Accelerating generative AI innovation: A comprehensive guide for technology leaders

Discover how to build a foundation for generative AI success

aws

Table of contents

INTRODUCTION

Are you ready to capitalize on generative AI?

Generative artificial intelligence (gen AI) presents opportunities for organizations to reimagine customer experiences, boost employee productivity, unlock growth opportunities, and drive innovation. To realize this future, organizations need more than a single, powerful large language model (LLM) or chat assistant. They need a full range of capabilities to build and scale gen AI applications that are tailored to their business and use cases—including applications with built-in gen AI, tools to rapidly build their own gen AI apps, a cost-effective and high-performance infrastructure, and robust security controls.

While gen AI is accelerating, many organizations face challenges moving from proofs of concept to real-world applications. Another universal challenge is identifying and implementing the right use cases to align with specific business goals. This guide is designed to help you unlock the value of gen AI quickly, scale securely, and innovate with higher performance and lower cost.

More than 80% of enterprises will have deployed gen AI applications by 2026[1]

At least 30% of gen AI projects will be abandoned after proof of concept by the end of 2025 due to poor data quality, inadequate risk controls, escalating costs, or unclear business value[2]

[1] "Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026," Gartner, October 2023
[2] "Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025," Gartner, July 2024

BUSINESS OPPORTUNITIES

Drive real business value with generative AI

Gen AI has the potential to create $2.6–$4.4 trillion in business value globally.[3] It can help your organization in four fundamental ways:

Creating new experiences

Deliver innovative and engaging ways to engage with customers and employees.

Extracting insights

Surface valuable insights from large volumes of documents and data to improve decision making.

Boosting employee productivity

Empower employees to create content, code, and ideas more efficiently.

Enhancing creativity

Generate new content, ideas, conversations, and multimedia.

Gen AI has the potential to create

$2.6–$4.4T

in business value globally[3]

[3] "The economic potential of generative AI: The next productivity frontier," McKinsey Digital, June 2023

BUSINESS CONSIDERATIONS

Generative AI brings opportunities and challenges

As a technology leader, you play a pivotal role in driving the successful adoption of gen AI in your organization. However, you must navigate the complexities of this cutting-edge technology and strike a balance between seizing the opportunities and mitigating the risks.

Key challenges implementing generative AI

Data infrastructure modernization, integration, and scalability

Legacy systems inhibit advanced analytics and AI capabilities and bring substantial capacity constraints.

Regulatory, ethical, data sovereignty, and data residency considerations

Leaders must navigate the complex, uncertain, and increasingly ambiguous regulatory landscapes and comply with relevant laws.

Balancing costs while maintaining performance

Training, building, and deploying gen AI models is challenging with budgets that remain close to flat year over year.

Model selection

As the landscape of language models continues to evolve, technical leaders are challenged to identify the most suitable model for each use case.

USE CASES

Generative AI use cases for every business

Gen AI has the potential to transform nearly every industry and function within your organization. The key is identifying the right use cases that align with your specific business needs and goals. Ask yourself: What requirement is most critical to the organization's success?

Some of the high-impact areas where generative AI can drive value include:

Enhance customer experiences

- Chatbots and virtual assistants
- Agent assist and call analytics
- Hyper-personalization

Boost employee productivity

- Conversational search
- Code generation
- Automated report generation
- AI-generated marketing content, sales content, guidance, and enablement
- New product development

Optimize business processes

- Intelligent document processing (IDP)
- Data augmentation
- Supply chain optimization

RESPONSIBLE AI

A commitment to safe, secure, and transparent artificial intelligence services and tools

You want your gen AI applications to be built with safety, fairness, and security in mind so that you can, in turn, deploy AI responsibly. Our practical approach to transforming responsible AI from theory into practice, coupled with tools and expertise, helps you implement responsible AI practices effectively within your organization. We are investing in new capabilities to foster responsible gen AI innovation across a broad set of capabilities, including built-in tools, to help ensure fairness, transparency, and customer protection, as well as resources to combat disinformation.

Learn more about responsible AI →

Responsible AI

Responsible AI is the practice of designing, developing, and using AI technology in a responsible manner, with the goal of maximizing benefits and minimizing risks and unintended harms. At Amazon Web Services (AWS), we define responsible AI using a set of core dimensions that we assess and update over time as AI technology evolves.

