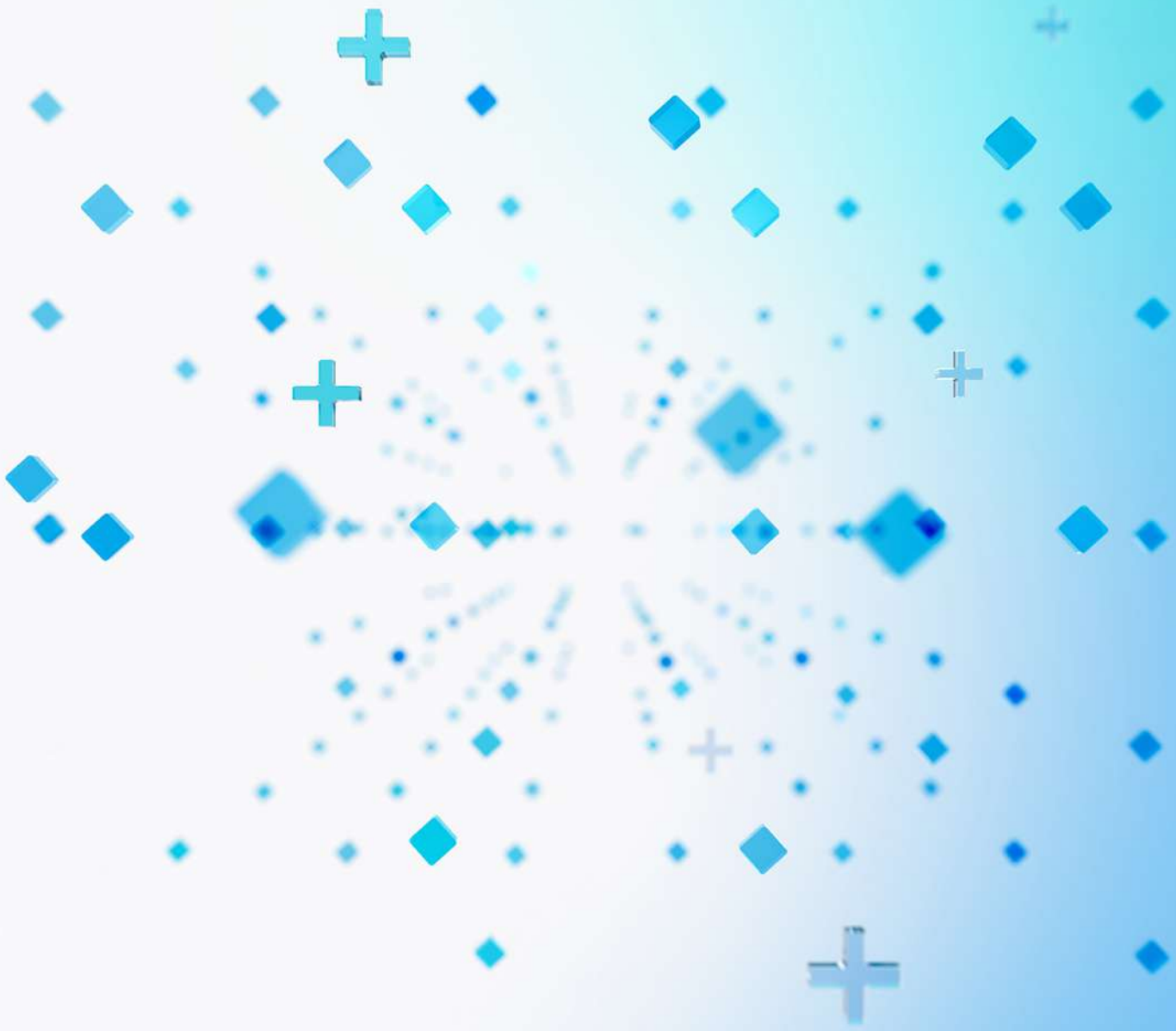
HCLSoftware

Agentic AI in BFSI (Part-2): Beyond the Sandbox

Scaling Agentic AI in BFSI through orchestration, governance and economic clarity



Executive Summary

In **From Friction to Flow**, the first whitepaper in this series, we explored how Agentic AI can unify Experience (X), Data (D) and Operations (O) - transforming siloed automation into orchestrated, context-aware execution across BFSI enterprises. That vision is now moving from promise to practice.

