Enterprise Strategy Group by TechTarget

SHOWCASE

AUGUST 2024

Simplifying, Accelerating, and Optimizing Enterprise Generative AI Adoption

Mike Leone, Practice Director and Principal Analyst

Abstract: Generative AI adoption is accelerating as organizations learn valuable lessons during proofs of concept (POCs) and turn those insights into concrete steps toward enterprise-wide production use cases. But that journey needs the guidance and advice of an experienced hand—a digital sherpa—to show the way. Teaming with a proven partner familiar with how to leverage large language models for generative AI is a smart move.

Introduction – Generative AI Exposure and Experimentation Promotes Wider Acceptance

As generative AI interest spikes and investment increases, organizations are taking stock of the progress of their initial forays into generative AI use cases. Those use cases are critical in helping to turn generative AI from a promising technology to a transformative real-world catalyst for innovation and efficiency. Topping the list of those use cases include automating business processes, supporting analytics tasks, increasing employee productivity, improving operational efficiency, and enhancing the customer experience, although there are many others fast emerging as exciting opportunities.[1]

Market Insight

More than a quarter (26%) of organizations have generative AI use cases in production, while another 24% have solutions in pilots or POCs.

Generative AI adoption is fast reaching critical mass, according to research from TechTarget's Enterprise Strategy Group. More than a quarter (26%) of organizations have generative AI use cases in production, while another 24% have solutions in pilots or POCs.[2]

Just as important, organizations increasingly see themselves as having embraced an AI culture: More than one-third (34%) of organizations said AI is fully embedded into their culture and operations, while another 27% said they are expanding AI more broadly across the business.[3]

Another sign of AI's growing importance is the active participation of C-suite executives and business stakeholders in generative AI purchasing and use case decisions, not just the technical teams from IT, data science, and data analytics. For instance, 30% of C-level/VP leadership teams said they contribute to shaping generative AI initiatives.

[1] Source: Enterprise Strategy Group Research Report, *Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns*, August 2023.
[2] Source: Enterprise Strategy Group Complete Survey Results, *Navigating the Evolving AI Infrastructure Landscape*, December 2023.
[3] Ibid.

Other vital organizational functions also noted they are involved in this process, including 24% of marketing/sales personnel, 23% of cybersecurity professionals, and 19% of risk/compliance experts.[4]

This highlights a growing trust and confidence in generative AI and a commitment to tackling more innovative projects and a broader set of use cases.

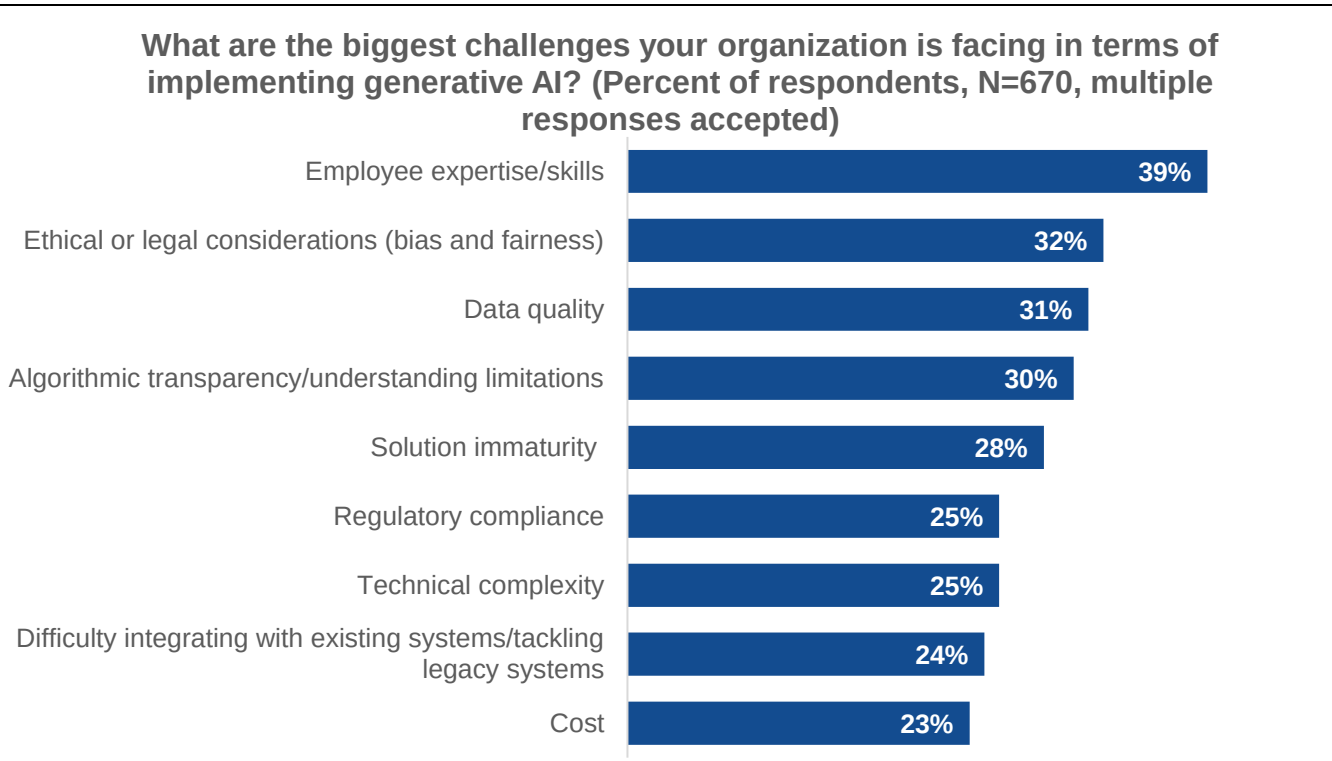
The Problem – Few Organizations Can Move From Concept to Production Without Help

