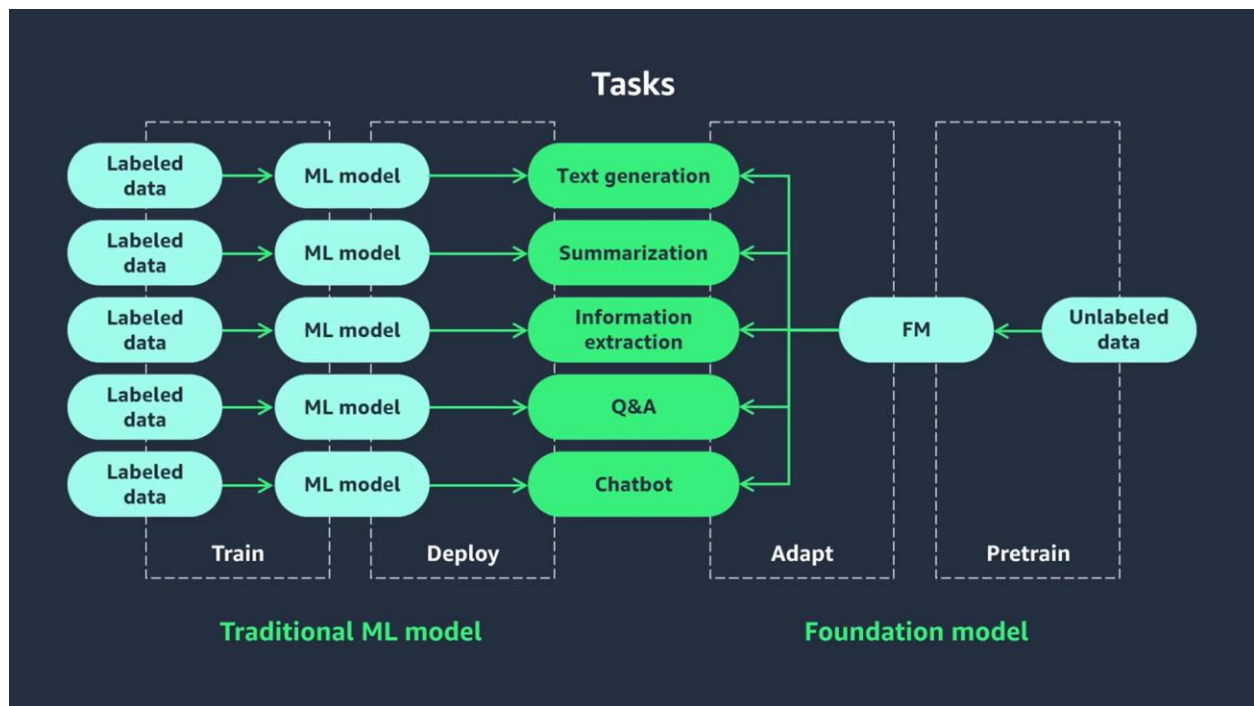
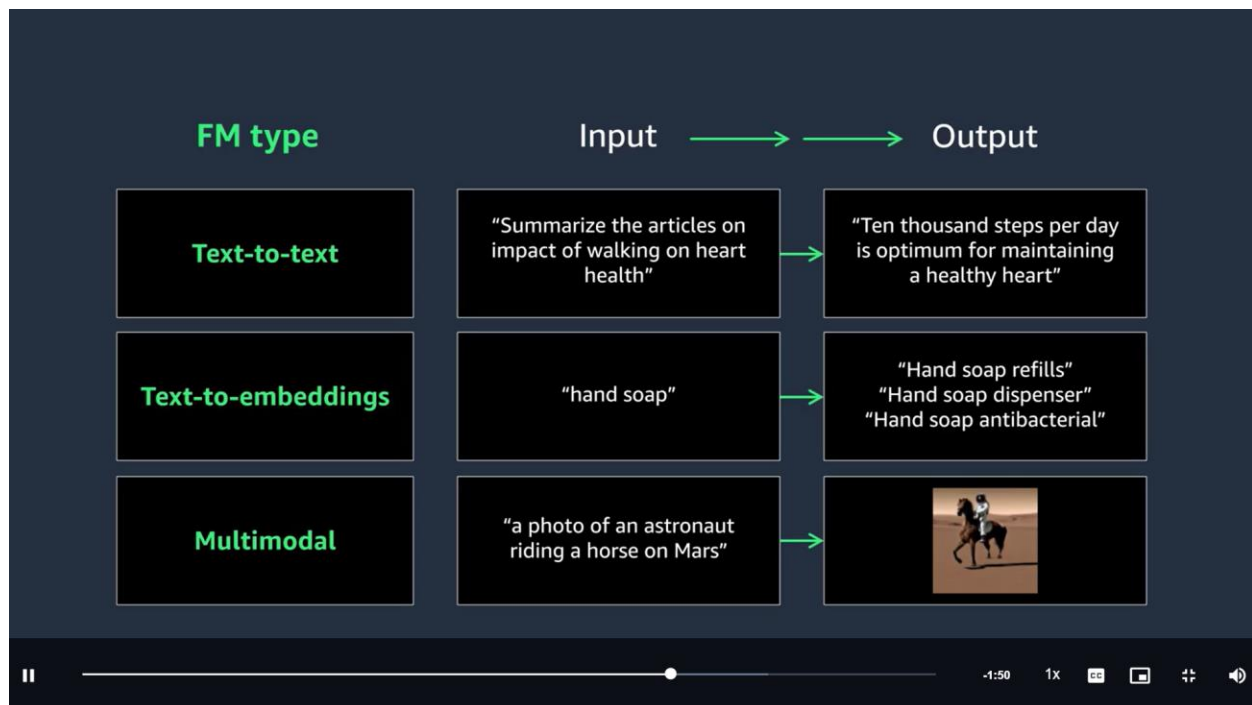