Banking, Financial Services and Insurance institutions are moving beyond generative AI experimentation toward agentic AI systems capable of planning, reasoning, coordinating actions and autonomously executing enterprise workflows. The potential is significant: accelerated operations, more adaptive customer and employee experiences, intelligent exception handling and continuous optimization across increasingly complex processes.

These themes formed the core of the second edition of the exclusive HCLSoftware–Business-World roundtable, where senior technology and risk leaders from banking, insurance, capital markets, development finance, and HCLSoftware came together to address a critical question: *Why do so many AI proof-of-concepts fail to scale into enterprise-wide production deployments in BFSI?*

While pilots often demonstrate promise, scaling AI into production remains a formidable challenge. Proofs of concept succeed because they operate within controlled, narrowly defined environments, often with significant manual oversight. Production, however, demands an entirely different level of rigor - where every AI-driven action must be explainable, every workflow governed, every handoff auditable and every deployment economically sustainable.

Key takeaways

- **Scale is an architectural challenge, not a capability gap:** Production environments require deterministic orchestration, audit trails and cross-system governance that most PoCs are not designed to address.
- **Economic sustainability determines board confidence:** Productivity gains are visible, but token consumption, lifecycle costs and monitoring overhead must be transparently modeled for enterprise expansion.
- **Agent sprawl is an emerging operational risk:** Without centralized governance, siloed AI deployments create redundancy, cost escalation and fragmented accountability.
- **Data and semantic readiness underpin trust:** Curated, sovereign and contextually structured data ecosystems are foundational to explainability, regulatory defensibility and safe scale.

Scaling Agentic AI in BFSI is not a race to deploy more agents. It is a disciplined journey toward orchestrated, governed, and economically sustainable intelligence. Institutions that align architecture, governance and human oversight will convert early experimentation into enterprise confidence and ultimately, into durable competitive advantage.

I. Why Agentic AI pilots plateau in BFSI?

Pilots prove capability. Production tests accountability.

Most agentic AI initiatives begin with promising pilots: a claims assistant that drafts responses, a service agent that retrieves policy details, a compliance copilot that summarizes obligations or an operations agent that routes exceptions. These pilots are useful, but they often stop short of production.

In production, an agent is not simply generating an answer. It may initiate work, trigger a workflow, call enterprise systems, influence customer outcomes, escalate exceptions or create an audit record. That changes the question from "Can the model produce a good response?" to "Can the institution stand behind the action?"

One executive reflected:

> *" In BFSI, you don't scale because something works. You scale because you can defend it - technically, financially and regulatorily. "*

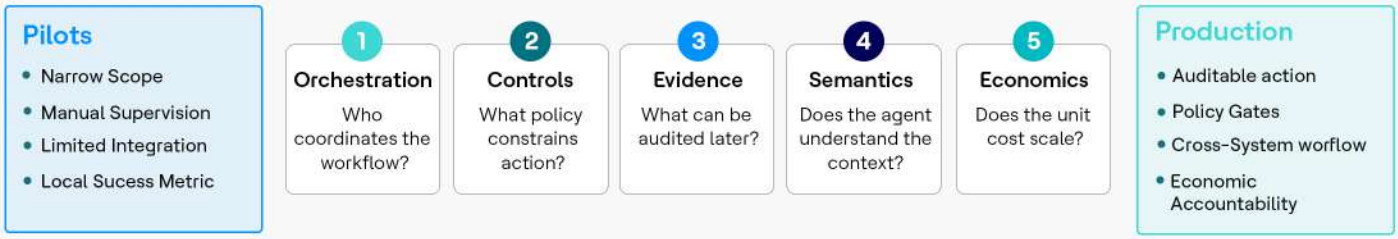


Figure 1. The pilot-to-production gap in BFSI Agentic AI.

