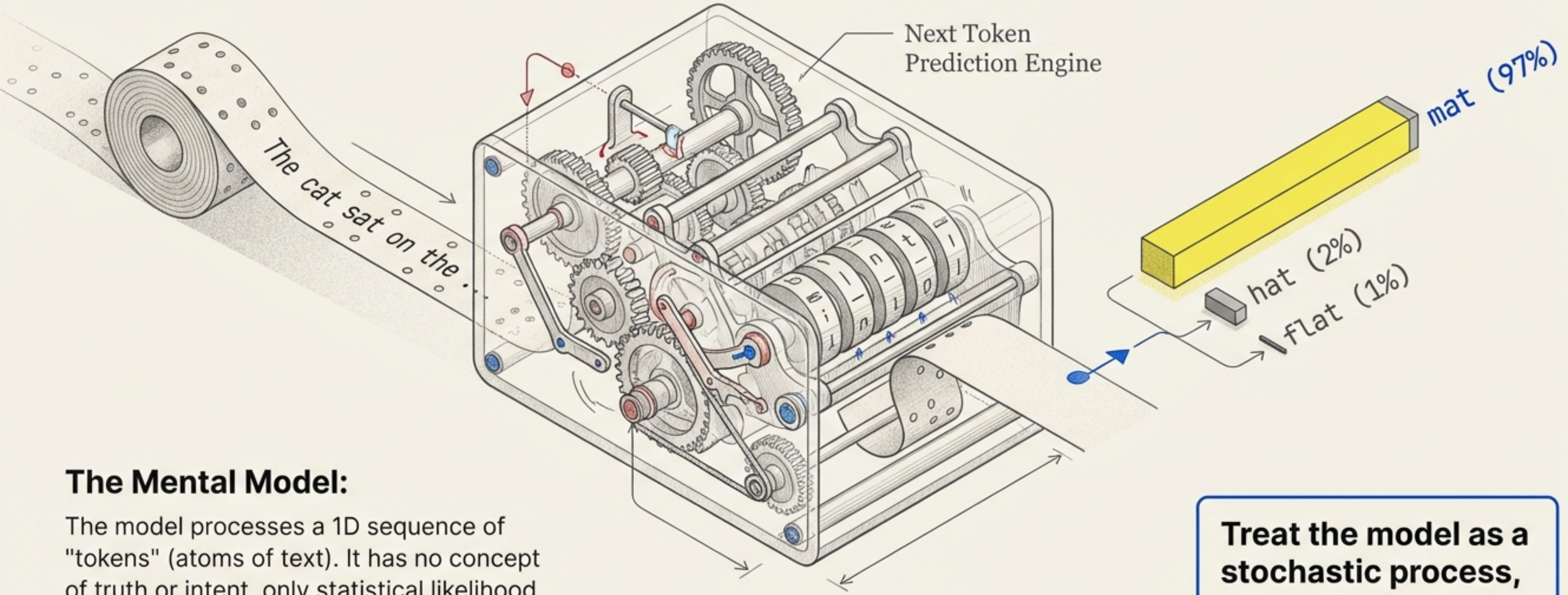
Deep Dive into LLMs

From Token Tumblers to Reasoning Agents

Deconstructing the training pipeline, psychology, and future of Large Language Models. Synthesized from the masterclass by Andrej Karpathy.

