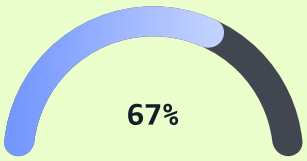
Build a foundation for success in the generative AI era

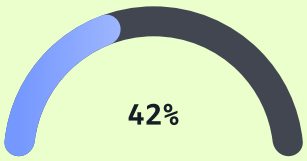
As a technology leader, you play a pivotal role in driving the successful adoption of generative artificial intelligence (gen AI). However, scaling beyond pilot projects presents numerous obstacles that can hinder your ability to fully leverage this transformative technology.

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.

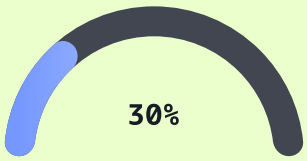
Discover how leading businesses are using the most comprehensive set of capabilities at every layer of the Amazon Web Services (AWS) gen AI stack to drive innovation.



of organizations surveyed said they are increasing investments in gen AI given the strong value seen to date[1]



of businesses said efficiency, productivity, and costs were top gen AI benefits[1]



At least 30% of gen AI projects will be abandoned after proof of concept by the end of 2025[2]

CLICK ANY LAYER TO LEARN MORE:

TOP LAYER
Applications to boost productivity

MIDDLE LAYER
Models and tools to build gen AI applications

BOTTOM LAYER
Infrastructure to build and train AI models

[1] "Now decides next: Moving from potential to performance," Deloitte, August 2024
[2] "Gartner Predicts 30% of Generative AI Projects Will Be Abandoned After Proof of Concept By End of 2025," Gartner, July 2024

TOP LAYER
Applications to boost productivity

Leaders are investing in gen AI in the workplace to:

- Boost workplace efficiency
- Accelerate software development

Employees see an average of **30% in productivity gains** with gen AI[3]

MIDDLE LAYER
Models and tools to build gen AI applications

BOTTOM LAYER
Infrastructure to build and train AI models

AMAZON Q

Transform how work gets done with generative AI–powered assistants

Amazon Q is the most capable gen AI–powered assistant for accelerating software development and leveraging your company's internal data. With built-in privacy and security, Amazon Q makes gen AI securely accessible to everyone in your organization.

- Amazon Q Business helps employees find information, gain insight, and take action faster
- Amazon Q Developer helps developers and IT pros with all of their tasks across the software development lifecycle

$260M

With Amazon Q Developer, Amazon has saved $260M and 4.5K developer-years of work[4]

[3] "Four GenAI Use Cases for the Digital Workplace," Gartner, October 2023
[4] "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

TOP LAYER
Applications to boost productivity

MIDDLE LAYER
Models and tools to build gen AI applications

Businesses are building gen AI apps to deliver:

- Better customer experiences
- Operational efficiency
  Revenue growth

>80% of enterprises will deploy gen AI apps by 2026[5]

BOTTOM LAYER
Infrastructure to build and train AI models

AMAZON BEDROCK

Build and scale generative AI apps with robust models and tools

Amazon Bedrock is a fully managed service that offers a choice of high-performing foundation models (FMs) from leading AI companies through a single API, along with a broad set of capabilities you need to build gen AI applications with security, privacy, and responsible AI.

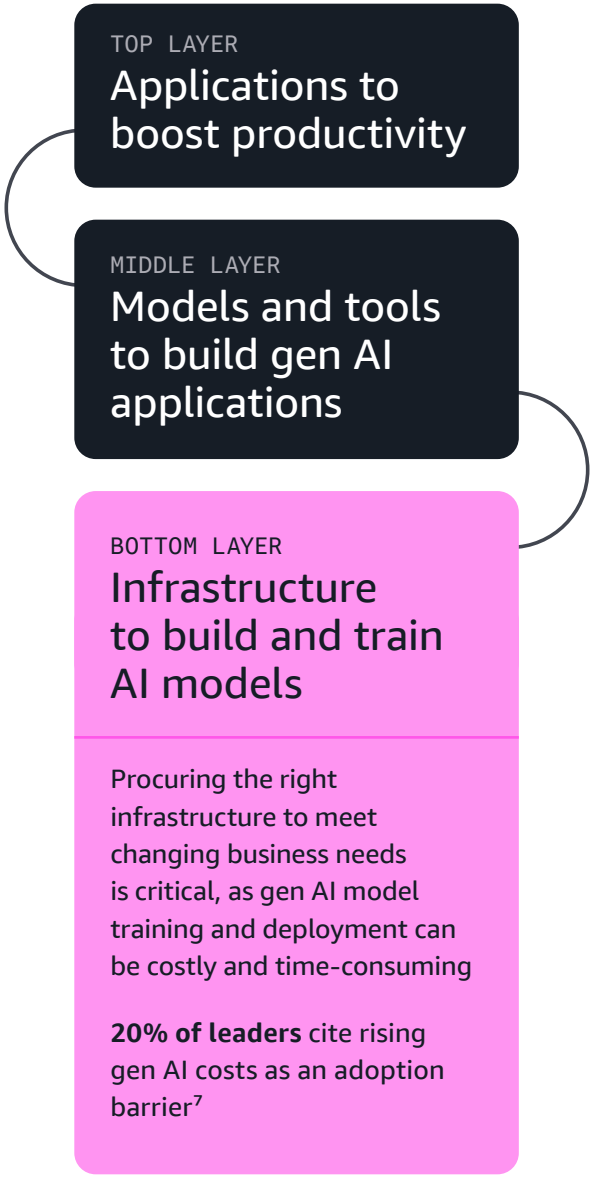
Amazon Nova is a new generation of FMs with industry-leading price performance. Generate high-quality images and videos and lightning-fast text. Exclusively available on Amazon Bedrock.