Another added:

> *" Our PoCs are impressive in a sandbox. But production is not a sandbox. It is a regulated ecosystem. "*

Why the plateau happens?

- Workflow ownership is fragmented across business, operations, technology, risk, legal, compliance, data and security teams.
- Controls are applied after the fact rather than embedded into agentic workflow design.
- Data access improves, but semantics, lineage, entitlement and business meaning remains inconsistent.
- Economic assumptions are based on productivity anecdotes rather than total cost per successful controlled outcome.
- Audit evidence is not captured at the level needed for model risk, operational resilience, or customer-impact review.

From probabilistic insight to governed execution

The key shift in agentic AI is not only from prediction to generation, or from generation to action. It is from probabilistic assistance to governed execution. BFSI firms can tolerate probabilistic drafts in low-risk contexts, but they cannot allow unbounded autonomy in regulated workflows.

This makes 'determinism' the wrong goal. The production goal is bounded autonomy: agentic behavior that is observable, policy-constrained, reversible where appropriate, auditable by design and accountable to named owners.

One banking executive noted:

> *“ It's not about whether the model is right most of the time. It's about whether we can explain its decision every time. ”*

Several leaders at the roundtable reflected on this subtle but critical shift. In experimentation, AI is evaluated on performance metrics. In production, it is evaluated on institutional defensibility.

Another participant from insurance industry observed:

> *“ In our world, every automated output must survive audit, compliance review, and public scrutiny. That standard is fundamentally different from a demo. ”*

Production design principles

- Define the action boundary: what the agent may recommend, draft, initiate, approve or never do.
- Treat human oversight as workflow design, not as a generic fallback.
- Log intermediate reasoning artifacts, retrieved context, tool calls, approvals and final actions, where appropriate.
- Continuously test for policy violations, drift, hallucination, data leakage and degraded business outcomes.

II. BFSI Agentic AI scale readiness model & use cases

Scale readiness is a portfolio capability, not a single-use-case milestone.

A production-ready agentic AI program requires maturity across six dimensions: orchestration, governance, auditability, economics, semantics and operating model. Weakness in any one dimension can prevent promising pilots from scaling, even when the model output is strong.

Figure 2. BFSI Agentic AI scale readiness model.

How leaders should use the model?

- Assess each current pilot against the six dimensions before prioritizing production investment.
- Separate model-readiness from enterprise-readiness; strong model output does not guarantee operational fit.
- Use the readiness gaps to build a production backlog across data, controls, architecture, monitoring and governance.
- Make readiness visible to executive sponsors so that scale decisions are made with the same evidence across use cases.

Production Agentic AI needs a control pane that coordinates action, policy, evidence and resilience

Agentic AI at enterprise scale requires more than a collection of copilots or agents. It requires an orchestration control plane that coordinates specialized agents, enterprise systems, human checkpoints, policies, audit trails and monitoring into one governed execution fabric.

The control plane does not replace enterprise systems. It coordinates work across them. It determines which agent or system acts next, applies policy gates, routes exceptions, records evidence, and exposes operational visibility to business and risk stakeholders.

- The control plane must provide: Stateful workflow management across agents, systems, tools and human decisions.
- Policy enforcement at decision points, not only after-the-fact review.
- Observability across agent actions, workflow status, exception queues, cost and risk signals.
- Audit-ready evidence capture for inputs, prompts, retrieved context, tool calls, approvals and outputs.
- Resilience patterns such as fallback, retry, rollback, manual takeover, incident escalation and dependency monitoring.