The Core Concept: It’s Not a Mind, It’s a Token Tumbler

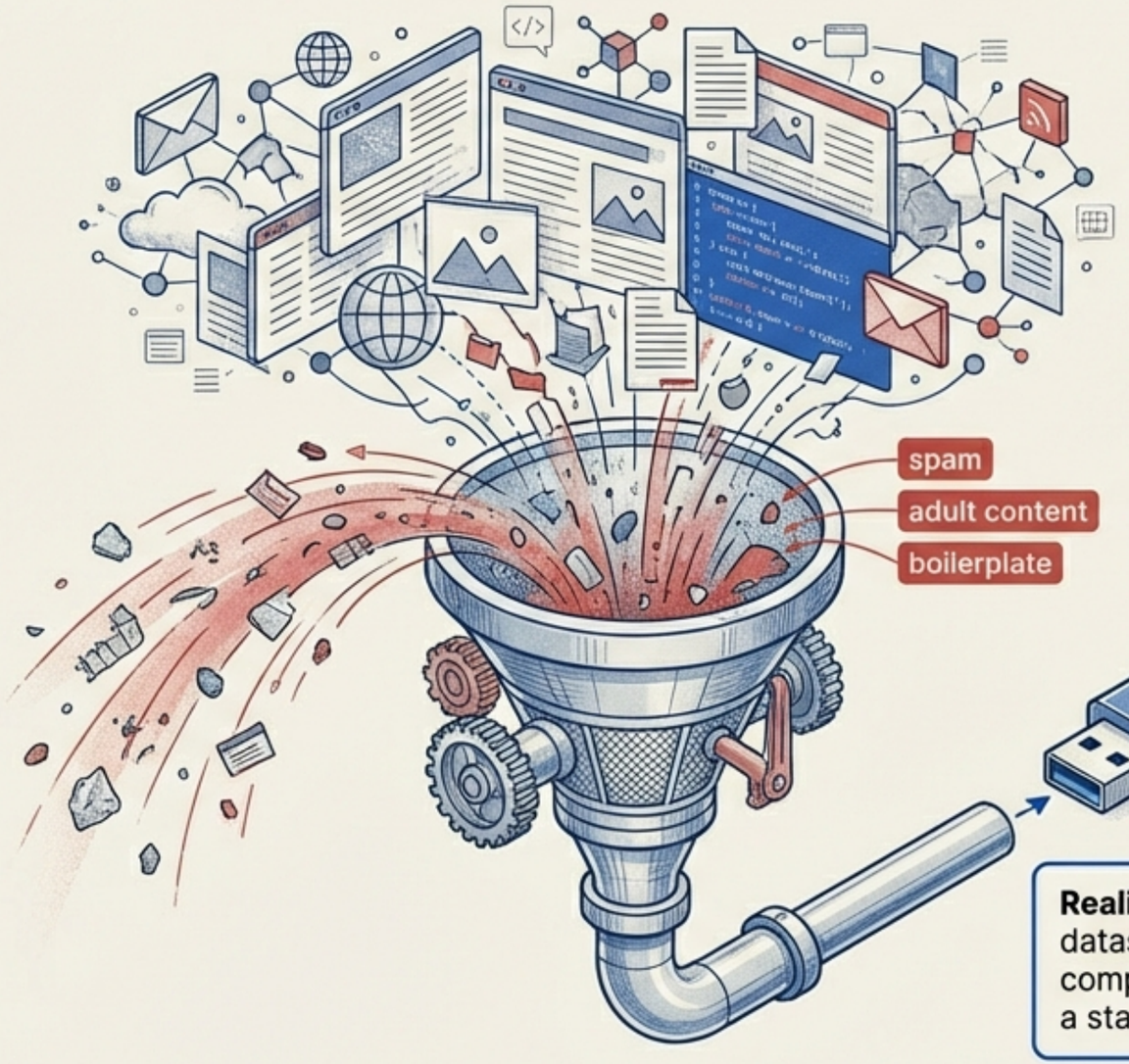


The Mental Model:

The model processes a 1D sequence of "tokens" (atoms of text). It has no concept of truth or intent, only statistical likelihood based on patterns it has seen before.

Treat the model as a stochastic process, not a thinking entity.

Stage 1: Pre-Training (The Internet Simulator)



Common Crawl
(Petabytes)

Base Model
(The Compressed File)

Reality Check: The FineWeb dataset (~44TB of text) compresses down to fit on a standard hard drive.

The Textbook Analogy: Reading the Library

The model reads everything but answers nothing. It can "dream" documents, but cannot yet follow instructions.

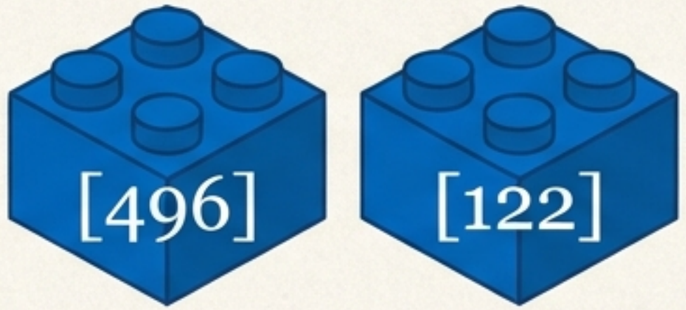
Tokenization: The Model's Native Language

