

A WHITEPAPER · MAY 2026

The Enterprise *Logic Layer*

*Platform, Playbooks, and The CPG Brain:
An Architecture for the Next Wave of Enterprise AI*

GURNOOR DHILLON · SANDEEP RAMESH · AXEL WEHR · EMEKA KALU-UMA

Co-founders, CalvinBall

With a personal reference from the Lead Architect

EXECUTIVE SUMMARY

The Enterprise *Logic* Layer

The first three years of enterprise AI have delivered a large volume of deployed tools and a strikingly small volume of changed decisions. The Global 2000 has bought copilots, retrieval systems, and agent frameworks at scale. None of these has altered, in any systematic way, the decisions a Fortune-500 company actually makes.

This paper argues that the failure is architectural rather than technological, and that the category that succeeds it is already taking shape.

We call the emerging category **the Enterprise Logic Layer**. It is not a better copilot and it is not a foundation model. It is the missing tier between a company's data and its business decisions — the place where business judgement is encoded, governed, and executed as a first-class system asset. We have been building it for two years, and the paper lays out four arguments that shape what we have built.

First, the category is best understood as platform plus playbooks. The platform is a three-tier technical architecture — IRIS for ingestion, MISL for reasoning, Tapestry for orchestration. The playbooks are industry-specific decision workflows that sit on top of it. The platform strategy makes the economics work. The playbooks are what the customer buys.

Second, the architecture is private by design. At the core of MISL sits a Federated Supervisor that enforces complete separation between enterprise data and the system's own compounding intelligence. The system becomes smarter over time on anonymised interaction patterns, not on enterprise data. This is not a contractual promise. It is the piece of the architecture that makes leaking enterprise context a violation of the system's architecture, not merely a violation of its policy.

Third, the moat is neither the platform nor the playbooks in isolation. The moat is the industry context — the decision grammar of a specific industry, encoded through deep embeds with anchor clients and governed by a council of industry leaders who have put their own capital behind the work. Horizontal AI players will never build this, because the economics of horizontal businesses forbid depth. Vertical AI players who do build it will compound for a decade.

Fourth, the right way to price and sell in this category is services-as-software. The product is not a tool used by a human team; it is the output the team used to produce. Sequoia named this pattern in mid-2024. We have been building against it since before the name existed.

The paper that follows lays out the structural failure of the copilot wave, the three-tier platform architecture, the private-by-architecture design, the construction of the first industry brain — the CPG Brain — through two years of stealth co-creation with Unilever, LVMH, Colgate, and CPG industry Architects, the four shipped playbooks (Brief, Ask, Why, Signal) with live enterprise evidence, the diagnostic framework enterprise leaders can use to evaluate their own AI portfolio, and the roadmap across adjacent industries.

We publish this paper as the founding technical document of the category. We welcome stress-testing in public and intend to do the same with our own work over the coming year.

The Structural *Failure* of the Copilot Wave

1.1 The forty-two tools problem

In Q1 2026 we were invited to review the AI portfolio of a top-five global consumer goods company. The company had, by their own count, forty-two AI tools in production across functions. A retrieval copilot in every department. An agent framework under evaluation. A generative-AI centre of excellence with real budget and a real headcount plan.

We asked the sitting Chief AI Officer a single diagnostic question: how many decisions has your company made differently in the last twelve months because of any of it? The honest answer was that no one could name more than three with confidence. Two of those three were arguable. The real answer was zero or close to it.

This pattern is not unusual. It is the norm. In the past nine months we have repeated this exercise with Chief AI Officers, CFOs, and board members at seven Global 2000 companies across four continents. None has been able to name five. Several have been unable to name one.

The Global 2000 is about to collectively write down a meaningful fraction of its AI investment. Tens of billions of dollars of copilot and RAG infrastructure currently on the books will be materially impaired in the next twenty-four months. This is not a bearish call on enterprise AI. It is a directional call on the current architecture of enterprise AI — and, more importantly, the early signal of a much larger opportunity the first wave has failed to address. We seize that opportunity in §1.3.

