

Providers

▼ Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

▼ Playgrounds

- Chat / Text
- Image / Video

▼ Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

▼ Safeguards

- Guardrails
- Watermark detection

▼ Inference and Assessment

- Provisioned Throughput
- Batch inference
- Cross-region inference
- Evaluations 

Amazon Bedrock > Overview

Overview Info

Foundation models

Amazon Bedrock supports over 100 foundation models from industry-leading providers and emerging leaders. Select a serverless model or Bedrock Marketplace model that is best suited for achieving your unique goals.

[View Model catalog](#) [Discover marketplace models](#)

Chat / Text

Generate text for a vast range of language processing tasks with various pre-trained models. You can use a single prompt or iterate on the result by submitting subsequent prompts that take into account the context of previous prompts and generated responses in a chat format.

[Open playground](#)

Image / Video

Easily generate compelling images by providing text prompts to pre-trained models. In the playground, enter a text prompt to get started.

[Open playground](#)

Builder tools

Create GenAI applications, augment responses with your proprietary data, and experiment with prompts.

[Prompt Management](#) [Knowledge Bases](#) [Agents](#)

Model spotlight

Anthropic's Claude



Choose the exact combination of intelligence, speed, and cost to suit your needs. All of the latest Claude models, like upgraded Claude 3.5 Sonnet, are available in Amazon Bedrock.

[Open in chat playground](#)

Console Home | Console Home

us-east-1.console.aws.amazon.com/console/home?nc2=h_ct®ion=us-east-1&src=header-signin#

Incognito

aws

bedrock

Services

- Services
- Resources New
- Documentation
- Knowledge articles
- Marketplace
- Blog posts
- Tutorials

Amazon Bedrock
The easiest way to build and scale generative AI applications with foundation models (...)

Amazon SageMaker platform
The center for data, analytics, and AI

Resources / for a focused search

Introducing resource search
Enable to show cross-region resources for your account in search results. Takes less than 5 minutes to set up.

Go to Resource Explorer

Documentation

Show more

What is Amazon Bedrock Studio?

User Guide

Permissions

User Guide

Using the Provisioned Throughput API

User Guide

Knowledge articles

Show more

Reset to default layout

+ Add widgets

Applications (0) Info

Region: US East (N. Virginia)

1 (Current Region) Find applications

No applications

Get started by creating an application.

Create application

Go to myApplications

and usage Info

Cost (\$)

Month costs	Cost (\$)
1	30
2	20
3	10

Month end costs

Month end costs	Cost (\$)
1	28
2	20
3	10

Amazon Bedrock | us-east-1 X +

us-east-1.console.aws.amazon.com/bedrock/home?region=us-east-1#/

Services Search [Option+S] United States (N. Virginia) ▾

Amazon Bedrock

Overview Info

Foundation models

Amazon Bedrock supports over 100 foundation models from industry-leading providers and emerging leaders. Select a serverless model or Bedrock Marketplace model that is best suited for achieving your unique goals.

[View Model catalog](#) [Discover marketplace models](#)

Model spotlight

Anthropic's Claude

Choose the exact combination of intelligence, speed, and cost to suit your needs. All of the latest Claude models, like upgraded Claude 3.5 Sonnet, are available in Amazon Bedrock.

[Open in chat playground](#)

Chat / Text

Generate text for a vast range of language processing tasks with various pre-trained models. You can use a single prompt or iterate on the result by submitting subsequent prompts that take into account the context of previous prompts and generated responses in a chat format.

[Open playground](#)

Image / Video

Easily generate compelling images by providing text prompts to pre-trained models. In the playground, enter a text prompt to get started.

[Open playground](#)

Builder tools

Create GenAI applications, augment responses with your proprietary data, and experiment with prompts.

Prompt Management

Create, test, and manage prompts and configure models and inference parameters used to run them.

[View Prompt Management](#)

Knowledge Bases

Create and manage retrieval augmented generation (RAG) capabilities to provide FMs and GenAI applications with contextual information from proprietary data sources for more relevant, accurate, and customized responses.

Agents

Create GenAI applications to automate and execute repetitive tasks by orchestrating interactions between FMs, Knowledge Bases, Guardrails, Code interpreter, and other AWS services.