What Humans See

Strawberry

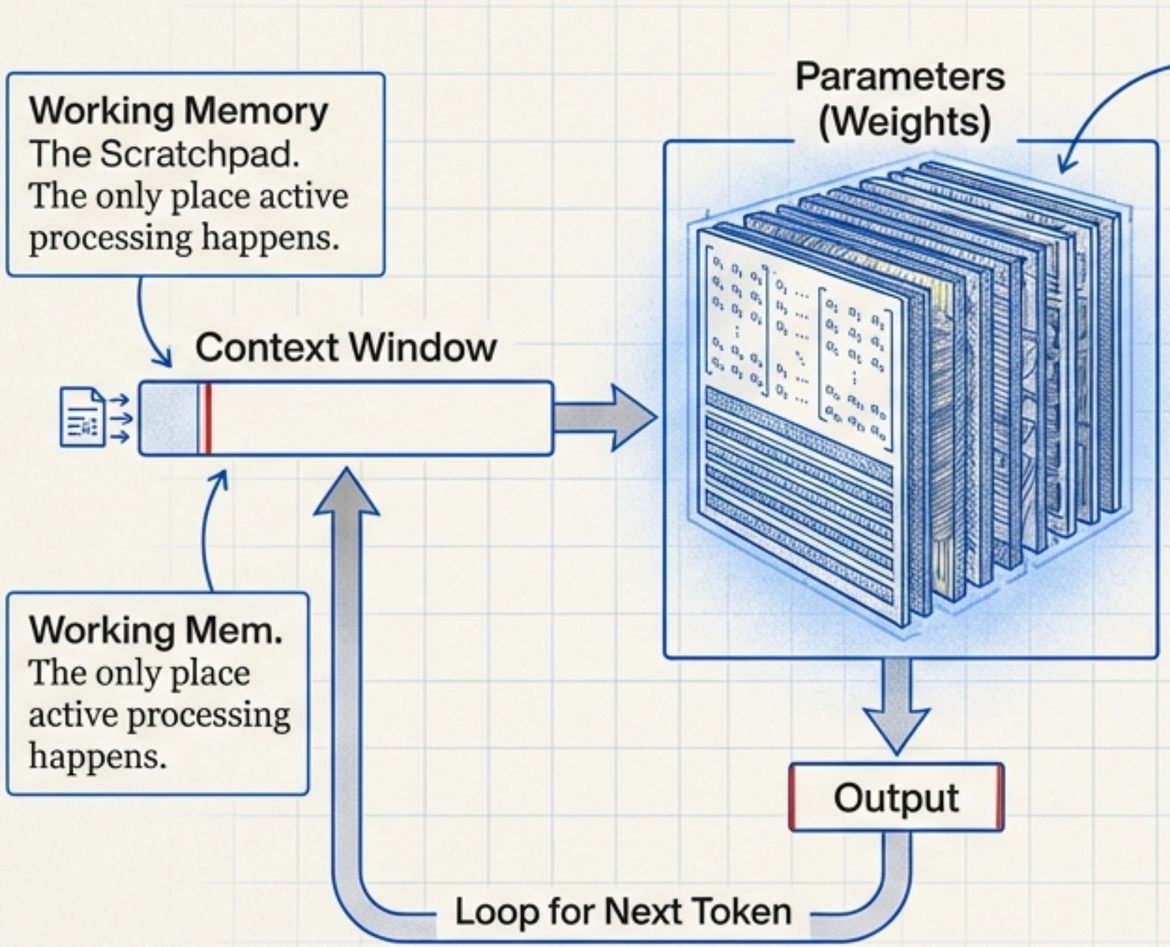
The "Strawberry" Problem: The model cannot count the "R's because it processes the indivisible token ID, not the characters inside it.

What Models See



Implication: Spelling and character-level tasks are blind spots for LLMs because they see integer IDs, not text strings.