Generative AI on AWS Essentials (Business)

- Define generative AI.
- Explain how generative AI works.
- Describe the benefits of using AWS for generative AI.
- Describe key AWS generative AI services.
- Qualify potential generative AI solutions to support customers' business outcomes.
- Identify potential customer use cases for AWS generative AI.

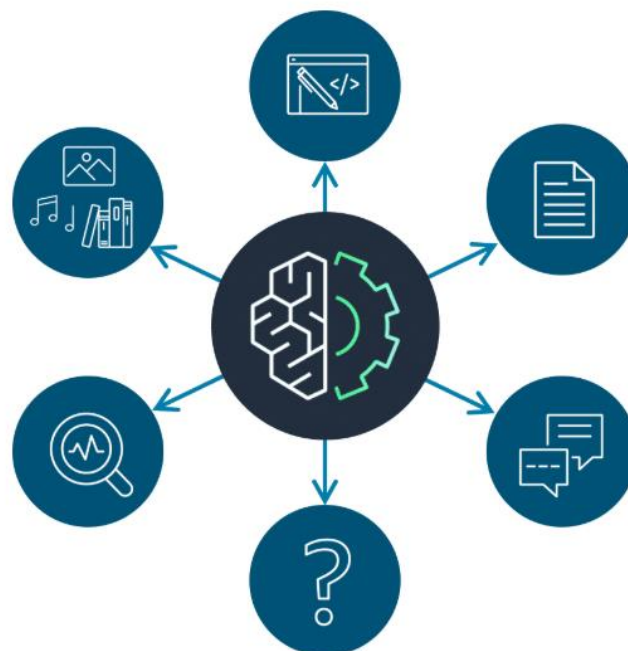
Introduction to Generative AI - Lesson 2 of 9





What is generative AI?

Generative AI is a type of artificial intelligence that can create new content and ideas, including conversations, stories, images, videos, and music.



What are foundation models?

Generative AI is powered by *very large* machine learning models that are pretrained on *vast* collections of data and are commonly referred to as foundation models (FMs).

What types of foundation models are currently in the market?

There are currently three main foundation model types in the market.

- **Text-to-text:** These natural language processing models can summarize text, extract information, respond to questions, and create content such as blogs or product descriptions. An example is sentence auto-completion.
- **Text-to-embeddings:** These FMs compare user search bar input with index data and connect the dots between the two. The result is more accurate and relevant results.
- **Multimodal:** These emerging foundation models can generate images based on a user's natural language text input.

How experienced is Amazon with ML innovation?

