



Priya Arora

GLOBAL HEAD OF THE
GENERATIVE AI CENTER OF
EXCELLENCE FOR AWS
PARTNERS

Andre Boaventura

GLOBAL GENERATIVE AI SPECIALIST

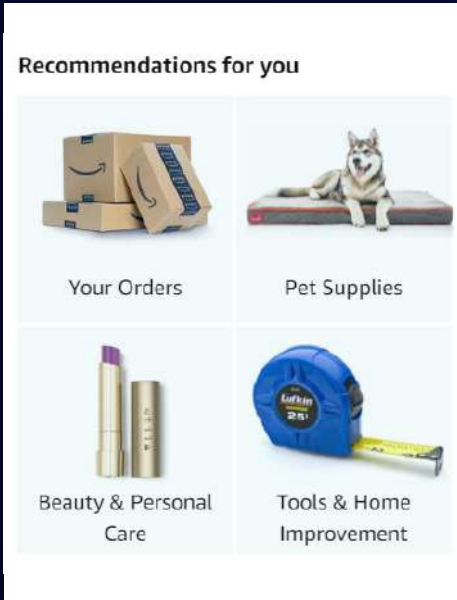
AMAZON WEB SERVICES

Innovation can
transform industries



GENERATIVE AI

ML innovation is in Amazon's DNA



**4,000 products
per minute** sold
on Amazon.com



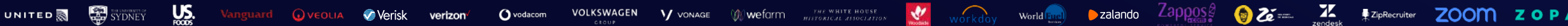
1.6M packages
every day



Billions of Alexa
interactions each week



First Prime Air delivery
on **December 7, 2016**



The tipping point for **Generative AI**



MASSIVE PROLIFERATION
OF DATA

AVAILABILITY OF
SCALABLE COMPUTE
CAPACITY

MACHINE LEARNING
INNOVATION

Generative AI is transforming all industries





**Enhance Customer
Experiences**

**Boost employee
productivity & creativity**

**Optimize business
processes**

Unlocking the potential of generative AI



Easiest way to build

with Foundation Models



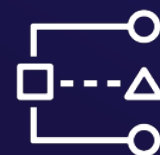
The most performant

infrastructure for generative AI



GenAI-powered

applications on AWS



Flexibility to build

your own Foundation Models

NOW GENERALLY AVAILABLE

Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models (FMs)



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



Choose FMs from Amazon, AI21 Labs, Anthropic, Cohere, Meta, and Stability AI to find the right FM for your use case



Privately customize FMs using your organization's data

Agenda

Bedrock Overview

Foundation Model updates

On-demand and Provisioned throughput

Pricing model

Feature Highlights

Demos

Building generative AI applications is challenging



Accessing multiple FMs
and newer versions



Customizing FMs
is not easy



Data privacy
and security



Getting FMs
to execute tasks



Connecting to
data sources



Difficult to manage
infrastructure

Amazon Bedrock

THE EASIEST WAY TO BUILD AND SCALE GENERATIVE AI APPLICATIONS WITH FMS



Access a range of leading FMs via a single API



Privately customize FMs with your own data



Enable data security and compliance



Build agents that execute complex business tasks by dynamically invoking APIs

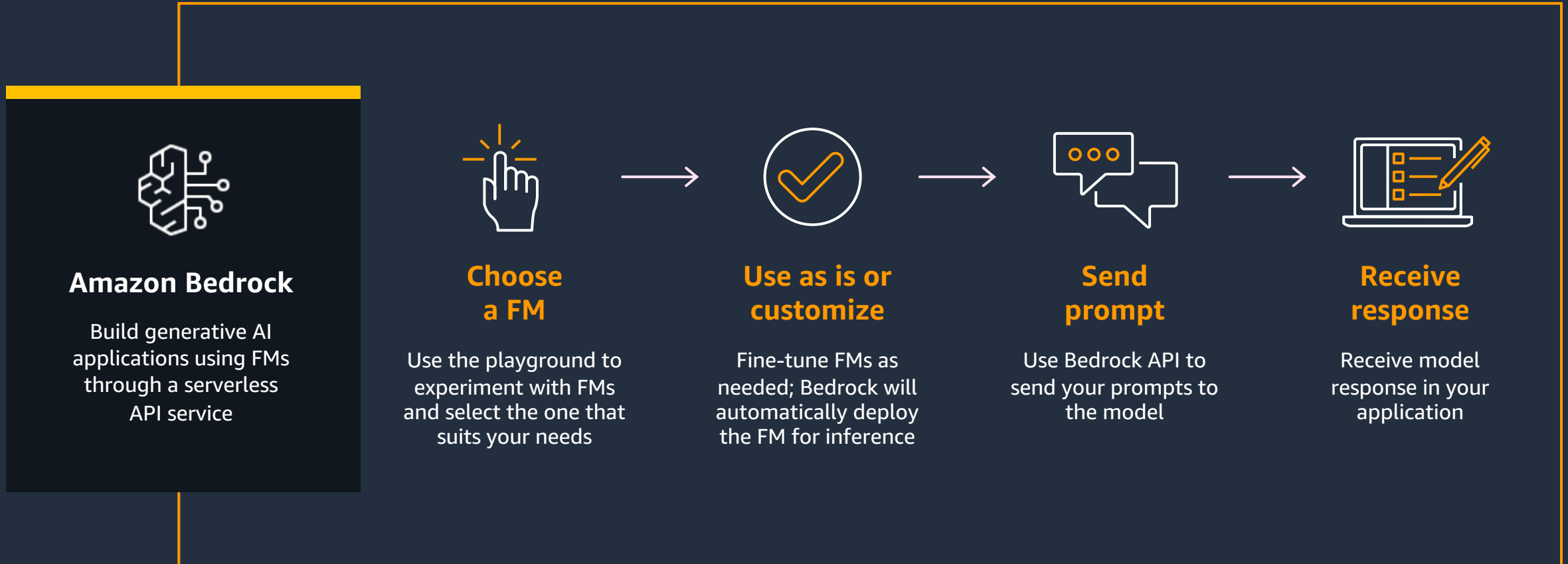


Extend the power of FMs with your data using retrieval augmented generation (RAG)