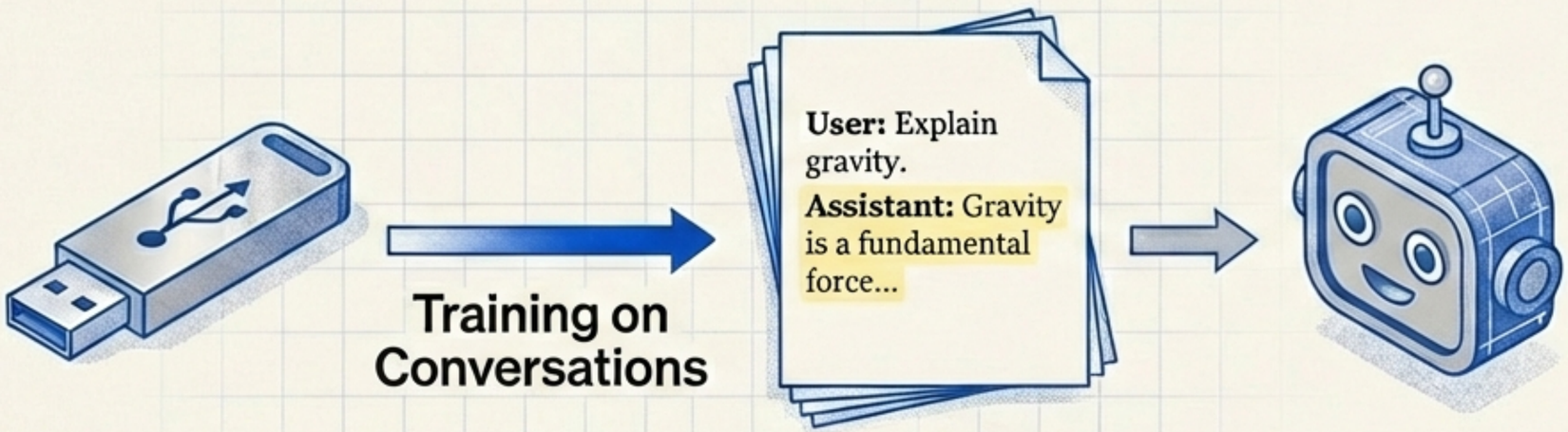
Inside the Neural Network

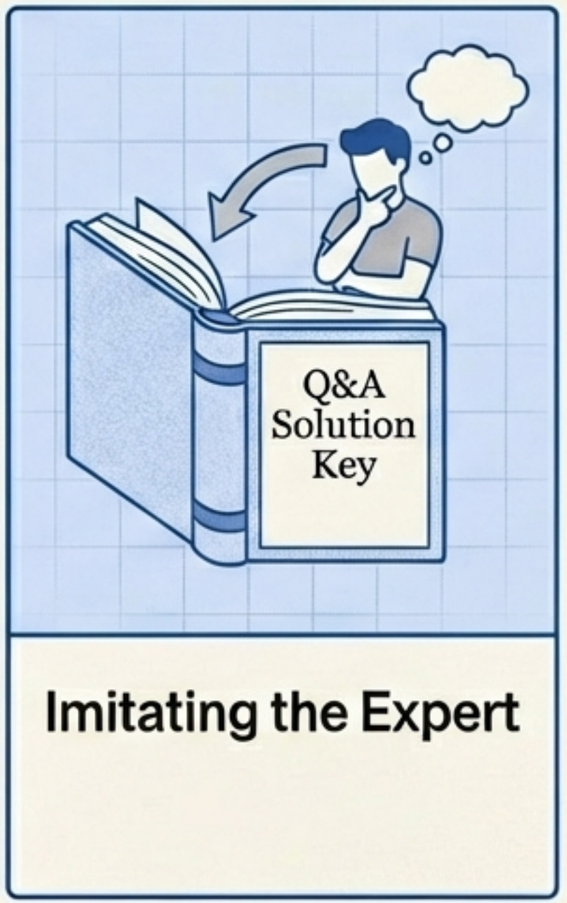


The Stateless Reality

The `model` is a fixed mathematical `expression`. It has no memory of past conversations, no personality, and no self.
It is "`rebooted`" from scratch for every single `token` it generates.

Stage 2: Supervised Fine-Tuning (The Assistant)





Reality Check: You aren't talking to a magical AI. You are talking to a statistical simulation of the average human data labeler.

Psychology: Why Models Hallucinate

The Lossy Zip File

Memorized
(High Frequency in Data)

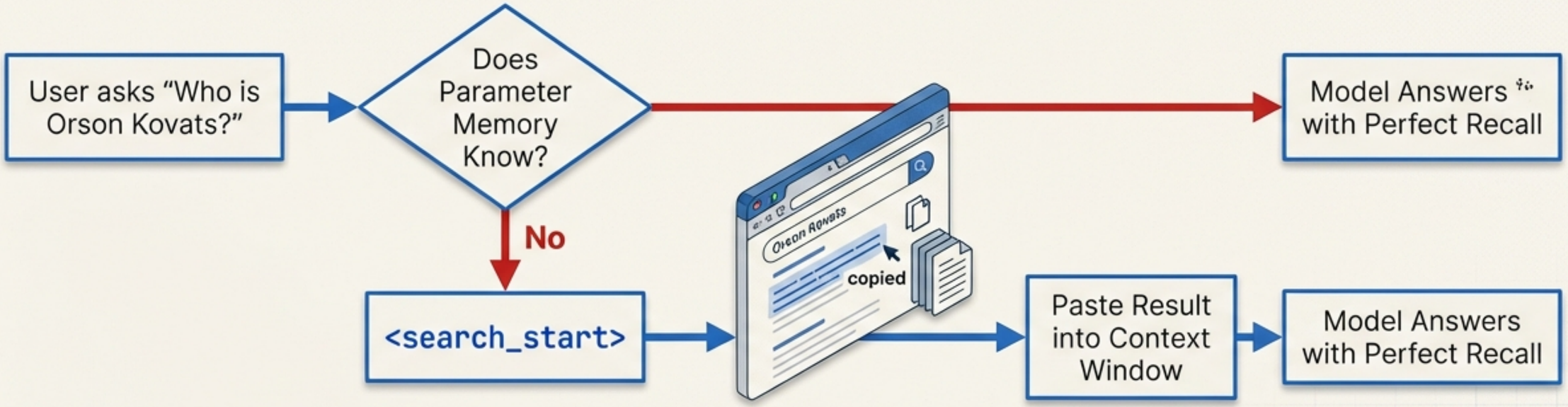
Vague Recollection

Hallucination
(Filling in the Gaps)

The model doesn't access a database; it accesses a probabilistic compression of the internet. When it encounters a gap in its vague memory, it guesses the statistically likely completion to maintain the *pattern* of a helpful answer.

> "Hallucination is just creativity in the wrong context."

