SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Token Economics:

What Enterprise AI Is Costing

*Per-token prices fell **99%** since 2022. Enterprise bills tripled anyway.*

READ

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Only 28% Can Prove It's Working

28%

report measurable AI value

Deloitte, Jan 2026

60%

generate no material AI value

BCG, 2026

Invoices are climbing.
Outcomes aren't.

Token economics: what AI produces per token, not just what it costs

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Where the Waste Lives

This is what happens when agents run in production, unsupervised against live systems.

Agents calling LLMs directly consume **50–500x** more tokens than a single-turn interaction, with no proportional gain in output

Retrieval pipelines attaching entire documents to every model call.

Background monitoring agents burning tokens against every system event, **including events with no human reviews.**

No **per-workflow** measurement of token cost against any defined business outcome.

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Four Fixes. One Discipline.

Match task complexity to the AI model cost, not the first engineer's default.

Swap full documents for trimmed context and cut spend 60 to 90%.

Track cost per resolved query, not just total token spend.

Set ROI thresholds and chargebacks before deployment, not after the fact.

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Not Every Task Needs a Frontier Model

The cost spread between cheapest and frontier models runs to 4,500x. Routing correctly closes most of that gap

The task split

Every workflow breaks into discrete tasks: retrieval, classification, reasoning, validation and summarisation.

The right-size match

Each task is matched to the model sized for its accuracy requirement, nothing larger

The single route

Every task routes once, directly, at the lowest cost that still clears the bar.

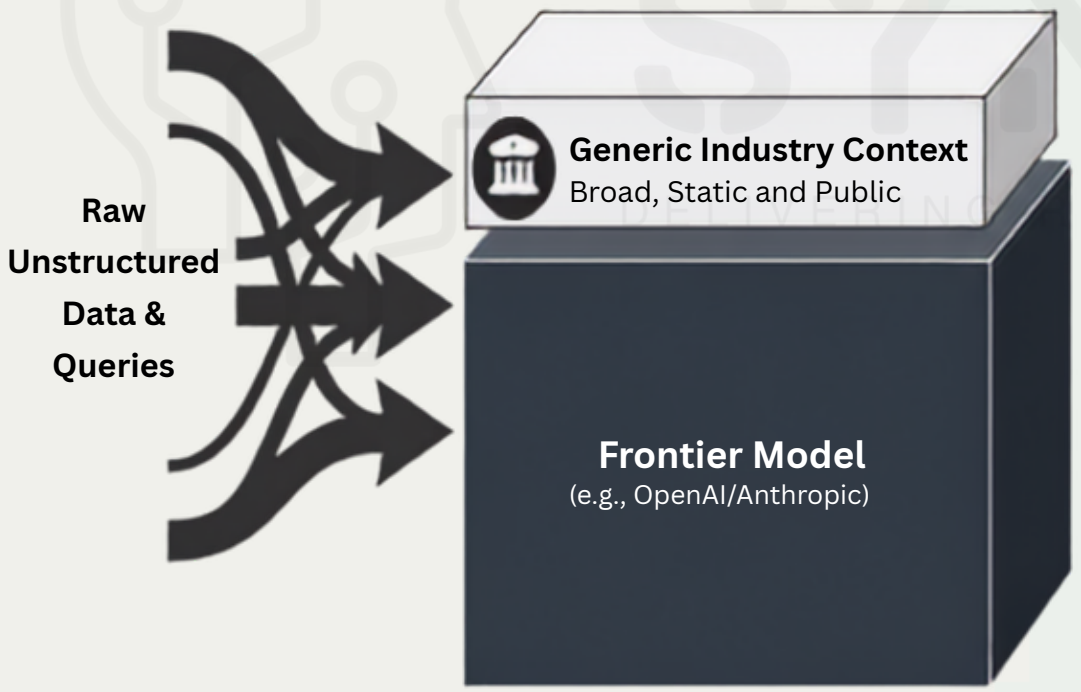
SynaptyX built this discipline into Lattice, the engine behind every engagement we run.

