

Discovery Report

August 2024

2024 Global Trends in AI

S&P Global
Market Intelligence

Commissioned by



©Copyright 2024 S&P Global. All Rights Reserved.

Table of contents

Introduction	3
Key findings	4
Defining AI leaders	7
AI maturity and adoption	9
AI project scaling challenges	11
Rise of generative AI	15
GPU availability	18
Environmental impact	21
Conclusions	24
Methodology	25
About the authors	27



Introduction

The 2024 Global Trends in AI report delves into the underlying trends surrounding AI adoption. In last year's Global Trends in AI report, we explored the divide between organizations that were successfully running AI in production and those that were not. In this year's study, we revisit this AI leadership theme, drawing on some key practices that leading organizations are doing differently — while deep-diving into the value drivers, infrastructure decisions and environmental practices that are shaping AI strategies. To develop this study, S&P Global Market Intelligence surveyed more than 1,500 global AI decision-makers and engaged in 1:1 interviews with senior IT executives about their AI projects and initiatives.

Key findings

1. AI applications are now pervasive in the enterprise; investments in product quality and IT efficiency are top priorities.

AI adoption continues at breakneck speed, with the technology increasingly viewed as an embedded and strategic capability.

- **AI initiatives are maturing rapidly:** In the last year, reported levels of AI maturity have undergone a radical shift. In 2023, survey respondents were still largely experimenting with AI or had isolated deployments in small pockets of their organizations. This year, the majority of respondents report that AI is “currently widely implemented” and “driving critical value” in their organization.
- **Product improvement and operational effectiveness are key investment drivers:** Organizations are increasingly applying AI to enhance top-line revenue and competitive differentiation, with improving product or service quality (42%) the most popular objective, and with many targeting increased revenue growth (39%). Simultaneously, organizations recognize the potential to boost their operational effectiveness by improving workforce productivity (40%) and IT efficiencies (41%), along with accelerating their overall pace of innovation (39%).

2. Many AI projects fail to scale — legacy data architectures are the culprit.

AI projects are challenged by weak data foundations. Legacy data architectures are impeding broader deployment.

- **Achieving scale remains a challenge:** Organizations are facing significant challenges in achieving the desired reach of their AI projects. The average organization has 10 projects in the pilot phase and 16 in limited deployment, but only six deployed at scale.
- **Availability of quality data is a major obstacle:** Data quality is the greatest challenge to moving AI projects into production. The challenge for project teams is not so much about identifying relevant data, but its availability; organizations are struggling to build a consistent, integrated data foundation for projects.
- **Modernizing data architectures is critical to success:** Given this, it is unsurprising that the greatest proportion of respondents (35%) cite storage and data management as the primary infrastructure issues hindering AI deployments — significantly greater than compute (26%), security (23%) and networking (15%).



3. Generative AI has rapidly eclipsed other AI applications.

Generative AI has gained significant traction in a short time. AI trailblazers are realizing concrete benefits and are poised to compound their competitive advantage.

Generative AI trailblazers are the 24% of organizations that suggest that generative AI is an integrated capability across workflows within their business.

- **Generative AI is the focus:** An astonishing 88% of organizations are now actively investigating generative AI, far outstripping other AI applications such as prediction models (61%), classification (51%), expert systems (39%) and robotics (30%). Dedicated budgets for generative AI as a proportion of overall AI investments are growing as organizations begin to recognize the potential benefits of integrated generative AI capabilities.
- **Generative AI adoption is exploding:** Despite only being in the market for a relatively short time, 24% of organizations say they already see generative AI as an integrated capability deployed across their organization. Just 11% of respondents are not investing in generative AI at all, and the majority of organizations are actively in the process of turning this investment into scaled-up, integrated capabilities.
- **Generative AI trailblazers are expected to compound their competitive advantage:** Organizations that have already integrated generative AI across their organization plan to continue increasing their investments: They expect generative AI budgets to reach 47% of their total AI budget in the next 12 months, far outpacing less “AI-mature” organizations. The majority of these generative AI trailblazers are seeing a significant positive impact from the technology across the full range of targeted benefits. Those benefits are likely to compound their competitive advantage given that those still in the experimentation phases of their generative AI projects are not seeing the same increases in organizational innovation, new product development and time to market.

4. GPU availability continues to be constrained, shaping infrastructure decision-making.

Access to GPUs is a major concern for organizations; GPU clouds may offer a scalable solution.

- **Accessing GPUs continues to be a challenge:** Four in 10 organizations surveyed suggest access to AI accelerators is a leading consideration in their infrastructure decision-making, and 30% cite GPU availability among their top three most serious challenges in moving AI models into production.
- **Regional pressures persist:** In some geographies, particularly in Asia-Pacific, lack of access to GPUs is restricting organizations from deploying AI; 38% of organizations in India see accelerator access among their top three challenges to moving AI projects into production.
- **Hyperscaler and GPU clouds serve as key channels for companies to access GPUs:** The need for accelerators has driven 46% of surveyed organizations to leverage hyperscale public clouds for model training, as well as — increasingly — specialist GPU cloud providers (32%).

5. Concerns about AI's environmental impact persist but are not slowing AI adoption; sustainable AI practices offer opportunity to mitigate emissions.

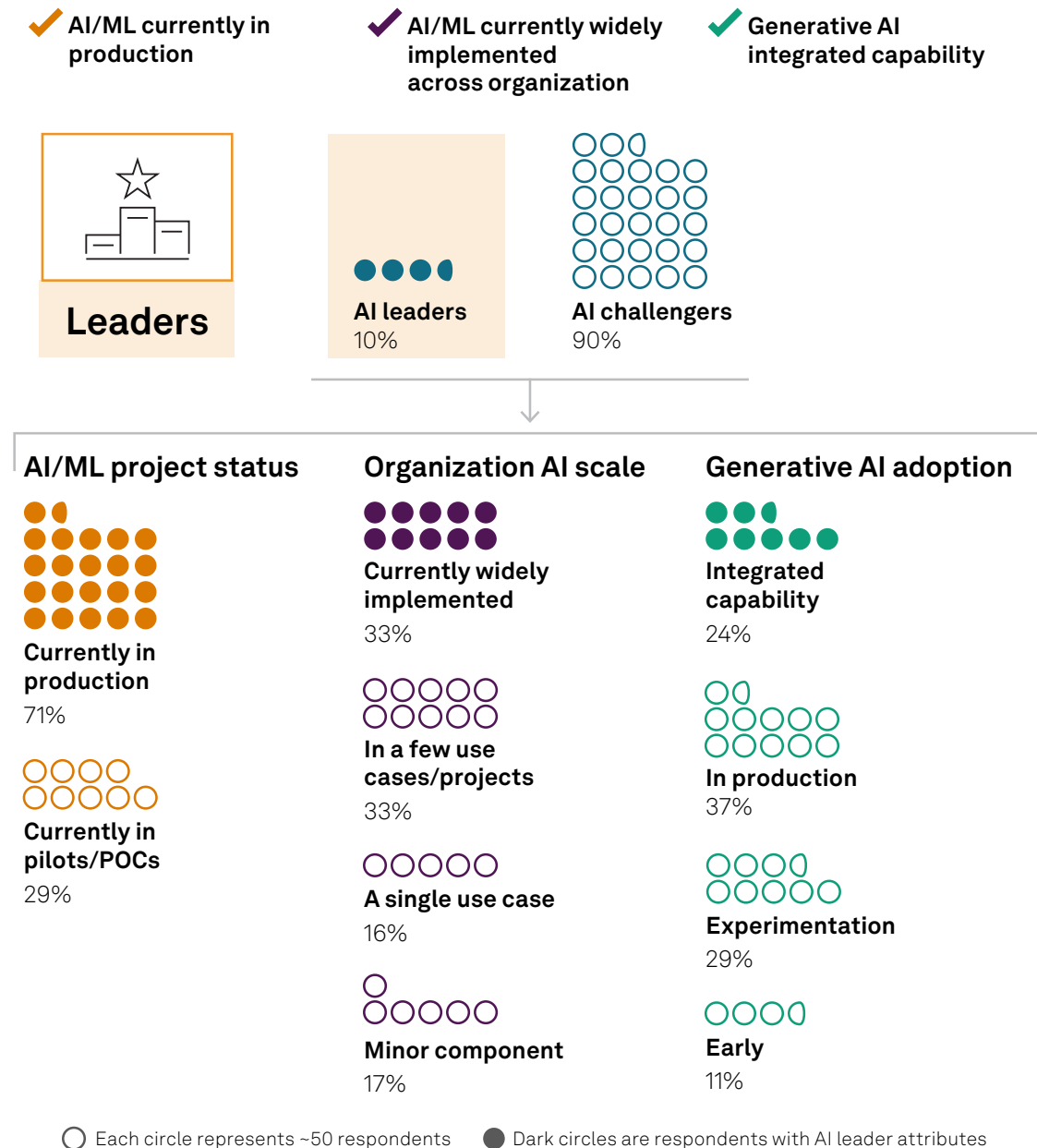
AI's environmental and energy impact is still a prominent concern for many organizations, but it is not slowing the decision to invest in AI projects. With many organizations seeing sustainability practices deliver meaningful impacts, there is a clear opportunity to address the emissions challenge.

