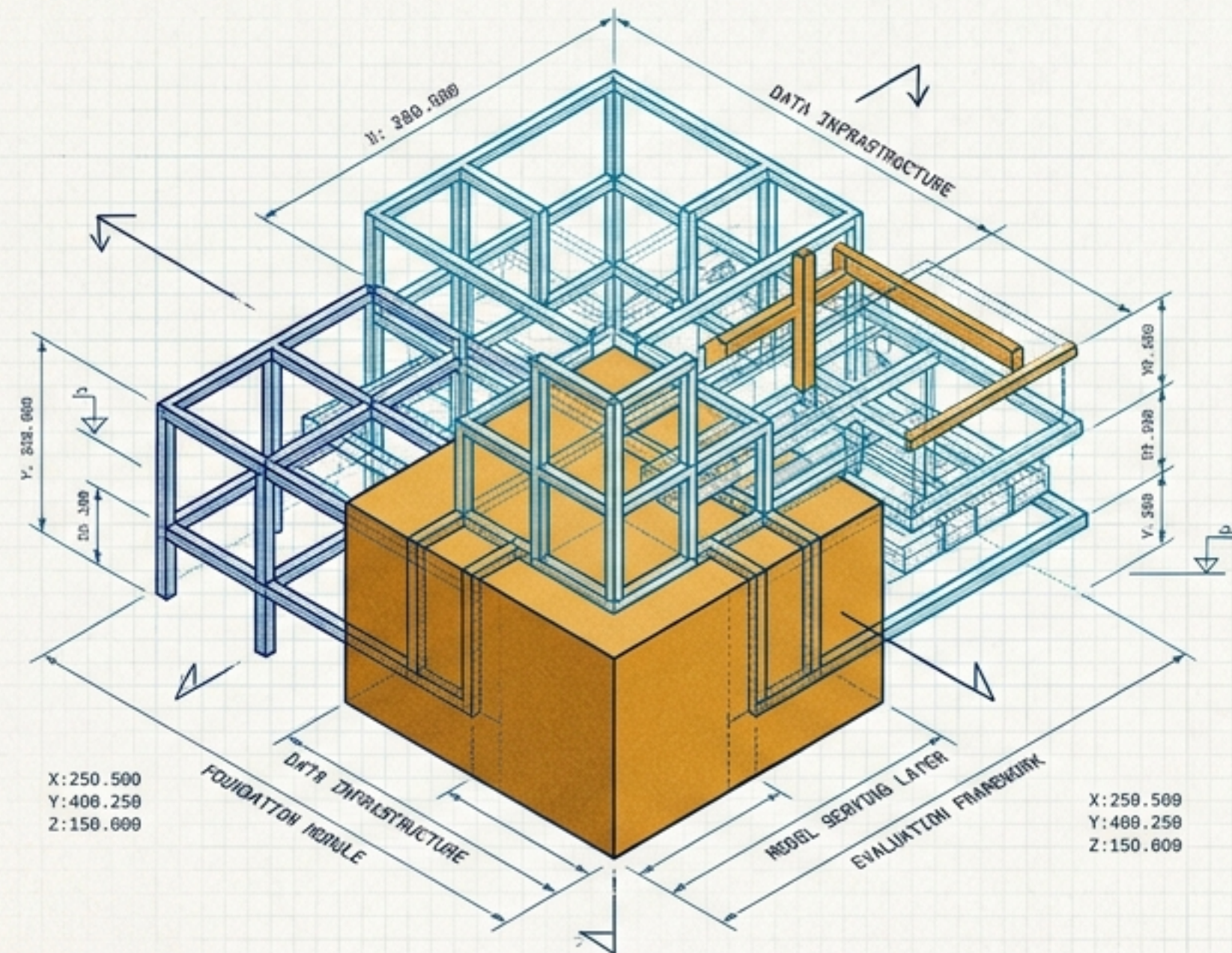


Architecting the AI Era

