S&P Global
Market Intelligence

451 Research Market Insight Report Reprint

AI infrastructure strategies evolve amid widespread data challenges

Highlights from VotE: AI & Machine Learning

July 18, 2025

by Greg Macatee

AI infrastructure remains a critical component driving organizational outcomes, with over 94% of respondents in our VotE: AI & Machine Learning, Infrastructure 2025 survey saying their infrastructure choices create competitive advantages. IT decision-makers continue to familiarize themselves with AI technologies but grapple with insatiable demand for compute resources and workload acceleration, as well as rapidly expanding and increasingly unwieldy data volumes.

Introduction

While AI technology users continue to advance their capabilities, persistent challenges — notably including infrastructure limitations — impede the achievement of more widespread benefits. 451 Research's Voice of the Enterprise: AI & Machine Learning, Infrastructure 2025 study provides an update on AI infrastructure buyer behavior based on an online survey of 704 mid- and senior-level IT decision-makers from the US, UK and India. This study explores AI workload execution venues, tools, purchasing strategies and decisions, and it offers fresh insight into organizational AI infrastructure budgets, as well as workload creation and movement patterns.

THE TAKE

AI infrastructure remains a critical component driving organizational outcomes. According to our VotE: AI & Machine Learning, Infrastructure 2025 survey, 94% of AI infrastructure buyers believe their choices create competitive advantages for their organizations. These buyers continue to familiarize themselves with AI technologies but report notable challenges throughout their infrastructure stacks. Insatiable demand for compute resources and workload acceleration remains an overarching trend — a chronic condition that has dominated the market since its inception — while rapidly expanding and increasingly unwieldy data volumes represent a more acute pain point. Issues of data volume and management span the entire data life cycle, driving customer demand for supportive infrastructure management, security and privacy capabilities. Where to deploy AI infrastructure is another major strategic consideration, with many leveraging cloud and other third-party environments in tandem with on-premises infrastructure near central operations. While AI deployments at the edge remain nascent, the edge is viewed as a critical workload execution venue of the future.

Summary of findings

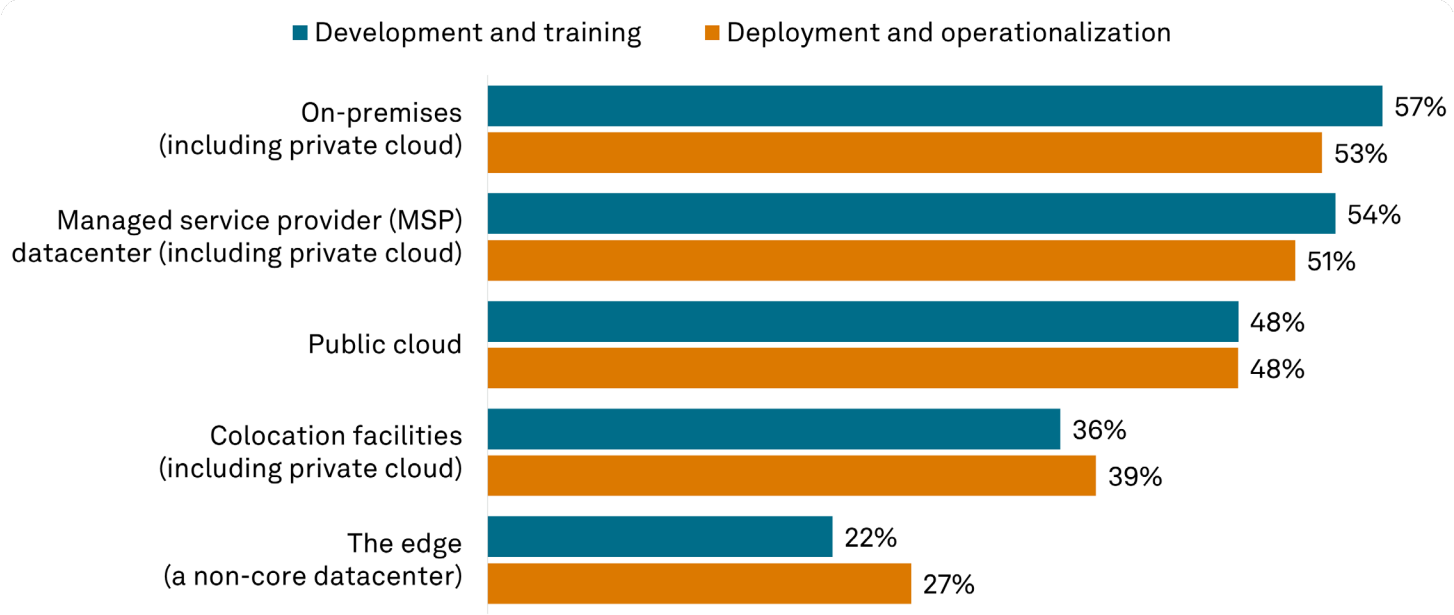
Organizations' AI capabilities are progressing. In last year's Voice of the Enterprise: AI & Machine Learning, Infrastructure 2024 survey, 73% of organizations reported some level of investment in generative AI (GenAI), while only 18% had implemented it across their businesses. These figures jumped in the latest survey to 89% of organizations investing in GenAI and 48% reaching widespread adoption. This 30-percentage-point increase in the share of organizations reporting advanced implementation — and thus experience using underlying technologies — helps to explain respondents' increased confidence in the ability of infrastructure to handle future workload demands. Deepening experience likely also contributes to declining AI/ML project failure rates, with this year's respondents reporting an average of 24% of projects abandoned prior to production, versus 31% in 2024.