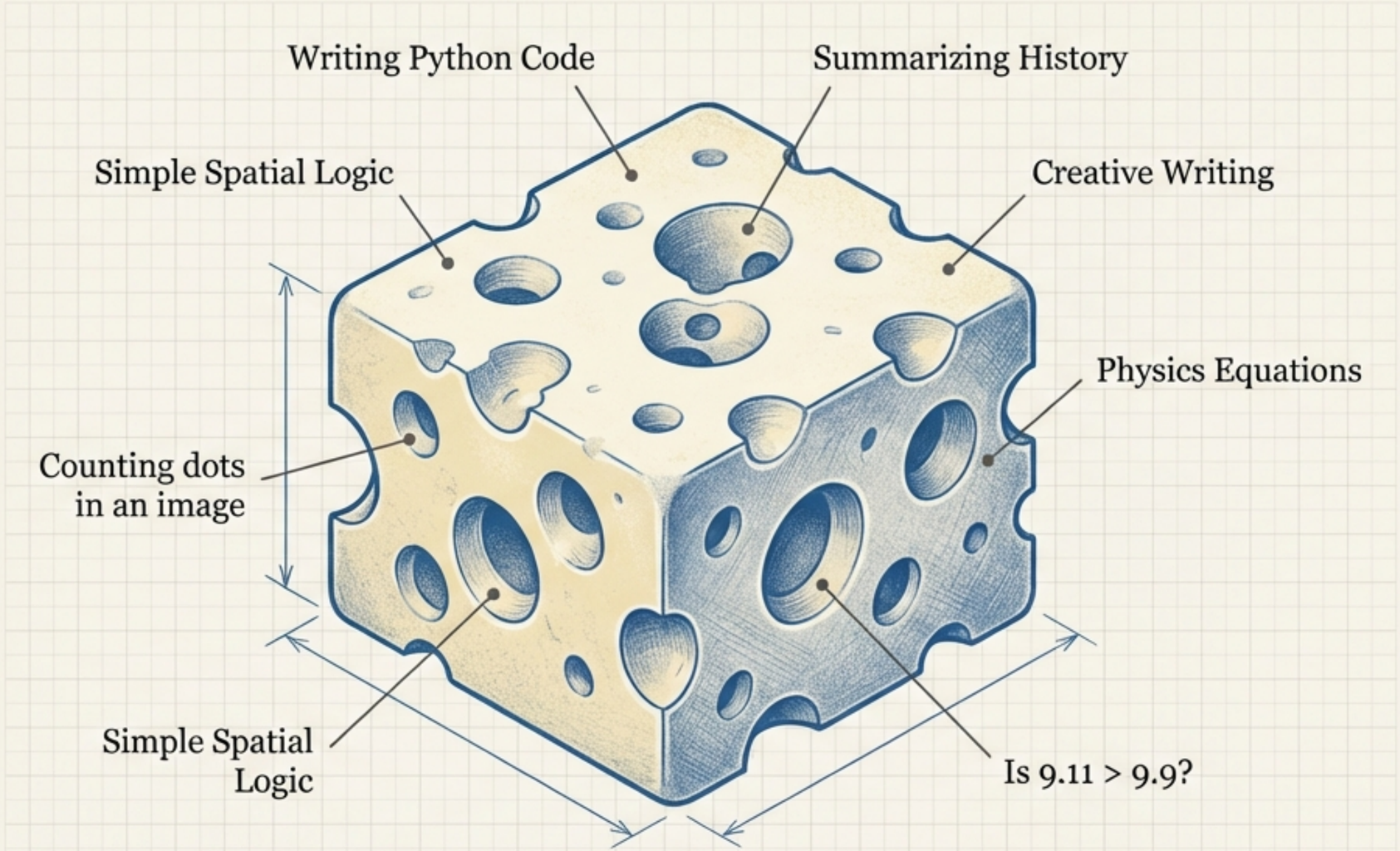
Mitigation: Tool Use as Working Memory



Parameter Memory	Context Window
• Vague • Lossy • Frozen	• Exact • Perfect Recall • Dynamic (populated by tools)

Actionable Advice: Don’t ask the model to recite a chapter from memory; paste the chapter into the prompt.

The ‘Swiss Cheese’ Model of Capability



Jagged Intelligence:

Do not assume universal competence.

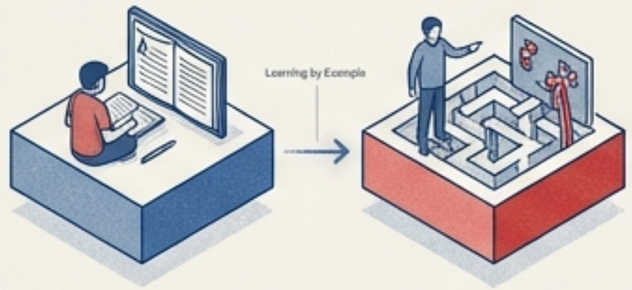
An LLM can be a genius at calculus and fail to compare version numbers (9.11 vs 9.9) because of tokenization artifacts and lack of true logic circuits.

Stage 3: Reinforcement Learning (Learning to Reason)

