UST

Scaling AI without slowing it down



Creating the foundation for enterprise AI

ust.com

By Aravind Nandanan,
General Manager
UST Telecommunications

Getting started with enterprise AI is no longer the main hurdle. Most large organizations now have access to strong models, cloud AI services, and more tools for building AI assistants, agents, and applications. The real challenge is turning these separate AI projects into a consistent capability across the whole company.

As more teams use AI, different parts of the organization move at their own pace and choose different technologies. For example, when solving a complex business problem spanning several departments, operations might use AI to identify issues and suggest fixes. Customer service could use an assistant to understand the impact and decide what to do next. Engineering might build predictive models to prevent the same problems from recurring.

Each team's AI project can add value, but they may choose different models, methods for measuring results, or technical setups. Over time, this can lead the company to support multiple engineering methods and overlapping technologies for similar problems.

It becomes harder to track costs, security checks vary by team, and data access decisions must be made for each new app. With only a few AI experts, resources get spread thin, and leaders struggle to see which projects really help the business. The aim is to keep AI fast and owned by the business, while also creating enough consistency to scale it responsibly.

This situation calls for a new kind of AI Cloud Center of Excellence (CCoE). Rather than serving as an innovation lab, approval center, or delivery team for every AI project, the CCoE should establish a shared foundation that enables business teams to build and manage AI quickly and consistently. The focus should be on centralizing the basics while allowing business units to retain control of their AI products and results.



AI value is local, but the foundations for scale should not be

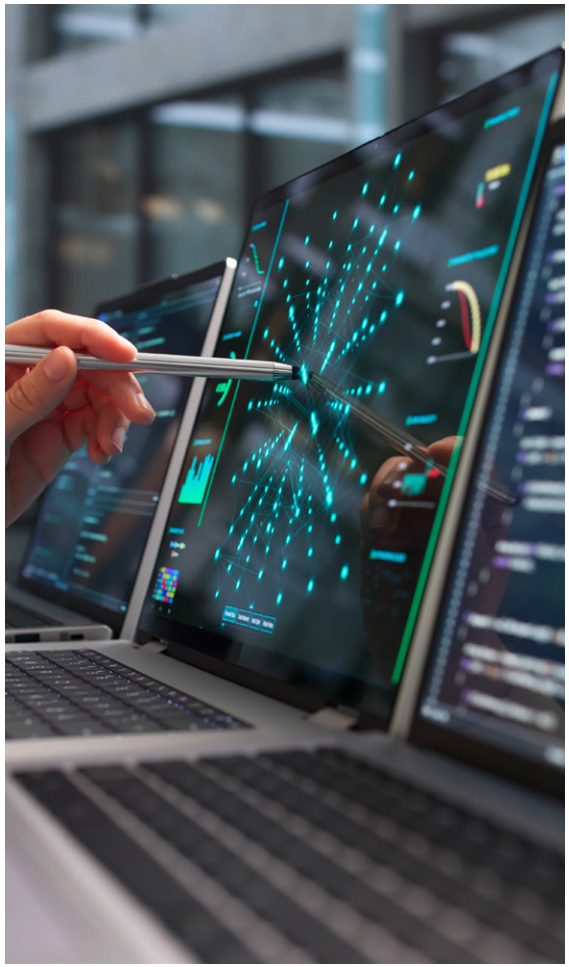


The best AI ideas usually come from people who know the business problem well. For example, network operations teams know where better predictions can prevent outages or speed up recovery. Finance teams understand their own processes and controls. Supply chain leaders know which manual steps add extra costs.

Teams should own the AI products they build and be responsible for the results. They shouldn't have to rebuild the technical and governance setup every time they launch something new. This is often where enterprise AI efforts run into trouble.

UST has identified five recurring challenges as organizations move beyond experiments:

- Turning AI investment into measurable business value
- Conflict between speed, cost, and risk
- Fragmented adoption across business units
- Duplication across data, platforms, and engineering stacks
- A shortage of specialized AI talent



These challenges show the gap between where AI creates value and what's needed to scale it. While individual business units can see benefits, long-term success depends on standards and capabilities that span the entire organization.

Gartner's research on cloud-based AI platforms highlights a related issue. It views the CCoE as the strategic governance function while positioning infrastructure and operations as the operational enabler. This approach aligns cloud governance with the discipline required to make cloud-based AI platforms effective at scale.

For organizations that already have an AI CCoE, much of the groundwork may already be in place. The next step is to build on it, going beyond cloud governance to help teams develop, launch, and manage AI more consistently across the company.

The AI CCoE should enable delivery, not own all of it



Building a center of excellence has its risks. When expertise and decisions are centralized, business teams might rely on it too much. More projects are piling up in the central queue, requiring more specialists. What was meant to speed up adoption can end up slowing delivery.

