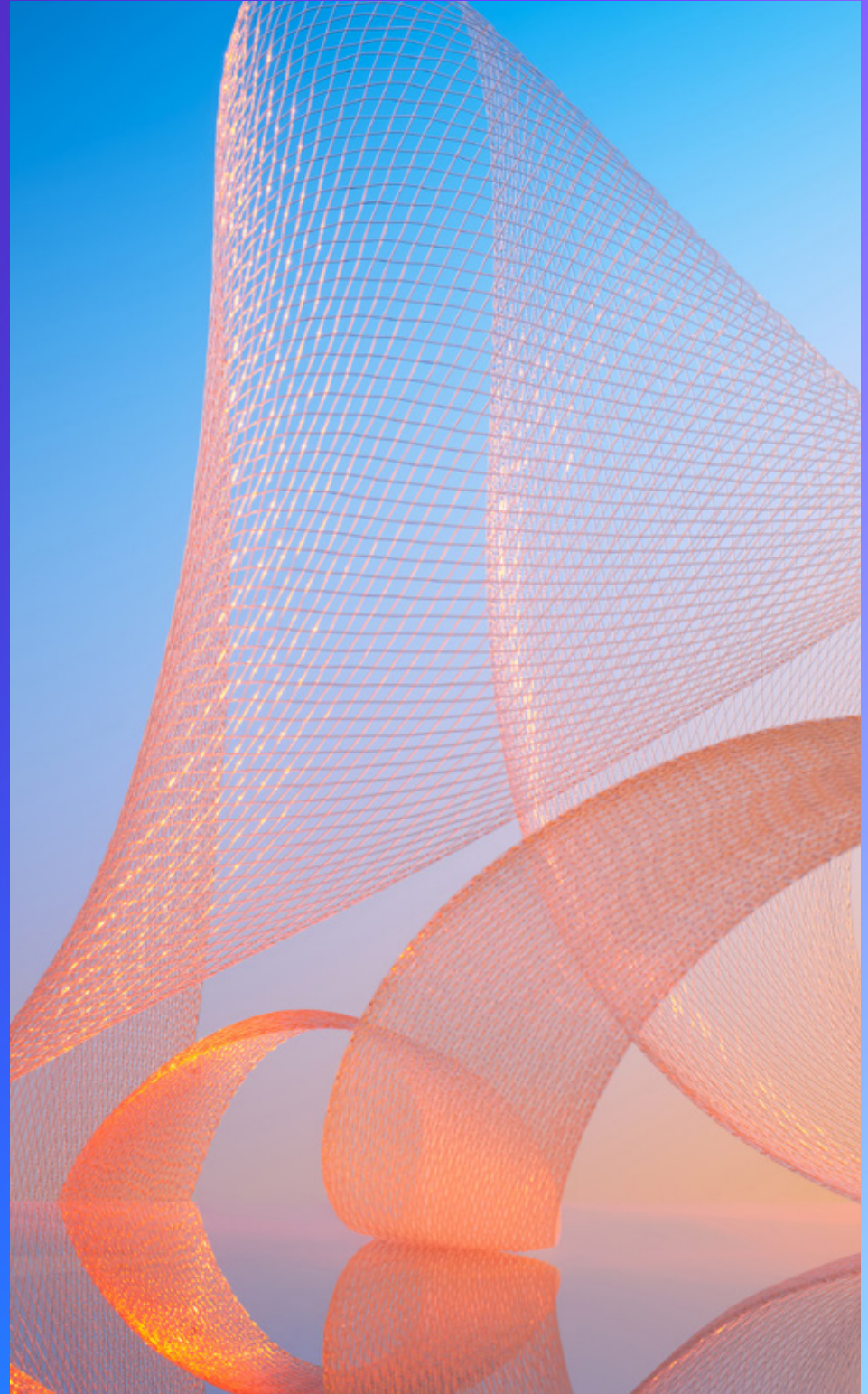




BUSINESS INNOVATION BRIEF

Maximize the business value of generative AI

Harness generative AI's full potential



Executive summary

A surge of excitement across consumers and businesses—combined with broad, easy access to the technology—has led us to an inflection point for generative AI. Today, all organizations should evaluate the impact of generative AI and how they can leverage its transformative value.