Get the best price performance without managing infrastructure

How Amazon Bedrock works



Benefits



Choice of leading foundation models



Easy model customization



Fully managed agents to execute tasks



Native support for RAG



Security and compliance

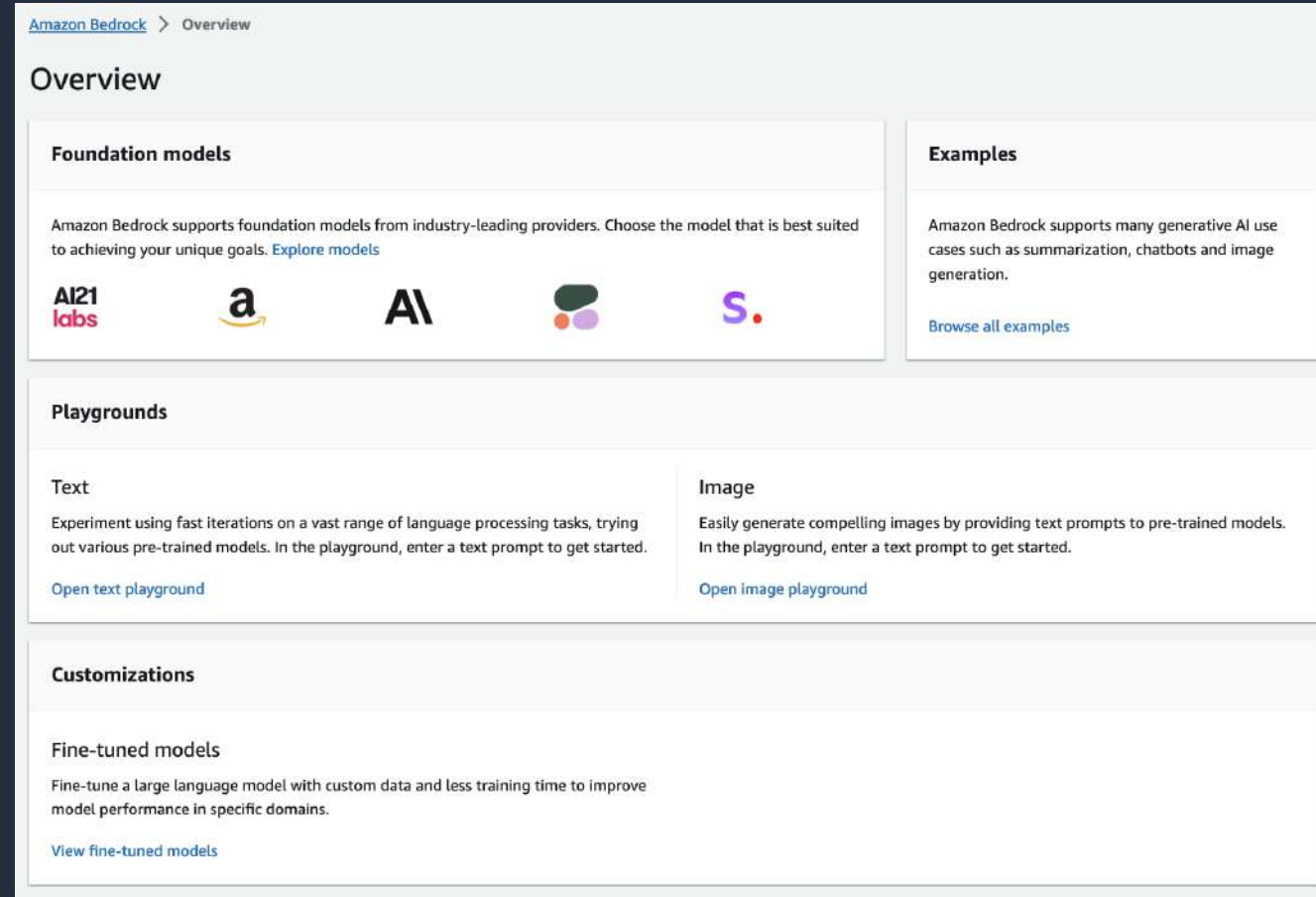


Work with a range of leading FMs

Use a single API access to FMs from Amazon and leading AI companies

Privately customize FMs for a particular task without annotating large volumes of data

Scale with managed infrastructure



Foundation model updates



Amazon Bedrock supports leading foundation models



Amazon Titan

Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings and search

AI21labs

Jurassic-2

Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch

ANTHROPIC

Claude 2

LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems



Command and embed

Text generation model for business applications and embeddings model for search, clustering, or classification in 100+ languages



COMING SOON

Llama 2

Fine-tuned models ideal for dialogue use-cases and language tasks

stability.ai

Stable Diffusion

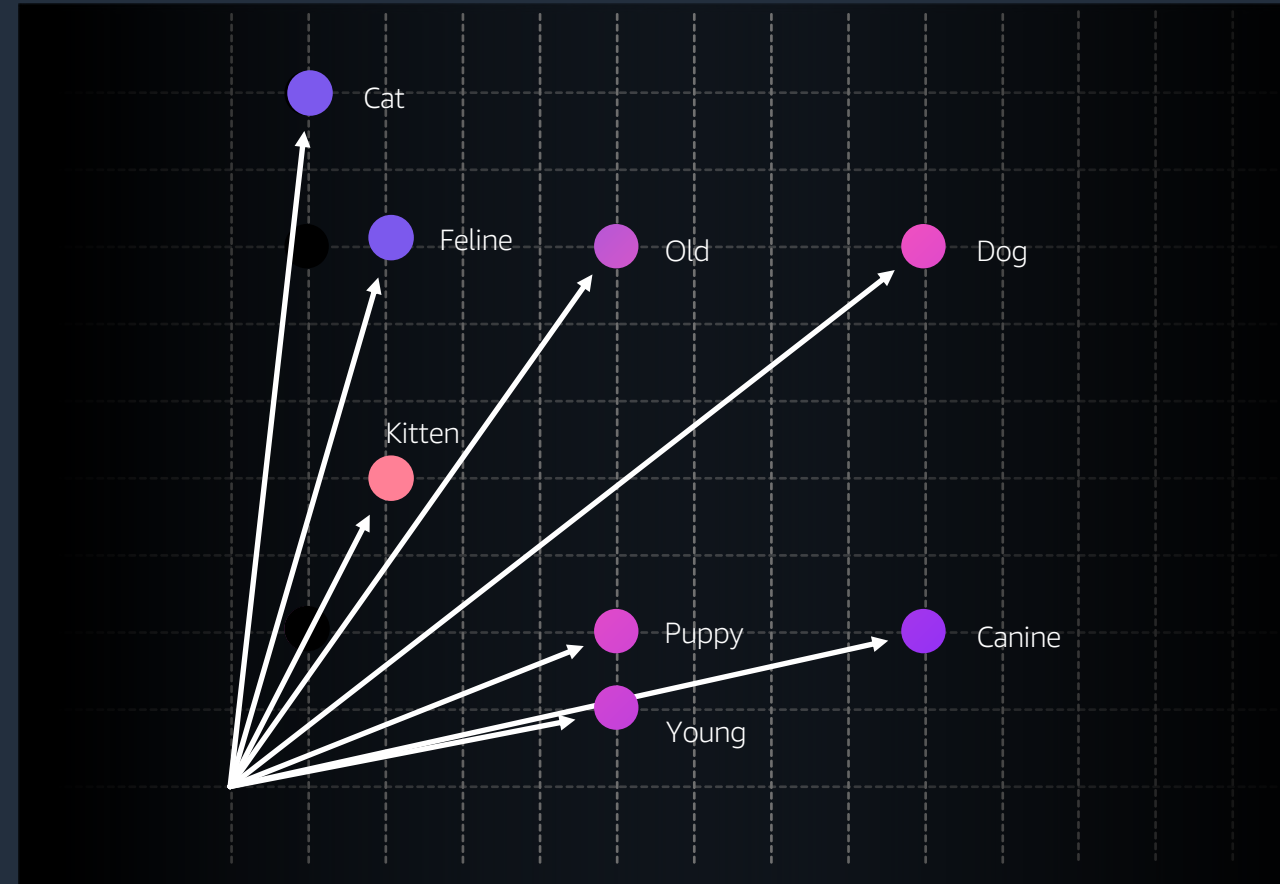
Generation of unique, realistic, high-quality images, art, logos, and designs

Tools for storing and retrieving **vector embeddings**

HUMAN TEXT

You may return or exchange your order within 30 days of when the order was shipped. Full refunds will be processed back to the original payment method. After 30 days, we offer a store credit only. Customers will be responsible for the cost to ship returns or exchanges after 30 days.

Words as **vectors**



WITHOUT SEMANTIC SEARCH

Result 1

10 results

1

Not applicable ?

text: Genuine FOOTJOY Shoes Bag Zipped Sports Bag Shoe Case
-Black Color. FOOTJOY Golf Shoes Bag & Sports Bag
Descriptions: FOOTJOY Golf Shoes Bag Material: PU
Color: Black Size: 35 (L) x 11 (H) (cm) Foldable type.



2

Not applicable ?

text: NIKE Genuine Zipped Shoe Bag Golf Shoe
Sports Gym Pouch NIKE Shoes Bag Information: Model:
golf shoe bag (Shoes Pouch) Color: Black Size: 36 x 23
x 15 (cm) Material: Polyester



3

Not applicable ?

text: 1 Pairs S... Stretcher for Golf Shoes. This pair
of shoes stretch your golf shoes in style! It is made
of high quality solid and durable in use.
used inside the shoe to preserve its shape, stop it from



4

Not applicable ?

text: Nike Cont... Golf Shoes (3 Y, White/Metallic
Silver/Bright C... #1.5 Swoosh design trad
embroidery at l... chest and Nike Athletic Department logo
clip label at front lower right hem. Dark grey heather jersey at



5

Not applicable ?

text: 100Pcs 70... Plastic Golf TEES.
Made of plastic strong;Packed in reseal able bag for
ease of use;Inn... sign, four corners to support;Flat end
even can clean up the mud on shoes or golf;Material: ABS



WITH SEMANTIC SEARCH

Result 2

10 results

1

Not applicable ?