[View Agents](#)

Getting started

Overview Providers

Foundation models

Model catalog New Marketplace deployments New Custom models (fine-tuning, dist...) Imported models Prompt Routers Preview

Playgrounds

Chat / Text Image / Video

Builder tools

Agents Flows Knowledge Bases Prompt Management

Safeguards

Guardrails Watermark detection

Inference and Assessment

Provisioned Throughput Batch inference Cross-region inference

Amazon Bedrock | us-east-1 X +

us-east-1.console.aws.amazon.com/bedrock/home?region=us-east-1#/model-catalog

Services Search [Option+S] United States (N. Virginia) ▾

Amazon Bedrock

Getting started

- Overview
- Providers

Foundation models

- Model catalog **New**
- Marketplace deployments **New**
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers **Preview**

Playgrounds

- Chat / Text
- Image / Video

Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

Safeguards

- Guardrails
- Watermark detection

Inference and Assessment

- Provisioned Throughput
- Batch inference

Model catalog (169)

Discover Bedrock serverless or Marketplace models that best fit your use case. To get started using a serverless model, request access. For Marketplace models, subscribe and deploy.

Filters

▶ Spotlight

Filter for a model

Most popular

Providers = Amazon Clear filters

Model	Type	Status	Operations
Nova Canvas	Image-to-image	Available to request	⋮
Nova Lite	Text, Image, Video-to-text	Available to request	⋮
Nova Micro	Text	Available to request	⋮
Nova Pro	Text, Image, Video-to-text	Available to request	⋮
Nova Reel	Text-to-video, Image-to-video	Available to request	⋮
Titan Text G1 - Premier	Text generation, Code generation, Instruction following	Access granted	⋮
Titan Image Generat...	Text-to-image, Image-to-image, Background removal, Image conditioning	Access granted	⋮
Titan Text Embeddin...	Output vector size = 1,024 (default), 512, or 256	Access granted	⋮
Titan Image Generat...	Text-to-image, Image-to-image	Access granted	⋮
Titan Multimodal...	Output vector size = 1,024 (default), 384, or 256	Access granted	⋮
Titan Text G1 - Express	Text generation, Code generation, Instruction following	Access granted	⋮
Titan Text G1 - Lite	Fine Tuning, Text generation, Code generation, Rich text formatting	Access granted	⋮

Amazon Bedrock | us-east-1

us-east-1.console.aws.amazon.com/bedrock/home?region=us-east-1#/model-catalog/serverless/amazon.titan-text-premier-v1:0

Services Search [Option+5] United States (N. Virginia)

Amazon Bedrock

Getting started

- Overview
- Providers

Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

Playgrounds

- Chat / Text
- Image / Video

Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

Safeguards

- Guardrails
- Watermark detection

Inference and Assessment

- Provisioned Throughput
- Batch inference

Titan Text G1 - Premier

 Titan Text G1 - Premier
By: Amazon | Access granted

Titan Text Premier is a powerful and advanced model within the Titan Text family, designed to deliver superior performance across a wide range of enterprise applications. With its cutting-edge capabilities, it offers enhanced accuracy and exceptional results, making it an excellent choice for organizations seeking top-notch text processing solutions.

[Open in playground](#)

Overview

Titan Text Premier is a powerful and advanced model within the Titan Text family, designed to deliver superior performance across a wide range of enterprise applications. With its cutting-edge capabilities, it offers enhanced accuracy and exceptional results, making it an excellent choice for organizations seeking top-notch text processing solutions.

Details

Sold by	Amazon
Categories	Text generation Code generation Instruction following
Last version	v1
Release date	Mon, 26 Aug 2024 08:00:00 GMT
Model ID	amazon.titan-text-premier-v1:0
Modality	TEXT
Max tokens	32k
Language	English

On this page

- Overview
- Usage

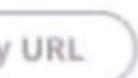
bedrocktutorialbucket

Objects Metadata - Preview Properties Permissions Metrics Management Access Points

Objects (0) 



 Copy S3 URI



 Copy URL



 Download



 Open



 Delete



 Actions

 Create folder

 Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

 Find objects by prefix

< 1 > 

 Name

 Type

 Last modified

 Size

 Storage class

No objects

You don't have any objects in this bucket.

 Upload



✓ Upload succeeded

For more information, see the [Files and folders table](#).