UST's AI CCoE model aims to prevent this issue. Capabilities requiring enterprise consistency, such as cloud infrastructure, responsible AI controls, FinOps and reusable engineering assets, are managed centrally and made available across the organization. Business units and shared services keep ownership of their priorities, products, data stewardship, adoption, and business results.

The main task is to decide what should be shared across the company and what should stay with each business unit. A three-layer model helps strike this balance, letting teams meet their own needs without rebuilding what everyone uses.

EXECUTION LAYER	ROLE IN THE AI OPERATING MODEL
Central AI platform team	Builds and manages the governed platform, common engineering patterns, security controls, model access, and observability
Shared AI engineering pool	Brings scarce AI architecture, engineering, evaluation, data, and security expertise into domains where needed
Embedded product teams	Combine engineering and domain expertise within business units and shared services, retaining ownership of products, adoption, and business outcomes

This setup gives platform engineering a new role compared to a traditional central delivery team. Instead of building every AI app, it spots the tools and features teams need again and again and makes them available for everyone to use safely and efficiently.

The same idea works for data and talent. Teams that work with the data should keep managing it, while the CCoE sets clear standards for access, security, and upkeep. AI experts can help when needed, but they should also share what they learn so other teams don't have to solve the same problems again.



Turn governance into something engineers can use



Most organizations understand what responsible AI should be. The tough part is making those principles work in practice, without forcing every AI project to work through new decisions and approvals. If teams must answer the same questions about security, risk, cost, and governance each time, progress slows, and teams end up making different choices.

A better way is to build these requirements into the AI development environment itself. UST used this approach with a large U.S. healthcare payer, setting up a shared AI environment on AWS. This gave teams standard ways to access models, apply security and privacy controls, check AI performance, and move apps to production. Instead of making each team set this up on their own, the company created reusable development paths with built-in standards from the start.

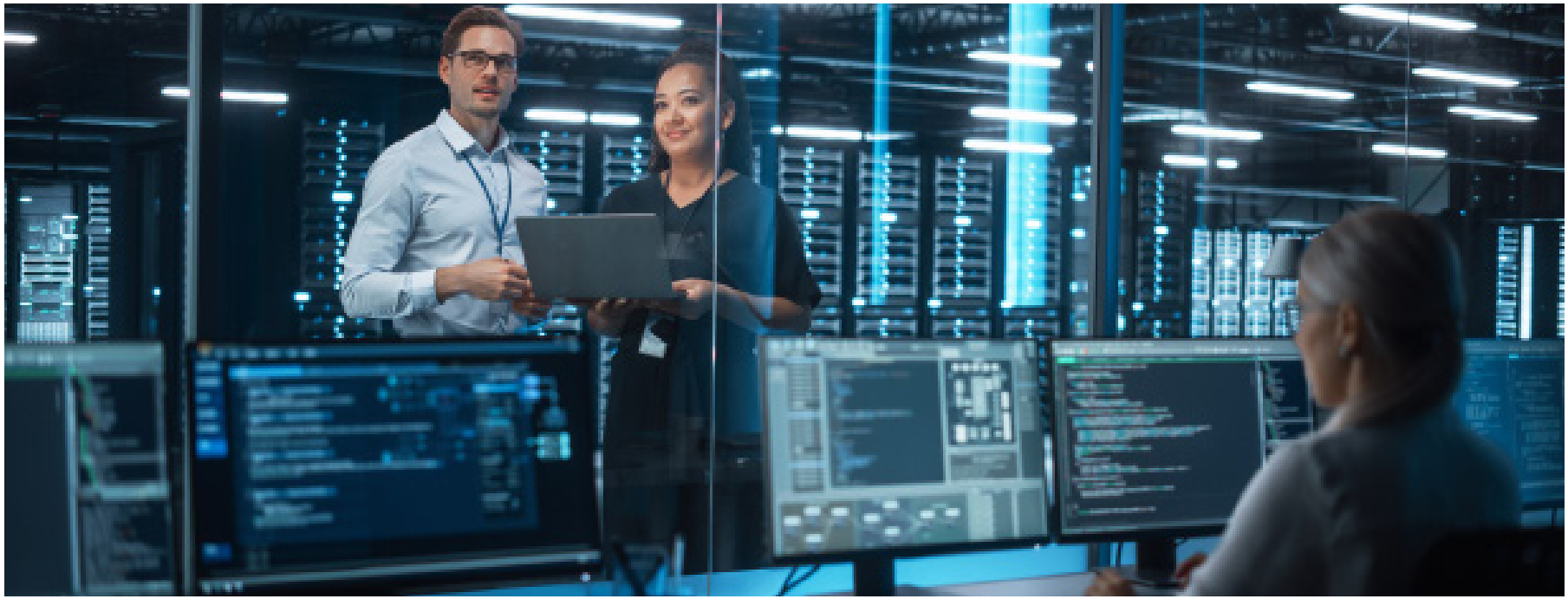
Business teams could focus on their own use cases while working in the same environment. Engineering teams used this shared approach in areas like claims, prior authorization, and member services. Lessons learned were added back into the platform as reusable tools. In 12 months, the company saw a 15% boost in engineering productivity, a 25% drop in MTTR, and $2 million in annual savings, with 70% of new use cases built on standard development paths.