FootJoy ladies Summer Series Golf Shoes 98847 White/Tan
The FootJoy Womens Superlite Spikeless golf
designed to provide versatility and style on and off
the course for all lady golfers. This new spikeless sole design



2

Not applicable ?

FootJoy TJ54252-13 Wide Contour Casual Mens Golf Shoes
n - 1 Wide. Pro Am Golf has specialized in the sale of
products for over 35 years Our products are designed to
that it helps you to improve the game Utmost



3

Not applicable ?

FootJoy Superlites Athletic Spikeless Golf Shoes 2016
White Grey/Black Medium 8.5. Pro Am Golf has specialized
le of golfing products for over 35 years Our products
ned to make sure that it helps you to improve the game



4

Not applicable ?

FootJoy Superlites Athletic Spikeless Golf Shoes 2016
Dark Blue/Lime Medium 10.5. Pro AM Golf has specialized
le of golfing products for over 35 years Our products
ned to make sure that it helps you to improve the game



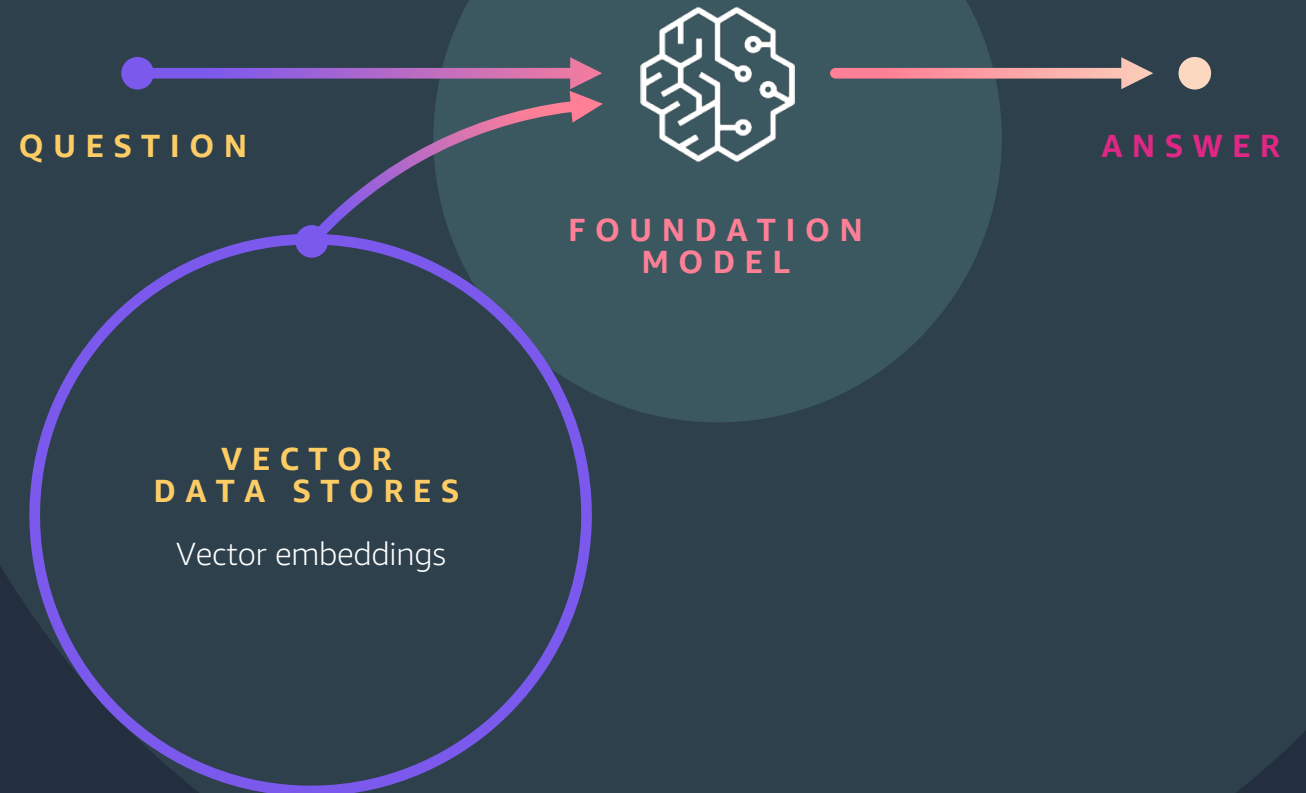
5

Not applicable ?

FootJoy Superlites Athletic Spikeless Golf Shoes 2016
White Grey/Black Medium 11.5. Pro AM Golf has specialized
le of golfing products for over 35 years Our products
ned to make sure that it helps you to improve the game



Vector embeddings support RAG



Titan Text Embeddings



Titan Text Embeddings V2.0

NEW

Translates text inputs (words, phrases) into numerical representations (embeddings). Comparing embeddings produces more relevant and contextual responses than word matching.

Max Tokens: **8,000**

Output Vectors: **1,536**

Language: **Multilingual** (25 languages)

Model ID: *amazon.titan-embed-g1-text-02*



Highlights

- Titan Text Embeddings offers fast, cost effective, high-performance, accurate embeddings in 25 languages.
- Optimized for text retrieval tasks, semantic similarity and clustering.
- Applications of this model includes semantic search and personalization.

Understanding Amazon Bedrock inference rates, limits, and throughput

Summary of Amazon Bedrock Inference options



On demand

Pay-as-you-go; no usage commitments

- Pricing based on input and output token count for LLMs
- Great for prototyping, POC's, small workloads with more relaxed requirements for throughput and latency
- Requests per minute (RPM) and Tokens per minute (TPM) enforced



Provisioned throughput

Provision sufficient throughput to meet your application's performance requirements

- Guaranteed throughput at a fixed cost
- Higher throughput available
- Flexible commitment term of 1 month or 6 months
- Pay hourly rate, discounted for extended commit
- Great for production workloads, or inference on custom models

Understanding on-demand RPM and TPM limits



On demand

- Hard limits on runtime quotas enforced
- Supports variety of workload types
 - High number of requests with low number of tokens
 - High number of tokens with low number of requests
 - And anywhere in between

Provisioned Throughput (PT) model units



Provisioned throughput

- Model Units provide throughput for a specific model, measured in terms of:

Max Input TPM | Max Output TPM

- Workloads use both input and output throughput
- Workloads use model units to achieve either one of those maximums at a time, or some portion of each
- Customers buy the number of model units that match their requirements
- Customers can flexibly use on-demand and multiple model units in the same account

Model pricing



Pricing dimensions



On demand

Pay-as-you-go; no usage commitments



Provisioned throughput

Provision sufficient throughput to meet your application's performance requirements



Model customization

Enable personalization and optimize solutions tailored to your business needs.