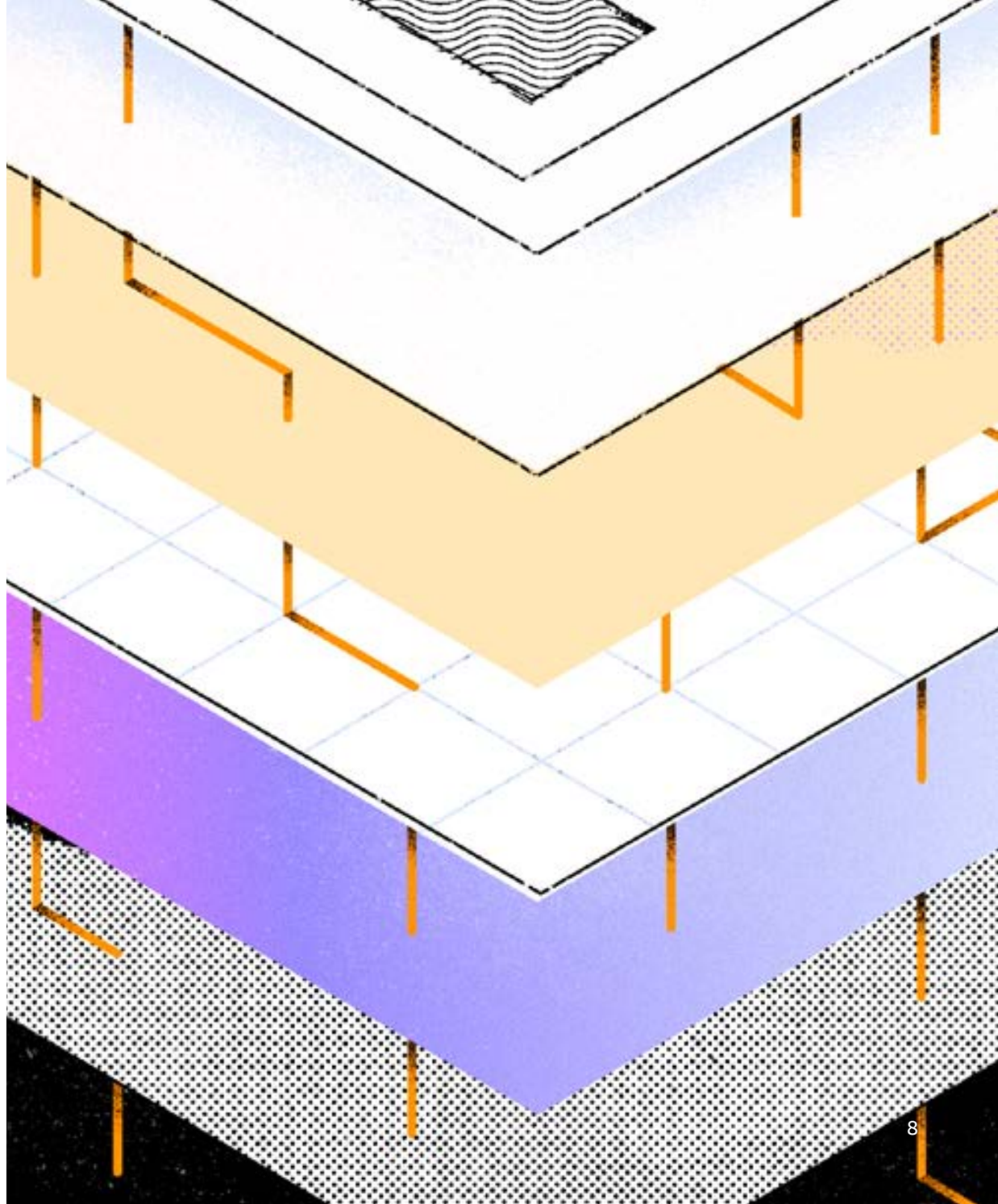
Core dimensions of responsible AI

- Controllability
- Privacy and security
- Safety
- Fairness
- Veracity and robustness
- Explainability
- Transparency
- Governance