- **Concerns about AI's energy and carbon impact remain prominent:** Nearly two-thirds (64%) of organizations say that they are concerned about the impact of AI/machine learning (ML) projects on their energy use and carbon footprint; 25% of organizations indicate they are very concerned.
- **Adoption of sustainable data infrastructure technologies is an area of focus:** Clearly, sustainability credentials from technology providers are becoming essential, with 42% of organizations indicating that they have invested in energy-efficient IT hardware/systems to address the potential environmental impacts of their AI initiatives over the past 12 months. Of those, 56% believe this has had a “high” or “very high” impact. Others have found that making changes in data infrastructure vendors (59%) and AI project scope (57%) have had a “high” or “very high” impact.
- **Sustainability is an important, but not the primary, factor in AI decision-making:** More than a quarter (30%) of organizations report that sustainability initiatives are a driver of AI adoption as they look to apply AI to improve energy efficiency and mitigate emissions. While this is notable, sustainability is, in fact, the least-mentioned driver overall. Even where energy-reduction initiatives are the goal, meeting sustainability targets can take a back seat to cost savings and improving operational efficiencies as the principal objective. In the context of all the issues that most strongly inform AI infrastructure decision-making, sustainability is mid-table: 37% of organizations are prioritizing it, but it is outranked by more prominent issues such as security (47%) and access to AI accelerators (44%).

Defining AI leaders

In the past year, the sentiment toward AI has rapidly changed, as has AI's strategic role. AI maturity has grown to the point that the vast majority of organizations have some form of AI in production, so a comparison of the “haves” and the “have nots” offers little analytical value. Instead, the emerging divide appears to be those that are able to harness the latest technology breakthroughs and deliver AI at scale, and those that cannot.

Figure 1: What makes an AI leader in 2024?



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





With the most significant AI breakthrough and arguably most important technological innovation in the past decade, generative AI as a strategic imperative is inescapable. Those that have quickly and efficiently tapped into this technological breakthrough are distancing themselves from the chasing pack.

Based on these assumptions, we have defined AI leaders to be those that have reported the following accomplishments:

- AI/ML projects in production environments driving **real-world impact** in critical operations.
- Implemented AI/ML **widely across their organization**, achieving far greater scale than limited and siloed AI projects.
- Capitalized on the most significant technological breakthrough of this decade (**generative AI**) and positioned it as an integrated capability across their business and workflows.
- The top 10% of the market engage in several **distinguishing practices** that set them apart from the bulk of organizations across the report’s five key themes.

Figure 2: Who are AI leaders in 2024?

10% of total (n=153)

 Industry	18% Healthcare	15% Government	15% Media/entertainment
 Company size	16% Enterprises (10,000-25,000)	14% Mid-sized (1,000-4,999)	12% Large (5,000-9,999)
 Region	16% North America	8% APAC	6% EMEA
 Business model	15% AI providers	10% Organizations	

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

There are several noticeable variances in the makeup and characteristics of AI leaders in 2024.

- **Industry:** Healthcare respondents (18%) have a greater proportion of AI leaders than other industries.
- **Company size:** Enterprises (16%) with greater access to capital, resources, AI skill sets and typically more mature digital transformation projects lead other company sizes proportionally as AI leaders.
- **Region:** North America (16%) has a significantly greater proportion of AI leaders than Asia-Pacific (8%) and EMEA (6%). Contributing factors may be greater access to AI talent from industry and educational institutions, and regional availability of venture funding and capital.
- **Business models:** AI providers (15%) are more likely to be considered AI leaders than other organizations. However, the difference is less pronounced than one might assume; in some instances, AI providers, despite building AI solutions for customers, have not necessarily fully implemented capabilities within their own company.

AI applications are now pervasive in the enterprise; investments in product quality and IT efficiency are top priorities

In the last year, AI's role has changed within many organizations, shifting from a minor component of an overall strategy to a critical embedded capability.



Key insights:

- The most common adoption status for AI has shifted from a **minor component of a broader strategy** in 2023 to **currently widely implemented and driving critical value** in 2024.
- Improving **product or service quality** and **generating cost savings from IT efficiencies** are the leading drivers for developing AI applications.
- Many strategic objectives such as **improving time to market** and **gaining product or service differentiation** are seeing greater focus.

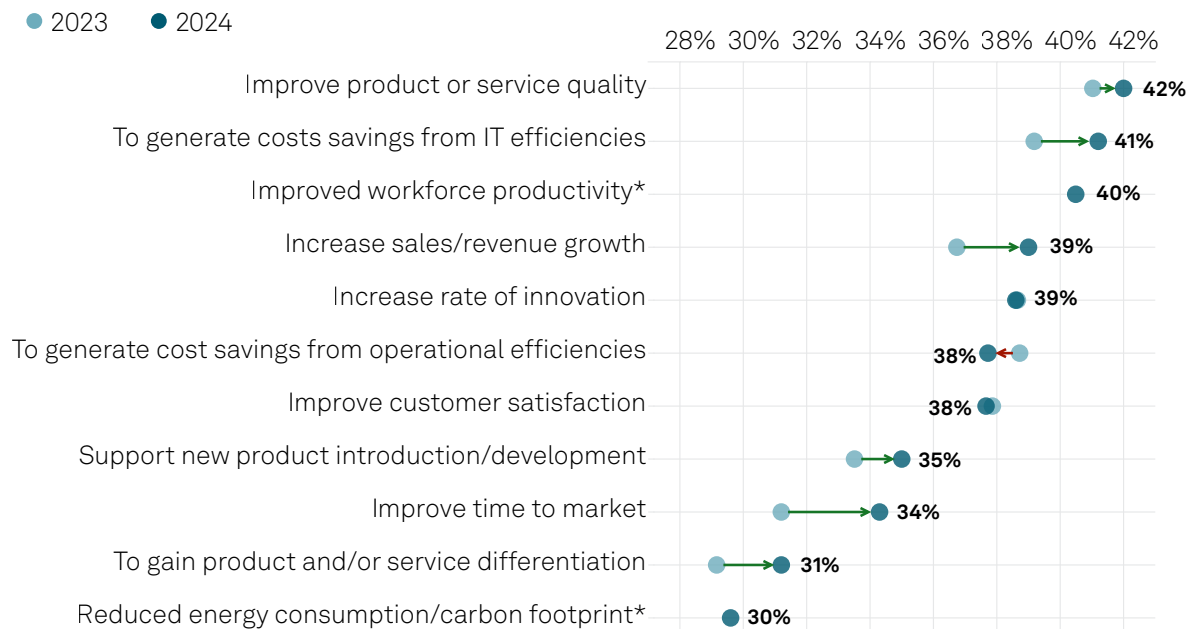
AI is becoming a fundamental aspect of many organizations' strategies, increasingly seen as both widely implemented and critical. The proportion of respondents who indicate that AI is a "minor component of a broader strategy" in their organization halved from last year's survey, while the proportion of respondents who see AI as "widely implemented, driving critical value" increased from 28% to 33%, becoming the most common answer. For North America respondents, it is even higher at 48%, in comparison to Asia-Pacific (26%) and EMEA (25%).

