

Rethinking infrastructure investments for AI success

Deploying and scaling AI workloads requires careful decisions around resource allocation, observability, cost, and security.



AI is rapidly transforming business processes, applications, and customer experiences. To prepare for the continued expansion of AI use cases, CIOs know they need to implement significant changes to their underlying IT infrastructure. Ninety-five percent expect to increase their infrastructure spending over the next 12 to 18 months to support AI demands, according to a new Foundry research survey conducted for UST.

Adapting infrastructure for AI workloads has become a business imperative, but to make the transition successfully, CIOs are carefully analyzing complex factors, including resource allocation, observability, costs, and security. The decisions they make now will shape the course of AI operations for years to come.

This white paper explores the survey's main takeaways and offers insights to help CIOs navigate a path to an agile, secure AI future.

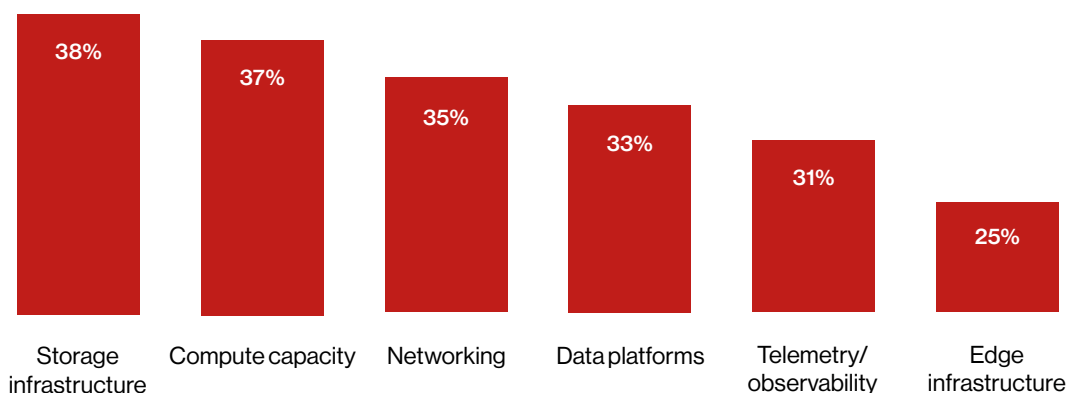
AI is driving increased cloud spending

AI workloads are becoming a core budget driver, consuming an average of 15% of IT infrastructure spend across industries, according to the study, which surveyed 105 US enterprise technology decision-makers in financial services, healthcare, manufacturing, retail, and consumer packaged goods.

Not surprisingly, the cloud is playing a central role in their plans. Nearly three-quarters of CIOs (73%) have increased cloud infrastructure to support AI workloads. They are expanding investments nearly equally across the cloud stack, including storage (38%), compute (37%), and networking (35%).

AI investment is expanding across the entire cloud stack

AI investment spans the full cloud stack as enterprises strengthen foundational capabilities.



Source: Foundry

Cloud flexibility provides rapid access to infrastructure for AI experimentation, which is easily scaled for deployment into production. Despite this advantage, CIOs say they must consider data privacy (43%) and security (39%) when deciding whether to host AI workloads in the cloud, on-premises, or a hybrid architecture. These are important factors to weigh, particularly for companies in heavily regulated industries such as healthcare or finance. The cost of scaling workloads in the cloud also puts pressure on ROI goals for AI initiatives.

Whichever way they proceed, CIOs should keep in mind that hosting AI applications and data in the same place may simplify security management, says Rick Clark, Global Head of Cloud Advisory at UST.

“If you have to secure apps and data in different locations, it creates more complex security risks,” Clark explains, “but may not be feasible for reasons such as data sovereignty requirements.”

AI delivers impressive results but increases costs

Initial investments in AI-enabling infrastructure are already starting to deliver tangible benefits. Nearly half of IT leaders (49%) said AI workflows have reduced the time it takes to launch products and services into production.

“Shortening the time to get from an idea to a product is one of the best things technology can do for a business,” Clark says. “It provides agility and enables the company to launch new products faster than the competition.”

As AI is applied to more business functions, efficiency and productivity benefits grow, freeing employees to focus on more strategic thinking. In another Foundry study, a majority of CIOs (68%) said AI is already starting to reshape operations.¹

But, according to the Foundry/UST survey, AI’s advances also come with trade-offs: heavy compute demand raises energy bills (41%) and cost per workload (36%).

Operational KPIs most impacted by AI workload deployments (% Top 2)



Source: Foundry

To succeed with AI, IT leaders must rethink observability

The survey underscores the importance of achieving observability and governance across infrastructure, applications, and LLMs – an essential capability for managing risks such as hallucinations, context overload, and model drift. To solve these problems, technology leaders are refining their observability strategies for the AI stack.

