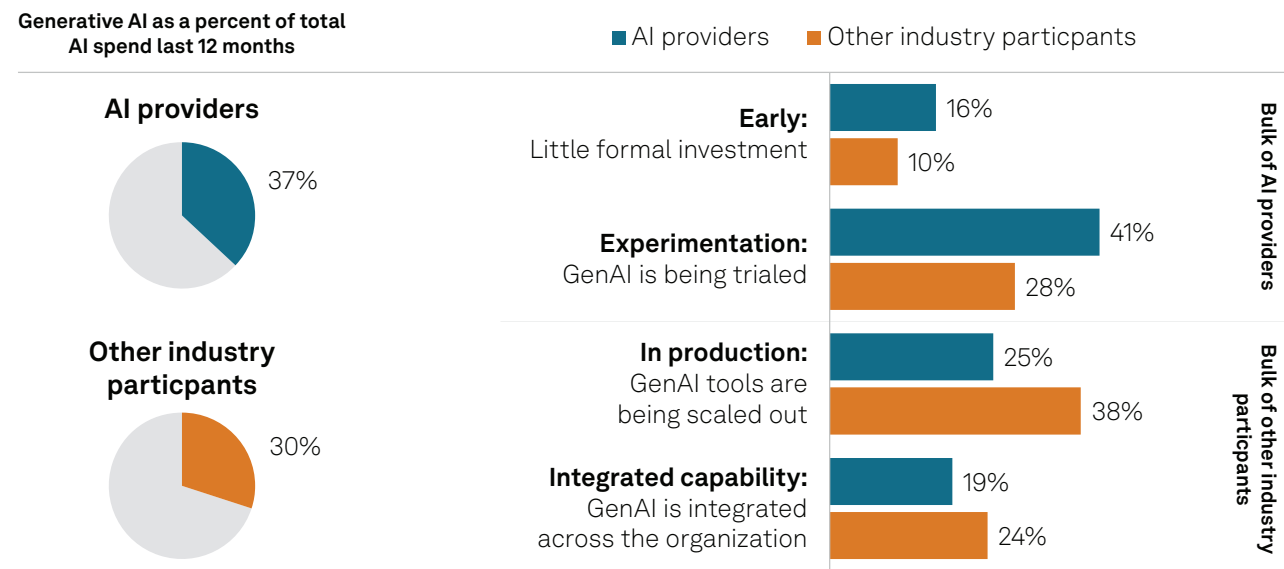


AI Providers Have Struggled To Keep up With Generative AI's Rapid Growth but Are Set To Make Big Strides

The Take

Generative AI's explosive emergence has driven rapid growth, with even seasoned AI players striving to grapple with its complexities and keep up. AI providers — organizations that provide AI technology, products or services — are diving headfirst into these new capabilities. However, only a small number have managed to turn their efforts into real-world applications and are doing so at rates consistent with or even slightly behind organizations in other industries. While 62% of industry respondents report being in the latter two stages of generative AI adoption (in production or fully integrated across the organization), the bulk of AI providers (57%) remain firmly in the preliminary stages of experimentation and exploration.

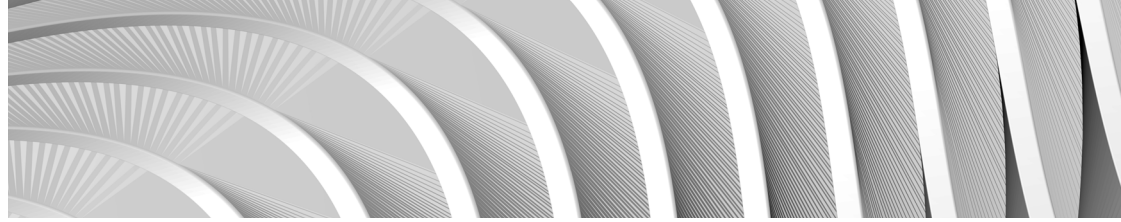
AI providers see lower levels of generative AI maturity despite higher levels of investment



Source: S&P Global Market Intelligence 451 Research Global AI Trends custom study, 2024.