1.2 Why retrieval was never going to be enough

For thirty years enterprise software has been built on a simple contract: the software holds the data, the human holds the logic. ERP systems, customer relationship management, data warehouses, business intelligence dashboards. All of them are elaborate machines for putting information in front of a human expert, who then applies judgement. The judgement lives in people's heads, in tribal knowledge, in the seventeen-years of experience category manager who knows which promotions work in Ramadan and which do not.

Large language models broke this contract in exactly one direction. They made the retrieval of information conversational, fast, and natural. They did not — and, in isolation, cannot — capture the logic that turns information into a decision. A foundation model knows the

internet and nothing about how a given company decides whether to accept a trade-spend increase from a distributor in Rio De Janeiro!

This is the structural limitation. Every AI tool built on top of a pure retrieval architecture inherits it. The copilot can find the right document; the copilot cannot make the decision the document exists to inform. The gap between retrieval and decision is the work. The work is where the money is. And the work cannot be captured by retrieval alone regardless of how sophisticated the retrieval becomes.

1.3 The ~\$1 trillion human middleware

Enterprises already know this, even if they have not named it. The gap between raw data and executive decision is currently bridged by a global workforce we estimate at approximately \$1 trillion in annual spend — a figure triangulated from public data across three categories. Global management consulting represents approximately \$450 billion. IT services and systems integration add a further \$500 billion. The fully-loaded cost of internal BI, analytics, and planning headcount across the Global 2000 contributes another \$300–400 billion. Aggregated conservatively, the midpoint across these categories is \$1 trillion; the true figure is almost certainly higher. Approximately \$400B of this spend is internal — the BI analysts, insight teams, planning functions, pricing committees, and trade-approval workflows that translate system-of-record data into executive decisions. Approximately \$600B is external — the management consultants, systems integrators, and agencies that enterprises pay to do the same work when the internal capacity falls short.

This is human middleware. It is the single largest category of spend the generative-AI wave promised to disrupt and has, so far, entirely failed to move. The reason is that copilots make the human middleware faster. They do not make it unnecessary. A twenty-percent productivity gain on a \$1 trillion spend is a meaningful number. A category that replaces the spend entirely is a trillion-dollar opportunity in present value. Our view is that the trillion-dollar opportunity is the actual prize.

SECTION 2

The Category: *Logic*, Not Knowledge

2.1 Two categories, not one

Every enterprise AI tool sold today falls into one of two categories, and the industry has spent three years pretending there is only one.

Knowledge tools answer the question “where is the information?” Enterprise search, RAG-based copilots, document Q&A, most “agents” as currently demoed. Their output is a retrieval: a passage, a summary, a surfaced document. They are useful. They are also not what anyone actually means when they say “AI transformation.”

Logic tools answer the question “what should we do?” They do not retrieve; they reason. They encode how a company’s best people actually make decisions — the trade-offs, the thresholds, the sequence, the escalations — and they execute that reasoning at machine speed against live data.

The first wave of enterprise AI is knowledge tools dressed up as logic tools. That is why the pilots don’t graduate.

2.2 The three properties a logic tool must have

A logic tool, to deserve the name, must satisfy three minimum properties:

- **Reasoning-native, not retrieval-native.** Given a situation, it must execute an inspectable chain of judgement and produce a defensible answer. The chain is not a black box. It is versionable, testable against historical cases, and correctable without re-training a model.
- **Composable across decision classes.** The same substrate that runs trade promotion must also run demand planning, new-product launch, and pricing exception. Forty-two copilots with private reasoning is forty-two projects. One logic layer with graphs over it is one asset that compounds across decision classes.
- **Accountable.** Every decision produced must be traceable to the reasoning that produced it, to the data that informed it, and to the human expert whose judgement that reasoning encodes. Without this, regulated industries will never adopt at scale, and unregulated industries will discover within eighteen months why they should have behaved as if they were regulated.

