UST + AWS | AI-NATIVE PLATFORM ENGINEERING - ISSUE 1

AI agents are moving into operations

But who operates the operators?

AI-native platform engineering: The control plane for agentic AI

ust.com

UST

As enterprises move to AI-first applications, AI-native software delivery, and AI-driven operations, they need a control plane traditional cloud platforms were not built to provide.

AI agents require many of the same management disciplines as APIs, such as versioning, routing, throttling, blast radius control. But agents introduce something APIs never had: non-determinism. They drift over time. Performance that looks strong early on may not hold in production. Cloud-native platforms manage APIs well, but they were not designed to manage agent behavior. It has no mechanism for detecting when an agent's performance declines or deciding when a task should be escalated to a person.

What's needed is a control plane for agentic AI.
A layer that sits between your applications and your infrastructure, managing agent behavior the way API gateways manage services.

The shift: From assistants to operators

AI agents are no longer just writing code and summarizing documents. In a growing number of enterprises, they're managing incidents, orchestrating operational responses, and making real-time reliability decisions with minimal human involvement. Early adopters are targeting 60% or more of operational incidents handled autonomously by AI agents.

But a pattern is emerging among organizations furthest along this journey.
AI agents that outperform in pilot routinely lose accuracy in production. Without proactive monitoring, 91% of machine learning models experience performance degradation. The cause isn't the model. It's the platform. Data changes, environments shift, and new edge cases appear. Most platforms are not built to detect that decline. By the time it becomes visible to the business, the impact is already growing .

Governance adds another layer of complexity . Most organizations enforce responsible AI through periodic reviews and blanket policies applied uniformly across models and data sources. That approach worked when AI was treated as a separate workload . When AI agents make thousands of operational decisions each hour, governance must run continuously and be built into the platform. These policies should adapt to different models, data sensitivities, and business contexts.

Agents can make incorrect operational decisions over time, and those issues are not always easy to reverse. The platforms enterprises built for cloud-native workloads were never designed for this. As AI agents multiply across operations, the same management discipline applied to APIs a decade ago is now overdue. For organizations in the middle of modernization, this layer helps turn infrastructure investment into intelligent operations.

Ravi Julapalli
Senior Director
AI-native Platform Engineering,
UST

UST + AWS PERSPECTIVE

AI-native platform engineering:
The control plane for agentic AI

As AI agents move from experimentation into production operations, a clear platform challenge emerges.. These agents need a control plane. They need the same management discipline enterprises once applied to APIs: routing, versioning, monitoring, and impact containment.

At UST, we deliver this through UST PACE, the Agent Control Plane for AI-Native Platform Engineering, built on Amazon Bedrock and AgentCore.

But a pattern is emerging among organizations furthest along this journey.

THE FOUR CAPABILITIES

1. **AI gateways (agent + model)**
   Manage agents the way API gateways manage services: traffic routing, throttling, versioning, and observability. The agent gateway manages requests in, and the model gateway manages responses out. Every agent becomes a managed endpoint.
2. **Capability boundaries and semantic isolation**
   This provides content-aware isolation. Attribute-level policies govern what each agent can access, use, and decide. Cloud-native IAM controls resources. PACE controls context.
3. **Runtime governance and drift detection**
   This provides governance while agents run, not just at deployment . Continuous evaluation that detects accuracy degradation before it surfaces as a business problem. Responsible AI controls can be applied by model, data source, and business context.
4. **Intelligent orchestration and human escalation**
   The Orchestrator acts as the Supervisor Agent. It routes tasks to domain and service agents, selects the right model for each task, and escalates when confidence falls below the threshold. This is where most operational incidents can be resolved autonomously, with people stepping in when needed.

For enterprises modernizing on AWS, UST PACE is the AI-native platform layer that turns infrastructure investment into intelligent operations.

- AI-Native Software Engineering: Intelligent Delivery
- UST PACE: Agent Control Plane [coming soon]
- Intelligent Operations [coming soon]

Since 1999, UST has worked side by side with the world's best companies to make a powerful impact through transformation. Powered by technology, inspired by people, and led by our purpose, we partner with our clients from design to operation. Our digital solutions, proprietary platforms, engineering, R&D, products, and innovation ecosystem turn core challenges into impactful, disruptive solutions. With deep industry knowledge and a future-ready mindset, we infuse expertise, innovation, and agility into our clients' organizations—delivering measurable value and positive lasting change for them, their customers, and communities around the world. Together, with 30,000+ employees in 30+ countries, we build for boundless impact—touching billions of lives in the process.

ust.com

Version 0101-20260330