Most organizations recognize the importance of this moment and the need to quickly develop a strategy to explore generative AI. But for many, questions remain, including:

- What is generative AI?
- How is generative AI different from previous generations of artificial intelligence (AI)?
- What are the main business use cases?
- Which customers are already using it?
- How should I start?
- What are the risks, and how can they be mitigated?

This business innovation brief provides an overview of generative AI, outlining its capabilities, use cases, and business value. It also offers valuable insights from Amazon Web Services (AWS) subject matter experts, leveraging our extensive knowledge and experience in AI and machine learning (ML) technologies.



Who is this content for?

This innovation brief is designed for business leaders seeking to better understand generative AI—and learn how they can leverage it to improve business outcomes.



Table of contents

Introduction: A new world of intelligence.....	4
Understanding generative AI.....	7
Business capabilities of generative AI.....	9
Business considerations for generative AI	10
Executive insights on generative AI	12
Common generative AI use cases by industry	13
How AWS can help you succeed with generative AI.....	16
Next steps	19



INTRODUCTION

A new world of intelligence

Consumers and businesses alike are fascinated by generative AI's ability to create sophisticated content, generate code, answer questions, and more—all from simple natural language prompts, often within seconds.

While a lot of attention has been given to how consumers are using generative AI, there is an even bigger opportunity in how businesses will use it to deliver amazing experiences for their customers and employees. The true power of generative AI goes beyond a search engine or a chatbot and will transform every

aspect of how companies and organizations operate.¹ Gartner estimates that by 2025, 30 percent of outbound marketing messages from large organizations will be synthetically generated.²

Awareness of the technology is spreading at an unprecedented pace. According to Fishbowl, which ran a survey among nearly 4,500 professionals in large organizations in the US, 27 percent of professionals have already used generative AI to assist with work-related tasks.¹

Seizing the opportunity

Organizations across industries are racing to seize the economic opportunities that generative AI presents. If leading financial projections prove accurate, the rise of generative AI is likely to usher in a new era of the global economy.

According to research by Goldman Sachs, generative AI could increase global GDP by as much as 7%, or roughly **\$7 trillion**, over the next 10 years.³

These lofty financial projections are not driven by consumer interest alone. The potential of generative AI to improve business productivity and output accounts for just as much, if not more, of the excitement and enthusiasm surrounding the technology.

For businesses of every size and across every industry, generative AI is a revolutionary technology that is beginning to drive considerable value—and has the power to fundamentally transform the business landscape.