Any tool that does not satisfy all three is a knowledge tool. Most of what is sold today as “decision intelligence” is a knowledge tool with a deterministic post-processing step. The category is not being built there.

2.3 Harvey as precedent, not exception

The vertical-AI thesis is not ours. Harvey proved it for law. They did not build a better legal-research copilot; they built the product that replaces the work of the junior associate, priced it against the associate’s salary, and sold it into firms whose partners had already exhausted their patience with generalist tools. They made a dent in the category not because they had better underlying models — they did not — but because they had depth a horizontal player cannot afford to build.

The question for every other industry is who builds the depth first. The firms who still believe a horizontal copilot will eventually understand their industry resemble, in a precise way, the firms who in 2015 believed an in-house data team could keep pace with a cloud data warehouse. They were not wrong about the capability of their team. They were wrong about the category they were in.

The *Platform*: IRIS, MISL, Tapestry

3.1 Why platform-first matters to the economics

CalvinBall is a platform, not a product. This distinction is not cosmetic. A platform amortises its development cost across every playbook built on top of it. A product does not. The reason the unit economics of CalvinBall improve with scale — and the reason our fastest deal has been four months against an industry average of eighteen — is that every customer benefits from, and contributes to, work done for prior customers on the same platform.

The platform has three tiers.

3.2 IRIS: the ingestion tier

IRIS takes in the messy middle of enterprise data — the PDFs, Excel files, working drafts, WhatsApp screenshots, category-manager notes — alongside the structured feeds from ERP, distributor management systems, Nielsen, Kantar, Google, Meta, Amazon. It produces a usable semantic representation of the enterprise that downstream tiers can reason over.

The observation that drove IRIS's design is that the \$1T human middleware exists, primarily, to do this ingestion work by hand. When enterprise AI fails in production, it almost always fails at the ingestion boundary. A retrieval copilot wired to a curated document store performs impressively in a demo. The same copilot wired to the actual mess of a Fortune-500's operational data environment performs poorly, because the mess is the job.

IRIS is explicitly designed around the mess. It is permissive about input formats, deterministic about output schema, and governed by MISL-defined dimensions and metrics so that no downstream component hallucinates a relationship that does not exist.

3.3 MISL: the Machine Intelligence Semantic Layer

MISL is the custodian of truth. Every dimension, fact, and metric the enterprise uses is defined once in MISL and served deterministically to every downstream engine. Dimensions are typed (product, distributor, channel, region, period). Metrics are derived from typed columns using parameterised SQL rather than model-generated SQL. Business logic — the conditions under which a promotion is eligible, the thresholds above which approval is required, the precedents that override a default — is encoded as versioned logic graphs owned by domain experts.

MISL is deliberately hybrid. It is symbolic where decisions must be auditable: the logic graphs that determine trade-promotion eligibility are inspectable, editable by a business

owner, and testable against historical cases. It is neural where pattern-matching adds value without compromising auditability. The division is architectural, not incidental. Pure-neural systems are not auditable at the level that regulated or high-stakes enterprise decisions require, and will not be for some time. Pure-symbolic systems are auditable but brittle and expensive to author. MISL is the hybrid that clears both tests.

Four engines sit on top of the MISL core:

- **The Fetcher** handles deterministic high-speed reporting — known-entity lookups, standard metrics, pre-registered queries. No LLM guessing.
- **The Agent** handles ad-hoc exploratory reasoning through a constrained ReAct loop with a schema guardrail that forbids the model from inventing tables or relationships.
- **The Comparative Engine** solves the year-over-year arithmetic problem that breaks generic LLMs, using GROUPING SETS semantics to handle multi-level rollups and full outer joins across periods in a single pass.
- **The SQL Functions library (“Playground”)** lets partners build hard-coded logic once and register it as a reusable skill the Agent can call, avoiding per-query regeneration of business logic.