Figure 3. Reference architecture for governed agentic execution.

III. High-value BFSI use-case portfolio

The best production candidates are not necessarily the most impressive demos. They are workflows with clear boundaries, measurable outcomes, repeatable decision patterns, strong data context, manageable risk, and well-defined human escalation points. Below mentioned use-case portfolio lens prioritizes use cases by business impact and governance complexity – not demo appeal.

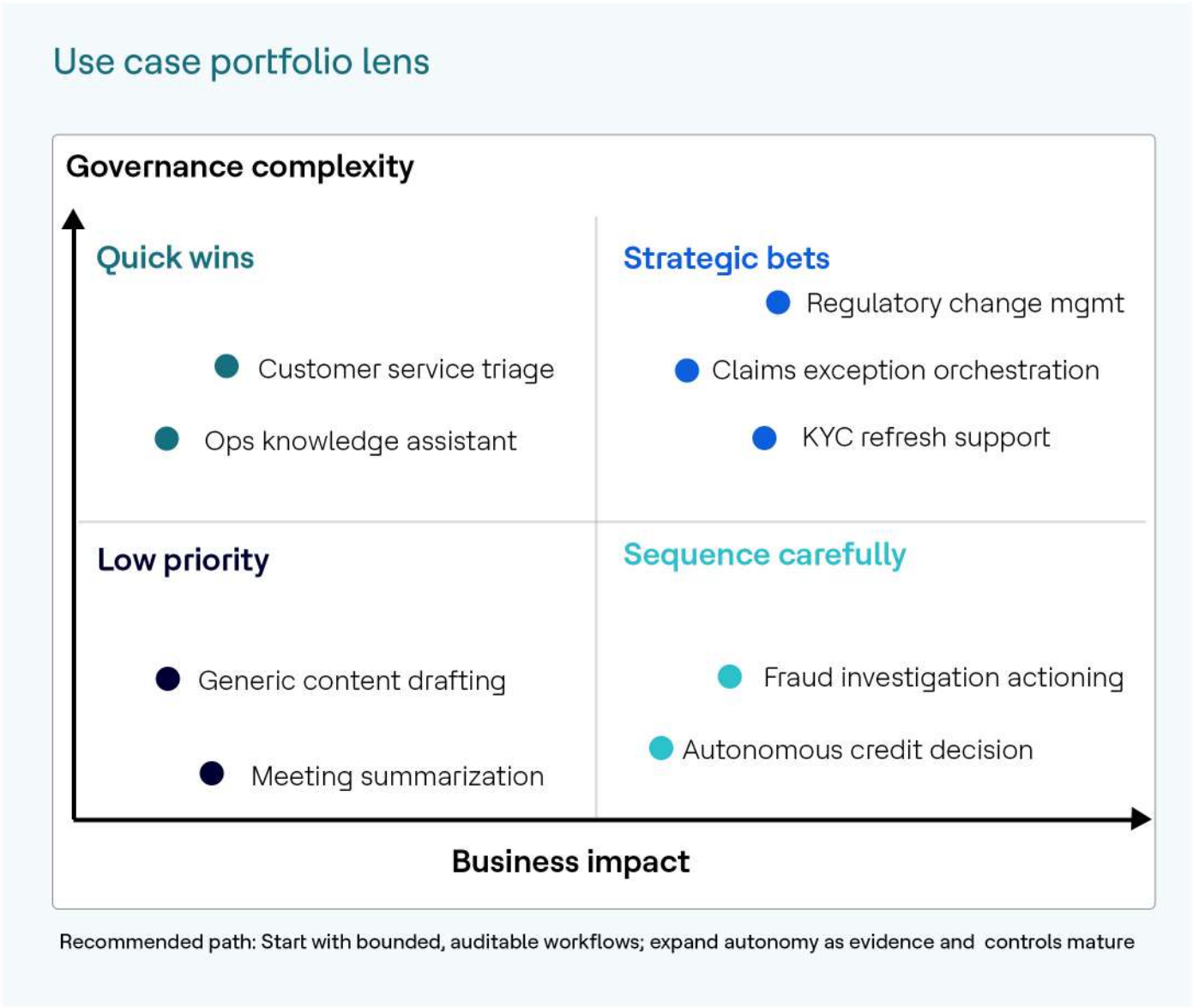


Figure 4. Use-case portfolio lens for BFSI Agentic AI.