RESPONSIBLE AI

Tools for building responsibly with artificial intelligence

- Amazon Bedrock Guardrails helps users implement safeguards tailored to gen AI applications and aligned with responsible AI policies
- Model evaluation on Amazon Bedrock helps users evaluate, compare, and select the best foundation models (FMs) for their specific use case based on custom metrics, such as accuracy, robustness, and toxicity
- Watermarking in Amazon Nova Canvas, Amazon Nova Reel, and Amazon Titan Image Generator models is designed to help reduce the spread of disinformation by providing a discrete mechanism to identify AI-generated images from these models
- AWS AI Service Cards are a resource for enhancing transparency by providing users with a single place to find information on use cases and limitations, responsible AI design choices, and performance optimization best practices for our AI services and models



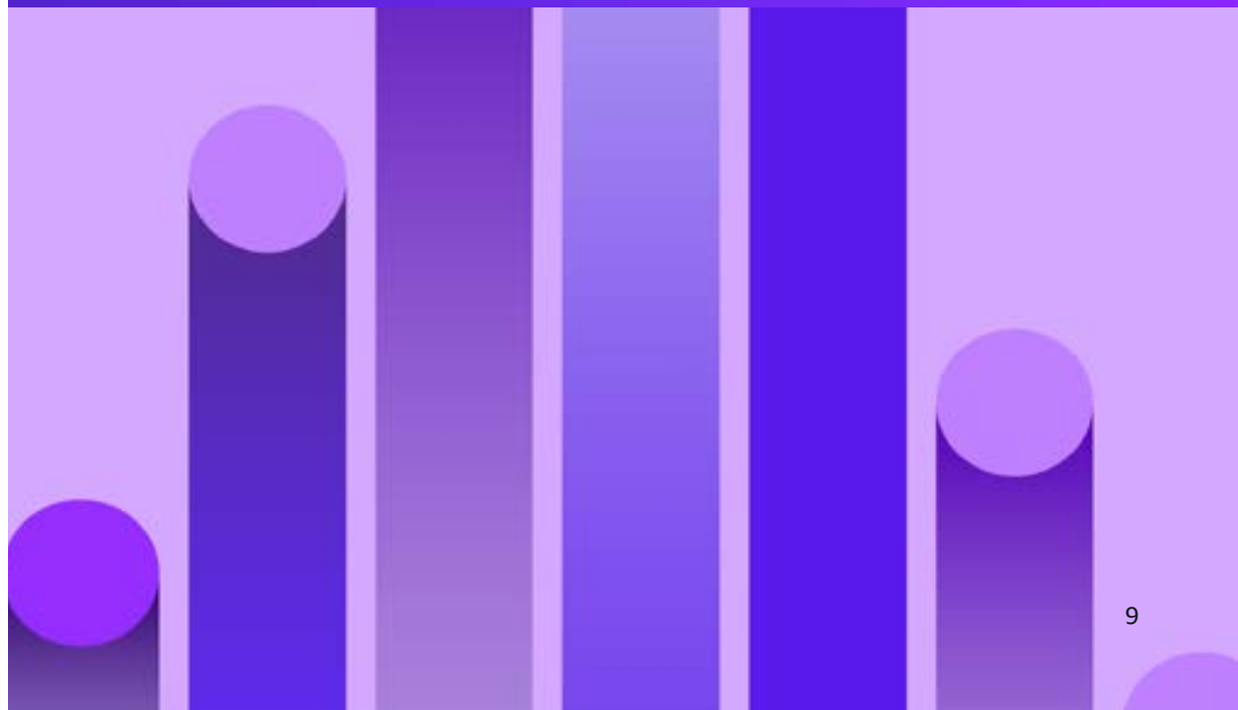
UNLOCK THE VALUE OF YOUR DATA

Make data your differentiator

Your data is the differentiator between a generic application and a gen AI application that knows your business and your customer. AWS makes it easy to use your data to customize and augment FMs to offer personalized experiences to your customers. AWS also offers a comprehensive range of data services to store, process, manage, and govern your data. The next generation of Amazon SageMaker addresses the challenges of harnessing all of organizational data—regardless of where it lives—for analytics and AI through unified data access and governance. It enables teams to securely find, prepare, and collaborate on data assets and build analytics and gen AI applications through a single platform, accelerating the path to AI-ready data.