A typical MISL deployment contains 40–200 logic graph nodes per decision class. In our current CPG deployments, 14 of 22 in-production logic graphs are owned and can be edited by business users — not by CalvinBall, not by the customer’s engineering team. Business-user ownership is the leading indicator of a deployment that compounds. Engineering-owned deployments stall at the second decision class.

3.4 Tapestry: the orchestration tier and the deterministic interface

Tapestry is the deterministic interface to the entire platform and the tier that makes everything above it defensible to an auditor. Every decision Tapestry produces is traceable, versioned, and reconstructable — the full chain of reasoning, the data inputs, the logic-graph version, and the policy state at time of decision are all captured as a single audit record.

Tapestry binds IRIS and MISL into deployable playbooks and composes those playbooks into end-to-end decision flows. A new-product launch decision touches demand planning, trade promotion, pricing, and supply. Each of those has its own logic graph, its own owners, its own cadence, its own data sources. Tapestry is the wiring harness that routes context and logic across all four, handles human-in-the-loop escalation, reconciles conflicts between recommendations, and enforces the organisation’s policies about which decisions must be reviewed by whom under what conditions.

The practical effect of the three tiers together is that a new business problem — for example, extending a deployed playbook from trade promotion into pricing-exception management — does not require a new vendor, a new integration project, or a new tool. It requires a new logic graph in MISL, drawing on the context IRIS already supplies, orchestrated through Tapestry patterns that already exist. This is the compounding property that copilot portfolios structurally cannot have.

3.5 Tapestry Studio: configuration as control

The authoring surface of the platform is Tapestry Studio — a no-code console where business users define metrics, dimensions, and logic in one place, with instant propagation to the Fetcher, the Agent, and the Playground. A metric defined once in Studio updates every downstream engine; there is no synchronisation step. Security is enforced at the namespace level with row-level security injected via AST rewriting, and forbidden-rule constraints the SQL agent must obey. This is how enterprise-grade multi-tenancy is done in practice.

Private by *Architecture*: The Federated Supervisor

4.1 The problem every Chief Artificial Intelligence Officer (CAIO) is actually worried about

Every CAIO we have worked with has the same private worry, and they are right to have it. They have read enough fine print to know that “private endpoint” and “we will not train on your data” are two very different claims. Most enterprise AI deployments, even behind a private endpoint, can leak enterprise context into the vendor’s training pipeline in ways that do not become visible until an audit eighteen months later. The distance between the contractual promise and the architectural reality is where enterprise AI deployments quietly fail their compliance review.

The current market answers this worry with legal language. Terms-of-service commitments, data-processing addendums, SOC-2 attestations. These are necessary and they are insufficient. A commitment is a promise. Architecture is a constraint. For regulated industries, and increasingly for any enterprise handling consumer data, only the constraint is credible.

4.2 The Federated Supervisor AI agent—architectural, not contractual

At the core of MISL sits the Federated Supervisor. It is the component that enforces complete separation between enterprise data and the system’s own compounding intelligence. Its guarantees are three:

One. Enterprise data never leaves the enterprise boundary. Customer data — documents, decisions, logic graphs specific to the customer’s operations, sensitive operational inputs — is processed, stored, and reasoned over within the customer’s own environment or a dedicated per-tenant deployment. It does not flow to CalvinBall’s shared infrastructure. It does not flow to any foundation-model vendor we integrate with. It is addressable only within the customer’s namespace.

Two. The system’s compounding intelligence learns on anonymised interaction patterns, not on enterprise data. The reinforcement-learning loop that makes CalvinBall better over time — the reason each customer benefits from the platform’s accumulated experience — operates on patterns of how the system was used: the shape of a

query, the structure of a correction, the distribution of human overrides. These patterns are decoupled from the data that produced them. They are anonymised, aggregated across the platform, and the Supervisor is the piece of the architecture that enforces the decoupling as a rule the code cannot violate.