³ "Generative AI could raise global GDP by 7%," Goldman Sachs, April 2023

⁴ "Generative AI Market," Polaris Market Research, January 2023

Global generative AI market

Market forecast to grow at a CAGR of 34.2%⁴





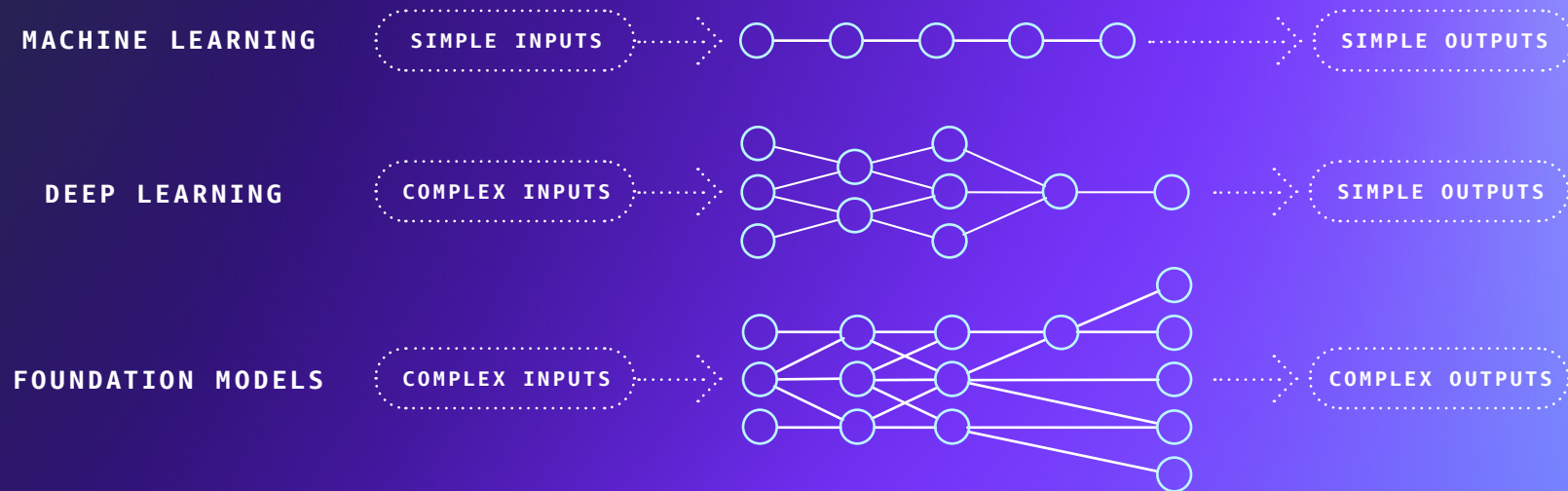
The strategic imperative

Organizations across industries and around the world are looking to leverage generative AI to reimagine customer experiences, boost employee productivity, ignite creativity, and optimize business processes.

For most businesses, however, [the path](#) to achieving these benefits remains unclear.

Many acknowledge the need to invest in generative AI—and to do it quickly, lest their competitors gain the advantage. However, few have succeeded in developing strategies for how they will adopt the technology, where they will put it to use, or how they will achieve and measure their results.

Read on to learn how your organization can start realizing the business value of generative AI today—so you can keep pace with the market and leapfrog your competition.



Understanding generative AI

Before your organization can fully unlock the business value of generative AI, it's important to have a fundamental understanding of how the technology works.

Generative AI is a term used to describe algorithms that can create new content and ideas, including conversations, stories, images, videos, and music. Generative AI is powered by extremely large ML models that are pretrained on vast amounts of data. These are commonly known as **foundation models (FMs)**.

Traditional forms of ML allowed us to take simple inputs, like numeric values, and map them to simple outputs, like predicted values. With the advent of deep learning, we can take complicated inputs, like videos or images, and

map them to relatively simple outputs, for example, if the image contains a cat or not. With generative AI, we can leverage massive amounts of complex data to capture and present knowledge in more advanced ways—mapping complicated inputs to complicated outputs, like summarizing a long document and extracting the key insights.

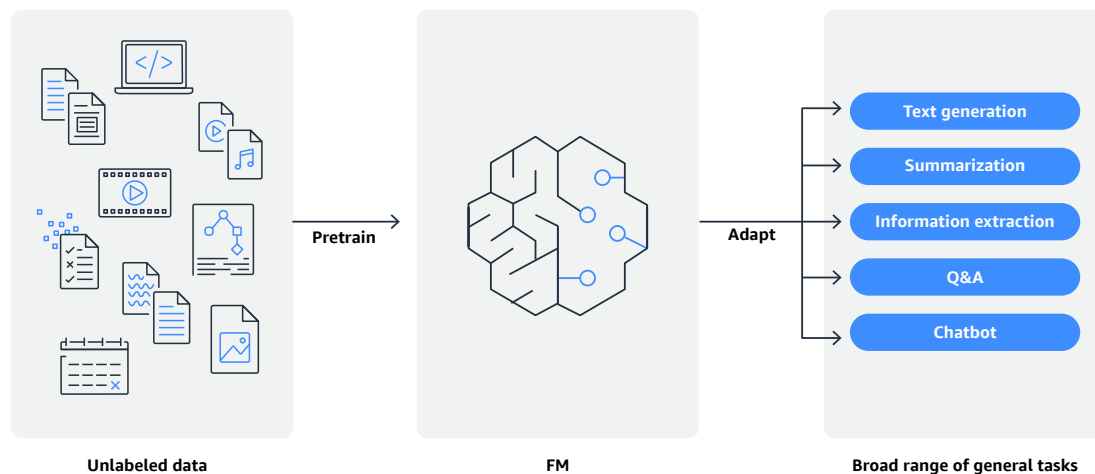
Text-based generative AI systems use a specific type of FM called a **large language model (LLM)**. LLMs can perform a wide range of tasks that span multiple domains, like writing code, solving math problems, engaging in dialogue, and analyzing the information in documents to answer questions.