AI's reach is not limited to the breadth of implementation but includes the technology's strategic impact. Historically, AI's value proposition has been closely associated with reducing costs. Previous AI advancements in robotic process automation, for example, were closely aligned with objectives such as headcount reduction or reducing outsourcing costs. It is not so much that the cost-reduction opportunity presented by AI is being crowded out — indeed, generating cost savings from IT efficiencies is the second most popular objective for AI — rather, cost drivers are being paired with more strategic objectives. For example, more than a third (39%) of respondents in our survey see revenue growth as a key driver of their AI initiatives. As Figure 3 illustrates, companies are not just trying to achieve more with AI than they were last year, but they also see a clearer alignment with revenue drivers. They are significantly more aware of the opportunity for AI to be used to gain product differentiation and drive improved time to market than they were last year.

AI seeing wider deployment in 2024

AI organizational scale	2024	2023
Widely implemented	33%	28%
Few use cases	33%	27%
Single use case	16%	16%
Minor component	17%	30%

Figure 3: The year-over-year change in drivers for AI application development



*These response options were not available for selection in the 2023 edition of the survey.

Q. What are the primary driver(s) for developing AI/ML applications at your organization? (2023) Base: All respondents (n=1,516).

Q. What are the primary driver(s) for developing AI/ML applications at your organization? (2024) Base: All respondents (n=1,519).

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



What are AI leaders doing differently?

Leaders perceive a wider range of objectives as driving their AI strategies. This spread of objectives helps better inform where AI could be most impactful. It also sets the basis for a stronger business case for investing in AI, helping build a narrative that can appeal to a wider spread of stakeholders.

“I started with these experiments a year and a half ago. And then it took us a year to build... we now have about 5-10 use cases in production.”

CIO, transportation/logistics/warehousing,
1,000-5,000 employees, US

Many AI projects fail to scale; legacy data architectures are the culprit

AI's growing strategic importance is driving a significant increase in initiatives across businesses. Broad experimentation and education are key, and organizations would be remiss not to encourage it. However, the opportunity is being throttled by projects that lack a clear pathway to recognizing value, hampered by data challenges. AI projects risk stalling in a limited deployment purgatory, costing a company money, time and resources, while not seeing desired levels of use. Initiatives are becoming snagged on data siloes, poor data quality and ineffective data and model pipelines.



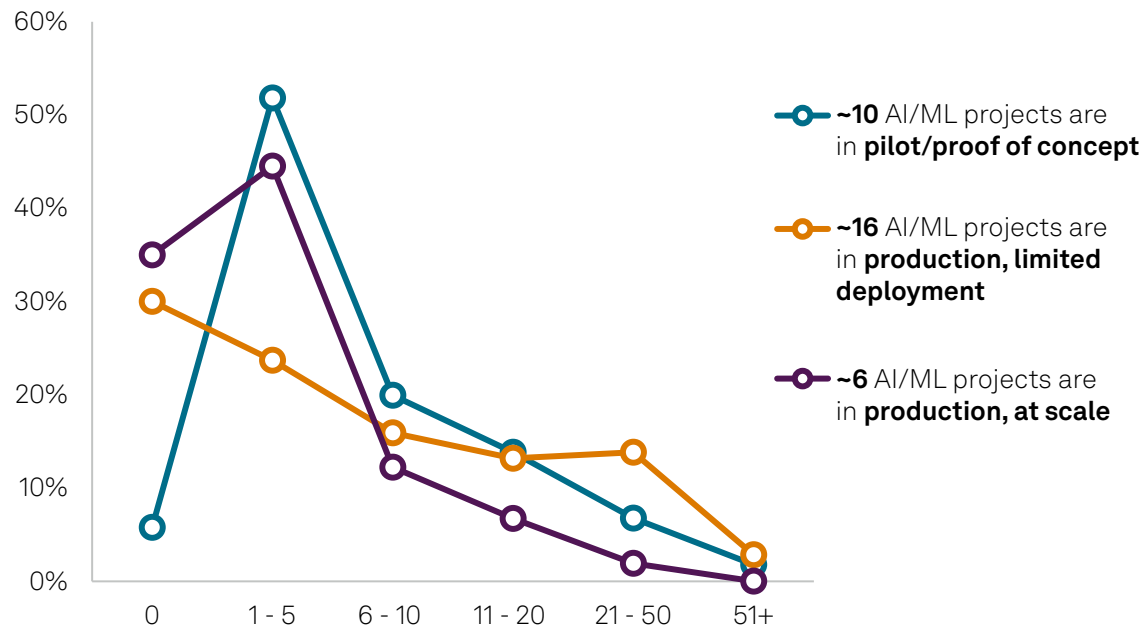
Key insights:

- In the average organization, **51%** of AI projects are in production but not being delivered at scale.
- **Data quality** is the greatest inhibitor when moving AI projects into production environments.
- **Storage and data management** are the most common infrastructural inhibitors to AI initiatives, identified by **35%** of organizations; however, those that have AI widely implemented feel these challenges less keenly.

As organizations invest to apply AI to an ever-growing set of objectives, a kink is emerging in organizations' project pipelines. While more initiatives are funneled toward AI project teams, there is a buildup of initiatives that have only been partially deployed. As Figure 4 illustrates, respondent organizations, on average, have more projects classified as being in production with a limited deployment than scaled-up capabilities. In chasing new initiatives, many organizations may fail to maximize the value of their existing investments. The crux of the problem appears to be data quality and availability, with legacy data architectures causing this pipeline stoppage in many organizations.

Data quality is the most frequently identified challenge as organizations move their projects from pilots to production. As Figure 5 illustrates, data quality concerns — identified by 42% of organizations as among their top three barriers — are even more significant than skills shortages (32%) and budget limitations (31%). Organizations in media and entertainment (59%), higher education (53%), and aerospace and defense (48%) feel the data quality challenge particularly keenly.

Figure 4: Many projects fail to graduate from limited deployment to delivering at scale



Q. How many AI/ML projects do you currently have: In pilot/proof of concept; in production, limited deployment; in production, at scale?
Base: All respondents (n=1,519).

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

The data quality challenge is not a lack of data to build performant models but, rather, that the data is not set up in such a way that project teams can take full advantage of it. When asked specifically to rank the primary data challenges to move projects to production environments, respondents indicated that availability of quality data is a more notable impediment than identifying relevant data. With 34% of organizations perceiving availability of quality data as a top three data challenge, outranked only by data privacy concerns (35%), it is clear that many organizations are poorly set up for effective data management.