Three. The compounding asset lives outside any individual enterprise boundary. The platform-level learning — the part of CalvinBall that becomes more valuable every month, the moat CalvinBall is building over time — is governed by the Supervisor and exists as a shared asset across customers. No customer’s data is in it. No customer’s data contributed to it. The asset exists because of the aggregate interaction pattern across many customers, none of which can be reconstructed from the asset itself.

The design principle is that trust is not a promise the vendor makes. Trust is a property of the CalvinBall architecture.

An enterprise security team that wants to audit the Supervisor can do so at the code level. The guarantees hold because the code enforces them, not because we have committed to behave.

4.3 Why this matters commercially

Two practical consequences follow, both of which shorten the sales cycle and widen the top of the funnel.

First, the compliance conversation gets shorter. A regulated-industry CFO — in pharma, financial services, healthcare, regulated consumer — can approve CalvinBall on a timeline measurably faster than the procurement cycle for a horizontal copilot of equivalent scope, because the architectural proof replaces most of the contractual negotiation.

Second, and more strategically, it takes an entire category of objection off the table. When a horizontal copilot sales team says “we don’t train on your data”, the sophisticated buyer’s next question is “prove it.” When we say the same thing, the answer is “audit the Supervisor.” The conversation ends in a place where we win by construction rather than by negotiation.

4.4 Why a horizontal player cannot copy this

The Supervisor is difficult to replicate for horizontal players for a non-obvious reason: their business model depends on the leak. A horizontal copilot improves, disproportionately, by learning across every customer’s data. To commit to complete data isolation is to surrender the compounding mechanism that makes them valuable. They can commit in law (and many do). They cannot commit to architecture without giving up the economics that make them horizontal.

We are able to commit in architecture because our compounding mechanism is different. We learn from interaction patterns, not from content. That is not a constraint we accepted

reluctantly; it is the design principle we chose because it is the only principle that makes the category work for regulated industries, and it is the one the incumbents structurally cannot match.

The *Moat*: How The CPG Brain Was Built

5.1 Why the moat is not the platform

We want to be precise about where the moat sits, because this is the question every investor and every serious operator asks us.

The platform is not the moat. The platform is the minimum competitive ticket. A determined, well-funded horizontal player could build a comparable platform in eighteen to twenty-four months. We have an advantage of eighteen months there; we do not have a decade there.

The moat is the industry context — the decision grammar of a specific industry, encoded in logic graphs and governed by the people who actually run the industry. Call it, in our case, The CPG Brain. It has two components, built deliberately and in stealth over two years.

5.2 Component one: co-creation with anchor clients

The first component is the decision grammar co-created through eighteen-to-twenty-four-month deep embeds with three anchor customers: Unilever, LVMH, and Colgate. Not pilots. Not proof of concept. Embedded work in which the customer's own category managers, insight leads, commercial heads, and supply planners helped prioritise & stress-test the logic graphs we had built first infused with domain expertise that now sit in MISL.

Two observations from this phase shaped everything that followed.

First, the logic of how a company decides is not discoverable from its historical decisions. The historical record is saturated with past exceptions, dead market conditions, political overrides, and one-time special cases that would be actively wrong to replicate. It took, in the first CPG deployment, four months of repeated interviews, structured elicitation, and parallel-run disagreement analysis to separate the logic from the artefacts. Organisations attempting this work with a light-touch elicitation approach will fail. We did, on our first pass, and we rebuilt.

Second, logic-graph ownership must transfer to business users or the deployment does not compound. In our first four graphs, ownership stayed with the CalvinBall team for too long. When we finally pushed ownership to category managers — with appropriate tooling, review

processes, and versioning discipline — the rate of improvement on those graphs tripled within a quarter. This is the lesson that produced Tapestry Studio in its current form.

The anchor-client co-creation is the expensive, slow half of the moat. It cannot be cloned by a horizontal player without that player also choosing to run eighteen-to-twenty-four-month embeds in the industry, at which point they have become a vertical player and have lost their horizontal economics. This is the trap Harvey identified first, for law. We have reproduced it, deliberately, for CPG.