Nearly half of CIOs (45%) said they have increased observability spending by 20% to 30%. But spending alone isn't the answer. Many organizations have acquired an array of point solutions that fail to provide the kind of information needed to reduce risk, making it especially difficult for developers to uncover and resolve issues.

“What we mostly see is monitoring and logging solutions,” Clark says. “They make developers’ work harder, not easier. An enterprise production environment could generate millions of logs a day, making it nearly impossible for development teams to find meaningful information. The result is AI chaos and hallucinations.”

To succeed with AI, Clark explains, CIOs need to stop treating observability as a set of disparate tools. Instead, they must make it an architectural layer that integrates into the engineering platform and feeds development models.

“An integrated observability layer is the nervous system for AI. Without it, AI can only guess. It cannot detect anomalies or predict failures without correlated data across dependencies.”

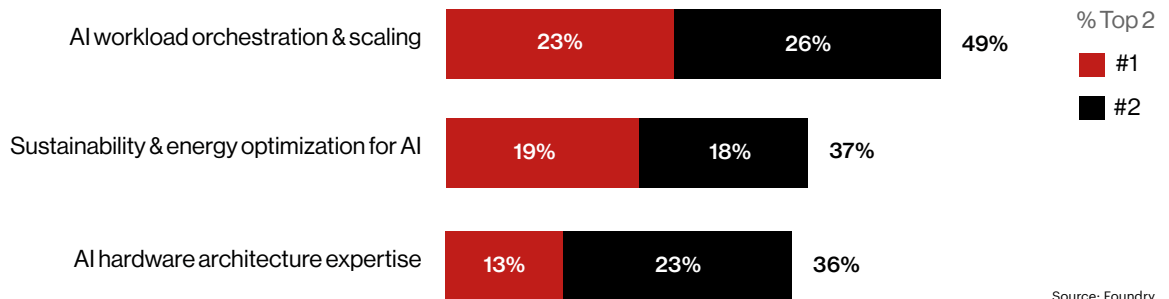
With a fully integrated observability layer, development teams can interrogate system behavior to discover unknown events. For example, instead of guessing at a problem, a developer could issue a specific command, such as “Show me all the times the application had a memory error in this function over the last two days.” The system will provide an actionable response in seconds.

An observability layer also means IT teams and software engineers can monitor operational signals in real time, generating insights that help them optimize resources and costs.

Top IT skills and capabilities needed

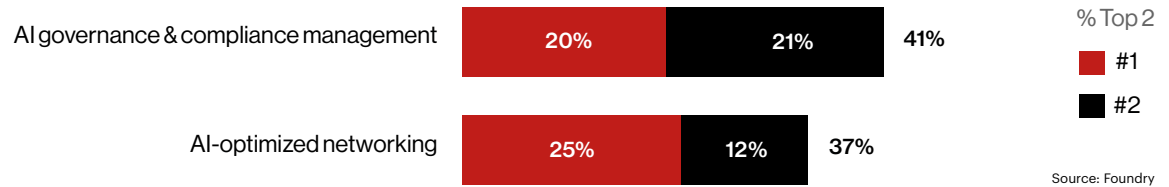
As AI use cases expand and CIOs implement advanced solutions such as agentic AI, improving IT team skills is critical. CIOs rated employee expertise in workload orchestration and scaling as a top factor for successful AI management, with 49% ranking it first or second. Other valuable competencies cited include sustainability and energy optimization (37%) and expertise in AI hardware architecture (36%).

Most important operational skills for managing AI infrastructure, (pt1)



The ability to apply AI solutions that enhance IT efficiency and performance is another area of focus. AI-based governance and compliance management topped CIOs' desired capability list, with 41% ranking it in first or second place. In addition, over a third of CIOs (37%) expressed a need to implement AI-optimized networking to address increasingly complex and distributed networks.

Most important operational skills for managing AI infrastructure (pt2)



Many CIOs lack expertise in these areas themselves, as the technology is new and rapidly evolving. In a recent Foundry survey, 98% of IT decision-makers said they are having difficulty implementing AI solutions, and their top challenge is a lack of in-house expertise.² To build and scale their AI practice effectively,

as well as leverage AI for IT operations, technology leaders must plan for continued investment in employee development.

Building a successful AI future

Today's CIOs tread a fine line as AI workloads grow in quantity and scale. While they need to quickly modernize infrastructure, they must also avoid costly missteps.

To establish a firm and lasting foundation for AI, leaders must build observability, guardrails, and agility into IT infrastructure as well as development platforms. "Leaders who build the right foundation will gain a compounding advantage, and those who skip these steps will be forced to rebuild under pressure," Clark says.

UST can help by providing domain knowledge and technical expertise to build transformative AI solutions.

To learn more, visit us at ust.com

1 Foundry, "[State of the CIO 2025](#)," February 21, 2025

2 Foundry, "[AI Priorities Study](#)," February 4, 2025