Take foundation models further with customization

FMs can be customized with your organization's proprietary data to deliver a more accurate output compared to an "out-of-the-box FM"—and allow you to create new proprietary products and services. For example, a large grocery chain that tracks shopper preferences can customize an FM to produce a better recommendation engine that is highly differentiated from competitors' offerings.

Organizations can also use customized FMs to easily create unique content that embodies their brand's tone and style. For instance, a financial firm that needs to auto-generate a daily activity report for internal circulation can customize an FM with proprietary data, including past reports. The FM could then learn how these reports should read and what data was used to generate them.

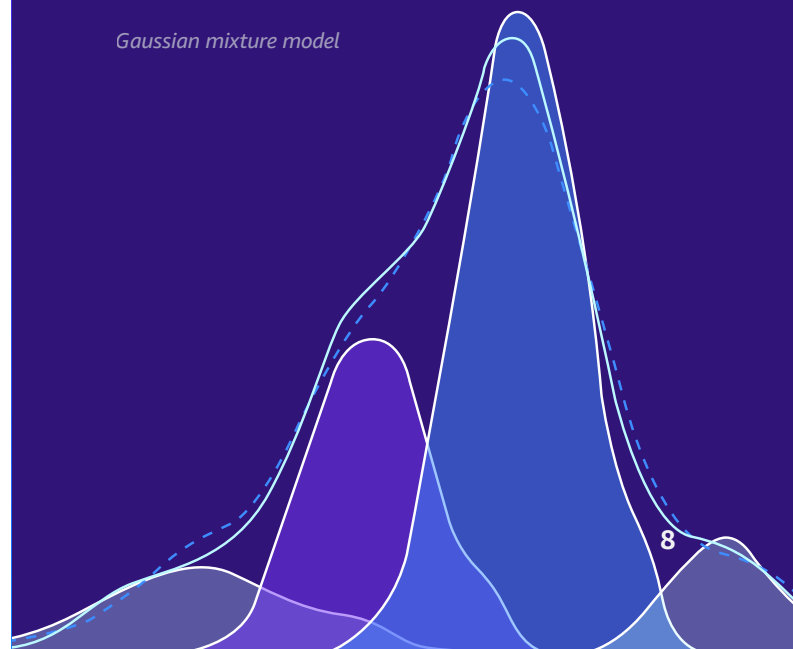
Now that you have a general understanding of how the technology works, let's begin exploring how you can put generative AI to work for your organization.



Moments in generative AI history:

Today's FMs that are used to create generative AI applications are built atop a long history of AI innovation. Two of the earliest models with generative AI capabilities are the hidden Markov model (HMM) and the Gaussian mixture model (GMM), both developed in the 1950s. HMMs use known data to make educated guesses about unknown data (for example, predicting whether a card player is cheating based on their results). GMMs can examine a group of data (such as a music playlist) and subgroups within that data (for example, genres) to infer unknown information (such as "this is a rap song"). Both are still used today.

Gaussian mixture model



Business capabilities of generative AI

Organizations in many industries are using generative AI to enhance productivity and create business value in many ways, including:



Code generation

Improve developer productivity by 57% with AI coding companion Amazon CodeWhisperer⁵



Virtual assistants

Enhance customer experience with human-like responses



Contact center analytics

Summarize and extract insights from customer calls



Conversational search

Extract insights from all your corporate information



Personalization

Improve personalized recommendations and generate tailored content



Content generation

Create text, images, videos, and music



Design and creativity

Get suggestions, generate prototypes, and explore innovative concepts

Moments in generative AI history:

Another early example of generative AI is ELIZA, a chatbot (or “chatterbot,” as they were previously known) developed by an MIT professor from 1964 to 1966. Like its namesake, Eliza Doolittle of *Pygmalion* and *My Fair Lady*, the program grew more sophisticated by “learning” from human interactions. ELIZA was most famously used to mimic the behavior of a therapist conducting an initial psychiatric interview, with the user playing the role of the patient.