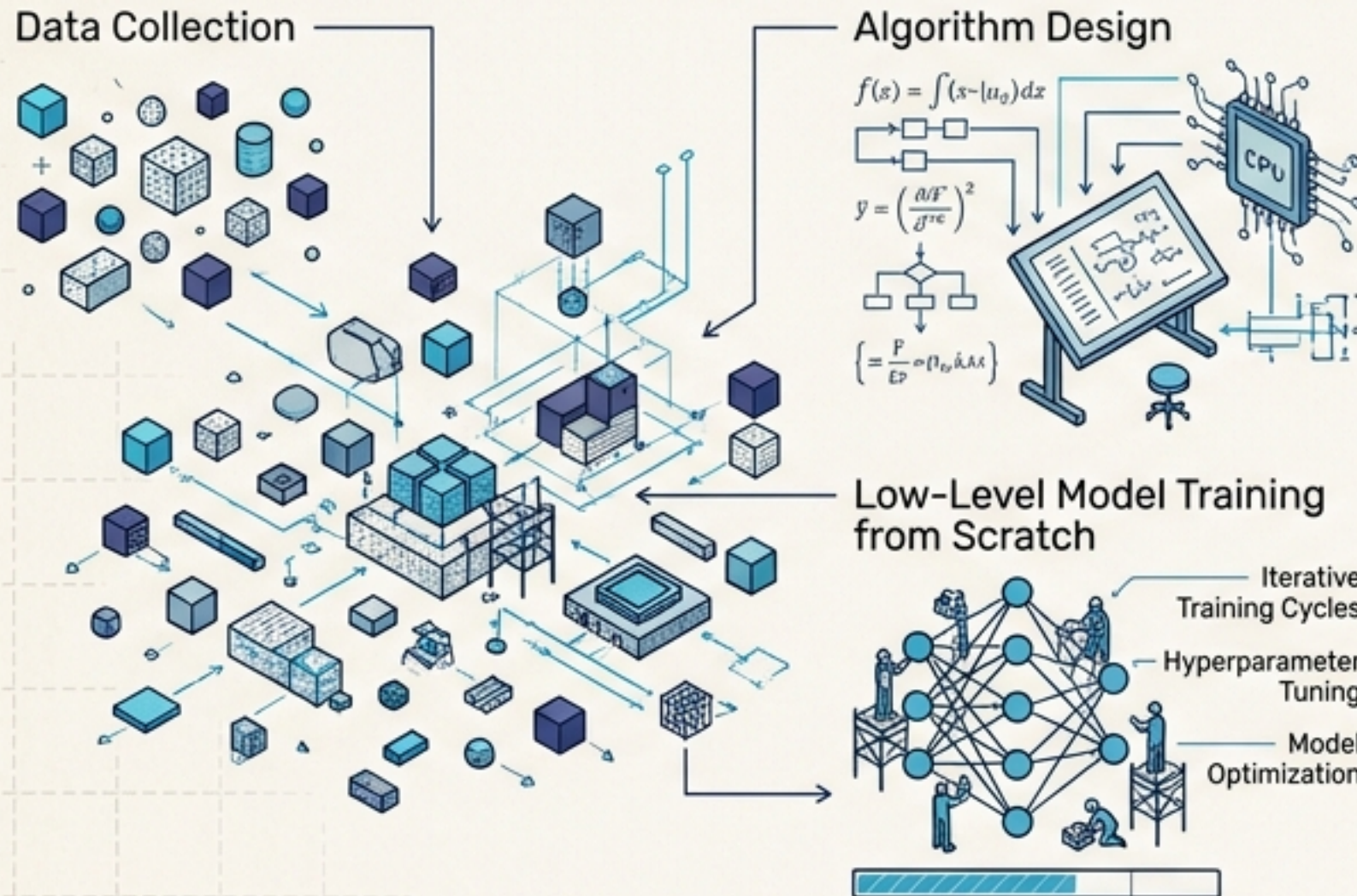
Core Frameworks from Chip Huyen's AI Engineering