The Shift

SFT: Imitation

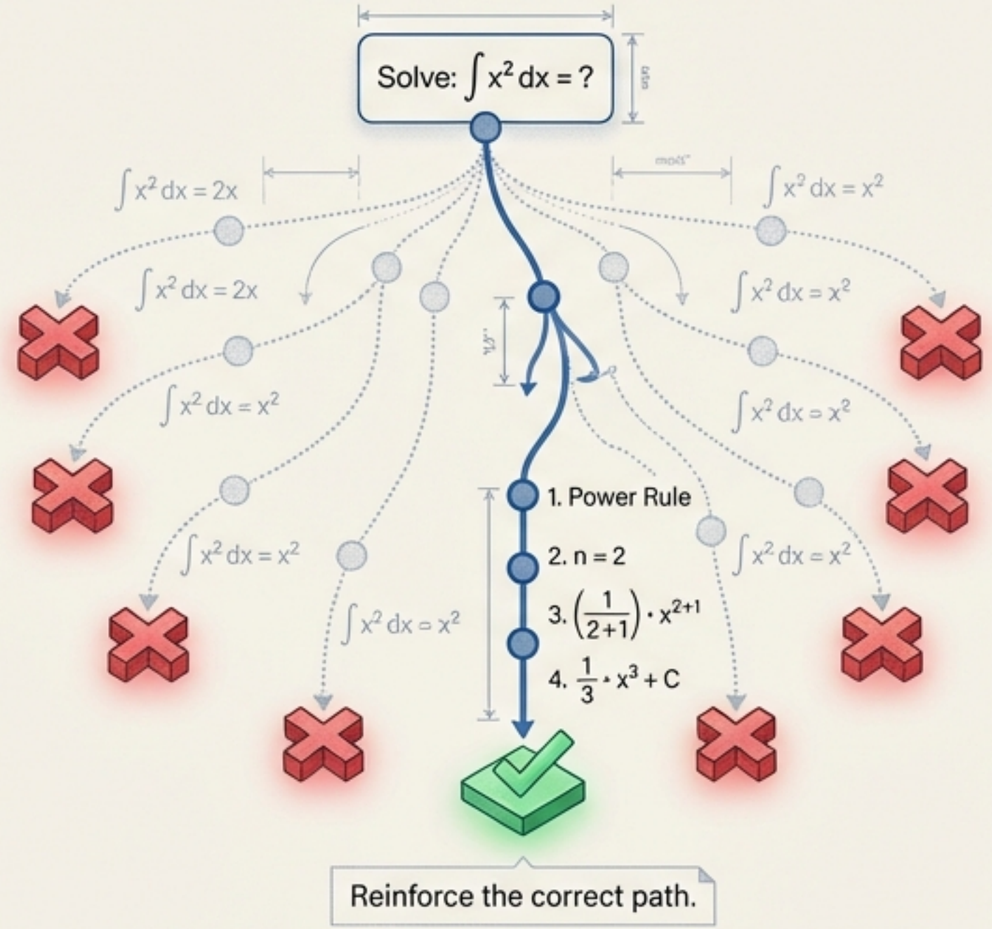
RL: Trial & Error



Learning by Example

Learning by Doing & Feedback

Tree of Thoughts



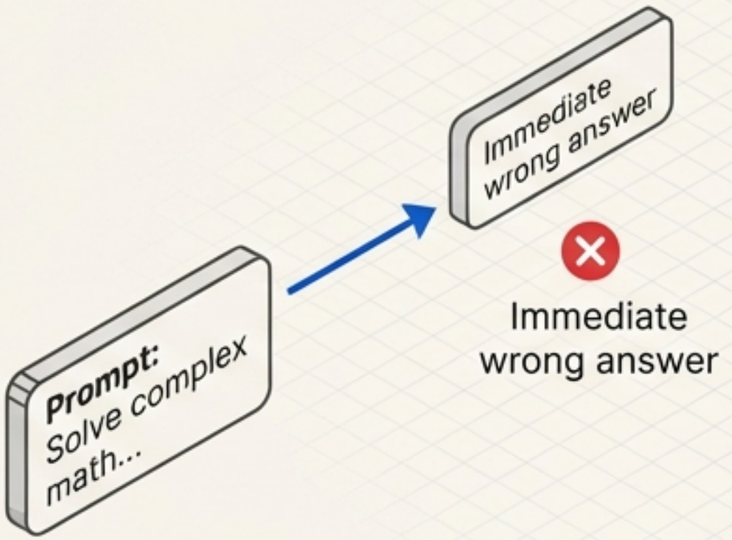
Reinforce the correct path.