Discover how a data foundation built on AWS gives you a strategic advantage when it comes to gen AI.

52% of CDOs and data leaders say they don't have the right data foundation to implement gen AI[4]



[4] "Scaling Generative AI for Value: Data Leader Agenda for 2025," AWS, 2024

EXECUTIVE INSIGHT

Future-proofing your generative AI foundation

With a rapidly evolving technology like gen AI, it's important to make informed decisions and build on a technical foundation that will allow for choice and flexibility.

As gen AI capabilities are adopted more widely, your ability to apply proprietary data on top of your FMs to differentiate your business will be a key competitive advantage. And, to scale your early wins into enterprise-wide adoption of gen AI, you should make sure that you are building with proper security, privacy, and operational resilience considerations from the start.

"This is not something you can just bolt-on later. Building on the right foundation gives you confidence and agility to create long-term value."

Ishit Vachhrajani, Global Head of Enterprise Strategy, AWS

How AWS can help your business build a foundation for generative AI success

For long-term success in gen AI, every organization needs access to a comprehensive set of tools and infrastructure to meet their unique needs now and in the future. We like to think of this set of tools as a three-layer stack.

Discover how leading businesses are using the most comprehensive set of capabilities at every layer of the AWS gen AI stack to drive innovation.

AWS generative AI stack

Top layer
Applications to boost productivity

Middle layer
Models and tools to build and scale secure, reliable, and responsible generative AI apps

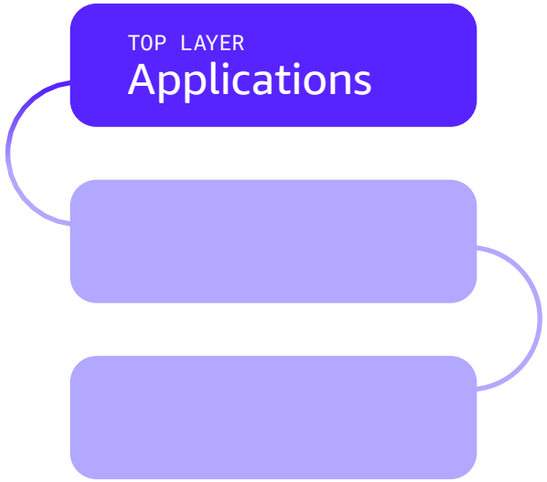
Bottom layer
Infrastructure to build and train AI models

With AWS, you can easily move throughout the stack to find the right tool or service for the job to be done. You don't have to feel stuck in the infrastructure or application layer and can quickly get started at the level that best meets you where you are in your gen AI journey. Many customers start at one layer and advance to using multiple—if not all—layers of the stack to scale gen AI use cases.

AWS GENERATIVE AI STACK

Top layer: Applications to boost productivity

To quickly take advantage of the productivity benefits of gen AI, you will need gen AI–powered applications and assistants that leverage LLMs and other FMs.



$260M

Amazon Q has saved Amazon $260M and 4.5K developer-years of work[5]

SPOTLIGHT:

Meet Amazon Q, the most capable gen AI assistant

Amazon Q helps accelerate software development and leverage your internal data. With built-in privacy and security, Amazon Q makes it easier for companies to use gen AI safely.

Amazon Q Business is a gen AI–powered assistant that lets users ask complex questions, find comprehensive answers, and execute actions—all based on your company's content, data, and systems. Within Amazon Q Business, employees can use Amazon Q Apps to quickly create and share lightweight AI apps instantly with natural language prompts—turning their ideas into apps based on your enterprise data.

Amazon Q Developer helps developers and IT professionals (IT pros) with all of their tasks across the software development lifecycle—from coding, testing, and upgrading to troubleshooting, performing security scanning and fixes, optimizing AWS resources, and creating data engineering pipelines.