75%

Amazon Nova is up to 75% more cost-effective than the best-performing models in their respective intelligence classes[6]

[5] "Gartner Says More Than 80% of Enterprises Will Have Used Generative AI APIs or Deployed Generative AI-Enabled Applications by 2026," Gartner, October 2023
[6] "Introducing Amazon Nova: A New Generation of Foundation Models," Amazon Press Center, December 2024

aws



AWS INFRASTRUCTURE

Most performant, cost-effective infrastructure for generative AI

From the highest performance NVIDIA GPU–based Amazon Elastic Compute Cloud (Amazon EC2) instances to continued investments in 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 the fully managed infrastructure, tools, and workflows of Amazon SageMaker AI, you can build, train, and deploy FMs at scale.

50%

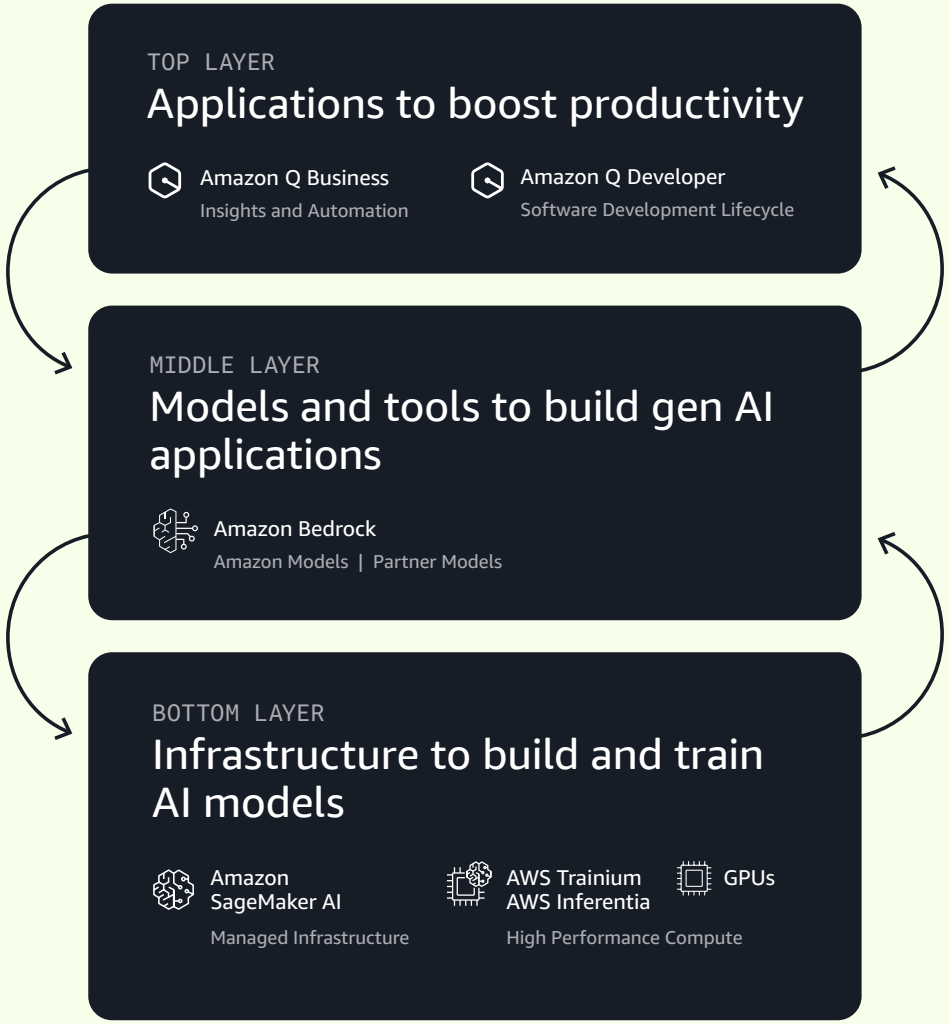
Enterprise customers like Salesforce and Workday are deploying their FMs using Amazon SageMaker AI, making on average more than 80B inference requests per day and reducing deployment costs by 50% on average

[7] "Now decides next: Generating a new future," Deloitte, January 2025

aws

Fuel generative AI breakthroughs with AWS

Drive innovation with the most comprehensive gen AI capabilities on AWS. Find the right tool or service with solutions that span the full stack and easily advance as your requirements evolve.



Unlock agility with AWS

Implement gen AI your way. With AWS, you get the flexibility to:

- **Access tools quickly:** Move throughout the stack to find the right tools or services for the job to be done
- **Start anywhere, move freely:** Begin your gen AI journey at any layer that suits your current needs
- **Scale on demand:** Expand across the stack as your initiatives grow

Get started now

RETURN TO FULL STACK

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.