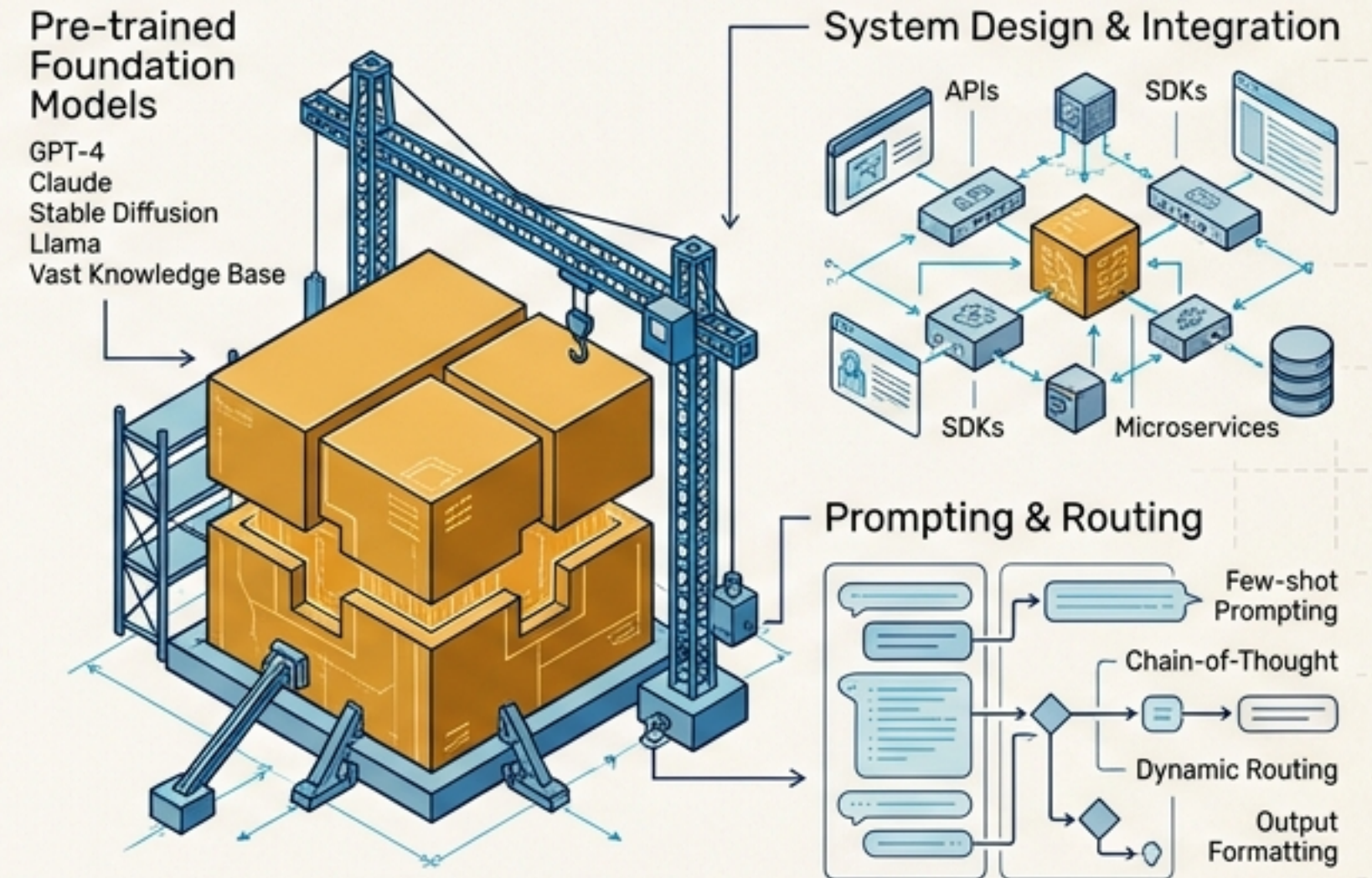
Based on the book review by Mayuresh Shilotri