IV. Governance, risk and economic alignment

Agentic AI should inherit the rigor of model risk, operational resilience, data governance and customer-impact controls.

Production Agentic AI in BFSI should be treated as a governed operational capability, not as an experimental AI feature. The governance model should draw from existing practices in model risk management, operational resilience, cyber risk, data governance, third-party risk, conduct risk and customer-impact review.

The practical test is simple: can the institution explain what happened, why it happened, who owned the decision, what controls applied, what evidence was captured and how the issue would be detected or remediated?

Sanjay Chopra, VP at HCLSoftware noted:

> *“ The risk is not building too few agents. It is building too many without clarity on control. ”*

Governance artifacts to design before production

- Agent and model inventory with owners, dependencies, data sources, risk rating and approval status.
- Validation pack with scenario tests, adversarial tests, control tests, fallback tests and acceptance criteria.
- Decision traceability plan covering inputs, retrieved context, prompts, tool calls, approvals, outputs and final business action.
- Human oversight model with escalation thresholds, authority, training, queues, and override procedures.
- Monitoring dashboard for quality, drift, policy exceptions, failure modes, cost, cycle time and customer-impact signals.

Without governance, scale becomes fragmentation. Without economic clarity, enthusiasm becomes hesitation.

The board-level case must measure successful, governed outcomes - not isolated task savings.

A credible agentic AI business case must go beyond labor savings. Production cost includes model calls, retrieval, orchestration, integrations, monitoring, validation, human review, exception remediation, audit preparation, operational resilience and ongoing change management.

This changes the ROI conversation. A workflow that looks efficient in a demo may become expensive if it creates high escalation volume, requires heavy human review, fails policy checks, or produces evidence gaps that increase audit effort.

One executive added:

> *“ Every initiative must show money saved or money earned. Without a clear narrative, it remains a pilot. ”*

Economic questions BFSI leaders should ask

- What is the fully loaded cost per successful, compliant outcome?
- Which steps are being automated, and which are merely being shifted to human review queues?
- How often does the workflow escalate, retry, fail, or require exception remediation?
- What is the incremental cost of auditability, monitoring, and control testing?
- Which value metric matters most: cost reduction, speed, customer experience, risk coverage, or audit effort reduction?

V. The next advantage is orchestration

Agentic AI will scale in BFSI when autonomy is designed around control.

Agentic AI is now entering a defining phase - one where experimentation must evolve into enterprise-grade execution. In the BFSI sector, this transition cannot be driven by enthusiasm, isolated pilots, or model sophistication alone. Success will depend on the strength of the underlying architecture, governance frameworks, data semantics, operational controls, economic viability, and accountability mechanisms that support AI at scale.

For BFSI institutions, scaling Agentic AI is ultimately a leadership and operating model decision. These organizations are not cautious because they lack ambition; they are deliberate because they understand the cost of failure in environments where trust, compliance, resilience, and regulatory scrutiny are non-negotiable. Before intelligence can be scaled for efficiency, it must first be engineered for transparency, control, and trustworthiness.

In this context, **HCL UnO** can serve as a universal orchestration layer for enterprise AI adoption — bringing together AI agents, workflows, human decision-makers, robotic automation, enterprise systems, data, and governance controls into a unified execution fabric. This positioning should remain closely aligned with validated product capabilities and approved enterprise messaging.