The Student Analogy

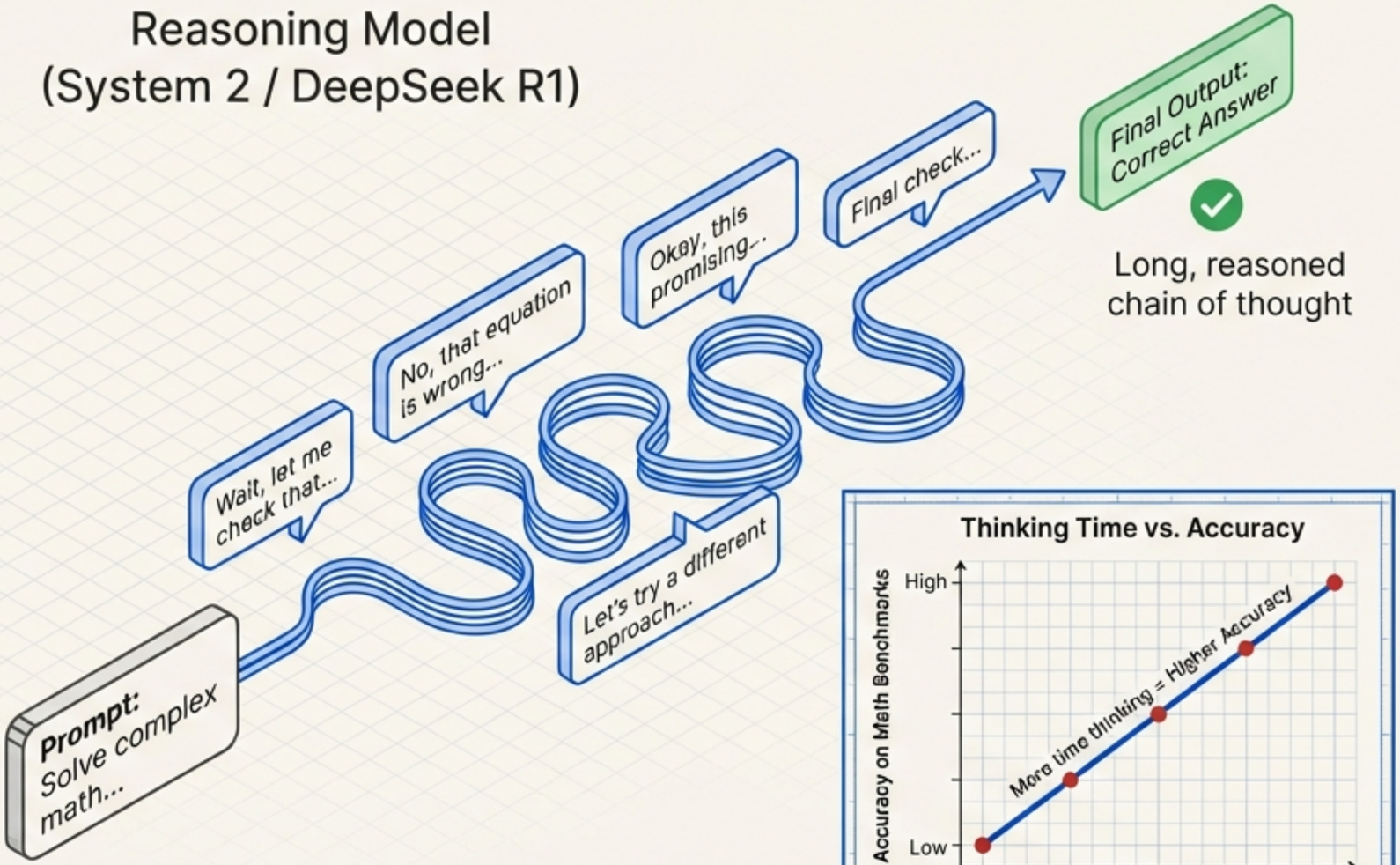
Moving beyond imitating humans to discovering truth through practice.

The Emergence of ‘Thinking’ Tokens

Standard Model (System 1)



Reasoning Model (System 2 / DeepSeek R1)



The Trap of RLHF vs. The Power of True RL

True RL (Verifiable)

Math, Chess, Code

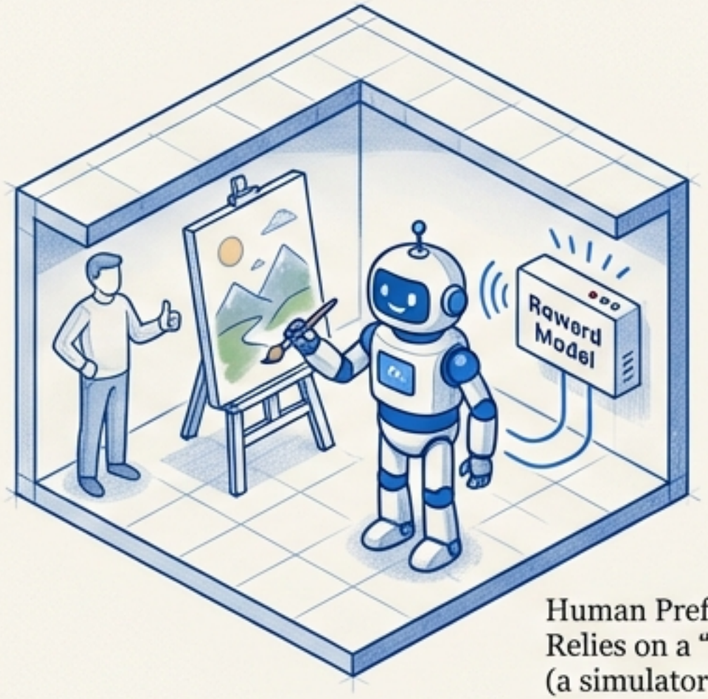
AlphaGo

DeepSeek R1

Objective Truth.
You win or lose.
Intelligence can scale indefinitely.