On demand pricing for text inference

Modality: Text and Text Embeddings

Model Provider	Model Name	Context Window	Prompt [per 1M tokens]	Completions [per 1M tokens]
AI21	J2 Mid	8,191	\$12.50	\$12.50
AI21	J2 Ultra	8,191	\$18.80	\$18.80
Amazon	Titan Text – Lite	4,000	\$0.30	\$0.40
Amazon	Titan Text – Express	8,000	\$1.30	\$1.70
Anthropic	Claude Instant	100,000	\$1.63	\$5.51
Anthropic	Claude 2	100,000+	\$11.02	\$32.68
Cohere	Command	4,096	\$15.00	\$20.00

Model Provider	Model Name	Context Window	Output vector	Input [per 1M tokens]
Amazon	Titan Embeddings – Text	8,000	1,536	\$0.10

On demand pricing for image inference

Modality: Image generation

Model Provider	Model Name	Image Resolution	Standard Quality [per 1,000 images]	Premium Quality [per 1,000 images]
Stability AI	SDXL 0.8	512 x 512 or smaller	\$18.00	\$36.00
Stability AI	SDXL 0.8	Larger than 512 x 512	\$36.00	\$72.00

Pricing for custom model training & inference

Modality: Text

Model Provider	Model Name	Price to train [1M tokens]	Price to infer from a custom model for 1 mu/hr.* [No commitment]	1 mo. commitment \$/hr./mu	6 mos. commitment \$/hr./mu
Amazon	Titan Text – Lite	\$0.40	\$7.10	\$6.40	\$5.10
Amazon	Titan Text – Express	\$0.80	\$20.50	\$18.40	\$14.80

Price to store each custom model per month: **\$1.95**

* No-commit option of 1 model unit is only available for inferencing from custom models

Data Privacy & Security



Data privacy



You are always in control of your data

- Customer data is not used to improve Amazon Titan models for other customers, and is not shared with other foundation model providers
- Customer data (fine-tuned models) remain in the region where they are created

Data security



You are always in control of your data

- Support for **AWS PrivateLink** so customers can establish private connectivity between virtual private clouds (VPCs) and the Bedrock service using VPC Endpoints
- Integration with AWS Identity and Access Management Service (IAM) to manage inference access, deny access for specific models, and enable Console access
- You can use CloudTrail to monitor API activity and troubleshoot issues as you build solutions
- Fine-tuned (customized) models are encrypted and stored using service managed keys; only you have access to your customized models through an endpoint
- Support for **Customer Managed Keys (CMK)** so customers can create and control keys to encrypt fine-tuned models
- Support for VPC configuration of fine-tuning jobs

Additional feature highlights



Model access

Fine grained model access

Customers can now easily subscribe to specific foundation models that best fit their use cases

For Claude models, customers will continue to 'Request' access by completing *Anthropic's models usage details* form.

Edit model access

Request access to models to use in Bedrock by selecting from the list below. For certain models, you may first need to submit some usage details before you are able to gain access. Note that only users with required IAM permissions can manage model access for this account. For more information about each model, see the [Providers](#) page.

Base models (7/7) ↻				
<input checked="" type="checkbox"/>	Models	Access status	Modality	EULA
<input checked="" type="checkbox"/>	<input type="checkbox"/> AI21 Labs			
<input checked="" type="checkbox"/>	Jurassic-2 Ultra	<input checked="" type="checkbox"/> Access granted	Text	EULA
<input checked="" type="checkbox"/>	Jurassic-2 Mid	<input checked="" type="checkbox"/> Access granted	Text	EULA
<input checked="" type="checkbox"/>	<input type="checkbox"/> Amazon			
<input checked="" type="checkbox"/>	Titan Embeddings G1 - Text	<input checked="" type="checkbox"/> Access granted	Embedding	EULA
<input checked="" type="checkbox"/>	Titan Text G1 - Express Preview	<input checked="" type="checkbox"/> Access granted	Text	EULA
<input checked="" type="checkbox"/>	<input type="checkbox"/> Anthropic			
<input checked="" type="checkbox"/>	Claude	<input checked="" type="checkbox"/> Access granted	Text	EULA
<input checked="" type="checkbox"/>	Claude Instant	<input checked="" type="checkbox"/> Access granted	Text	EULA
<input checked="" type="checkbox"/>	<input type="checkbox"/> Stability AI			
<input checked="" type="checkbox"/>	Stable Diffusion XL Preview	<input checked="" type="checkbox"/> Access granted	Image	EULA

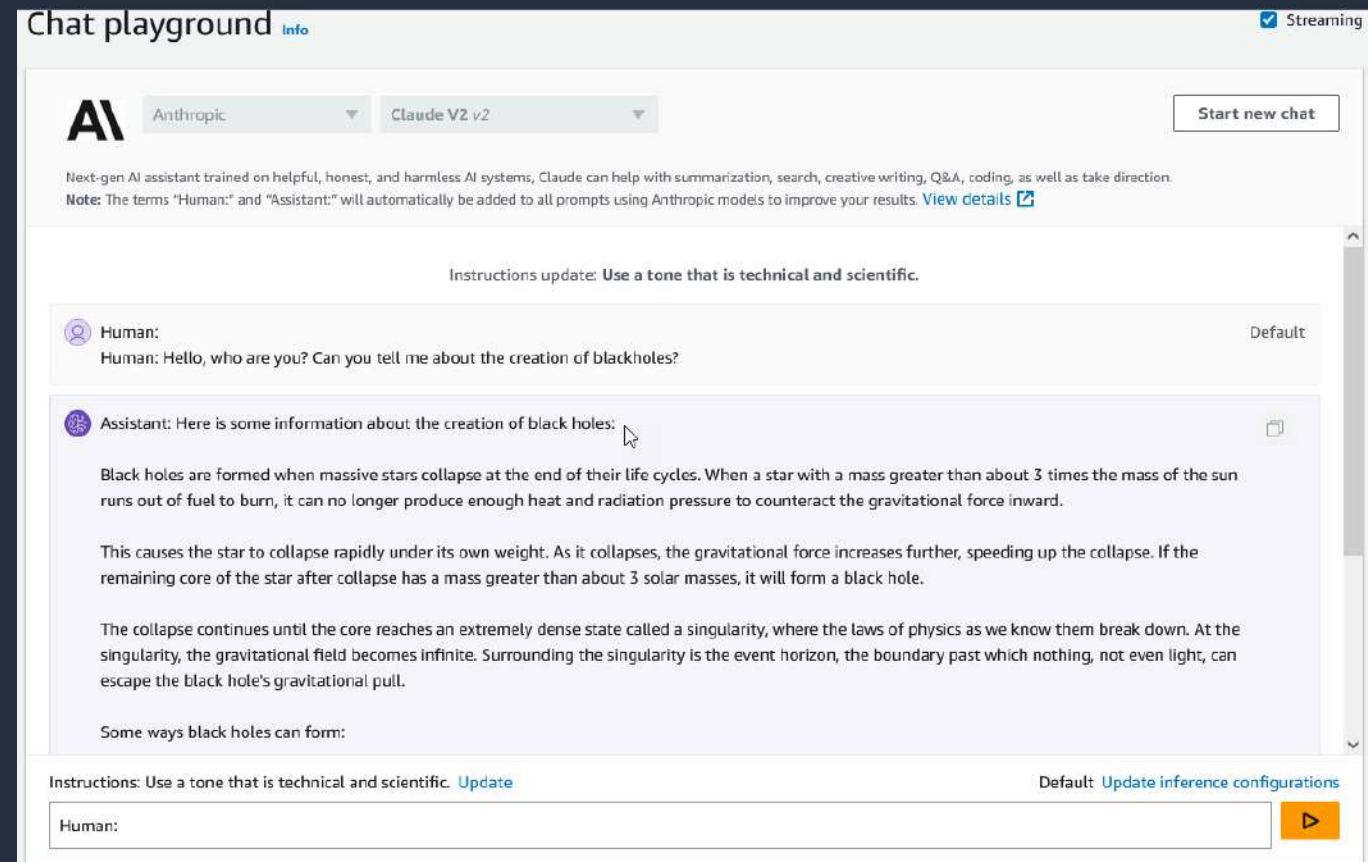
i By clicking Save changes, you are requesting access to the selected third party models through the AWS Marketplace. By doing so, you agree to the seller's pricing terms and End User License Agreements (EULA), and the [Bedrock Service Terms](#). You also agree and acknowledge that AWS may share information about this transaction with the respective sellers, in accordance with the [AWS Privacy Notice](#). AWS will issue invoices and collect payments from you on behalf of the seller through your AWS account. Your use of AWS services is subject to the [AWS Customer Agreement](#) or other agreement with AWS governing your use of such services.