UnO as an orchestration fabric for Agentic AI

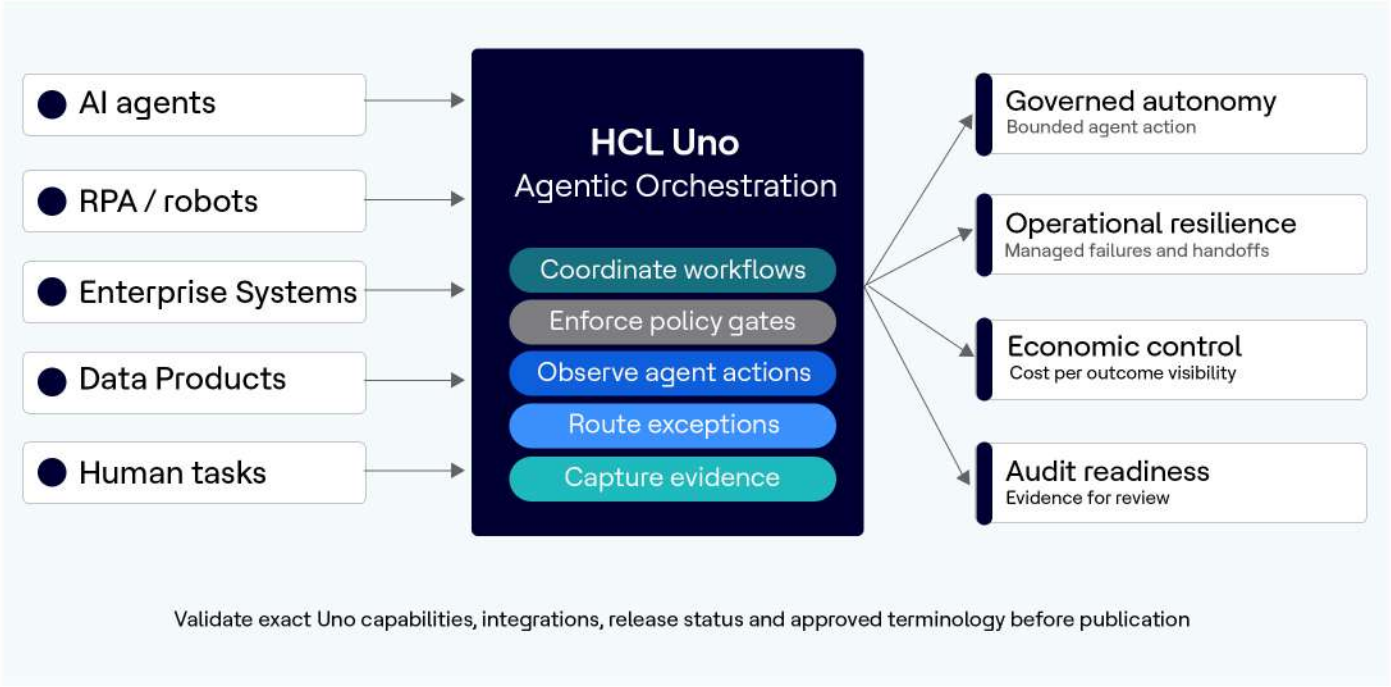


Figure 5. HCL UnO – Orchestrating every process, automating every outcome.

The institutions that will lead the next era of AI transformation will not necessarily be those deploying the highest number of agents. They will be the ones that successfully embed orchestration, governance, and economic clarity into their operational fabric - transforming Agentic AI from an experimental capability into a trusted enterprise standard.

The path forward is not about launching more pilots. It is about building scale readiness with intent and discipline. A focused 90-day readiness sprint can help BFSI leaders identify high-value use cases, uncover operational and governance bottlenecks, design the enterprise control plane, and establish a credible investment case for production deployment. The objective is clear: to emerge not with another proof of concept, but with a board-ready roadmap for enterprise-scale AI adoption.

HCLSoftware

hcl-software.com