AI and ML have been a focus for Amazon for over 20 years. We have played a key role in democratizing ML and making it accessible to anyone who wants to use it, including more than 100,000 customers in many industries. Many of the capabilities that customers use with Amazon are driven by machine learning. In fact, Amazon uses ML to power the following:

- Ecommerce recommendations engines
- Fulfillment center optimizations
- Supply chain, forecasting, and capacity planning
- Billions of weekly Alexa interactions
- Prime Air drone delivery
- Computer vision technology in Amazon Go

Why Customers Build with AWS - Lesson 3 of 9

Delivering for customers

- **Broadest choice of foundation models**

Challenge: Customers may need access to a variety of models and model sizes. They need the flexibility to optimize for cost, performance, and latency for different types of use cases and applications.

AWS solution: AWS was the first to recognize the importance of offering a broad choice of leading FMs from companies including Amazon, AI21 Labs, Anthropic, Cohere, Meta, Mistral AI, and Stability AI via a single API.

- **Easy to use tools for every type of customer**

Challenge: Customers may want to build applications with existing FMs.

AWS solution: Amazon Bedrock allows the customer to select the best FM for their use case and customize applications with their proprietary data in a secure and private environment. We've also built applications with integrated generative AI to help customers without ML expertise quickly take advantage of generative AI and increase employee productivity.

- **Purpose built infrastructure to help you scale**

Challenge: Whether customers are training their own models, customizing models, or running ML applications, they need to train and run inference at scale with an infrastructure that's purpose-built for ML.

AWS solution: AWS invested heavily in purpose-built accelerators, AWS Trainium and AWS Inferentia, to provide the most performant and low-cost infrastructure for generative AI.

- **A comprehensive set of data services**

Challenge: Many customers may have data locked in departmental silos or legacy applications. Their privacy controls may be ad-hoc, and data can be inaccurate and messy.

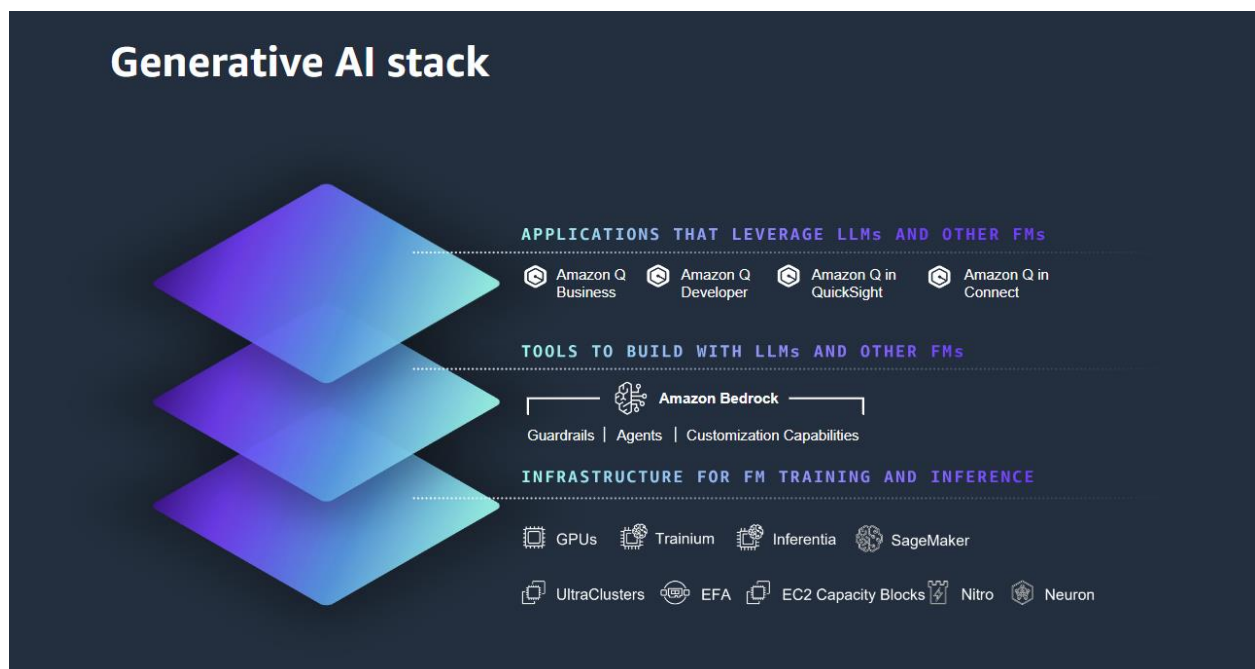
AWS solution: AWS provides an extensive set of services for managing data. This includes specific-purpose databases, data movement, metadata management, and data

governance. With the most comprehensive set of data and AI services, a customer can use its organization's data to securely customize an FM on AWS. This functionality allows them to build differentiated applications that know their business, data, and customers.