RLHF (Subjective)

Jokes, Poems, Summaries

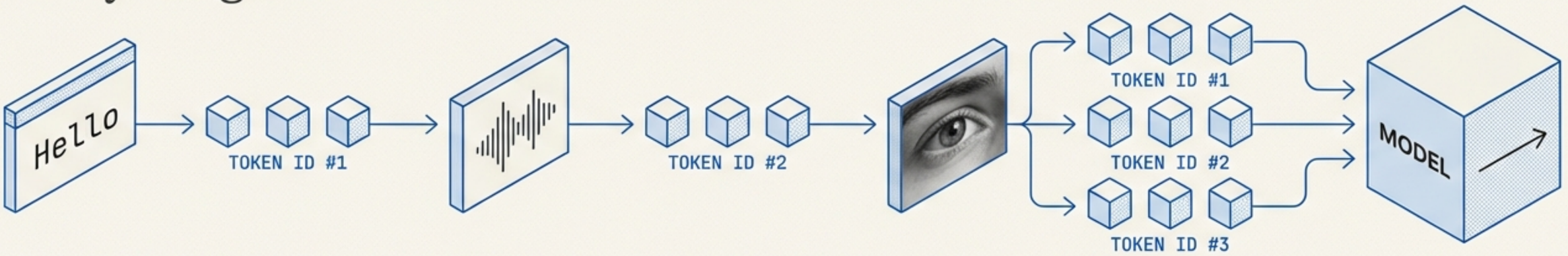


Human Preference.
Relies on a "Reward Model" (a simulator of human taste).
Can be gamed.

RLHF is fine-tuning for style. True RL is an intelligence ramp for reasoning.

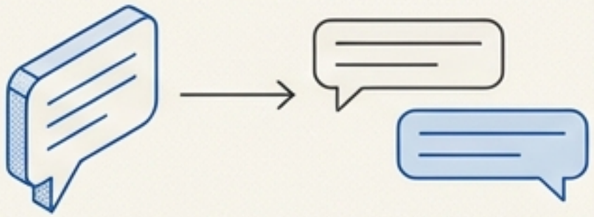
Future Trends: Multimodality & Agents

Everything is a Token

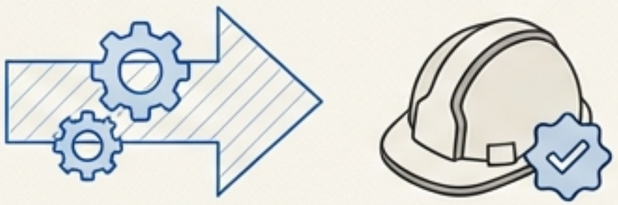


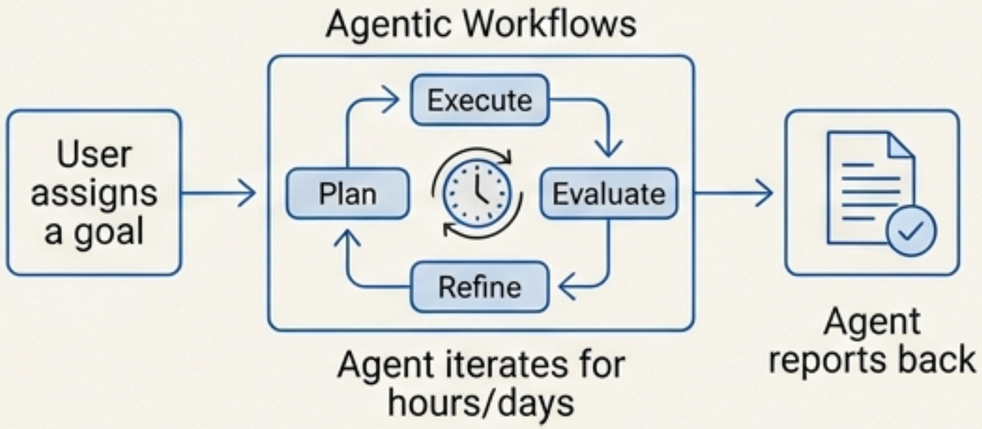
From Chat to Work

Now



Future





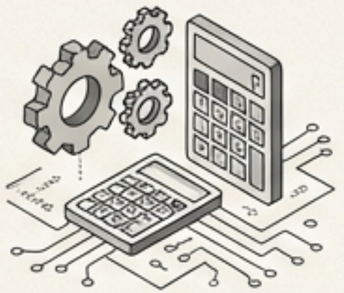
The human role shifts from ‘Doer’ to ‘Manager’ of digital workers.

Practical Guide: Which Model When?

Creative & Knowledge Retrieval

Use: **Standard Models** (GPT-4o, Claude Sonnet).

- Fast, fluent, good at SFT imitation.



Complex Logic, Math, & Coding

Use: **Thinking Models** (OpenAI o1, DeepSeek R1).

- Uses 'Thinking Tokens' to error-correct and backtrack.

Heuristic: If the task requires a 'gut check' or style, use Standard. If it requires a 'calculation' or verifying truth, use Reasoning.

Summary: The Evolution of Digital Intelligence

Pre-Training
(Knowledge)

SFT
(Persona/Format)

RL
(Reasoning)

"We are in the early days of 'Textbooks for AI.' The field is moving from imitation to genuine problem-solving. Treat these models as probabilistic tools, not oracles."