Amazon Bedrock Playground

Quickly evaluate Foundation models using the Chat playground

Provide instructions to assume a role throughout the conversation in its entirety

Prompts and responses are not retained once the browser session is complete, or by clicking on 'Start new chat'



Knowledge base for retrieval augmented generation (RAG)

PREVIEW

- Connect FMs to data sources including vector engine for Amazon OpenSearch Serverless, Pinecone, and Redis Enterprise Cloud
- Enable automatic data source detection
- Provide source attribution



Amazon Bedrock > Knowledge base > Create Knowledge base

Step 1

[Provide details](#)

Step 2

Set up data source

Step 3

[Review and create](#)

Set up data source [Info](#)

Set up your data source by specifying the S3 location of your data, choosing an embeddings model to convert the data, and providing details for a vector database in which Bedrock can store, manage, and update your embeddings.

Data source

Specify the S3 location of your data. The vector database will ingest your data and convert it into an embedding.

Data source name

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 50 characters.

S3 URI



[View](#)

[Browse S3](#)

KMS Key - optional

Provide the KMS key to allow Bedrock to decrypt and encrypt your data.

KMS Key for transient data storage - optional

Provide a KMS key to store the transient data while we are in process of converting your data into embeddings.



[Create an AWS KMS key](#)

Embeddings model

Select an embeddings model to convert your data into an embedding. Pricing depends on the model.



Titan G1 Embeddings - Text v0.02

Vector database

Select your previously created database to allow Bedrock to store, update, and manage embeddings. [Learn more](#)

Select an existing database [Info](#)

☐ **Vector engine Amazon OpenSearch Serverless**
If you are a first time user, visit [OpenSearch](#) to create a vector database.



☐ **Pinecone**
If you are a first time user, visit [Pinecone](#) to create a vector database.



☐ **Redis Enterprise Cloud**
If you are a first time user, visit [Redis Enterprise Cloud](#) to create a vector database.



[Cancel](#)

[Next](#)

Agents enable generative AI applications to complete tasks in just a few clicks

PREVIEW



1

**SELECT YOUR
FOUNDATION MODEL**



2

**PROVIDE BASIC
INSTRUCTIONS**



3

**SELECT RELEVANT
DATA SOURCES**



4

**DEVELOPER SPECIFIES
LAMBDA FUNCTIONS**

| Breaks down and orchestrates tasks |

| Securely accesses and retrieves company data |

| Takes action by invoking API calls on your behalf |

| Provides fully managed infrastructure support |

Custom models

Create fine-tuned models using APIs

Fine-tuned models can be invoked the same way as base models, via the playground, or through the API on Provisioned throughput

Currently, fine-tuning is supported for Titan Text models. Support for other models will be enabled over time.


Customers with access to Amazon Titan Text models will now have access to fine tuning **NEW**

Tagging lets you easily manage, identify, organize, search for, and filter fine tuning jobs and models

[Amazon Bedrock](#) > Custom models


Custom models

Fine-tune and incrementally train a base model with your own data to customize the model for your domain.



Step 1. Fine-tune a model [Info](#)

To adapt a pre-trained model to specific tasks, fine-tune the model by providing your own labeled training data. This process can produce accurate models with smaller datasets and less training time.



Step 2. Purchase provisioned throughput [Info](#)


After a custom model is created, you need to purchase provisioned throughput to load custom models in playground.

[Purchase provisioned throughput](#)

Models

Training Jobs


Models



Actions ▼

Fine-tune model

Models that you have fine-tuned and have had their jobs successfully completed will appear here.

< 1 > 

Custom model name	Source Model	Provider	Creation time
No custom models			

Model invocation logging

Use model invocation logging to collect metadata, requests, and responses for all model invocations in your account.

Supports both S3 and CloudWatch logs

PREVIEW

Amazon Bedrock > Settings

Settings

These are account level settings

Model invocation logging [Info](#)

Use model invocation logging to collect metadata, requests, and responses for all model invocations in your account.

☒ Model invocation logging
Enabling model invocation logging will start publishing invocation logs.

Select the data types to include with logs - *Optional*
Select the data types for requests and responses for all model invocations.

☒ Text
☒ Image
☒ Embedding

Select the logging destinations

☒ S3 only ☐ Cloudwatch Logs only ☐ Both S3 and Cloudwatch Logs

S3 configurations

S3 bucket

[View](#) [Browse S3](#)

[Cancel](#) [Save settings](#)

Bedrock Demo



Applications built on top of Bedrock



Generative AI can be used for a wide range of use cases

Chatbots

Virtual assistants

AI-powered Contact Center

Personalization

Conversational search

Summarization

Code generation

Data to insights

Writing

Media design

Modeling

Document processing

Process optimization

Cybersecurity

Data augmentation

**Enhance
customer
experience**

**Boost
employee
productivity**

**Creativity
and content
creation**

**Improve
business
operations**

Demo: Intelligent Document Processing (IDP)

Improve employee productivity and make faster decisions with intelligent document processing

Organizations in various sectors, including healthcare, finance, legal, retail, and manufacturing, often handle a substantial volume of documents critical to their daily operations.

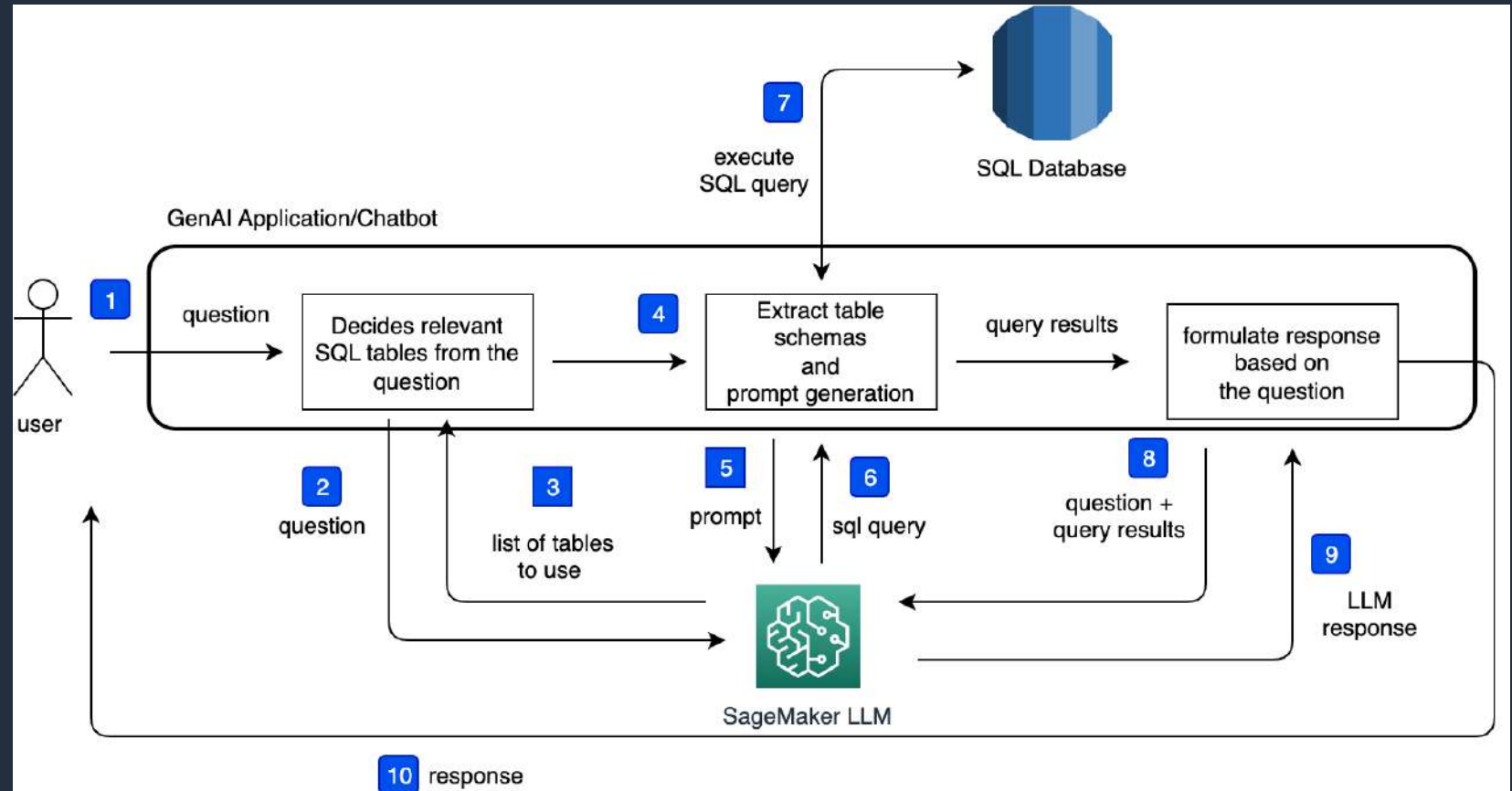
IDP with AWS AI services empowers organizations to automate information extraction from diverse document types and formats swiftly and accurately, eliminating the need for machine learning expertise, thereby enabling timely, cost-effective, and high-quality decision-making.