5.3 Component two: the Architects

The second component is the Architects — a dozen Fortune-500 consumer-goods leaders who serve as the governing expertise layer of the CPG Brain. Each Architect is either an investor in CalvinBall or has contributed personal expertise to the design of the platform. Publicly named members include:

- **Antoine de Carbonnel**, former Chief Commercial Officer, Gojek; former senior leader, McKinsey (Lead Architect, and the author of the reference letter published alongside this paper)
- **Dharnesh Gordhon**, former CEO, Nestlé Indonesia; former CEO, Godrej International
- **Francois Walewski**, Strategy Director, Bacardi
- **James Lafferty**, four-time Forbes Top 100 CEO (P&G, BAT, Coca-Cola)
- **David Tan**, Regional Director, Sales & Marketing, Panasonic Avionics Corporation
- **Vittoria Gambirasi**, Managing Director, BrandHack.ai; ex-Nestlé

The Architect roster and a personal reference from the Lead Architect are published alongside this paper.

The Architects serve three functions.

First, it contributes to the decision patterns the co-creation work does not surface. The political overrides that never make it into a deck. The board-level escalations that follow unwritten rules. The conditions under which a written rule gets broken, on purpose, for reasons only an ex-CEO can explain. These are encoded into MISL as governance-tier logic, versioned against named contributions, and reviewed by the Architects quarterly.

Second, it opens doors. The Architects make introductions. CalvinBall's fastest enterprise close is four months, against an industry average of eighteen, almost entirely because our top-of-funnel conversation is an Architect-sponsored introduction rather than a cold-email motion.

Third, and critically, majority of the Architects have put their own capital into CalvinBall as a lead angel investor. The distinction between an advisor and an angel matters. An advisor offers an opinion; an angel backs it with their own money and their own reputation. The Architect Reference — a personal reference letter from the Lead Architect, written in his own words, describing why he invested and what he sees is published alongside this paper.

5.4 Why a horizontal player will never build this

There is a tempting argument made by several foundation-model labs in private conversations with CPG CIOs ; that horizontal capability will eventually make vertical context unnecessary. The argument is wrong on three counts.

First, the CPG Brain is not made of text a foundation model could train on. A substantial fraction of its value is in the logic graphs authored by domain experts who worked with CalvinBall at its inception, named category managers and former CEOs, bound to specific industry conditions, versioned against specific historical decisions. This is not content that lives on the open web. It is content that exists, today, in the monthly minutes of a Chatham-House-rule council and in the change log of a governed MISL deployment. No training corpus contains it, and the people who produced it are not in the business of publishing it.

Second, even if the content existed, the economic trap holds. A horizontal player who commits to the depth required in CPG cannot also commit to the depth required in law, pharma, energy, corporate finance, industrial distribution, and healthcare. The horizontal economics that make them horizontal forbid it. They will pick one or two verticals at most; in every other vertical, they will cede the category to whichever vertical player has been compounding for three years by the time the horizontal notices.

Third, the governance structure — the Architects — is not reproducible by a check. It is reproducible only by twenty-four months of credible commitment to an industry. A horizontal's quarterly priorities do not survive that length of commitment. A vertical's entire existence is that commitment.

SECTION 6

The *Playbooks*: Brief, Ask, Why, Signal

6.1 Products are outcomes, priced against the work they replace

CalvinBall does not sell features. We sell playbooks (specific user workflows) — units of industry expertise packaged as software, priced against the human headcount they displace, delivered as services-as-software in the precise sense Sequoia described in 2024.

The naming convention is deliberate. Playbooks are named for the outcome they deliver in the voice the executive already uses — Brief, Ask, Why, Signal — not for the function they perform. When a CMO says to her team “did you read the Brief?” the product has earned its place in the executive vocabulary. The feature work, and the category noise, disappears behind the outcome.

Four playbooks are in production today. Each has live enterprise evidence. Each replaces a specific human workflow, priced against that workflow’s headcount cost.