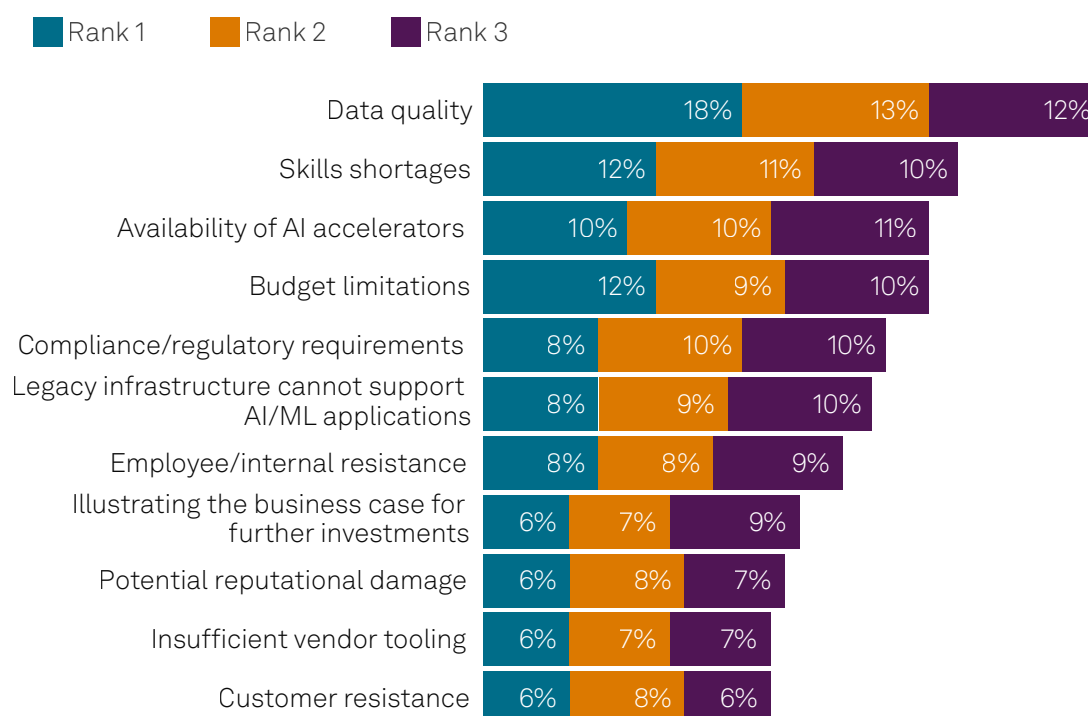
Legacy data technologies seem to be a leading cause of these data management shortcomings. Data management and storage are most commonly seen as the infrastructure components that inhibit AI application development. More than a third (35%) of respondents see them as a more serious issue than security (23%), compute (26%) and networking resources (15%).

Tellingly, organizations that are most effectively scaling AI initiatives are less constrained by these data management and storage components. Just 28% of respondents who reported that AI is widely implemented within their organization perceive storage and data management challenges as their greatest inhibitor; instead, they feel greater pressure from networking or compute resources. This compares to 42% of respondents who perceive AI as being limited to a few use cases or projects within their organization. Organizations that are delivering AI at scale appear to have focused on investing in upgrading the systems and technologies used to store or manage data.

“We still have challenges with master data. Branches had different SKUs for inventory; if I take that siloed data and put it into a model, we’ll get the wrong results. Cleaning up this data is our focus.”

CIO, transportation/logistics/warehousing
1,000-5,000 employees, US

Figure 5: Top three impediments to organizations moving an AI/ML application from pilot to production environments



Q. What are the primary challenges or impediments to moving an AI/ML application from proof-of-concept/piloting stages to production environments?

Base: All respondents (n=1,519).

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

This investment appears to be key because data management and storage shortcomings are filtering through into AI project life cycles, with organizations struggling to effectively prepare data for model building and deployment. Many organizations report that the most challenging aspects of AI initiatives are the data preprocessing stages (see Figure 6). Despite the growing number of organizations indicating that AI has been widely implemented within their organization over the past 12 months, there has been no meaningful improvement year over year in terms of performance against these data preprocessing steps. Bringing AI projects live but limiting their value or extensibility with weak data foundations sets a poor precedent for the next wave of initiatives in the early stages of exploration.

“The first thing I’ve done is doubled down on data strategy, effectively building a data platform and governance and capabilities around that. This gives us more control over our data. You’ll probably find in a lot of companies that they bolted on many of these acquisitions and have a lot of disparate systems, which means disparate data – that’s a challenge.”

CIO, manufacturing/food and beverage
1,000-5,000 employees, UK

Figure 6: Organizations find the early data steps of the AI life cycle as challenging as model building

Proportion of respondents that identify AI life cycle stage as “most challenging”



Q. What stages of the AI/ML application life cycle are currently the most challenging (Rank 1)?

Base: All respondents (n=1,519).

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

Immature data management toolsets are a worrying backdrop for the increasingly data-hungry AI strategies many organizations are embarking upon. More than three-quarters (80%) of respondents forecast an increase over the next 12 months in the volume of data they use to develop their AI models, and just less than half (49%) are forecasting growth in data volumes of more than 25%. Perhaps more fundamentally, though, is that the challenge for organizations that underinvest in data management may come with new data-related pressures — in particular, the mix of data types organizations are employing for model training. The proportion of organizations using unstructured rich media and text data for AI initiatives has increased since 2023, and outdated data management technologies may prevent organizations from delivering these projects meaningfully.



What are AI leaders doing differently?

Leaders are significantly less likely to see storage and data management as their primary inhibitors, presumably because these companies have already prioritized modernizing their data architectures. By building a solid data foundation at the outset, AI leaders have ensured that valuable pilots have a clear path to deliver at scale.

Generative AI has rapidly eclipsed other AI applications

Organizations have rushed to invest in generative AI, with interest outstripping longer-standing forms of AI. As the dust settles on this explosion of investment, a small cohort of generative AI trailblazers has emerged. These organizations have more widely integrated capabilities and are seeing significant competitive benefits from the technology around new product development, enhanced innovation and faster time-to-market. These competitive advantages are likely to grow as generative AI trailblazers set out to establish a significant gap between themselves and others, shaped by their investment and infrastructural advantages.



Key insights:

- **88%** of organizations are actively investigating generative AI.
- **24%** already see generative AI as having graduated to an integrated capability across their organization.
- The majority of these generative AI trailblazers see a “high” or “very high” impact from generative AI initiatives on increasing the **rate of innovation (79%)**, supporting **new product introduction (76%)** and **improving time to market (76%)**, among other areas of competitive differentiation.

Generative AI is the force driving enterprise AI strategies in 2024. The vast majority of organizations (88%) are actively investigating generative AI models to create net-new data or content. This interest outstrips much longer-standing forms of AI such as prediction models (61%), classification (51%), expert systems (39%) and robotics (30%). Considering that awareness of generative AI only blossomed in late 2022 and enterprise-grade solutions are still undergoing development, this is a testament to the perception of its transformative potential. Most of the survey respondents also expressed an interest in hypothetical artificial general intelligence — models that can outperform humans across all cognitive tasks — suggesting that many organizations have their eye on AI’s evolving horizons.

This interest is being converted into investment. Generative AI budgets on average are set to grow from 30% of total AI budgets to 34% over the next 12 months. Many senior executives are keenly aware of the implications for the technology and see a need for an accelerated investment roadmap.

“People can now see it and interact with it. Our board members are fairly traditional manufacturing folks. When they see ChatGPT, it becomes very real to them. We’re now looking at field service maintenance, our technicians who go out there to service machines. How can they use GenAI to quickly access historical maintenance records and ask support questions in real time?”

**SVP/chief digital officer, manufacturing/
industrial products**
5,000-10,000 employees, US



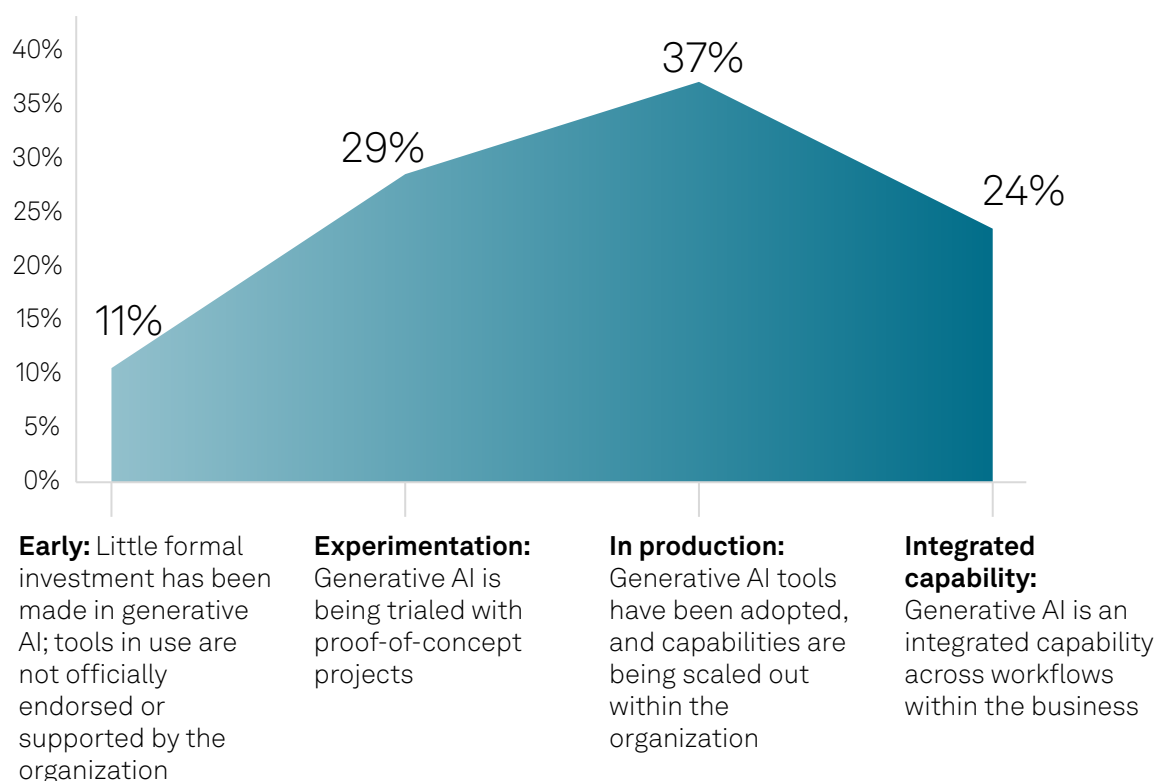
What are AI leaders doing differently?