AI providers cite common adoption challenges such as increasing demands on IT infrastructure and data requirements with equal frequency compared to other segments. However, AI providers may have to apply stricter standards for AI project implementation and evaluation because they are responsible for managing and supporting AI development across their customers, as opposed to organizations managing a single deployment.

AI providers face unique challenges — such as deployment scale and meeting product and service excellence requirements — that come with delivering AI offerings. These challenges can outweigh some advantages in AI expertise. AI providers (15%) are more likely to reference hardware challenges such as access to accelerators than other enterprises (10%) as the primary impediment to moving projects into production. Managing this AI infrastructure with intensive customer demand and at scale could continue to strain AI providers dealing with these sourcing challenges. Additionally, they more frequently cite reputational challenges (14% compared with 5% elsewhere) due to the customer-facing nature of their generative AI investments.



There is, however, a much more bullish perception coming from this segment of the market that suggests many are poised to emerge as leaders once they overcome these initial resource barriers. AI providers are investing in generative AI at markedly higher rates, dedicating on average 37% of their AI budget to generative AI, compared with 30% for non-AI providers. They also show greater confidence in the surrounding AI ecosystem, distributing AI workloads across a wider swath of IT venues and readily adopting commercial licenses (54%) and open-source models (41%). Non-AI companies, meanwhile, show a higher propensity to build in-house and more conservatism in spreading around data and jobs.

Business impact

AI providers exhibit exceptional proficiency in deriving value from their generative AI investments despite early adoption levels.

These respondents cite an average impact score of 3.7 across 11 financial and business outcomes, compared with 3.4 for non-AI providers. Two areas where AI providers are far outperforming the average are in using generative AI to accelerate innovation and boost workforce productivity.

AI providers are less hindered by IT infrastructure limitations and continue to invest heavily. Not only are AI companies already better distributed across various IT venues (e.g., public cloud, edge locations) than their industry counterparts, but they also expect to upgrade and expand these existing investments in the next 12 months. They project the average number of IT workload environments in use will grow from an average of 3.5 to 3.9 for AI training and 3.1 to 3.6 for inferencing, prioritizing flexibility, reliability and sustainability.

GPU clouds are emerging as a major opportunity for AI providers looking to bridge the computational capability gap: Almost half (48%) of AI providers are actively investing in GPU clouds for model training compared with 31% of non-AI providers. Scalability and sustainability are the principal factors driving interest in this emerging service area. The availability of GPUs is among the most significant challenges AI providers face in scaling up their initiatives — and GPU clouds are well-positioned to address this concern.

Providers lead in plans to increase data volume to build and train AI models. Twenty-two percent of AI providers expect to increase the data used to build or train their models by 50% or more in the next 12 months. Only 12% of non-AI providers say the same. This data-centric approach is fundamental to achieving quality and reliable model outputs.

AI providers are highly environmentally conscious. Seventy-three percent of AI providers are concerned about the energy impact of AI, a 10-percentage-point difference between providers and the broader population. AI providers directly experience — and are incentivized to optimize — the immense computational demands of training and running AI models. This awareness, coupled with growing public scrutiny, creates a strong incentive to take sustainability seriously.

Looking ahead

Generative AI has emerged as a transformative force, redefining industries and challenging the status quo with its potential to create new products, services and businesses. However, realizing this potential calls for significant investment as well as technological and business process change. Experts in AI are grappling with the same challenges as industry players; this highlights the criticality of cultural and resource constraints that limit generative AI's impact at this stage in the market's maturity and the challenges of achieving scale and ensuring reliability.

However, by leveraging their existing resources, ongoing investment and deep expertise, AI providers are setting themselves up to reap significant rewards as the market evolves. Higher levels of capital investment in generative AI in conjunction with a sustained focus on critical infrastructure and data quality and management fundamentals are the keys AI providers plan to leverage to unlock progress. A continued commitment to sustainability is increasingly essential as the world seeks technological solutions to pressing environmental challenges.



WEKA helps enterprises and research organizations achieve discoveries, insights, and outcomes faster by improving the performance and efficiency of GPUs, AI, and other performance-intensive workloads.

[WEKA Turbocharges AI Workloads with an AI-Native Data Platform](#)

[WEKA Drives GPU Acceleration](#)

[WEKA for Generative AI](#)