6.2 Brief: the executive briefing

Outcome: the leadership team walks into the 8 a.m. review with a pre-assembled, auditable briefing that replaces the weekly work of the BI analyst cohort.

Replaces: the multi-person reporting team (15–25 analysts in a typical Global 2000 CPG) whose work is consumed in data extraction, deck preparation, and pre-read distribution.

Evidence: a global CPG giant deployed Brief to replace the work of a 17-person reporting team. Outcome: 10× faster decision-making on recurring review cadences, material reduction in meeting-preparation overhead, and a strategic-focus shift the leadership team attributed directly to the playbook.

6.3 Ask: the concierge

Outcome: any executive can get a precise, schema-disciplined answer to any business question in natural language, in seconds, with the reasoning attached.

Replaces: the ad-hoc analyst queue — the cohort of insight leads and junior analysts who field one-off questions from the business and return answers in days.

Evidence: a \$3B Indonesian food & beverage leader, ingested 700M+ data records into CalvinBall's stack and saved 4,000+ human hours per month across its sales intelligence function, with ability to ask cross-context questions for sales performance and unlocking real-time anomaly detection across channels.

A global beauty brand using Ask for D2C operations reported zero dedicated D2C headcount against an operation saving 100+ hours per week.

6.4 Why: the diagnostic

Outcome: when a metric moves, the system returns a causal explanation the executive can act on — traced across SKU, channel, region, and period back to the data and the logic that produced it.

Replaces: the root-cause-analysis function typically performed by an external consultant, a regional analyst, or the category's head of insights.

Evidence: India's #2 FMCG (\$10B revenue, 22 categories, 64 data environments), is replacing 170 BI analysts' manual data work with CalvinBall's diagnostic stack. Outcome: quarterly distribution interventions converted into a real-time dynamic root cause at a node level, an estimated \$10M of value creation over three years, projecting 5× decision velocity, 4.62× ROI in year one, insight resolution from seven days to one hour.

6.5 Signal: the early-warning system

Outcome: always-on monitoring of the metrics that matter, tuned by the client's own logic, surfacing anomalies before they become quarter-end problems.

Replaces: the manual anomaly-detection workflows typically performed by regional managers, agency partners, or internal performance teams.

Evidence: in a CPG Quick Commerce deployment, Signal moved OOS and ROAS by monitoring competitor availability, share-of-shelf, and promo intensity in real time — issuing preemptive alerts that name the SKU at risk, the cause, and the corrective action. Detection compressed from weekly review to live; recovery shifted from retrospective to in-flight.

6.6 The arc of an executive day

The four playbooks are not independent. They are the rhythm of a leadership team's day — the Brief in the morning, the Ask during the working day, the Why when something moves, the Signal always on in the background. When a future playbook ships — Plan, Act, Price, Launch — it will extend the same vocabulary, because the principle is that playbook names are verbs the executive already uses.

6.7 Unit economics

We do not disclose pricing in this paper, but the shape of it is straightforward. A playbook is priced against the fully loaded cost of the human workflow it replaces, typically in a band of 25–50% of that cost, recurring, with margin uplift over time as The Brain becomes more efficient at the customer's edge cases. This is the services-as-software unit economics Sequoia identified. It is the only pricing structure in which the buyer is motivated, the CFO is defended, and the vendor compounds.

SECTION 7

Enterprise *Traction* and Architect-Led Distribution

CalvinBall operates today with:

- **\$440K in 2025 invoiced revenue**, a 6× year-on-year growth rate, on essentially zero outbound GTM investment.
- **\$10M+ in qualified ARR pipeline** across Hershey's, Mondelez, bp, Indofood, Telkomsel, Wipro, KPMG, and additional names under NDA.
- **NVIDIA Inception Program** membership, delivering technical validation and preferred access to tools.
- Channel partnerships with KPMG and BrandHack.AI.