Leaders are investing in a broader range of AI types with a greater likelihood of engaging with robotics, expert systems and classification models alongside generative AI. This expanded portfolio equips them to identify solutions that can meet diverse business needs and fosters a more holistic approach where these technologies can be brought together.

Generative AI adoption is progressing rapidly. As Figure 7 illustrates, a set of trailblazers — 24% of organizations — have already graduated generative AI investments into scaled-up production capabilities. In contrast, 11% of companies have not invested in generative AI, 29% are still experimenting with the technology and 37% have generative AI in production but not yet scaled. This is a remarkable level of uptake for a technology that only made it into the public consciousness with the launch of ChatGPT in November 2022.

Figure 7: Generative AI maturity and levels of investment

Proportion of respondents at each stage:



Q. Which of these statements most accurately reflects the use of generative AI at your organization?

Base: All respondents (n=1,519).

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

Organizations that have integrated and broadly deployed generative AI experience a wide range of benefits. Importantly, these benefits are commonly seen in areas that deliver competitive advantage. More than three-quarters (79%) of these trailblazers see generative AI as having a “high” or “very high” impact on their rates of innovation, 76% on their time to market, 76% in supporting new product introduction, 74% on improvements to their product or service quality, and 67% on their product and/or service differentiation. These levels outstrip organizations that are less “AI-mature” and suggests that relative adoption of generative AI may shape industry winners and losers. Organizations that fail to rapidly institute meaningful generative AI projects could end up losing out to those that can.

Future levels of investment appear to be setting the scene for these trailblazers to amplify their advantage. Generative AI trailblazers have invested heavily to lead the pack. The average organization that successfully graduated generative AI to an integrated capability had invested 44% of its AI budget into generative AI, a significantly higher level of investment than other organizations. Companies at earlier stages of generative AI maturity invested 26% on average. Trailblazer companies are set to expand generative AI budgets further, continuing to out-invest less AI-mature organizations.

Generative AI trailblazers are more sophisticated in their enabling infrastructures and strategies. They use a wider array of venues for AI model training and inference. More fundamentally, however, they take far more factors into account when it comes to AI infrastructure planning. They are more likely to plan their infrastructure considering security, AI accelerator access, data privacy, scalability, customer support, and access to AI tools and frameworks than organizations that have not invested to the same degree. The only factor these trailblazers are less likely to account for than those experimenting with generative AI is up-front costs, which they see as less important than longer-term operating expenditures. By looking at these considerations at the outset of infrastructure decision-making, these organizations are ensuring these issues do not emerge as projects progress.

“We created a chatbot for internal usage, such as employees searching for the ideal health plan. We are trying to extend that to publicly available information — earnings and annual reports — to look at our last five years and understand our strategy. If these models are working well, then we can go to more unstructured data.”

CIO, transportation/logistics/warehousing

1,000-5,000 employees, US



What are AI leaders doing differently?

Compared to other organizations investing in generative AI, AI leaders have prioritized generative AI initiatives that boost innovation rates and enhance IT efficiencies. By prioritizing these areas, they create a virtuous cycle where increased innovation facilitates further use of generative AI, and streamlined IT processes ensure sustainable and effective delivery.

GPU availability continues to be constrained, shaping infrastructure decision-making

AI accelerators play an important role in optimizing the performance of AI. These specialized hardware devices — GPUs being the most prominent example — are designed to accelerate model training and inference; they are faster and more efficient than CPUs for AI workloads. Organizations can face challenges accessing GPUs, and that scarcity is elevating their position in infrastructure planning and is encouraging uptake of specialist AI cloud computing platforms.



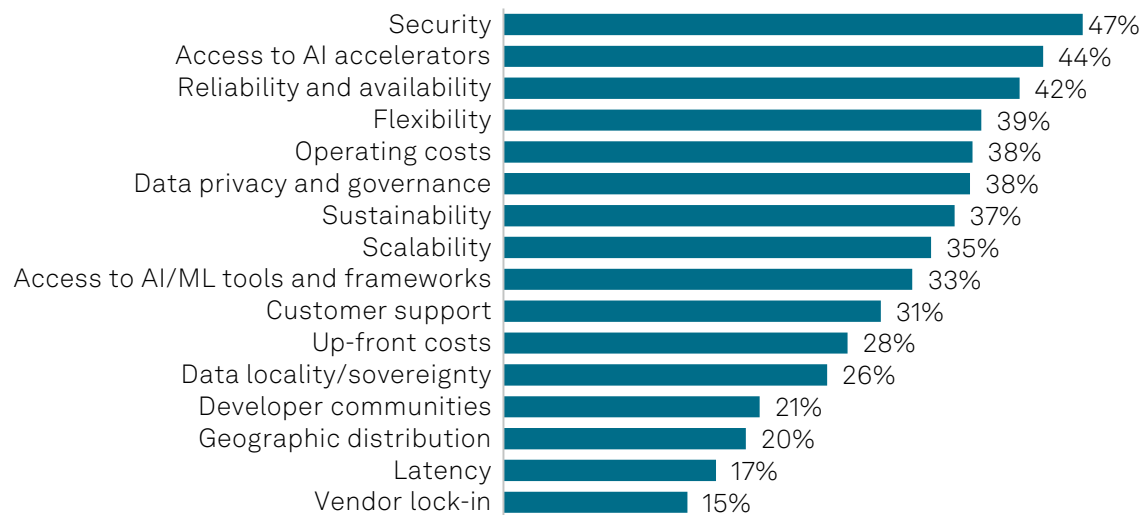
Key insights:

- After security, the leading factor in infrastructure decision-making is **accelerator availability**, identified by **44%** of organizations.
- Hyperscaler public clouds are one pathway to GPUs, but many are also turning to specialist AI clouds. GPU clouds are emerging as a key venue for both training — employed by almost a third, **32%** of organizations — and inference, **31%**.
- In some geographies, particularly in Asia-Pacific, lack of access to AI accelerators is already limiting organizations from moving models into production.

The leading infrastructure decision factors relate to security, AI accelerator access, and reliability and availability. As Figure 8 illustrates, access to AI accelerators ranks highly, outstripping even long-standing areas of concern such as operating costs and flexibility. Telecommunications companies (53%), higher education (53%) and manufacturing organizations (51%) prioritize this access particularly strongly.



Figure 8: Security and access to AI accelerators leading infrastructure decision-factors



Q. Which factors most influence the AI infrastructure decisions made at your organization?

Base: All respondents (n=1,519).