AI infrastructure purchasing and decision-making authority predominantly resides with IT. Despite the growing number of personas involved in the AI infrastructure decision-making process (more than 20 roles were cited), IT is the most frequently cited area of the business, and in many cases has the final say in the choices being made. Specifically, 70% of organizations report that IT management groups (e.g., CIOs) are involved in the process, while 60% involve IT infrastructure managers and administrators. Information security management groups including CISOs and CSOs (47%), executive management including CEOs and company boards (46%), and chief data and analytics officers (32%) round out the top five sets of roles involved. Just under half of organizations' IT groups (combining executive and other management and administrators) have primary authority, more than any other area within the business. Additionally, nearly 20% of organizations' AI infrastructure decision-making authority resides with executive leadership.

AI infrastructure budgets will grow rapidly and come from a variety of sources. Organizations widely anticipate growth in spending across AI infrastructure technologies, including computing devices (e.g., AI PCs), servers (on- and off-premises), accelerators (on- and off-premises), storage (on- and off-premises), and networking (on- and off-premises), with more than 70% of organizations intending to increase spending in all these technology categories. More than a third of organizations expect to increase spending by 25% or more over current levels, and roughly one in 10 will increase spending by 50% or more. AI infrastructure spending typically represents 10%-24% of respondents' current IT budgets, with 44% of organizations reporting spending levels in this range, while an additional 33% of respondents say AI infrastructure spending accounts for 25% or more of their IT budgets. AI infrastructure purchases are often funded at least partially by budgets outside of IT, with more than two-thirds of organizations receiving at least 10% of their funding from other sources.

Organizations show diversity in workload venue preferences when training and operationalizing AI/ML models. Notably, respondents express a slight preference for deploying these workloads in on-premises environments over public cloud: 57% perform model training on-premises versus 48% in public cloud, while 53% deploy production models on-premises versus 47% in public cloud. More than half of organizations train and deploy models in managed service provider environments, and more than a third do so in colocation facilities. Nearly a quarter (22%) of organizations train models in edge locations (defined as a "non-core datacenter"), and 27% operationalize models at the edge (see figure below). Most organizations plan to maintain a hybrid approach: 63% intend to use a combination of on-premises and public cloud environments in the next year when creating new AI/ML workloads, compared to 18% planning for on-premises only and 19% using only public cloud.

Venues used for training and deploying ML models



Q. In which locations does your organization locate its ML infrastructure used for developing and training (deployment and operationalization of) its models?

Source: 451 Research's Voice of the Enterprise: AI & Machine Learning, Infrastructure 2025.

AI/ML workload movement strategies hint at optimization rather than overhaul. To some extent, strategy likely correlates with level of adoption, as organizations with early-stage capabilities and most models in proof-of-concept will differ from advanced implementors with clear, developed strategies. Overall, when asked whether they will move AI/ML workloads between on-premises and public cloud environments, 69% of organizations anticipate only slight shifts or none at all, including 31% citing a slight net movement of workloads from on-premises datacenters to the public cloud. About one in six (17%) plan to only move AI/ML workloads into the public cloud from on-premises facilities, and 11% will exclusively move workloads on-premises from the public cloud.

Organizations report challenges throughout the AI infrastructure stack. Despite progress in advancing AI capabilities, roughly half of organizations say key technical aspects of supporting AI workloads — compute, storage and networking — are "somewhat" or "very" challenging. This statistic remains consistent across on-premises, public cloud and edge environments. About six in 10 organizations report challenges with data management and availability (57% on-premises, 56% cloud, 56% edge) and security and privacy (64% on-premises, 62% cloud, 60% edge). These challenges align closely with respondents' top cited AI infrastructure buying criteria, including security (55%), infrastructure reliability and availability (47%), and data privacy and governance (45%). (See 451 Research's Voice of the Enterprise: Data & Analytics, Data Management Practices 2025 for more insight on these topics.) When asked which infrastructure improvements would most improve AI/ML workload performance, buyers cite access to cloud-based accelerators (54%), followed storage performance (46%). This includes storage workload requirements such as extreme IOPS, throughput, scalability, and efficiency achieved by direct data access and improved memory management (e.g., key value [KV] caches) for more streamlined token delivery. On-premises GPUs (46%), enhanced networking (44%) and memory capacity (40%) round out the top five.

Sustainability considerations have improved slightly from a year ago. In our 2024 survey, just 33% of organizations cited sustainability as a key factor behind their AI infrastructure decisions. In this year's survey, that figure has grown to 39%, indicating welcome progress but still significant room for improvement. This trend may be attributable in part to organizations' increasing reliance on public cloud, managed services providers and colocation facilities, which customers may view as a way to offload responsibility for certain aspects of sustainability.

CONTACTS

Americas: +1 800 447 2273
Japan: +81 3 6262 1887
Asia-Pacific: +60 4 291 3600
Europe, Middle East, Africa: +44 (0) 134 432 8300

www.spglobal.com/marketintelligence
www.spglobal.com/en/enterprise/about/contact-us.html