

MARCH 2024

The Rapid Rise of Enterprise AI and What It Means to You

Mike Leone, Principal Analyst

Abstract: Enterprise-level artificial intelligence—particularly generative AI (GenAI)—is transforming how organizations accomplish critical goals, from driving revenue and profits to creating a superior customer experience. AI is soaring up the list of most organizations' top priorities, attracting more attention and support from the C-suite, the board, and lines of business (LOBs). AI has certain challenges and hurdles to overcome, but organizations are plowing ahead because of the technology's massive potential for change, improvement, and innovation—particularly at the client level. Here's why enterprise AI has captured everyone's imagination and why organizations need the help of technology partners.

The Exploding Demand for Enterprise Al Solutions

In a remarkably short period of time, AI has vaulted onto and up the lists of enterprise technology priorities. In fact, it now outpaces cloud computing as a key initiative and is rising rapidly, according to research from TechTarget's Enterprise Strategy Group. In fact, in a recent survey, 54% of enterprises said they intend to adopt GenAI within the

next 12 months.¹ The keyword here is *adopt*, not *consider* or *evaluate*. GenAl is an action item, often driven by LOB stakeholders who want to leverage Al for their most critical business goals across a wide variety of applications and use cases. Those requirements have the broad support of C-suite executives and board members, many of whom are asking *how* and *when* to make GenAl a key part of the operating plan—not *if*.

Market Insight



54% of enterprises intend to adopt GenAl within the next 12 months.

Al is widely seen as a highly leverageable force multiplier that does essential work faster and often more precisely than humans do. It also delivers additional value by freeing up staff to hand off routine tasks to Al tools and machine learning algorithms and focus instead on more innovative and creative work that opens up new business opportunities. At the same time, organizations must acknowledge some early-stage challenges associated with adopting and using GenAl in real-world enterprise settings. For instance, just 36% of organizations reported that they have a policy in place that limits the use of GenAl. It's also important to note that, according to the same survey, only 46% of organizations are working to create an acceptable use policy for Al.

Other important challenges exist, including bridging the significant and growing AI skills gap, identifying and overcoming emerging ethical and legal considerations, and ensuring and delivering pristine data quality. Of course, GenAI use cases—whether in the cloud, in an on-premises setting, or in a hybrid model—must account for the large, diverse, and always shifting regulatory compliance mandates, particularly those affecting data privacy.

¹ Source: Enterprise Strategy Group Research Report, <u>Beyond the GenAl Hype: Real-world Investments</u>, <u>Use Cases</u>, <u>and Concerns</u>, August 2023. All Enterprise Strategy Group research references and charts in this showcase are from this report, unless otherwise noted.



GenAl Opportunities in the Enterprise: High-value Use Cases Make It Essential

For any technology to flourish, real-world use cases are essential. Those use cases require the right infrastructure models, as well as powerful software that can be deployed in a variety of settings. In the fast-moving world of GenAl, use cases now abound. A key driver for the emergence of those use cases was the introduction of large language models (LLMs) for artificial intelligence. The good news: Even more LLMs are on the way, offering even more functionality, faster learning, and more useful inference engines. The *even better* news: As these models become more powerful, they are also becoming more precise and highly contextual. This results in practical, innovative solutions for dramatic—even unexpected—advances.

Market Insight



The top use case for GenAl today, cited by nearly a third of organizations (31%), is for data insights.

When use cases take hold, their sheer diversity encourages even more experimentation, more pilot programs, and the organizational commitment of major resources—budgets, staff, IT infrastructure, and industry partnerships. Research from Enterprise Strategy Group emphasizes that the single biggest category of use cases relates to data insights—that is, the ability to start with huge volumes of raw data and provide contextual awareness and relevance to transform them in

actionable insights. Nearly one-third of organizations (31%) said data insights was a top priority for their GenAl projects, making it the top area of emphasis.

What other use cases are garnering substantial attention from organizations? One such use case near the top of the list—and certain to continue to gain in importance—is cybersecurity threat analysis. Keeping up with the rapid pace of emerging threats is critically important to organizations, which are struggling to plug the growing cybersecurity skills gap. This is an area where organizations are using AI to level the field against attackers. Other important use cases include workflow automation, cybersecurity threat hunting, contact center knowledge base, incident response, and chatbot dialog. But perhaps the most telling prediction about the importance of GenAI in decision-making use cases was noted by Enterprise Strategy Group: "To better assess GenAI, organizations will turn to...GenAI."

The Need to Support Hybrid Al Environments With Client-based Infrastructure

Based on the diversity of enterprise use cases for Al—especially those implemented on client devices—a key decision must be made as to where those use cases should be deployed: in the cloud, on premises in the data center, on premises on high-performance PCs, or in a hybrid of multiple scenarios. In fact, organizations indicated to Enterprise Strategy Group that, while they tend toward public or hybrid cloud environments for GenAl use case deployments, 40% of those use cases will be centered in on-premises data centers, edge locations, or colocation facilities.²

These decisions are, of course, highly unique to each organization's priorities, which encompass compliance and reporting requirements, security and privacy considerations, latency and network access performance, budgetary constraints, and the need for high resilience. Additionally, as data footprints continue to expand dramatically, organizations are looking for a complementary approach to the public cloud with better cost predictability and to ensure availability of resources in real time. This leads to the inevitable need for hybrid deployment and operating models, using the best of the public cloud along with the best of client-based, on-premises infrastructure.