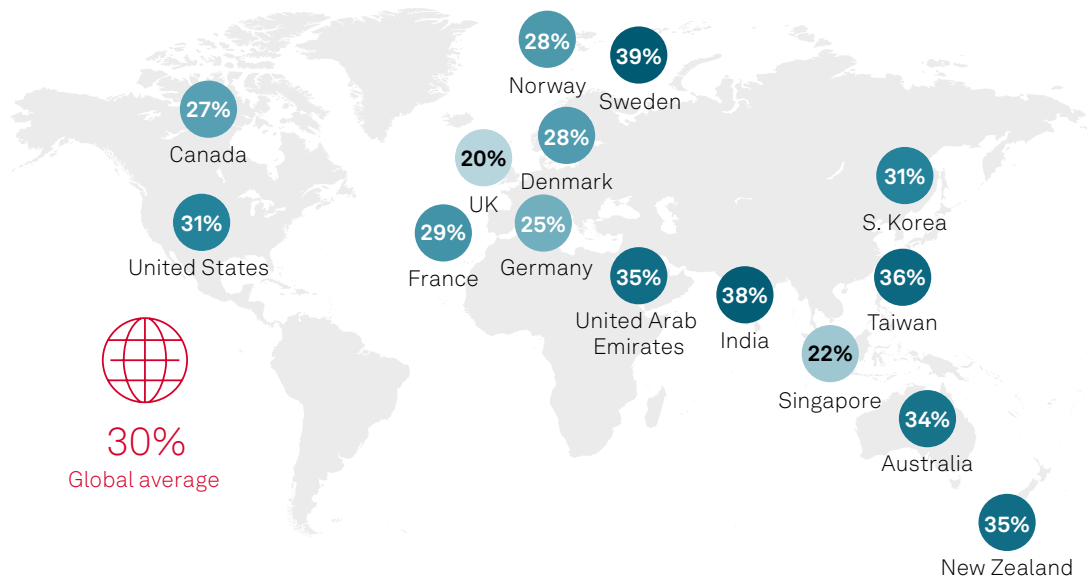
Source: S&P Global Market Intelligence 451 Research Global Trends in AI custom survey, 2024.

Hyperscaler public clouds offer an important avenue for organizations seeking GPUs, but they are not the only game in town. While hyperscaler cloud computing incumbents represent the most popular venue for AI training and inference, identified by 46% and 40% of organizations, respectively, specialist AI clouds have emerged as an auxiliary, or even alternative, venue. GPU clouds have seen an explosion in popularity that reflects the high demand for GPUs. Almost a third, 32%, of organizations that have invested in AI are executing training workloads using GPU clouds, and 31% for inference. These specialist cloud offerings are particularly popular with information technology and services companies, with 51% citing GPU clouds as a venue for training.

As the landscape of AI development and deployment widens, GPU clouds are poised to see further growth. Organizations are anticipating an increase in training and inferencing venues that will be employed over the next 12 months, and in this growth environment, GPU clouds are projected to grow to a 34% adoption rate for both inference and training. Higher education institutions appear to be a particularly fast-growing customer cohort. Our data highlights scalability as being the principal role organizations see GPU clouds as stepping into; the ability for organizations to easily and cost-effectively manage fluctuating AI workloads is a clear driver of adoption.

GPU availability challenges are felt keenly by organizations in some countries, including a number of major Asia-Pacific economies; India, Taiwan, New Zealand and Australia are more likely to rank GPU availability among their top three challenges to bringing a model into production. Sweden, where 39% identify it as a top three challenge, and UAE (35%) also stand out in this regard.

Figure 9: National differences in the level of impact GPU availability is having on bringing models into production



Q. What are the primary challenges or impediments to moving an AI/ML application from proof of concept/pilot stages to production environments? Ranks 1, 2 and 3.

Base: All respondents (n=1,519).

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



What are AI leaders doing differently?

Leaders are more likely to leverage GPU clouds for both training and inference and are particularly invested in how the technology can be used to reduce the time it takes to bring AI initiatives live. Leveraging specialist cloud-based services to expedite the development process by securing access to scarce GPU resources appears to be a clear avenue to drive through an AI advantage.



Concerns about AI's environmental impact persist but are not slowing AI adoption; sustainable AI practices offer opportunity to mitigate emissions

The story this year is about expansion: expanding scope of AI, expanding workload requirements, expanding data infrastructure. In this context, sustainability practices are key. With many sustainability practices delivering measurable value and the potential of applying AI to address energy consumption, there is a clear opportunity to address the emissions challenge.



Key insights:

- **Nearly two-thirds 64%** of respondents say that their organization is “concerned” or “very concerned” with the sustainability of AI infrastructure.
- Popular sustainability measures include investing in energy-efficient **IT hardware (42% of organizations)**, **increasing investment in AI governance (40%)**, **providing training and education on sustainability (37%)** and **establishing sustainability guidelines (35%)**.
- **Nearly one-third (30%)** of respondents say reducing energy consumption is a driver for AI/ML adoption at their organization.

Sustainability remains important to organizations, with almost two-thirds concerned about the impact AI/ML has on energy use and carbon footprints, including 25% that are “very concerned.” The topic is an influential factor in AI strategies, with 37% seeing sustainability as among the factors that most influence the AI infrastructure decisions made at their organization. As Figure 8 illustrates, this is comparable with operating costs, data privacy and scalability. This focus on sustainability is important in the context of expanding data demands, as well as venues for training and inference, forecast for AI workloads over the next 12 months.

The most identified (by 42% of respondents) sustainability approach that organizations have embarked on in the past 12 months is investing in more energy-efficient IT hardware. Notably, sustainability ranks as the second most common primary role for GPU clouds, just behind scalability. More efficient resource allocation and a commitment by many leading GPU cloud providers to employ more eco-friendly datacenters appears to align closely with this efficiency initiative.

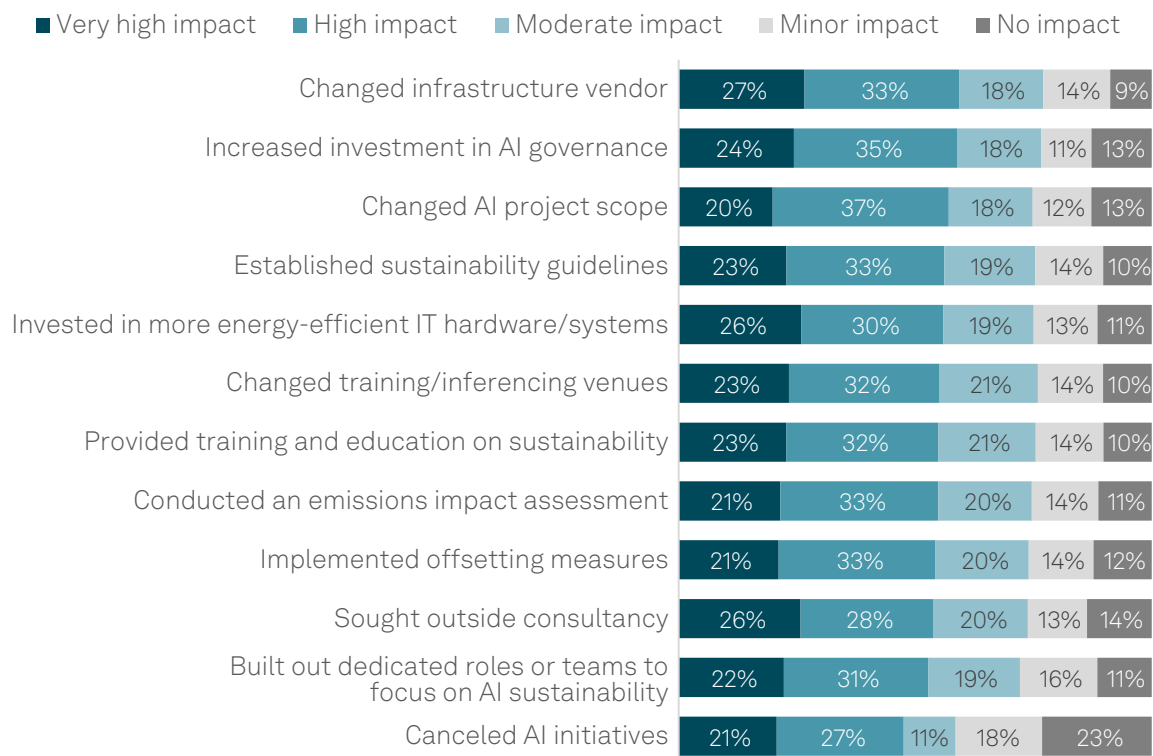
“In terms of ESG, it’s really around how we measure, drive and achieve our targets, through leveraging digital and then finance. It’s about end-to-end process efficiency and optimization. It is about how to automate things to drive back-office efficiencies ... it is top of mind.”