Demo: Gather business insights with Generative AI

Empowers business users to gain insights into historical, current and future states 24/7 without writing any code or SQL query.

- Use natural language to interact with enterprise data
- Customize the interface – text or speech
- Reduce dependency on database and programming experts
- Powered by an agent and a pipeline using LangChain
- Chatbot has memory buffer
- Cross-industry application



Demo: Avatar generation and diffusion moderation

Discover Stable Diffusion features & its implementation on AWS with Amazon Bedrock & Amazon SageMaker

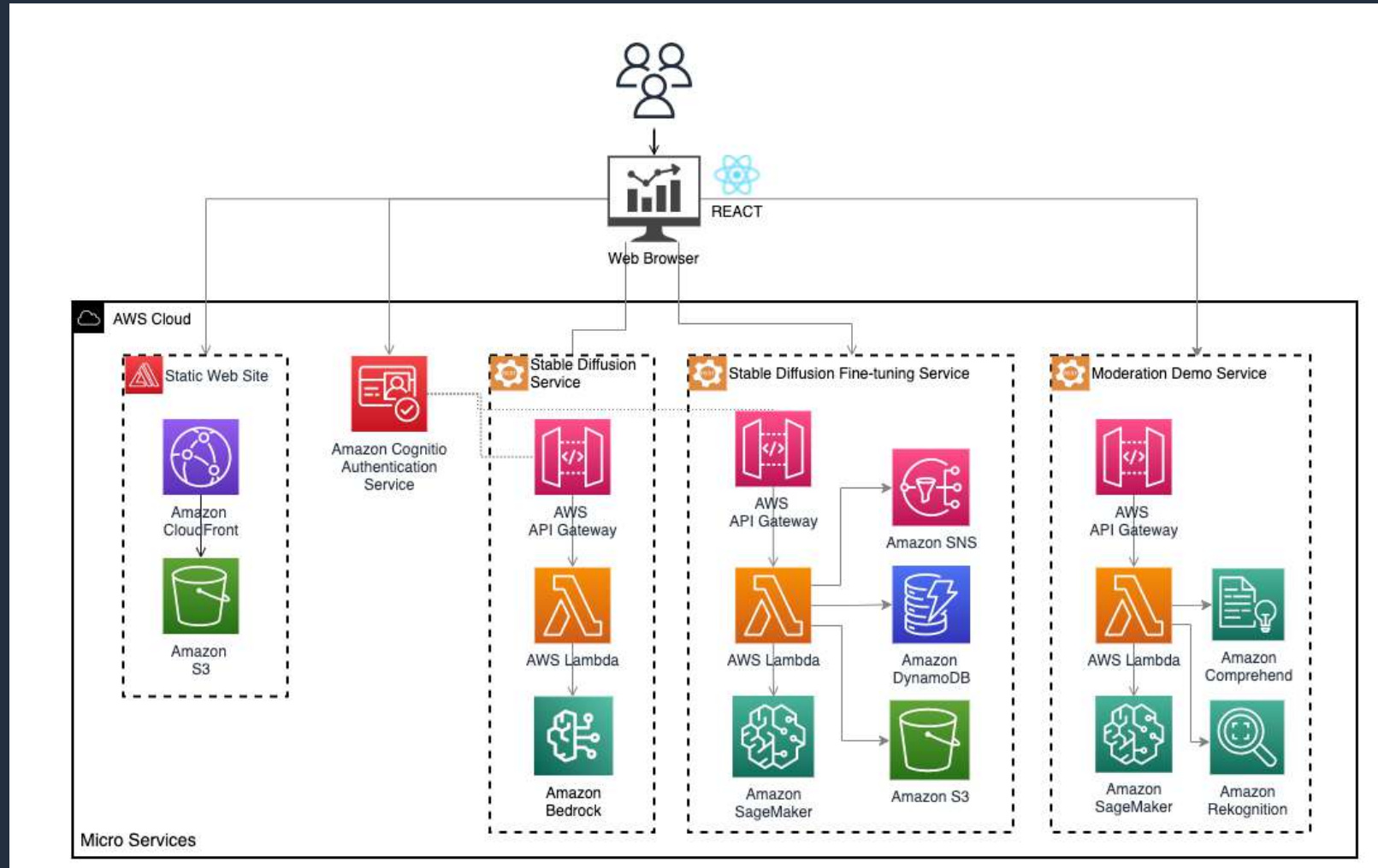
Features

- Text-to-image Generation
- Image-to-image Generation
- Masking / inpainting
- Fine-tuning
- Stable diffusion moderation



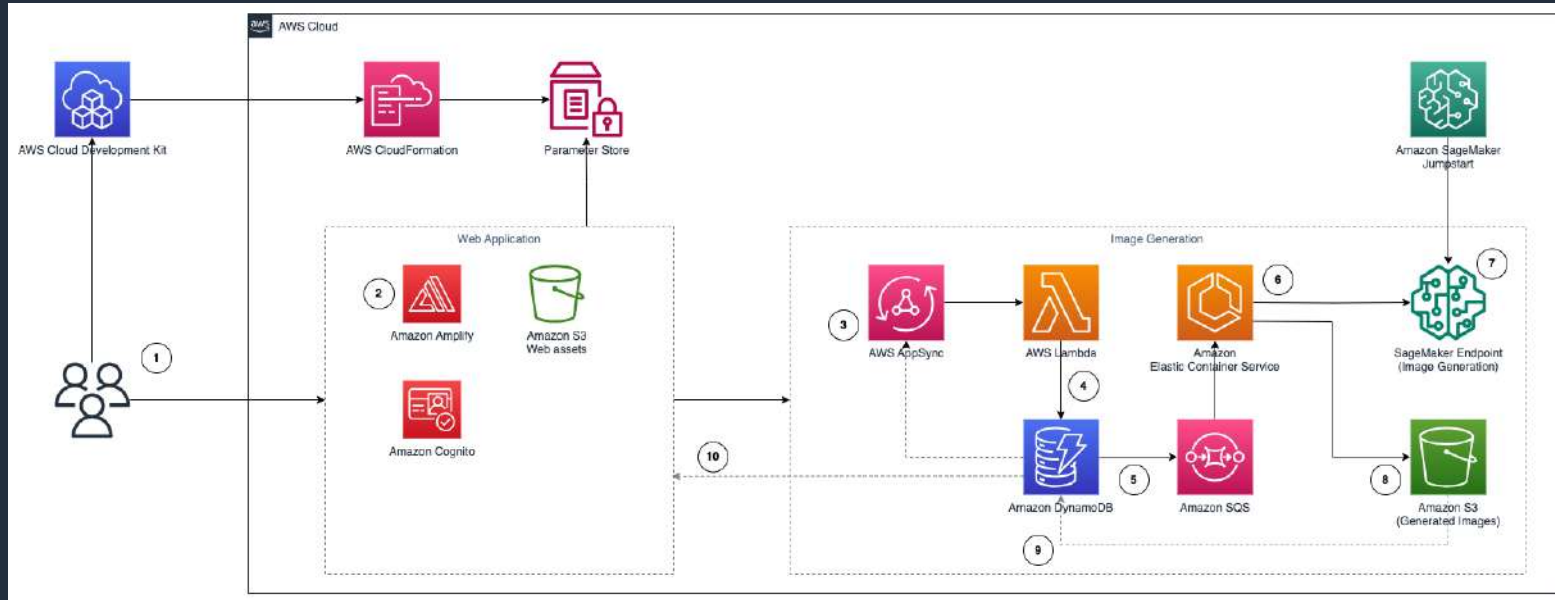
Use Cases

- Write marketing copy
- Ad campaign generation
- Content personalization
- Social media post generation