Business considerations for generative AI

As you work to identify the capabilities of generative AI that are most useful to your organization—and develop a strategy for implementing them into your business processes—you will need to determine which FMs to use in creation of generative AI applications.

You should also carefully consider the infrastructure you will be using to support your FMs. Your models will benefit from a cost-efficient infrastructure that meets your requirements for performance.

When evaluating FMs used to create generative AI applications, look for models that offer:

- Easy ways to build generative AI applications
- Cost-effective infrastructure
- Generative AI-powered applications
- Flexible ML tools



Responsible AI, security, and privacy

With their vast size and open-ended nature, FMs raise new issues in defining, measuring, and mitigating responsible AI concerns across the development cycle, such as accuracy, fairness, intellectual property (IP) considerations, hallucinations, toxicity, and privacy. For example, looking at the issue of fairness, can we ask an LLM to assign male and female pronouns at the same rate in reference to a doctor? Does that still apply if the prompt describes the doctor as having a beard? And should we do the same for other professions? What about the Women's National Basketball Association (WNBA)? You can see that simply defining fairness in the context of an LLM is challenging and requires new approaches and solutions.

Generative AI technology and how it is used will continue to evolve, posing new challenges that will require additional attention and mitigation. To tackle these challenges and foster innovation, academic, industry, and government partners are working together to explore new solutions and concepts to ensure that generative AI continues to evolve in a responsible, private, and secure way.

Data privacy and security are also critical to scaling generative AI responsibly. When it comes time to customize and fine-tune a model, organizations need to know where and how their data is being used. They need to be confident their private data is not being used to train a public model and that customer data remains private. Organizations need security, scalability, and privacy to be baked in from the start to be viable for their business applications.

Read the blog post *Responsible AI in the Generative Era*

[Learn more ›](#)



Moments in generative AI history:

In 2014, the development of the first generative adversarial network (GAN) marked one of the biggest breakthroughs in generative AI. In a GAN, two models (a “generator” and a “discriminator”) compete in a zero-sum game. The generator manufactures content that appears increasingly “genuine,” while the discriminator analyzes its opponent’s techniques to better identify fakes. This novel approach of using AI to train other AI proved revelatory, while GANs themselves unlocked a new era for digital imagery.

In the following section, we feature an AWS industry leader who shares his experiences and strategies for cloud best practices, cultural change, organizational agility, and transformation with generative AI.

Executive insights on generative AI

Like any new technology that becomes mainstream, generative AI has a learning curve. Executives are asking about where generative AI fits or doesn't, how to use it effectively, and about the non-technical considerations.

History teaches us that we see profound, positive change only when people, processes, skills, and culture are addressed alongside technology. Based on these learnings and from the thousands of customers we talk to, my advice for those intrigued by generative AI is simple.

First, be curious. Learn what generative AI is, why it has captured people's imaginations, and what problems it can solve. Dive deep into areas such as the security of your data when customizing models. And encourage others to learn these things, too, rather than delegating them to your IT team.

Second, think big and work backwards from the customer. This is a standard way of thinking here at AWS! Really understand the opportunities in your business—whether it's an opportunity to improve supply chain efficiencies, invent new services, or enhance customer service. Fall in love with the problem before you jump to a solution, looking for areas that reduce costs, increase resilience, or improve your top line. And think big about the opportunity; small thinking becomes a self-fulfilling prophecy.

Finally, start now. Most business initiatives take time to get traction, so start experimenting quickly. You will learn more from this than the endless planning and waiting for the hypothetical perfect time that is typical of many business technology adoptions.