AI Engineering is a Fundamentally Distinct Discipline

Traditional ML Engineering

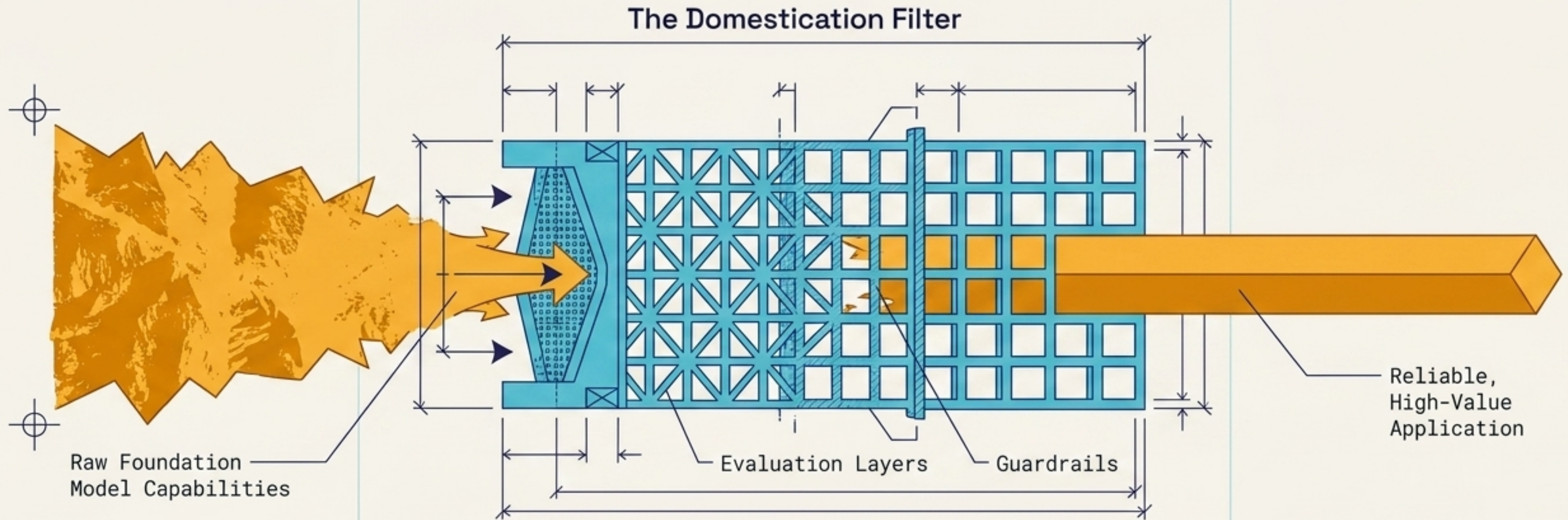


AI Engineering



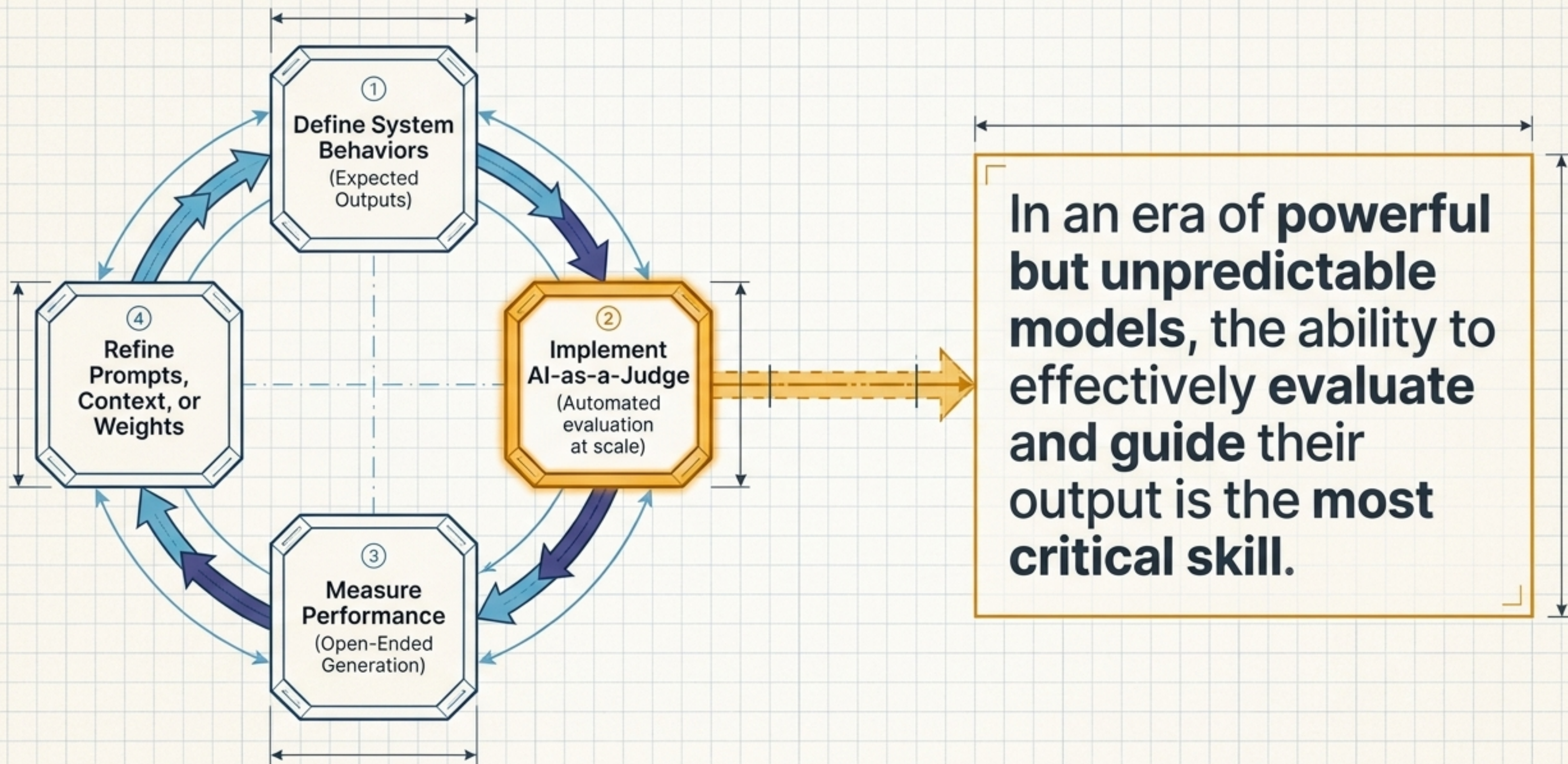
For most developers, the challenge is no longer about creating models, but about effectively harnessing the immense power of pre-trained foundation models.

The Challenge is No Longer Creating Power, But Harnessing It



To transform unpredictable generative models into production-ready software, engineers must build robust evaluation pipelines to domesticate raw capabilities.