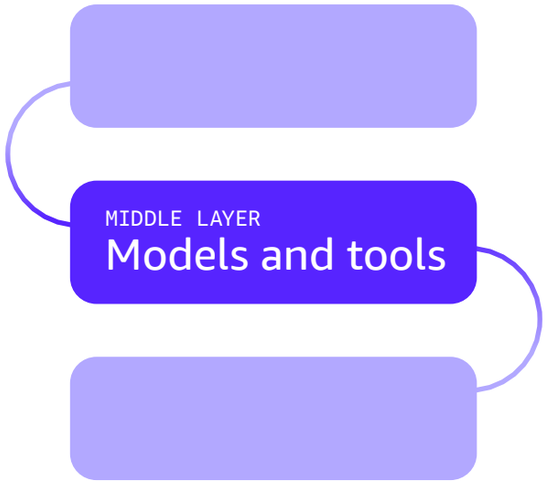
Amazon Q is also integrated into **Amazon QuickSight**, **Amazon Connect**, and **AWS Supply Chain**, bringing this capable assistant directly to your employees.

[5] "Amazon CEO Andy Jassy Says Company's AI Assistant Has Saved $260M And 4.5K Developer-Years Of Work: 'It's Been A Game Changer For Us'," Yahoo Finance, August 2024

AWS GENERATIVE AI STACK

Middle layer: Models and tools to build and scale secure, reliable, and responsible generative AI apps

For customers seeking access to the latest models, flexibility to select the appropriate model for their use case, and robust tools and security controls to swiftly build and scale trustworthy gen AI applications, we offer Amazon Bedrock. Amazon Bedrock provides a comprehensive solution designed to empower customers with cutting-edge technology, customizable options, and robust security measures.



Tens of thousands of customers from virtually every industry and of all sizes, like Intuit, Sony, PGA TOUR, Lonely Planet, Toyota, New York Stock Exchange (NYSE), KT Corporation, Siemens, Choice Hotels, Rocket Mortgage, and others, are using Amazon Bedrock to accelerate their gen AI initiatives.

SPOTLIGHT:

Build effective and responsible generative AI apps with Amazon Bedrock

Amazon Bedrock is a fully managed service that offers a choice of high-performing FMs from leading AI companies like AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, Stability AI, and Amazon through a single API, along with a broad set of capabilities you need to build gen AI applications with security, privacy, and responsible AI.

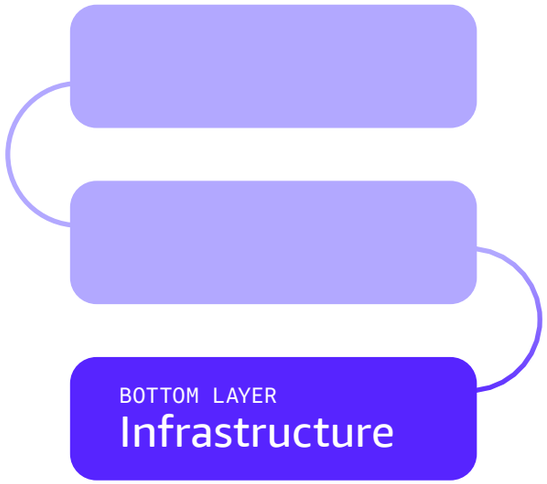
Amazon Nova is a new generation of foundation models with industry leading price performance

Generate high-quality images and videos, and lightning fast text. Exclusively available on Amazon Bedrock. Amazon Nova is up to 75% more cost-effective than the best performing models in their respective intelligence classes.

AWS GENERATIVE AI STACK

Bottom layer: Infrastructure to build and train AI models

To cost-effectively build, train, and deploy FMs at scale, you need purpose-built infrastructure and services.



50%

Enterprise customers like Salesforce and Workday are deploying their FMs using Amazon SageMaker AI, making more than 80 billion inference requests per day and reducing deployment costs by 50% on average.

SPOTLIGHT:

The most performant, cost-effective infrastructure for generative AI

From the highest-performance NVIDIA GPU–based Amazon Elastic Compute Cloud (Amazon EC2) instances to our purpose-built machine learning (ML) accelerators AWS Trainium and AWS Inferentia, AWS delivers the best price performance for training and deploying gen AI models at scale.

With Amazon SageMaker AI, data scientists and ML engineers can easily build, train, and deploy FMs at scale with purpose-built tools, including IDEs and notebooks backed by high-performance accelerated computing, purpose-built infrastructure for distributed training at scale, governance and MLOps tools, inference options and recommendations, and model monitoring and evaluation.