While many initial enterprise PC deployments were done in the cloud, organizations now embrace the wisdom and flexibility of deploying AI using a ground-up, client-based model. That puts a strong emphasis on having the right

² Source: Enterprise Strategy Group Research Report, Navigating the Evolving Al Infrastructure Landscape, September 2023.



compute infrastructure in place to support what are likely to be performance-intensive and security-sensitive use cases

Instead of simply "parking" AI infrastructure in the cloud, the smart strategy will be to use client-based PCs in order to complement cloud-based AI processing. AI on next-generation client hardware provides enhanced control, flexibility, and privacy. Endpoint-based solutions will enable organizations to maintain control over their data absolutely, which is especially useful when dealing with personal information or confidential business data. As shown in Figure 1, Enterprise Strategy Group research uncovered several factors influencing the location of AI environments.³

Figure 1. Factors Influencing the Physical Location of an Al Environment





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

How AMD Meets Enterprises' Needs for AI Infrastructure

At the heart of that AI infrastructure are the chip components enabling solutions that not only run in the cloud, but also on the edge and on high-powered endpoints. Those components include:

- Processors for both servers and endpoints.
- Accelerators for requirements such as Al inference and Al learning.
- Graphics cards to support visually driven AI use cases.
- System-on-a-chip (SoC) solutions.

One long-established technology partner that has made a pronounced commitment to chip-level solutions supporting AI use cases is AMD. The AMD commitment to AI solutions spans all the above-mentioned solution types, enabling AMD to work with a broad range of OEM partners and enterprise clients to help spread the benefits of AI.

The AMD focus on AI takes three forms:

3

³ Ibid.



- A broad and deep solutions portfolio.
- An extensive, open ecosystem of hardware and software partners.
- User experiences for AI use cases that can fit both demanding enterprise-level requirements and budget-limited workloads.

AMD AI infrastructure, designed to power real-world implementations of GenAI and LLMs, includes a number of hardware solutions. One of the key elements is AMD Ryzen™ processors with AMD Ryzen™ AI, which leverage native AI engines and help deliver such capabilities as accelerated multitasking and improved user productivity.

 Along with other AMD solutions such as Instinct[™] and Alveo[™] accelerators, EPYC[™] server processors, Versal[™] adaptive SoCs, and Radeon[™] graphics cards, Ryzen processors enable AMD to work with leading hardware and software partners to provide sophisticated AI solutions for enterprises.

AMD solutions help organizations go beyond high performance for training and inference requirements, offering superior scalability to deal with increasingly complex and demanding LLMs still to come.

Current and emerging AI use cases demand a wide range of features from the underlying infrastructure, including low latency, high throughput, privacy and security functionality, and cost efficiency to make AI feasible and affordable at the endpoint level. AMD has engineered its products to help those building AI solutions implement them across the full spectrum of architecture—from the local endpoint to the data center and from the edge to the cloud.

Conclusion: Making GenAl the Heart of Enterprise Business Strategy

Generative AI is much more than just an exciting, high-potential technology; it is already paying dividends for organizations in every industry and of all sizes. There is little debate that it is an optimizable asset that promotes innovation and delivers tangible business value. Organizations are also learning that client-based AI initiatives are invaluable to their overall AI strategies, bringing control, flexibility, and performance to PC-based AI use cases that complement both server-based and cloud-based AI programs.

Al in the enterprise still has challenges to meet and overcome. Those challenges, however, are business decisions on policies, priorities, the types of guardrails to be created on proper usage, and deployment options. The technology's utility and performance are undeniably appropriate for transformative use cases, supported by exciting advances in Al infrastructure from the client to the data center and from the edge to the cloud. A hybrid approach to enterprise Al—combining local Al solutions with those developed and deployed in the cloud—represents an important advancement in making GenAl a truly useful, cost-efficient business breakthrough.

Innovative technology leaders such as AMD and others are laying the groundwork for new client-based AI solutions. As a result, organizations now have the flexibility, control, and performance to use GenAI in ways that enhance security and help deliver new capabilities from the ground up.

For more information on how AMD helps enable enterprise AI at a client level and at the edge, please visit https://www.amd.com/en/solutions/ai.html.

©TechTarget, Inc. or its subsidiaries. All rights reserved. TechTarget, and the TechTarget logo, are trademarks or registered trademarks of TechTarget, Inc. and are registered in jurisdictions worldwide. Other product and service names and logos, including for BrightTALK, Xtelligent, and the Enterprise Strategy Group might be trademarks of TechTarget or its subsidiaries. All other trademarks, logos and brand names are the property of their respective owners.

Information contained in this publication has been obtained by sources TechTarget considers to be reliable but is not warranted by TechTarget. This publication may contain opinions of TechTarget, which are subject to change. This publication may include forecasts, projections, and other predictive statements that represent TechTarget's assumptions and expectations in light of currently available information. These forecasts are based on industry trends and involve variables and uncertainties. Consequently, TechTarget makes no warranty as to the accuracy of specific forecasts, projections or predictive statements contained herein.

Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of TechTarget, is in violation of U.S. copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact Client Relations at cr@esg-global.com.