Moving a generative AI project from POC to production is hard, and it will likely remain difficult even as adoption continues to grow and in-house teams become more familiar with generative AI large language models (LLMs).

Survey respondents told Enterprise Strategy Group that a significant problem area is the increasingly complex demands for regulatory compliance, security, and responsible-use standards. More than half (51%) of organizations said they face challenges associated with balancing competing goals of accuracy, performance, fairness, and ethics related to machine learning models.[5]

Asked further about challenges, organizations also voiced a number of concerns. Leading the list of top challenges include confronting and overcoming the large and growing generative AI skills gap. In fact, 39% of organizations told Enterprise Strategy Group that employee expertise/skills is a top challenge in implementing generative AI projects, making it the number-one speed bump cited (see Figure 1).[6]

Figure 1. Top 10 Challenges Faced Implementing Generative AI



Source: Enterprise Strategy Group, a division of TechTarget, Inc.

[4] Source: Enterprise Strategy Group Research Report, *Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns*, August 2023.
[5] Source: Enterprise Strategy Group Research Brief, *Navigating the AI Ethics Landscape: The Case for Governance*, January 2024.
[6] Source: Enterprise Strategy Group Research Report, *Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns*, August 2023.

Additionally, there is substantial pressure to shorten the time to value from generative AI investments: 28% of organizations said it took at least three months before they saw value from their AI initiatives.[7] Respondents noted a number of other key challenges:

- "Limited availability of quality data for models" is the number-one challenge organizations cited in AI implementations.[8] Growing popularity and endorsement of LLMs and generative AI puts even more pressure on organizations to keep up, both in terms of infrastructure investment and in bridging the skills gap as new use cases emerge.
- The battle for precious resources such as budget, infrastructure capacity, and in-house skills will only intensify.
- The lack of a modern data strategy and questions about data readiness in the cloud will add further uncertainty for organizations looking to derive tangible value from their generative AI investments.

The Right Approach – Don't Try to Go It Alone

Moving beyond POC to production is complex and potentially expensive and can take longer than necessary to generate positive return on investment (ROI). Organizations need solutions that deliver a smooth, reliable transition from POC to production, with guardrails for secure, appropriate usage.

Keeping in mind organizations' earlier concerns about the limited availability of quality data, organizations need to keep data preparedness and data quality as top priorities to ease the transition from a proof of concept to a production system. Additionally, organizations must focus on ensuring their generative AI architecture is built upon the importance of unsiloed data, all in a single spot.

Large language model operations (LLMOps) is essential in order to make implementation efficient, scalable, and consistent with best practices and organizations' key performance indicators. That means it is critically important for organizations to team with experienced third parties that have "walked the walk."