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

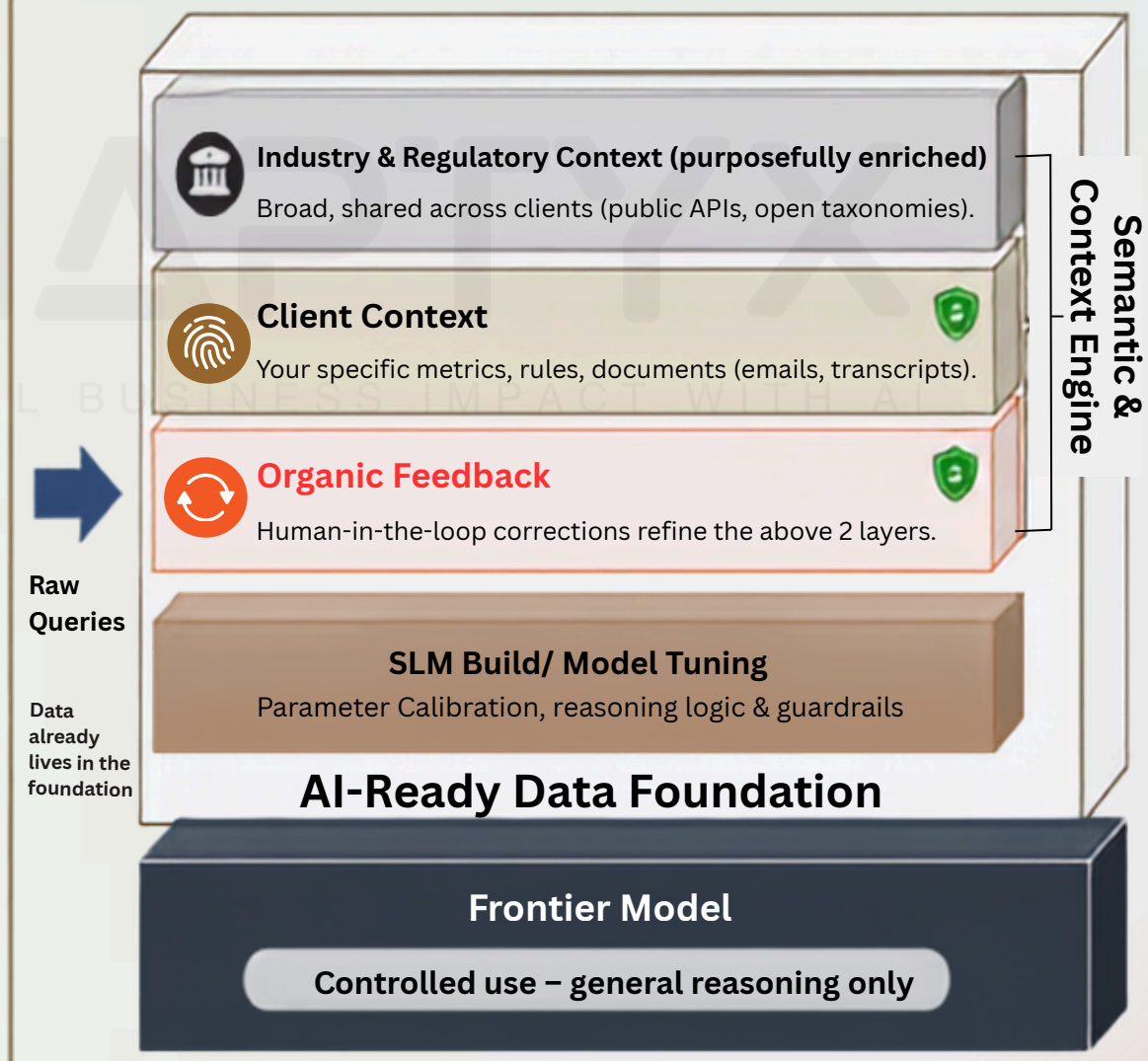
From Raw Pipe to Right-Sized Spine

The “Raw Pipe”: Inefficient Monolithic Design



- Heavy Token Burn
- High API Latency
- Vendor Lock-in
- Generic Indefensible Outputs

Your Right Sized AI Spine



- Token Efficiency/ API Cost Savings
- Faster Response
- Model Agnostic
- Deterministic Answers

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Every Task Gets Checked. Every Cost Gets Counted

SynaptyX builds this into Lattice: cost per outcome as a first-class metric, not a post-deployment audit finding

Pre-deployment testing:

A full pipeline test suite runs before every release, and reruns on every model or prompt change.

Async quality checks:

Safety and quality are scored on every response, without adding latency.

Inline correction:

A separate judge model catches failures and triggers an automatic retry.

Glass Box Guarantee

Every SynSights output is validated, traceable, and explainable.

Built on a 70/30 Foundation

70% people, process, governance & data. 30% technology.

SYNAPTYX

DELIVERING REAL BUSINESS IMPACT WITH GENAI

Ready to Turn Spend Into Outcomes?

The real cost is the spend producing nothing measurable and that's exactly what we turn into outcomes you can prove.

CORP@SYNAPTYX.AI

or Visit - www.synaptyx.ai