Phil Le-Brun
Director,
AWS Enterprise
Strategy

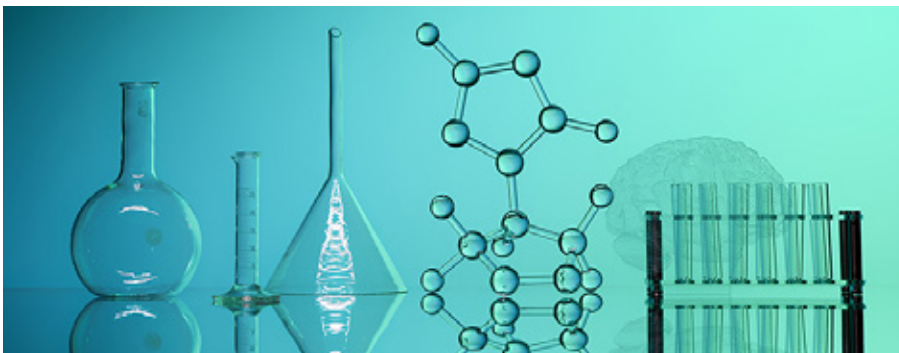
Joining AWS in 2019, Phil's experience implementing technology at scale—including a 25-year career spearheading digital transformation efforts at McDonald's Corporation—has allowed him to learn a variety of practical lessons. He shares this knowledge with enterprises to help them achieve their cloud-based technology goals, such as supporting organizational agility and increasing customer centricity.

Common generative AI use cases by industry

In previous sections, we've explored the capabilities of generative AI that can benefit all businesses and how to select the right models and infrastructure for your business' needs. Now, let's dive a bit deeper and examine generative AI use cases that are unique to specific industries.

Healthcare and Life Sciences

- **Accelerating innovations in pharmaceutical R&D:** Life sciences organizations can expedite the development of new, more effective drug candidates with greater efficiency. For example, researchers can predict the structure of proteins to better identify biological targets, generate novel amino acid sequences, and identify docking sites for specific targets to create precision therapeutics. Ultimately, this will result in more targeted and effective therapeutic options for patients.
- **Improving clinical engagement and patient experience:** Generative AI can be used to improve patient navigation, care coordination, and the productivity of healthcare providers. Through conversational interactions, clinicians can access data from electronic health records, research publications, and medical policies more easily. Generative AI can also accelerate digital transformation and personalized medicine by facilitating intelligent document processing (IDP), providing personalized guidance on medical and pharmacological practices, and helping to design effective clinical trial protocols to ensure safety and efficacy for new drug development.
- **Summarizing health and scientific data:** Researchers and clinicians can leverage generative AI to reduce time to insights and quickly find the data they need for more data-driven decision making. In healthcare, search and summarization can be used to enhance efficiencies by, for example, auto-generating chart note components and note summaries, simplifying workflows, and automating administrative tasks. The applications in life sciences are vast, including the identification of patterns in clinical trials, the ability to summarize findings from hundreds of thousands of research papers, and the identification of efficiencies across the value chain.



USE CASES BY INDUSTRY



Financial Services

- **Improving experiences:** Financial services firms can better serve customers and employees by deploying chatbots that resolve problems faster and personalizing products and recommendations, like leveraging a repository of customer interactions with summarization and queries; AI also enables the creation of knowledge articles from disparate internal sources.
- **Increasing knowledge-worker efficiency:** Knowledge workers at financial services firms can quickly draft investment research, loan documentation, insurance policies, regulatory communications, RFIs, and business correspondence; achieve deeper insights into customer behavior; improve collaboration; and increase the business value of unstructured content.
- **Analyzing market sentiment:** Through faster and more thorough analysis of social media, news articles, and financial data, financial services firms can surface market commentary, identify opportunities sooner, and proactively mitigate risks.
- **Building new products and automating business processes:** Financial services firms can create on-demand structured data products from large unstructured data sources, and generative AI can enhance productivity tools for end users and improve developer productivity with automatic code generation.



Automotive and Manufacturing

- **Improving product design:** Manufacturers can use AI to optimize the design of mechanical parts—or create entirely new material, chip, and part designs—improving quality and durability, lowering costs, and simplifying production.
- **Creating new in-vehicle experiences:** Virtual assistants and personalized route recommendations can enhance experiences for drivers and passengers.
- **Improving testing and maintenance:** AI can improve product testing by generating information missing from datasheets—and unlock new assisted maintenance use cases to better maintain and service machinery, including products in use by consumers.
- **Enhancing overall equipment effectiveness for factories:** Digitize and capture historical machine maintenance data, repair data, equipment manuals, production data, and potentially data from other manufacturers to generate suggestions for maintenance, repairs, or equipment parameters—and improve productivity, availability, and quality.