Demo: Generate Images for Advertising

Create high quality images for Ad campaigns using first party data and Generative AI

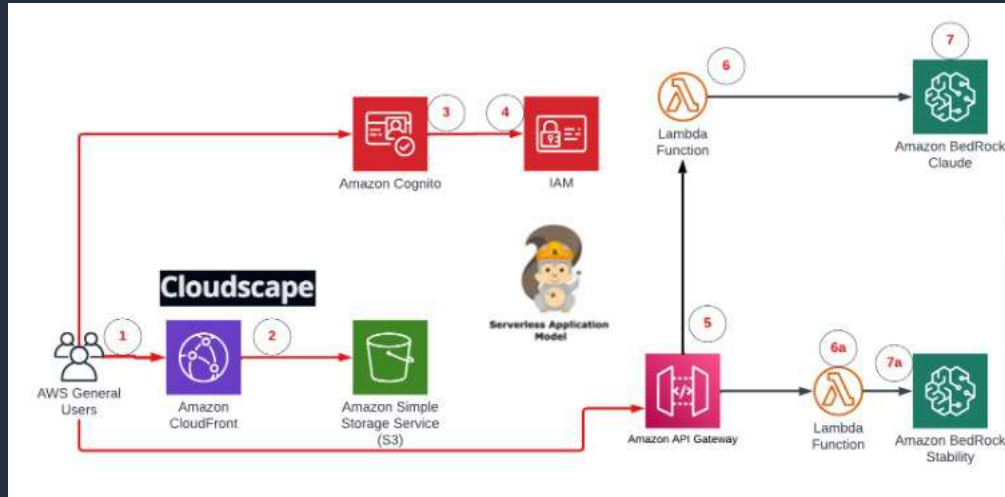


Metallic green colored car,
sunset time, color photo,
outdoors in pleasant landscape
on a mountain road.



Demo: Transform Advertising Text with Generative AI

Let your Imagination be your Best Advertising Tool



Best Mexican coffee 100% organic, is made with high quality beans, expertly roasted and immediately packed for freshness. It's a balanced coffee whose flavor resembles the heights regions from Colima, Chiapas, Guerrero



Creative Ad Tagline Generate ad tagline

Samples
Coffee MX

Product name
Coffee MX

Give a product description
Best Mexican coffee 100% organic, is made with high quality beans, expertly roasted and immediately packed for freshness. It's a balanced coffee whose flavor resembles the heights regions from Colima, Chiapas, Guerrero, Hidalgo, Jalisco, Nayarit, Oaxaca, Puebla, Querétaro, San Luis Potosí, Tabasco and Veracruz. Stimulate your senses with intense flavors and dry fruits notes.

Select gender
Not-specified

Select age group
50+

Result

Coffee MX

Here is a 29-word ad tagline for Coffee MX targeted to 50+ age group:

"Coffee MX: Savor the rich, organic flavors of Mexico in every sip."

Explanation:

The tagline focuses on the organic, high-quality Mexican coffee beans to appeal to the target 50+ demographic who may appreciate natural, authentic flavors. It evokes the sensory experience of drinking the coffee using words like "savor" and "rich". The reference to Mexico provides a sense of an exotic yet approachable origin. The brevity of the tagline makes it memorable while conveying the gourmet, flavorful nature of the product.

Other resources



SKILLBUILDER.AWS

600+ digital
courses



COURSES AND LEARNING PLANS SUBSCRIPTIONS CLASSROOM TRAINING AWS CERTIFICATION AWS PARTNER TRAINING

AWS Skill Builder

Your learning center to build in-demand cloud skills

FILTERS Search... NEWEST TO OLDEST

Featured content

Browse through the most popular courses offered in AWS Skill Builder

Cloud for CEOs

FREE

EN | 10m 00s ★ 5.0

Digital training

Build with Amazon MemoryDB for Redis

FREE

EN | 1h 00m ★ 5.0

Digital training

AWS Best Practices for Hybrid Cloud Adoption

FREE

EN | 1h 00m ★ 5.0

Digital training

AWS for Games Learning Plan: Cloud Game Development

FREE

12 courses | 11h 40m

Learning Plan

AWS Certified Developer - Associate Official Practice Question Set (DVA-C02)

FREE

EN | 45m 00s ★ 5.0

Digital training

Featured subscription content

Get the most out of your AWS Skill Builder subscription with these training offerings available only with the subscription.

Migrate a monolith web application to AWS using Application Migration...

Challenge AWS Game Skills: Basic

Exam Prep: AWS Certified Solutions Architect - Professional (SAP-C02) (wit...

Exam Prep: AWS Certified Advanced Networking - Specialty (ANS-C01) (with...

Industry Quest: Financial Services

Hands-on generative AI training opportunities

Learn the fundamentals
of generative AI for
real-world applications



The screenshot shows the Coursera interface for the course "Generative AI with Large Language Models". The top navigation bar includes the Coursera logo, a search bar, and a "Search" button. The left sidebar contains a "Course Material" section with a list of weeks (Week 1, Week 2, Week 3) and a "Grades" section. The main content area displays the course title, a search bar, and a "Search" button. Below this, the "Week 1" section is expanded, showing the course introduction, contributor acknowledgments, and a list of videos. A "Get started" button is prominently displayed. The bottom of the page features a row of five course cards, each with a title and a brief description.

Generative AI with Large Language Models

Search in course **Search**

Course Material

- Week 1
- Week 2
- Week 3

Grades

Notes

Messages 2

Course Info

Week 1

1h 56m of videos left 45 min of readings left 2 graded assessments left

Generative AI use cases, project lifecycle, and model pre-training

Show Learning Objectives

Introduction to LLMs and the generative AI project lifecycle 1 graded assessment left

Get started

- Course Introduction Video • 6 min
- Contributor Acknowledgments Reading • 10 min
- Introduction - Week 1 Video • 5 min
- Generative AI & LLMs Video • 4 min

[IMPORTANT] Have questions, issues or ideas? Join our Community!

Migrate a monolith web application to AWS using Application Migration...

Challenge AWS Game Skills: Basic

Exam Prep: AWS Certified Solutions Architect - Professional (SAP-C02) (wit...

Exam Prep: AWS Certified Advanced Networking - Specialty (ANS-C01) (with...

Industry Quest: Financial Services

Generative AI for Business Professional (3+ hours)

This specialized plan offers a strategic overview of generative models, enabling you to grasp the pivotal role they play in driving innovation across industries. From understanding the technology's impact on market dynamics to identifying opportunities for differentiation, this learning experience empowers participants to make informed decisions that harness the power of generative AI for business's success.






ABOUT THIS LEARNING PLAN		LEARNING PLAN STRUCTURE
	AWS Partner: Generative AI on AWS Essentials (Business) E-learning Duration 1h 5m	▼
	AWS PartnerCast - Generative AI Business Overview: Landscape and Market Opportunities (Business Focused) - Business E-learning Duration 1h 16m	▼
	AWS PartnerCast - Generative AI: The Future of Contact Center Automation and Modernization - Business E-learning Duration 44m	▼







[Link to enroll](#)

Generative AI for Technical Professionals (9+ hours)

Dive into the realm of AWS Generative AI with our specialized learning plan crafted for technical professionals, including professional services, solutions architects, data analysts and engineers, and Technical Pre-Sales Architects. This comprehensive plan offers a deep exploration of generative models, equipping participants with the skills to architect innovative solutions that push the boundaries of what's possible. From honing your understanding of foundations models to orchestrating seamless integrations of generative AI into complex architectures, participants will gain practical insights and hands-on experience.