Enterprise Strategy Group's research pointed out that AI infrastructure management is heavily weighted toward third parties (76%), and preconfigured AI tools and services are used by 35% of organizations in order to address skills gaps or help needed for AI initiatives.[9] Organizations need provable, deployable solutions that are based upon a solid, market-ready LLMOps foundation. They also need a well-structured and curated data ecosystem in order to properly train generative AI models at scale. This is essential in order to promote highly desirable characteristics in generative AI solutions, including:

- Observability.
- Scalability.
- Efficient data access.
- Preprocessing.
- Simplified management.
- Comprehensive data governance.
- Security.
- Integrated data architecture.

Market Insight

76% of organizations are turning to third parties to help manage AI infrastructure to some extent.

[7] Source: Enterprise Strategy Group Research Report, *Navigating the Evolving AI Infrastructure Landscape*, September 2023.
[8] Source: Enterprise Strategy Group Research Report, *Navigating the Evolving AI Infrastructure Landscape*, September 2023.
[9] Ibid.

Accelerating Generative AI Initiatives With Grid Dynamics and AWS

Grid Dynamics, in concert with its strategic partner Amazon Web Services (AWS), has a demonstrated track record in helping organizations bring their generative AI initiatives from ideas and concepts to production-class systems with tangible ROI.

Since its inception in 2006, Grid Dynamics has focused on building and delivering cloud-native solutions using a wide range of leading-edge technologies, including generative AI and LLMOps. For more than 15 years, Grid Dynamics has worked with AWS technologies on solutions such as cloud operations and DevOps and now has built a close relationship with AWS in taking organizations on their generative AI journey from initial ideas and goals to fully functioning solutions. Grid Dynamics holds a significant number of AWS consulting partner certifications across different industries, delivery methodologies, and AWS product competencies.

One of the innovative solutions Grid Dynamics has created and delivered for AWS customers looking to create real business value from their generative AI investments is its LLMOps Platform Starter Kit for AWS. This tool enables streamlined development, deployment, and operationalization of LLM projects in AWS environments, built on the following core concepts:

- **Data management.** Ensuring the accuracy, integrity, finetuning, and observability of data throughout its lifecycle.
- **Architectural design.** Developing a robust framework that supports the scalability and integration needs of the system, including implementing guardrails at various intervention levels (input, retrieval, generation, output) to monitor and control LLM behavior.
- **Retrieval-augmented generation (RAG).** Integrating RAG systems with LLMs to enhance accuracy and relevance by retrieving domain-specific information, enabling models to retrieve relevant, up-to-date information for a given context.
- **Deployment.** Rolling out models or systems to production environments efficiently, and managing prompts as a logical extension of experiment management for generative AI.
- **Data privacy and protection.** Safeguarding sensitive information against unauthorized access and ensuring compliance with legal regulations.
- **Ethics and fairness.** Addressing and mitigating biases, ensuring transparency, and maintaining fairness in data usage and algorithmic decision-making. Also, utilizing guardrails to reduce hallucinations, toxicity, and contradictions in LLM outputs, as well as implementing moderation, topical, and evidence-based guardrails to align LLM behavior with organizational guidelines.

This Grid Dynamics toolkit is an attractive approach for many organizations looking to overcome the challenges mentioned earlier in this paper, including the AI skills gap, overcoming technical hurdles in development and deployment, performance scalability, and the need for greater visibility and observability in managing LLMs.

Its relationship with AWS enables Grid Dynamics to help organizations make the LLM lifecycle journey more efficient and faster, as well as to improve security, system performance, and demonstrate faster time to value.

Conclusion

In order to address the many challenges associated with the still-evolving spectrum of generative AI use cases, organizations should look for a seasoned professional to help guide them from a POC to a production-class solution. Doing so only or mainly through internal teams is a recipe for inefficiency, at best, and for failed projects, at worst.

Organizations in AWS environments should investigate, evaluate, and link up with generative AI experts such as Grid Dynamics to make their generative AI journey a successful one. Grid Dynamics has a demonstrated track record of success in the still-nascent generative AI space, and its close relationship with AWS acts as a force multiplier in helping organizations get their generative AI projects off the ground, and on the path toward substantial ROI now and in the future.

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com
www.esg-global.com