The unusual feature of CalvinBall's distribution motion is that top-of-funnel is almost entirely architect-led rather than outbound-led. An Architect-sponsored introduction is the default entry into an enterprise conversation. The effect is measurable: CalvinBall's fastest enterprise close is four months, against the industry-reported average of eighteen for enterprise AI infrastructure sales.

SECTION 8

The *Diagnostic*: Five Questions for Enterprise Leaders

An interactive version of this diagnostic is available at calvinballtech.com/diagnostic.

Question one: Name five decisions. Can you name five specific business decisions your company has made differently in the last twelve months because of any AI tool in your portfolio?

Question two: Point to the reasoning. For any AI tool in your stack that claims to support a decision, can you point to the specific chain of reasoning it applied? Can a business owner edit that reasoning without calling engineering?

Question three: Test composability. Take two related decision classes — trade promotion and pricing exception, or credit approval and vendor onboarding. Can reasoning deployed for the first be reused, in part or whole, for the second?

Question four: Check ownership. What percentage of your AI tools are owned and maintained by business users rather than engineering, central IT, or an external vendor?

Question five: Audit the audit trail. Pull one decision your AI tools supported in the last 30 days. Reconstruct, from the tool's outputs alone, why it recommended what it did. If this takes more than 15 minutes, the tool is not production-grade for any regulated or high-stakes decision class.

If the honest answers to these questions make you uncomfortable, we recommend three concrete steps, independent of any vendor including us. One, stop buying new copilots for 90 days; freeze the portfolio and run the diagnostic against every tool in production. Two, identify one high-value decision class — ideally worth more than \$10M in annual P&L impact — and commit to a full logic-layer deployment for it. Three, separate Knowledge-tool evaluation from Logic-tool evaluation in your governance process.

Roadmap and Ask

9.1 Roadmap across industries

CPG is the beachhead. The Architects council in CPG is the template. Our working roadmap stands up the equivalent in two additional industries over the next twelve months. The candidate industries under active evaluation are financial-services corporate-credit decisioning, pharma commercial operations, and industrial distribution. Each would follow the same construction pattern: anchor-client co-creation, Architects council governance, and platform amortisation of the underlying IRIS/MISL/Tapestry stack.

9.2 The institutional conversation

CalvinBall is bootstrapped with \$2M+ of founder capital and backed by the Architects as lead angels. We are in active conversation with institutional investors focused on enterprise infrastructure and applied AI. Use of proceeds will deploy into the three vectors that matter: distribution (pipeline conversion, new-geography expansion, channel partnerships), application (accelerate playbook release to cover broader CPG workflows), and infrastructure (platform hardening, partner ecosystem enablement).

Direct correspondence welcomed at axel@calvinballtech.com.

9.3 The Architect conversation

For CPG leaders whose diagnostic came back uncomfortable, we open a small number of Architect-introduced conversations each quarter. The point of entry is the whitepaper and the Architect Reference. The conversation that follows is private.

Closing

We are in the early-middle of what will be a ten-year rebuild of enterprise software on AI-native foundations. The first three years were dominated by the retrieval revolution — the displacement of enterprise search, the widespread adoption of copilots, and the emergence of RAG as a default pattern. This was necessary work. It is also, essentially, complete.

The next phase will be dominated by the reasoning revolution — the emergence of industry-native decision layers that turn a company’s judgement into a composable, inspectable, editable asset, architected so that enterprise data stays within enterprise boundaries. The companies that catch this turn will compound their AI investments across decision classes and across industries. The companies that do not will spend the rest of the decade explaining to their boards why their copilot portfolio was the wrong bet.

We are building CalvinBall’s Enterprise Logic Layer for the industries that feel this pain most acutely — starting with consumer goods. The platform connects across industries. The moat compounds within them. The architecture ensures the data never leaves.

We publish this paper as the founding technical document of the category.

— Gurnoor Dhillon, Sandeep Ramesh, Axel Wehr, Emeka Kalu-Uma

CalvinBall · May 2026