USE CASES BY INDUSTRY



Education

- **Summarizing text:** Students and teachers can create concise summaries of research documents, lecture transcripts, and class notes to make them easier to search and browse.
- **Improving automation:** AI can transform information into sample test questions, accelerate grading, measure student performance across a wide range of factors, and provide personalized feedback and recommendations to teachers and students.
- **Personalizing learning environments:** Educators can create personalized learning pathways for student segments—or even individual students—and leverage simulations and virtual reality to make learning more engaging.



Media and Entertainment

- **Expediting content creation:** From concept design and storyboarding to post-production workflows, media and entertainment (M&E) companies can automate lower-level tasks to increase production speed and allow creative talent to iterate faster and realize the director's vision.
- **Enhancing music:** Artists can complement and enhance their albums with AI-generated music to create new genres.
- **Improving the media supply chain:** Generative AI applications can aid or automate tasks like localization, content moderation, artwork creation for promotional materials, and even content restoration.

Moments in generative AI history:

The 2017 introduction of a new type of deep learning model, the transformer, set the stage for modern generative AI. Unlike older models, which break down input data, process it, then put the pieces back together, transformers process the entire input all at once. This makes them ideal for natural

language processing (NLP), where understanding the full context of the input is critical. In 2018, OpenAI took the technology further by creating the first Generative Pre-trained Transformer (GPT). From there, OpenAI developed its GPT-2 engine in 2019—which it then used to power ChatGPT, introduced in late 2022.

How AWS can help you succeed with generative AI

You can unlock the full business value of generative AI for your organization with AWS. Reinvent your applications, create entirely new customer experiences, drive unprecedented levels of productivity, and ultimately transform your business.

Experience and expertise

One of the key advantages of AWS lies in a rich AI heritage built over two decades of focused investment. In fact, more than 100,000 customers currently use AWS for AI and ML.

Amazon, the driving force behind AWS, harnesses ML capabilities to power its ecommerce recommendations engine, optimize robotic picking routes in fulfillment centers, and much more. Further, ML informs Amazon's supply chain, forecasting, and capacity planning.

Deep learning is also employed in the Amazon Prime Air drone delivery system and the computer vision (CV) technology behind Amazon Go, the innovative retail experience that allows customers to select items and leave the store without traditional checkouts. And Alexa, which is supported by more than 30 different ML systems, helps customers with a wide array of tasks billions of times each week.

With thousands of dedicated ML engineers, AI and ML are deeply ingrained in the heritage of Amazon and AWS—continuing to shape the future.

More than
100,000
customers currently
use AWS for AI and ML



Why build with AWS?

Organizations of all shapes and sizes choose to build generative AI and other AI and ML applications on AWS for many reasons. Here are some of the top advantages of building on AWS, according to our customers:

The easiest place to quickly build, train, and deploy FMs into your applications and workloads

Amazon Bedrock is the easiest way for customers to build and scale generative AI-based applications using FMs. Bedrock makes **Amazon Titan** FMs and models from leading AI startups including AI21 Labs, Anthropic, Cohere, and Stability AI accessible via an API. Customers using Bedrock can leverage the benefits of AWS, which is architected to be the most flexible and secure cloud computing environment available today. **Agents for Amazon Bedrock** is a fully managed capability that makes it easier for developers to create generative AI applications that can deliver up-to-date answers based on proprietary knowledge sources and complete tasks for a wide range of use cases.