	AWS Partner: Generative AI on AWS Essentials (Business) E-learning Duration 1h 5m	▼
	AWS PartnerCast - Building Generative AI on AWS: Key Service Features and Demos (Technical Focused) E-learning Duration 1h 27m	▼
	AWS PartnerCast - Breaking the Limits with AWS Generative AI - Technical E-learning Duration 1h 2m	▼
	AWS PartnerCast - GenAI Series: Building Generative AI Applications with AI21 on AWS - Technical E-learning Duration 1h	▼
	Introduction to AWS Inferentia and Amazon EC2 Inf1 Instances E-learning Duration 15m	▼

	AWS PartnerCast - Building Generative AI Solutions with Best Practices and Design Choices - Technical (Level 300 Advanced ML) - Technical E-learning Duration 1h 30m	▼
	AWS PartnerCast - GenAI Series: Building Generative AI Applications with Hugging Face on AWS - Technical E-learning Duration 1h 30m	▼
	Introduction to Amazon SageMaker E-learning Duration 13m	▼
	AWS PartnerCast - GenAI Series: Distributed Training on AWS SageMaker - Technical E-learning Duration 1h 30m	▼



Generative AI for Developers (6+ hours)

Explore the cutting-edge realm of AWS Generative AI through our comprehensive learning plan tailored for developers. From understanding the fundamentals of generative models to mastering the intricacies of AWS tools such as CodeWhisperer, this learning plan offers hands-on experiences, real-world use cases, and expert guidance.

ABOUT THIS LEARNING PLAN		LEARNING PLAN STRUCTURE
	Amazon CodeWhisperer - Getting Started E-learning Duration 30m	▼
	AWS PartnerCast - Tech Talks - CodeWhisperer - Technical E-learning Duration 1h 30m	▼
	Introduction to AWS Inferentia and Amazon EC2 Inf1 Instances E-learning Duration 15m	▼
	AWS PartnerCast - Amazon CodeWhisperer: Machine Learning-powered code generator - Technical E-learning Duration 1h	▼
	AWS PartnerCast - Building Generative AI Solutions with Best Practices and Design Choices - Technical (Level 300 Advanced ML) - Technical E-learning Duration 1h 30m	▼
	AWS PartnerCast - GenAI Series: Building Generative AI Applications with Hugging Face on AWS - Technical E-learning Duration 1h 30m	▼

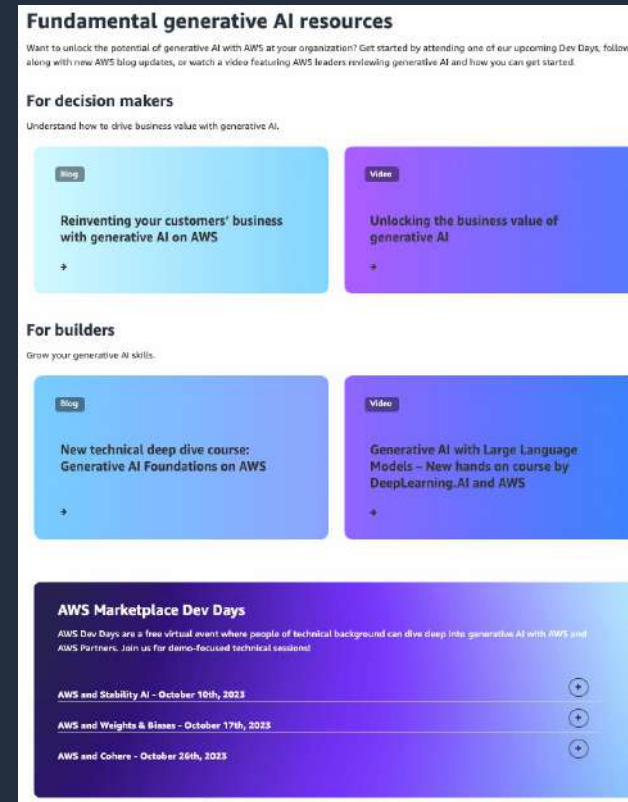
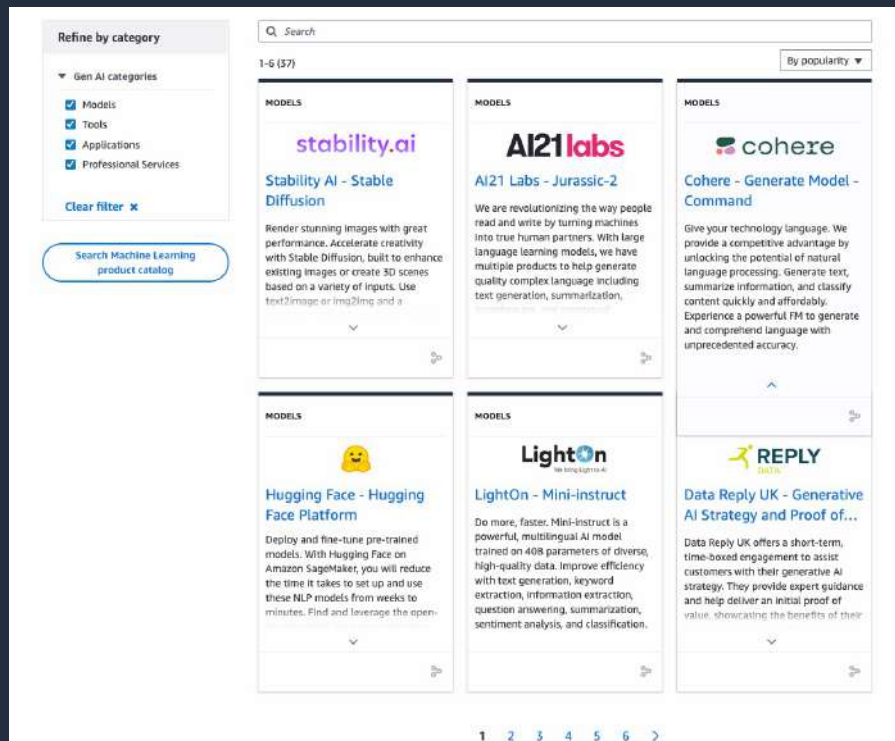


[Link to enroll](#)



AWS Marketplace

AWS Marketplace helps organizations discover, procure, experiment with, provision, and govern third-party software, data, and services. From FMs to tooling, applications to consultants, find a generative AI solution provider to kick-start your transformation journey.



<https://aws.amazon.com/marketplace/solutions/machine-learning/generative-ai>

Amazon Bedrock Workshop

 workshop studio

Amazon Bedrock Workshop X

- ▶ Prerequisites
- ▶ Prompt Engineering
- ▶ Text Generation
- ▶ Text Summarization
- ▶ Question Answering
- ▶ Conversational Interface - Chatbots
- ▶ Image Generation
- ▶ Code Generation
- ▶ Agents
- ▶ Entity Extraction

Amazon Bedrock Workshop

Amazon Bedrock Workshop

Welcome to "Amazon Bedrock" workshop!

The goal of this workshop is to give you hands-on experience leveraging foundation models (FMs) through Amazon Bedrock. Amazon Bedrock is a fully managed service that provides access to FMs from third-party providers and Amazon; available via an API. With Bedrock, you can choose from a variety of models to find the one that's best suited for your use case.

Within this series of labs, you will be taken through some of the most common usage patterns we are seeing with our customers for Generative AI. We will explore techniques for generating text and images, creating value for organizations by improving productivity. This is achieved by leveraging foundation models to help in composing emails, summarizing text, answering questions, building chatbots, creating images and generating code. You will gain hands-on experience using Bedrock APIs, SDKs, and open-source software for example LangChain and FAISS to implement these usage patterns.

This workshop is intended for developers and solution builders.



Workshop studio



Code in Github

Q&A





Thank you!