CIO, manufacturing/food and beverage
1,000-5,000 employees, UK

In addition to exhibiting a clear appetite to invest in better optimized technology, many organizations are increasing investment in AI governance (40%), providing training and education on sustainability (37%), and establishing sustainability guidelines (35%). Less popular are initiatives that would represent rolling back AI investments. Just 5% say they have canceled AI initiatives in the past 12 months, and 19% report they have changed project scope.

Organizations largely find sustainability practices to be effective. When identifying the projects that are more likely to deliver high impact than minor or no impact, we see that practices are more likely to deliver impact than not, including those that are sometimes overlooked (see Figure 10). While just 19% of organizations had engaged in changing AI project scope to address environmental concerns in the past 12 months, 57% of those that did see it as “highly” or “very highly” impactful. Another practice that may represent significant opportunity is changing infrastructure vendor — a step taken by 27% but considered the most impactful practice overall.

Figure 10: Sustainability steps assessed by impact



Q. What impact have you seen from the following actions on the environmental impacts of your AI projects?
 Base: Respondents from organizations that had invested in step over past 12 months, all respondents (n=1,519).
 Source: S&P Global Market Intelligence 451 Research Global Trends in AI custom survey, 2024.

A number of organizations are already applying AI to energy consumption — countering the resource-intensive nature of AI workloads with AI-driven efficiency benefits. Nearly one-third (30%) of organizations treat reducing energy consumption and carbon footprint as a key driver for developing AI/ML applications. While currently the least popular objective for AI/ML applications, it is notable that almost a third of organizations may already be applying AI to better assess and predict emissions or to inform efficiency improvements that could contribute to energy savings. As workload pressures expand, we expect this proportion to increase.

The drive to reducing energy consumption is not just about meeting sustainability goals. While 11% of respondents see meeting enterprise sustainability goals as their priority in reducing energy consumption, that was less than improved operational efficiency (13%) and achieving cost savings (12%). Concerns about regulatory compliance (9%) and customer expectations (9%) are also apparent. In some instances, this may give sustainability advocates the ability to address sustainability impacts while couching the value in the context of wider business goals.



What are AI leaders doing differently?

AI leaders are taking more steps to address the environmental impact of their projects and are more likely to see these steps as “highly” impactful. In particular, AI leaders are significantly more likely to consider changing training or inferencing venues over the next 12 months, establish sustainability guidelines and implement offsetting methods. By engaging with multiple steps in parallel, the value of each appears to be magnified. While an emissions impact assessment may be useful, it delivers more value when that assessment can be converted into decisions about choice of vendor partner or AI training or inferencing environments.

Conclusions

The 2024 Global Trends in AI report represents a very different AI-adoption landscape than the 2023 edition. AI is gaining more widespread implementation, with greater focus on delivering product and service quality improvements and revenue growth. The maturation of generative AI is a key driver of this transition. Yet challenges remain. Many organizations are struggling to shift investments into capabilities that they can deliver at scale, and they acknowledge there is pressure on the sustainability of business operations.

Five key action points for businesses:



Build a robust data architecture for AI success: Organizations must establish a clear pathway for scaling AI projects into production, ensuring efficient data management and storage. It is crucial to invest in a strong data foundation before committing to high volumes of pilot projects. This will help enable seamless AI value delivery.



Smart investment is key to generative AI success: Organizations benefiting from generative AI have redistributed their budgets to focus on these initiatives. Success hinges on sophisticated decision-making and robust infrastructure. To emulate this, organizations should ensure comprehensive purchasing practices and maximize GPU availability and utilization, including investigating specialized GPU and AI cloud services.



Explore generative AI-driven IT efficiencies: Generative AI can automate routine model development tasks and improve IT decision-making to drive more streamlined delivery. This self-reinforcing approach can underpin a more sustainable AI roadmap.



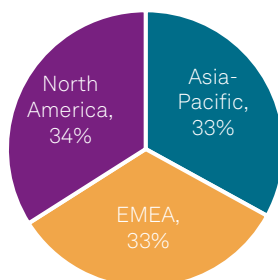
Expand sustainability practices: Changing infrastructure vendors or revising AI project scopes can have a meaningful impact on overall emissions. The value of sustainability measures is compounded when they are used together, so organizations should empower project teams to engage with a variety of approaches.



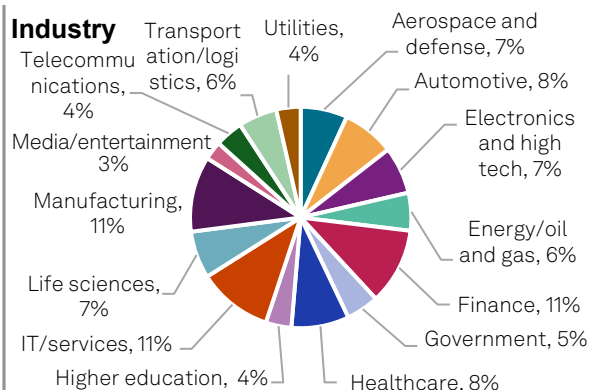
Build a holistic AI strategy: Generative AI offers significant opportunity, but organizations should create a well-rounded AI strategy. A narrow approach to AI — one that fails to investigate a mix of technologies — neglects the opportunity to bring together different types of models and closes off many high-impact use cases.