Generative AI with AWS - Lesson 4 of 9

AWS offers the following generative AI services for customers:

- **Amazon Q:** A generative AI–powered assistant designed for work that can be tailored to meet business needs.
- **Amazon Bedrock:** The easiest way to build and scale generative AI applications with FMs
- **Amazon Elastic Compute Cloud instances powered by AWS Inferentia and AWS Trainium:** The best price-performance infrastructure for training and inference in the cloud
- **Amazon SageMaker** Access and fine-tuning of a wide selection of FMs with Amazon SageMaker



Top layer

The top layer provides generative AI applications that customers can use to work with FMs without specialized knowledge or coding.

Middle layer

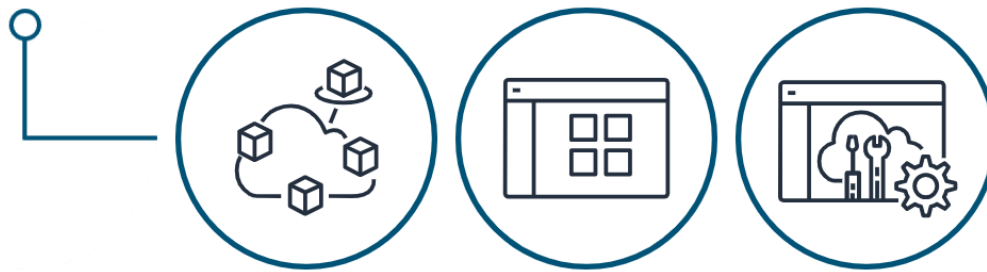
The middle layer provides tools, like Amazon Bedrock, that can be used to customize and train FMs for customers seeking to develop generative AI applications. This tier also assists customers in evaluating, selecting, and consuming these models through various distribution channels.

Bottom layer

The bottom layer provides solutions to customers looking to optimize the FM training and inference costs.

Amazon Q: is a generative AI-powered assistant that is designed specifically for work and can be tailored to a customer's business. Customers can get fast, relevant answers to pressing questions, generate content, and take actions – all informed by their information repositories, code, and enterprise systems. Amazon Q provides tailored information to employees to streamline common business tasks, quickly build applications on AWS, reduce time to business insights, provide better customer service, and plan and manage supply chain inventory more efficiently.

Key benefits



Get answers and guidance on AWS capabilities, services, and solutions

Ask Amazon Q to "Tell me about Amazon Bedrock Agents," and Q will give you a description of the feature plus links to relevant materials.

Choose the best AWS service for your use case and get started quickly

Ask Amazon Q "What are the ways to build a web app on AWS?" and it will provide a list of potential services like AWS Amplify, AWS Lambda, and Amazon EC2 with the advantages of each.

Diagnose and troubleshoot issues

Use Amazon Q if you encounter issues in the AWS Management Console, like Amazon EC2 permissions errors or Amazon S3 configuration errors. Choose the Troubleshoot with Amazon Q button, and it will use its understanding of the error type and service where the error is located to give you suggestions for a fix.

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available through an API. Customers can choose from a wide range of foundation models to find the model that is best suited for their use case.

Amazon Bedrock is the easiest way to build and scale generative AI applications with FMs.

Key benefits



Accelerate development of generative AI applications using FMs through an API, without managing infrastructure.



Choose from AI21 Labs, Anthropic, Stability AI, and Amazon to **find the right FM** for each use case.



Privately customize FMs using organizational data.



Enhance data protection using comprehensive AWS security capabilities.



Deploy scalable, reliable, and secure generative AI applications **with familiar AWS tools and capabilities**.

Amazon Bedrock supports leading foundation models

To learn more about the supported foundation models, choose each of the numbered markers.