The most cost-effective infrastructure for generative AI

For years, AWS has invested in developing silicon that delivers the highest levels of performance and cost optimization for AI and ML workloads. The results—**AWS Trainium** and **AWS Inferentia**—deliver the lowest costs for training models and running inference in the cloud. AWS has also developed **Amazon Elastic Compute Cloud** (Amazon EC2) instances to help you take advantage of these capabilities. For example, **Amazon EC2 Trn1** instances powered by

Trainium save you up to 50 percent on training costs,⁶ while **Amazon EC2 Inf2** instances powered by AWS Inferentia2 deliver up to 40 percent lower cost per inference.⁷

Greater flexibility

AWS offers you the flexibility to work with open-source models or build your own FMs. For example, in addition to models offered through Bedrock, **Amazon SageMaker JumpStart** allows you to discover, explore, and deploy open-source FMs—such as Llama 2, OpenLLaMA, RedPajama, Mosaic MPT-7B, FLAN-T5/UL2, GPT-J-6B/NeoX-20B, BLOOM/BLOOMZ, and more. You can also use **Amazon SageMaker** to create your own models with managed infrastructure, tools, and efficient distributed training.

Faster time to value with generative AI-powered applications

With generative AI built in, services such as CodeWhisperer, an AI coding companion, can help you improve productivity. In addition, you can deploy common generative AI use cases such as call summarization and question answering using AWS sample solutions that combine AWS AI services with leading FMs.

Further reading on responsible AI:

[AWS responsible AI resource hub ›](#)

[eBook: Democratized, Operationalized, Trusted: The 3 Keys to Successful AI Outcomes ›](#)



⁶ Over other comparable Amazon EC2 instances

⁷ Compared to prior generation AWS Inferentia-based instances

AWS generative AI services

Facilitate your generative AI applications with a range of AWS technologies, including:



Amazon Bedrock ›

Build and scale generative AI applications with FMs. Bedrock supports a variety of FMs, including:

- **Amazon Titan:** For text summarization, generation, classification, open-ended Q&A, information extraction, embeddings, and search
- **AI21 Labs Jurassic-2 Multilingual LLMs:** For text generation in various languages
- **Anthropic Claude 2:** LLM for thoughtful dialogue, content creation, complex reasoning, creativity, and coding based on Constitutional AI and harmlessness training
- **Stability AI Stable Diffusion:** Generates unique, realistic, high-quality images, art, logos, and designs
- **Cohere Command + Embed:** Text generation model for business applications and embeddings model for search, clustering, or classification in over 100 languages

AWS Trainium: Train models faster with up to 50% cost savings⁸ using this ML model accelerator⁹

AWS Inferentia2: Run high-performance FM inference with up to 40% lower cost per inference using this accelerator

Amazon CodeWhisperer: Enjoy 57% faster application development¹⁰ while helping to ensure security with this AI coding companion, which is at no cost for individual use

Amazon SageMaker: Build your own FMs with managed infrastructure and tools to accelerate scalable, reliable, and secure model building, training, and deployment

Amazon SageMaker JumpStart: ML hub that provides access to algorithms, models, and ML solutions so you can quickly get started with ML. With SageMaker JumpStart, ML practitioners can choose from a broad selection of **publicly available FMs**. ML practitioners can deploy FMs to dedicated SageMaker instances from a network-isolated environment and customize models using SageMaker for model training and deployment.

Next steps

Now that you have a better understanding of generative AI, what it can do, and its potential business benefits, the next step is to clearly define your objectives and identify use cases for leveraging it. It's best to start with smaller experiments and simple, precise goals. Once you've achieved some quick wins, you can begin scaling your efforts upward and outward.

Collaboration with experts is highly recommended to ensure you consider factors such as data availability, data quality, and ethical implications related to generative AI. Furthermore, infrastructure considerations should not be an afterthought, as they can significantly impact costs, scalability, and energy consumption. Engaging with AWS experts can provide valuable guidance throughout the decision making process and stages of implementation.

The time is now

The dramatic rise of generative AI brings us to a tipping point. FMs grow more sophisticated and powerful every day. For organizations, it's the power to transform business by creating entirely new customer experiences and driving unprecedented levels of efficiency and productivity.

All of which leads to an indisputable fact: To compete in this new era of profound technological advancement, every organization needs to consider making generative AI a part of its innovation road map.

With the most cost-effective cloud infrastructure for generative AI; a host of AI products, services, and solutions; and years of trusted AI expertise, AWS can help turn the promise of generative AI into results for your organization.



Partner with AWS to accelerate your generative AI journey today.

[Get started ›](#)