Teams shouldn't have to figure out how to apply AI policies for every new project. Whenever possible, security, governance, and engineering best practices should already be built into the development process.

This is even more important as companies move from AI assistants to agentic AI. An agent can understand context, pick tools, and take actions in business or tech settings. Organizations need to know more than just whether an app was approved. They must know which model was used, what data it got, what actions it could take, what it did, and if those actions can be undone.

Companies shouldn't have to figure out what happened only after something goes wrong. Information about how an AI system behaved, which model it used, how it was checked, what risks were found, and what was approved should be recorded as part of normal operations.

For telecom providers, the risks are higher because AI systems might take actions that affect live networks and critical services. This means more autonomy must be earned. Before an AI use case can do more on its own, teams need clear proof that it works reliably, stays within set limits, can be reversed if needed, and has the right approvals.



Make every AI initiative improve the economics of the next one



Reuse is often seen as an engineering benefit, but it's also an economic one. If several teams each build their own model-access layers, retrieval methods, evaluation tools, security controls, and integrations, the company ends up paying for the same thing more than once.

An AI CCoE can change this by treating each project to improve the shared foundation. In UST's model, a production project can add reusable parts like integrations, evaluation tools, agent features, data-access methods, and deployment guides to the shared platform. Once these are standardized and managed, other teams can use them instead of starting from nothing, saving time and money on future projects.

UST's model highlights this ongoing benefit. Each AI project should not only create business value but also add reusable assets that lower future costs and add value when used by others. This is a good way to measure AI maturity. If the tenth project costs about the same and takes as long as the first, the company might be growing its AI portfolio without getting better at delivering it.

Cost transparency is also important. The cost of AI workloads can vary widely depending on the model, token usage, context size, retrieval setup, infrastructure choices, and the number of model or tool calls required. UST's CCoE model includes features such as tagging, chargebacks, model routing, and budgets within the shared control system.

This lets teams choose models based on both cost and technical needs. The goal isn't always to pick the most powerful model, but the one that best balances cost and performance.

Measure whether the enterprise is getting better at AI



The easiest AI metrics to track are often not the most useful. Counting pilots, trained employees, or models tested shows activity, but not whether the company can scale AI. Better metrics show if each new round of AI delivery is faster, safer, and more cost-effective.

Gartner suggests using a variety of measures to assess whether AI is becoming easier to manage, scale, and use across the business. For an AI CCoE, the key question is whether the company is getting better at delivering AI over time. Are apps producing real business results? Are teams launching faster and reusing what others have built? Is production AI under the right controls, and does the company know what each workload costs? These measures show if the CCoE is making the business more capable, not just busier.

UST has seen this approach work in cloud operations. In one large CCoE project, UST supported hundreds of apps across multiple platforms and achieved a 12% increase in quarterly onboarding, five times more assets, three times fewer defects, and a 31% smaller team over eight years. These results are unique to that project and aren't benchmarks for everyone. But they show how standardization, reusable assets, and automation can help scale without needing a bigger central team.

The goal is to become a more capable enterprise, not a larger center



The companies that scale AI best usually don't have the biggest central AI teams. A better sign of maturity is when safe, cost-effective AI delivery becomes routine across the entire business.

To get there, responsibilities need to be clearly divided. Shared platforms, model access, security, responsible AI rules, evaluation, monitoring, FinOps, reusable engineering, and specialist skills should be managed at the enterprise level. Business priorities, product ownership, data management, process changes, adoption, and accountability should stay with the teams that create value.

An effective AI CCoE connects these areas and should keep improving the company's ability to deliver AI. Integrations become reusable components, evaluations become part of standard testing, security is automated, and engineering patterns from one team are shared with others. As models, vendors, systems, and rules change, the best investment is in a setup that lets the company adapt without having to rebuild its foundations each time.

For companies moving from single AI wins to scaling AI, this is the true job of the AI CCoE: building shared tools and guardrails so innovation stays close to the business, while the company keeps getting better at delivering AI.

Success isn't about how much AI the CCoE controls. It's about whether each new AI project is quicker to build, safer to operate, and cheaper to scale.

Together, we build for boundless impact

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 35+ countries, we build for boundless impact—touching billions of lives in the process.

ust.com

© 2026 UST Global Inc.

Version 0102-20260923

U ▪
S T