AI21 Labs

Jurassic-2

Contextual answers, summarization, and paraphrasing

Anthropic

Claude 3, Claude 2.1 & Claude Instant

Summarization, complex reasoning, writing, and coding

Cohere

Command & Embed

Text generation, search, and classification

Meta

Llama 3, Llama 2

Dialogue use cases and language tasks

Mistral AI

Mistral 7B, Mistral 8x7B

Text summarization, Q&A, text classification, text completion, and code generation

Stability AI

Stable Diffusion XL 1.0

High-quality images and art

Amazon

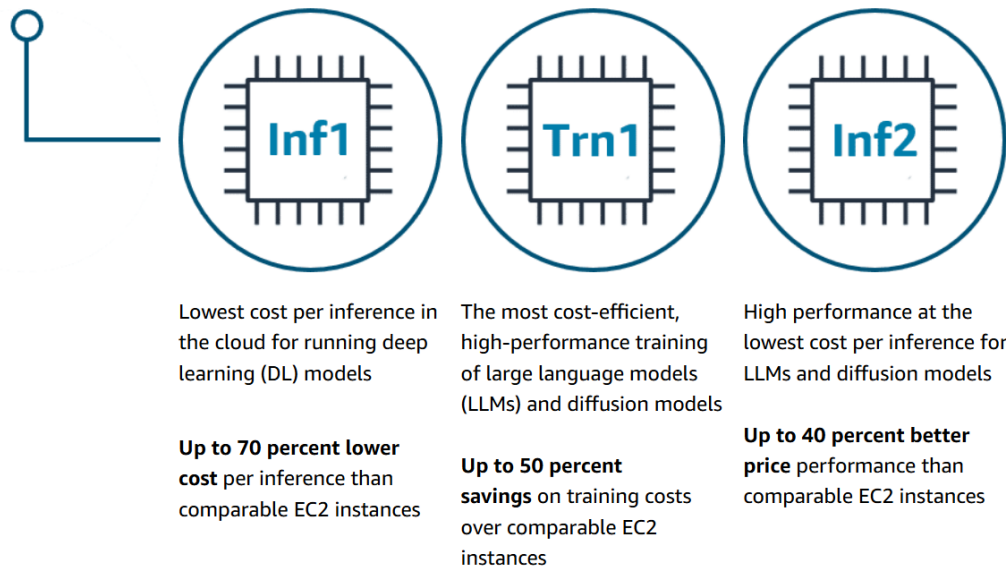
Amazon Titan

Summarization, image and text generation and search, and Q&A

AWS Inferentia and AWS Trainium EC2 instances: AWS Inferentia and AWS Trainium are purpose-built ML accelerators that AWS designed from the ground up.

The first-generation of AWS Inferentia delivers significant performance and cost-savings benefits for deploying smaller models. AWS Trainium and AWS Inferentia2 were built for training and deploying ultra-large generative AI models with hundreds of billions of parameters.

Key benefits



Amazon SageMaker: SageMaker offers the widest selection of FMs and access to the latest publicly available FMs for faster time to market. This customer choice is important because we believe no one model will rule them all. We are still in the early days of generative AI, and these models will continue to evolve at unprecedented speeds. That's why customers need the flexibility to use different models, at different times, and for different use cases.

Key benefits



Key Customer Personas - Lesson 5 of 9

Discovery questions for C-suite CMO and Marketing

When talking with this persona, you might focus on the following topics: accelerating content production, increasing reach with personalized ads, and quickly creating localized content. To learn more about C-suite marketing discovery questions, choose each of the following flashcards.

How often are you waiting for new media assets to be created to roll out new marketing campaigns?



How would you expand your brand's reach if you had the ability to deliver highly personalized ads?



How would your ability to enter new markets be accelerated if you had the ability to automatically generate marketing material?



With Amazon Rekognition Custom Labels, you can detect objects, scenes, and concepts in images and videos. Allowing automatic tagging of media assets to accelerate content production.



With Amazon Personalize and Amazon Pinpoint, you can develop campaigns that are tailored to your customers' interests.



With Amazon Translate, you can quickly translate text-based content into multiple languages. Enabling automatic creation of localized websites, apps, and marketing material.



Top of mind for key customer personas



Security and compliance: Customers place high priority on data security protections, regulatory compliance, and governance policies. Using secure infrastructure and following best practices helps mitigate risks.