Evaluation-Driven Development (EDD) is the New Cornerstone



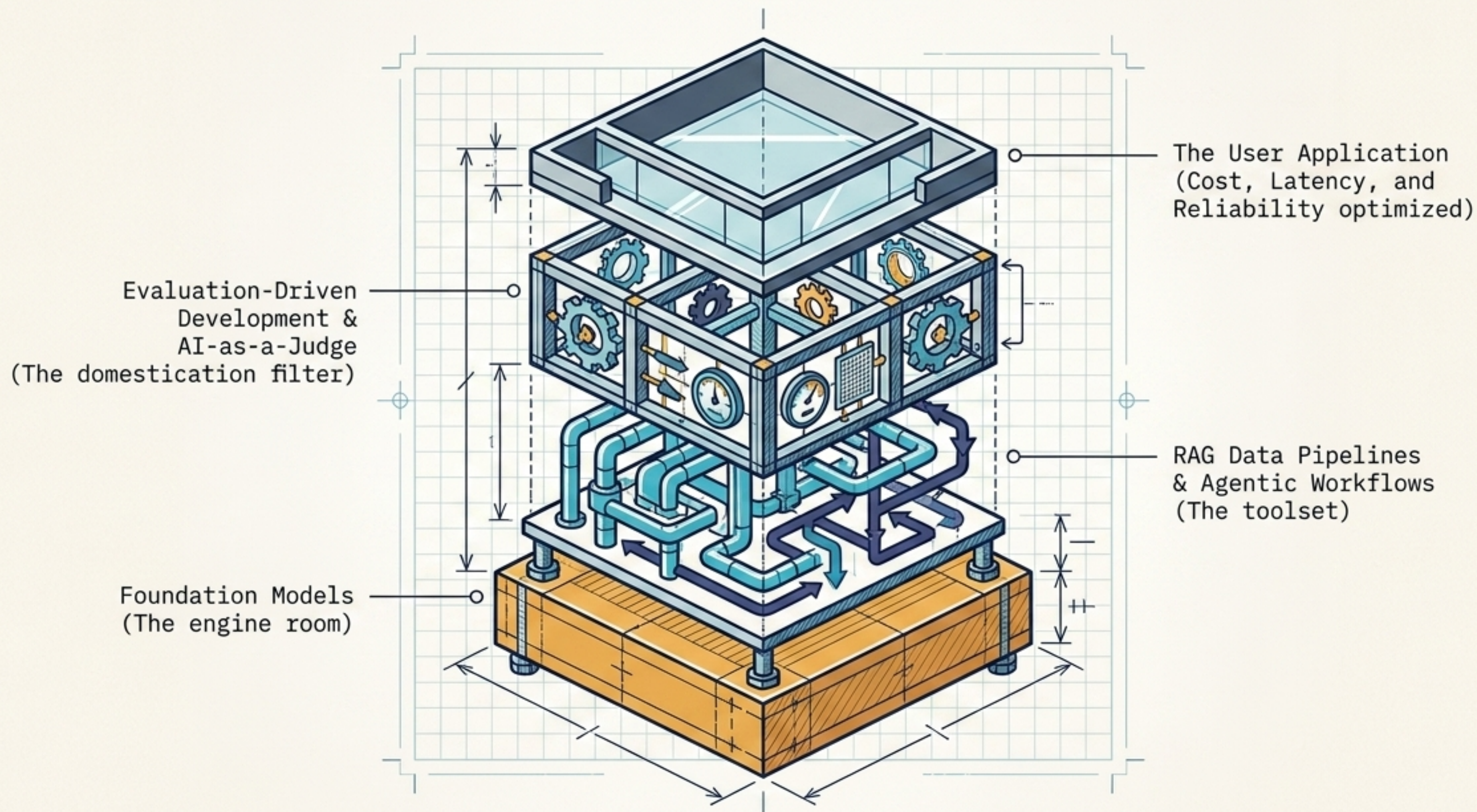
The Strategic Toolkit Requires Intentional Trade-Offs

	RAG (Retrieval-Augmented Generation)	Fine-Tuning
Primary Goal	Address Knowledge Gaps	Correct Behavioral Issues
Mechanism	Dynamically injecting external context at inference	Adjusting underlying model weights
Engineering Focus	Optimizing vector search and data pipelines	Curating high-quality training datasets



Modern builders must also integrate Prompt Engineering and Agents while continuously balancing cost, latency, and reliability.

The Blueprint of a Modern AI Application



The Definitive Guide for the Pragmatic Builder

READ THIS IF...

- ☑ You are a software engineer building applications leveraging LLMs.
- ☑ You are a PM or tech leader responsible for AI strategy and execution.
- ☑ You need to make informed architectural decisions (e.g., RAG vs. Fine-Tuning).

SKIP THIS IF...

- ☑ You are an AI researcher looking for novel model architectures or training algorithms.
- ☑ You are a beginner seeking step-by-step coding tutorials.

"A crystal-clear, strategic guide to the modern AI stack and a must-read for any engineer building in this space." — Mayuresh Shilotri