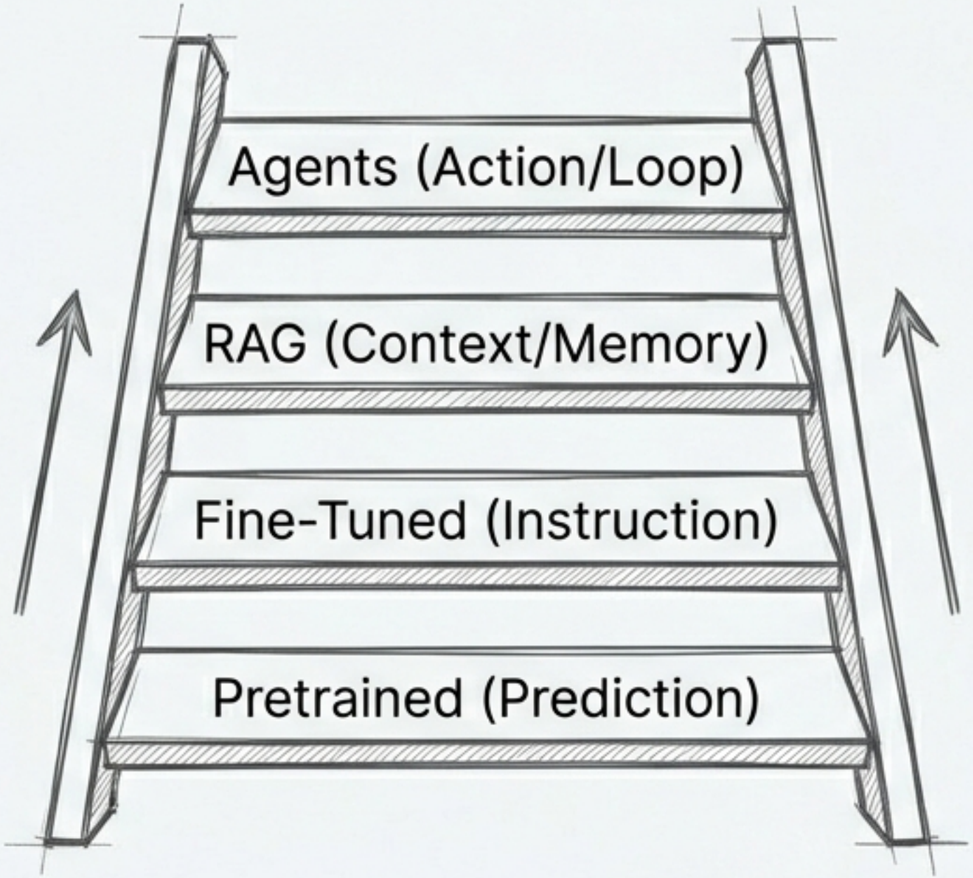
The Frontier: Self-Correction

The Challenge of Intrinsic Verification

- Models often hallucinate corrections.
- Risk of changing correct answers to incorrect ones.
- Future: External verifiers and Reward Models.

NotebookLM

Summary: The Evolution of Cognition



Paradigm Shift: “What comes next?” → “**What should I do?**”

NotebookLM