Methodology

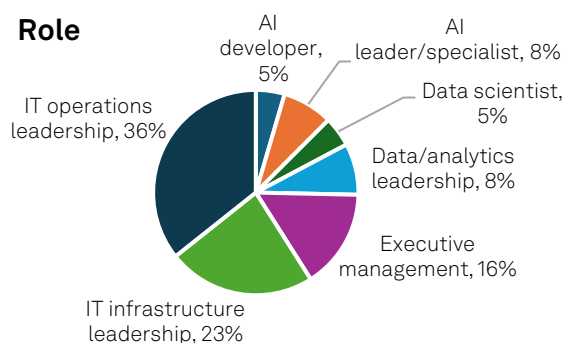
Region



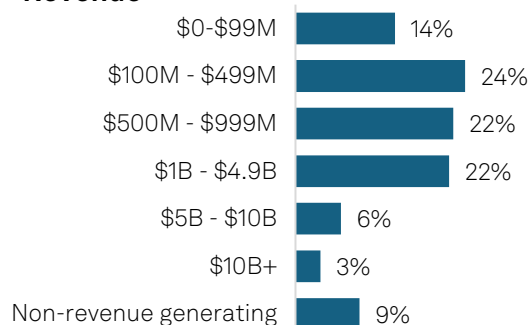
Industry



Role

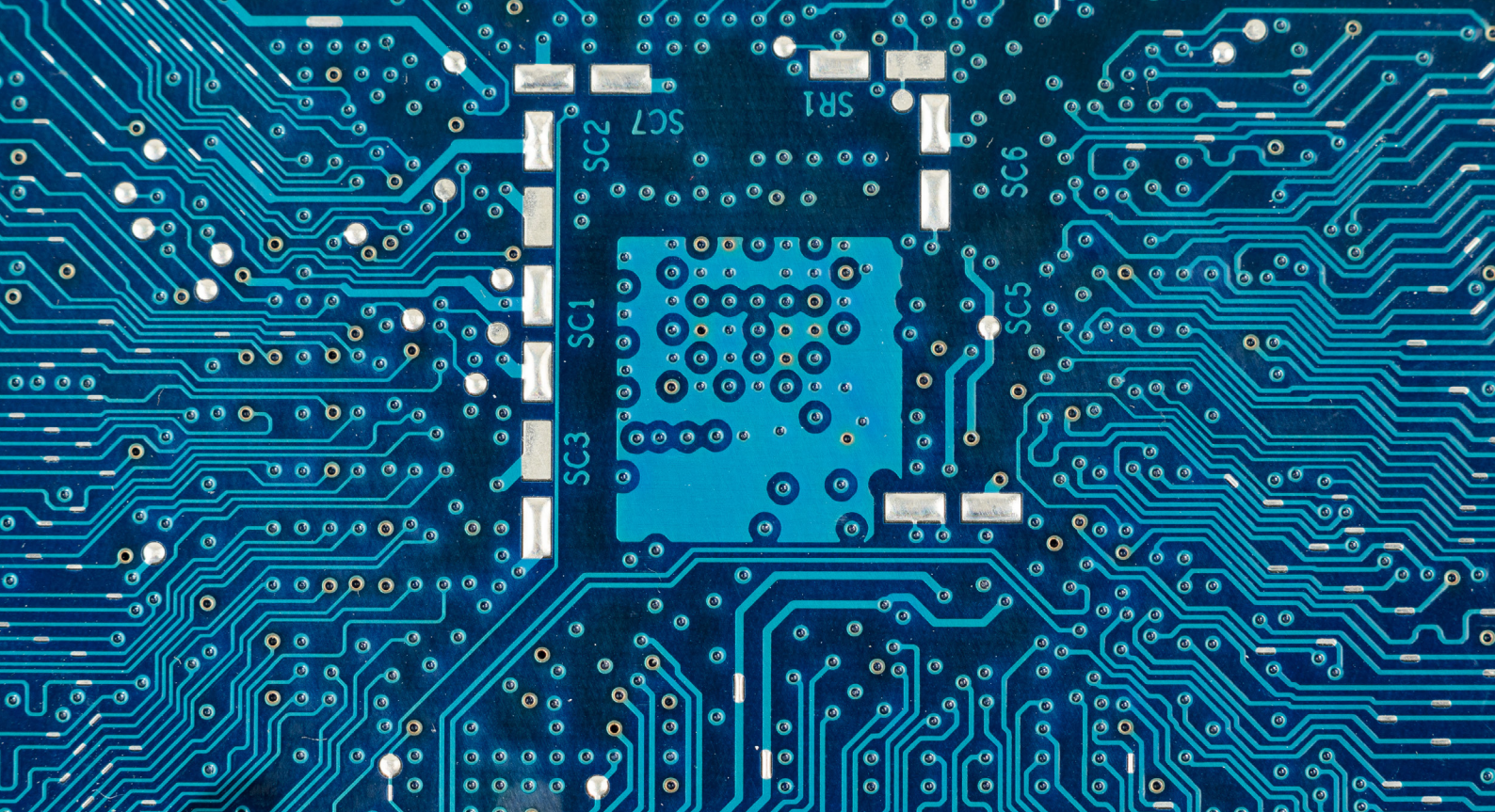


Revenue



The findings presented in this report draw on a survey fielded in North America, Europe, the Middle East and Africa, and Asia-Pacific in Q2 2024.

The survey targeted 1,519 AI/ML decision-makers/influencers, filtering for respondents with AI/ML deployed in pilots and production environments across the following industries: aerospace and defense, automotive, energy/oil and gas, finance, government, healthcare, higher education, IT and services, life sciences, manufacturing, media/entertainment, telecommunications, transportation and logistics, and utilities. The most common respondent job roles were IT operations leadership, IT infrastructure leadership and executive management. This report also draws on contextual knowledge of additional research conducted by S&P Global Market Intelligence 451 Research.



Brought to you by



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](#)

About the authors



Alex Johnston

Senior Research Analyst, Data, AI & Analytics

Alex Johnston is a senior research analyst on the 451 Research Data, AI & Analytics team at S&P Global Market Intelligence. He focuses on emerging technologies and how they can be applied in business contexts. Alex's primary coverage areas are artificial intelligence, distributed ledger technology and event stream processing. Alex's recent areas of concentration include monitoring the emerging generative AI market, tracking the evolution in blockchain use cases and investigating real-time architectures.



David Immerman

Consulting Analyst

David Immerman is a consulting analyst at S&P Global Market Intelligence. He is responsible for executing on a range of custom research initiatives and development of thought leadership across technology sectors including industrial IoT, digital transformation, edge computing, AI/machine learning and fintech, among others, and verticals such as manufacturing and automotive.

About this report

A Discovery report is a study based on primary research survey data that assesses the market dynamics of a key enterprise technology segment through the lens of the “on the ground” experience and opinions of real practitioners — what they are doing, and why they are doing it.

About S&P Global Market Intelligence

At S&P Global Market Intelligence, we understand the importance of accurate, deep and insightful information. Our team of experts delivers unrivaled insights and leading data and technology solutions, partnering with customers to expand their perspective, operate with confidence, and make decisions with conviction.

S&P Global Market Intelligence is a division of S&P Global (NYSE: SPGI). S&P Global is the world's foremost provider of credit ratings, benchmarks, analytics and workflow solutions in the global capital, commodity and automotive markets. With every one of our offerings, we help many of the world's leading organizations navigate the economic landscape so they can plan for tomorrow, today. For more information, visit www.spglobal.com/marketintelligence.

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

Copyright © 2024 by S&P Global Market Intelligence, a division of S&P Global Inc. All rights reserved.

These materials have been prepared solely for information purposes based upon information generally available to the public and from sources believed to be reliable. No content (including index data, ratings, credit-related analyses and data, research, model, software or other application or output therefrom) or any part thereof (Content) may be modified, reverse engineered, reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of S&P Global Market Intelligence or its affiliates (collectively S&P Global). The Content shall not be used for any unlawful or unauthorized purposes. S&P Global and any third-party providers (collectively S&P Global Parties) do not guarantee the accuracy, completeness, timeliness or availability of the Content. S&P Global Parties are not responsible for any errors or omissions, regardless of the cause, for the results obtained from the use of the Content. THE CONTENT IS PROVIDED ON "AS IS" BASIS. S&P GLOBAL PARTIES DISCLAIM ANY AND ALL EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OR USE, FREEDOM FROM BUGS, SOFTWARE ERRORS OR DEFECTS, THAT THE CONTENT'S FUNCTIONING WILL BE UNINTERRUPTED OR THAT THE CONTENT WILL OPERATE WITH ANY SOFTWARE OR HARDWARE CONFIGURATION. In no event shall S&P Global Parties be liable to any party for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including, without limitation, lost income or lost profits and opportunity costs or losses caused by negligence) in connection with any use of the Content even if advised of the possibility of such damages.

S&P Global Market Intelligence's opinions, quotes and credit-related and other analyses are statements of opinion as of the date they are expressed and not statements of fact or recommendations to purchase, hold, or sell any securities or to make any investment decisions, and do not address the suitability of any security. S&P Global Market Intelligence may provide index data. Direct investment in an index is not possible. Exposure to an asset class represented by an index is available through investable instruments based on that index. S&P Global Market Intelligence assumes no obligation to update the Content following publication in any form or format. The Content should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. S&P Global keeps certain activities of its divisions separate from each other to preserve the independence and objectivity of their respective activities. As a result, certain divisions of S&P Global may have information that is not available to other S&P Global divisions. S&P Global has established policies and procedures to maintain the confidentiality of certain nonpublic information received in connection with each analytical process.

S&P Global may receive compensation for its ratings and certain analyses, normally from issuers or underwriters of securities or from obligors. S&P Global reserves the right to disseminate its opinions and analyses. S&P Global's public ratings and analyses are made available on its websites, www.standardandpoors.com (free of charge) and www.ratingsdirect.com (subscription), and may be distributed through other means, including via S&P Global publications and third-party redistributors. Additional information about our ratings fees is available at www.standardandpoors.com/usratingsfees.