CUSTOMER SUCCESS STORIES

Customers across industries are enhancing customer experiences, boosting employee productivity, and optimizing business processes with generative AI on AWS.

FERRARI →

BAYER CROP SCIENCE →

AMAZON PHARMACY →

Ferrari

Ferrari redefines automotive luxury

Ferrari, a luxury Italian auto manufacturer, is using the broad model selection of Amazon Bedrock to apply gen AI to several use cases, from accelerating the vehicle design process to providing personalized services to its customers. Using LLMs in Amazon Bedrock, Ferrari developed a car configurator to make it easier and faster for customers to personalize their car, which increased sales leads and reduced vehicle configuration times by 20 percent. Ferrari also fine-tuned LLMs—including Amazon Titan, Claude 3, and Llama—on its own internal documentation to create a gen AI chatbot that helped their sales professionals and technicians enhance the after-sales experience.

"Amazon Bedrock has simplified our approach. We can connect to a single layer of APIs to quickly test, benchmark, and deploy different models."

Mauro Coletto, Head of Business Analytics and AI at Ferrari

Crop Science

Bayer Crop Science empowers data scientists to innovate faster

Bayer Crop Science uses Amazon Q Business and Amazon Q Developer to deliver “DSE Platform Intelligence,” which helps developers generate documentation and find code easily. By implementing Amazon Q Business, Bayer Crop Science reduced onboarding time by up to 70 percent, and Amazon Q Developer improved developer productivity by up to 30 percent.

UP TO
70%
reduced onboarding time

UP TO
30%
improved developer productivity

amazon pharmacy

Amazon Pharmacy improves customer care

Using Amazon SageMaker AI, Amazon Pharmacy created an LLM-based chatbot so that its customer care representatives could focus on improving the customer experience. The chatbot helps customers get more transparent pricing and receive their medications quickly while saving customer care agents significant amounts of time.

Amazon Pharmacy uses Amazon Bedrock and Amazon SageMaker AI to provide upfront pricing estimates for 99 percent of prescriptions and optimize operational efficiency, helping customers access needed medications faster.

"Using AWS, we can customize solutions for our industry while prioritizing security and privacy."

Alexandre Alves, Senior Principal Engineer, Amazon Pharmacy

Why customers are choosing AWS to succeed with generative AI

Choice

AWS empowers organizations with choice and flexibility, offering a comprehensive set of AI capabilities, from chips to FMs to applications, enabling rapid innovation. With access to the latest models, organizations stay at the leading edge with choice and flexibility.

Security

AWS is architected to be the most secure cloud computing environment available today for building and running gen AI applications, offering built-in security across AI services and robust security, compliance, governance, and privacy tools to help protect each layer of the AI stack.

Data

AWS helps you build trusted and differentiated gen AI experiences using your organizational data. With AWS, you can easily and privately leverage your data to customize FMs, tapping into vector capabilities in AWS databases, built-in governance controls, and customization techniques like retrieval augmented generation (RAG), fine-tuning, and pretraining.

Scale

AWS offers unparalleled scale and reliability for gen AI applications, leveraging decades of experience, vast global infrastructure, and seamless integration across diverse data sources and applications so organizations can deliver real, meaningful value to their customers and employees at scale.

AWS expertise

Take advantage of AWS experts and partners to innovate faster with gen AI.

AWS Generative AI Innovation Center

A program pairing customers with AWS science and strategy experts with comprehensive expertise spanning the entire gen AI journey. The AWS Generative AI Innovation Center helps customers build a strategic and holistic approach aligned with business goals, use cases, and operational challenges.

AWS Generative AI Competency Partners

AWS Generative AI Competency Partners drive the advancement of gen AI technologies with AWS customers to launch transformative applications across industries.

NEXT STEPS

Fuel generative AI breakthroughs with AWS

The future of business is being shaped by gen AI. AWS can help your business deliver tangible value and a competitive advantage, today and tomorrow. Join the thousands of organizations that are leveraging our secure, flexible, and scalable AI services to innovate boldly and continuously—all while maintaining the highest levels of trust, privacy, and security.

Start now →