Privacy: Customers want assurances that models do not disclose private data. Controls should prevent unauthorized data usage during model development.



Risk management: Customers are concerned about potential misuse when deploying generative AI applications. Proactive threat modeling and defense-in-depth security controls can help manage risks.



Balancing innovation with security: Customers want to experiment with new technologies but also manage risks and maintain security. Collaborative governance, review processes, and security guidance can enable innovation while upholding standards.



Content generation: Customers want assurances that AI-generated content aligns with brand messaging without hallucinations. Controls to prevent copyright infringement are also important.



Generating insights: Data analysts or marketing teams can use AI to analyze data and identify high-potential leads for campaigns to improve business outcomes.

As you are talking to key customer personas, the following example keywords can help you start conversations about generative AI opportunities.

- **Text:** Summarizing or automating content creation
- **Images:** Generating images, creating avatars
- **Audio:** Summarizing, generating, or converting text in audio
- **Video:** Generating or editing videos
- **Code:** Generating code, optimizing code
- **Chatbots:** Automating customer service and more
- **ML platforms:** Applications and ML platforms
- **Search:** AI-powered insights or vector search
- **Gaming:** Generative AI gaming studios or applications
- **Data:** Synthesizing, designing, collecting, or summarizing data

Prospecting with Qualifying Questions - Lesson 6 of 9

Understanding your customer's business

Which security and compliance requirements must your organization meet when implementing generative AI solutions?



What knowledge and experiences do you currently have with generative AI?



What are the specific features or capabilities that would be most beneficial to customers?



AWS supports 143 security standards and compliance certifications, including PCI-DSS, HIPAA/HITECH, FedRAMP, GDPR, FIPS 140-2, and NIST 800-171.



Understand the business outcomes your customer wants to achieve. Then work backward from those goals.



Center the conversation around who the end-users will be and how they will ultimately benefit.



Where can you enhance your customer's experience?



What are some of your more labor-intensive internal processes that may benefit from productivity or process improvements?



What are some creative processes that are already outsourced or not core to your organization's capabilities and mission?



Help your customer think about activities that drive better total customer value, foster cross-sell and up-sell, enhance conversion rates, or improve customer satisfaction metrics.









Dive into how your customer measures the efficiency and effectiveness of these internal processes.



The ability of LLMs to generate text and summarize documents, apply judgment, and simplify interactions will lead many customers to rethink their outsourcing models.



Evaluate business strategy and determine key business outcomes

<p>What existing data and cloud investments have you made as an organization, and what's your long-term roadmap?</p> 	<p>Where do you have concerns with the integration of generative AI capabilities into your current operating process and customer engagement activities?</p> 	<p>Did you achieve the intended business outcomes with previous AI & ML projects? If not, what were the lessons learned?</p> 
<p>Data residency and regulation will be a key consideration when customers are building their generative AI strategies and making vendor selections.</p> 	<p>Initiate discussion around change management needs when introducing generative AI solutions. How will they address employee concerns and provide training?</p> 	<p>Explore the maturity levels of your customer's: people, process, and tools.</p> 

Customer Use Cases - Lesson 7 of 9

Generative AI use cases can be grouped in three categories depending on who will be positively impacted. If you are meeting your customer to discuss their top business outcomes, generative AI can make a positive impact in the following areas:

- Reinventing the way customers interact with their company and enhancing their experience
- Boosting their employees' productivity
- Optimizing and augmenting their business processes, which can result in higher efficiencies and lower costs

1. Enhancing the customer experience

Business outcomes include the following:

- Improving customer engagement with seamless omnichannel access to virtual agents across different collaboration tools and channels
- Empowering customers to quickly find answers and complete transactions on their own by implementing conversational voice and text-based chatbots
- Discovering business insights out of real-time or recorded conversations to detect emerging trends

2. Boosting employee productivity

Business outcomes include the following:

- Improving productivity of internal teams that keep systems running
- Saving time and improving accuracy as the system automatically trains on each customer's information
- Creating faster time to market
- Reducing cost of service

3. Improving business operations

Business outcomes include the following:

- Improving employee productivity
- Increasing operational efficiencies
- Enhancing customer experiences

Use cases by industry Healthcare

- Life sciences
- Financial services
- Manufacturing
- Retail
- Media and entertainment