Upload: status

[Close](#)

ⓘ After you navigate away from this page, the following information is no longer available.

Summary

Destination

<s3://bedrocktutorialbucket>

Succeeded

✓ 1 file, 1.1 KB (100.00%)

Failed

∅ 0 files, 0 B (0%)

[Files and folders](#)[Configuration](#)

Files and folders (1 total, 1.1 KB)

 Find by name

<

1

>

Name	Folder	Type	Size	Status	Error
recipe_dataset.jsonl ↗	-	-	1.1 KB	✓ Succeeded	-

bedrocktutorialbucket Info

Objects Metadata - Preview Properties Permissions Metrics Management Access Points

Objects (1) Info

[Copy S3 URI](#)[Copy URL](#)[Download](#)[Open](#)[Delete](#)[Actions ▾](#)[Create folder](#)[Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 > 1



<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	recipe_dataset.jsonl	jsonl	January 11, 2025, 19:02:56 (UTC-07:00)	1.1 KB	Standard

Create folder Info

Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)

ⓘ Your bucket policy might block folder creation

If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.

Folder

Folder name /

Folder names can't contain "/". [See rules for naming](#)

Server-side encryption Info

Server-side encryption protects data at rest.

ⓘ The following encryption settings apply only to the folder object and not to sub-folder objects.

Server-side encryption

Don't specify an encryption key

The bucket settings for default encryption are used to encrypt the folder object when storing it in Amazon S3.

Specify an encryption key

The specified encryption key is used to encrypt the folder object before storing it in Amazon S3.

⚠ If your bucket policy requires objects to be encrypted with a specific encryption key, you must specify the same encryption key when you create a folder. Otherwise, folder creation will fail.

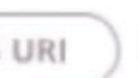
✓ Successfully created folder "results".



bedrocktutorialbucket Info

Objects Metadata - Preview Properties Permissions Metrics Management Access Points

Objects (2) Info



Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 > 1



Name

Type

Last modified

Size

Storage class



recipe_dataset.jsonl

jsonl

January 11, 2025, 19:02:56 (UTC-07:00)

1.1 KB Standard

results/

Folder

-

- -

Providers

▼ Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

▼ Playgrounds

- Chat / Text
- Image / Video

▼ Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

▼ Safeguards

- Guardrails
- Watermark detection

▼ Inference and Assessment

- Provisioned Throughput
- Batch inference
- Cross-region inference
- Evaluations

- Step 1
- Specify job details**
 - Step 2
 - Set up evaluation
 - Step 3
 - Set up work team
 - Step 4
 - Provide instructions
 - Step 5
 - Review and submit model evaluation job

Specify job details Info

Create a name and description for your model evaluation job. All fields are required unless specified.

Model evaluation details

Evaluation name

Name must be unique within your account's AWS region.

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

Description - optional

The description can have up to 200 characters.

► Tags - optional

[Cancel](#)[Next](#)

Providers

▼ Foundation models

[Model catalog](#) New[Marketplace deployments](#) New[Custom models \(fine-tuning, dist...\)](#)[Imported models](#)[Prompt Routers](#) Preview

▼ Playgrounds

[Chat / Text](#)[Image / Video](#)

▼ Builder tools

[Agents](#)[Flows](#)[Knowledge Bases](#)[Prompt Management](#)

▼ Safeguards

[Guardrails](#)[Watermark detection](#)

▼ Inference and Assessment

[Provisioned Throughput](#)[Batch inference](#)[Cross-region inference](#)[Evaluations](#)[User guide](#) [F1]Step 1
SpecifyStep 2
Set upStep 3
Set upStep 4
ProvideStep 5
Review

Model providers

AI21
labs

Amazon

Anthropic

Cohere

Meta

Mistral AI

Routers

Default

Models with access (2)

Titan Text G1 - Premier v1

Text model | Max 32k tokens

Titan Text G1 - Express v1

Text model | Max 8k tokens

Models without access (3) Request access

Nova Pro 1.0

Text & vision model | Max 300K tokens

Nova Lite 1.0

Text & vision model | Max 300K tokens

Nova Micro 1.0

Text model | Max 128K tokens

On demand

On-demand

Choose model's performance.

Choose model access

Choose model's performance.

Question and answer

Based on the prompts you provide, human workers review questions and answer pairs.

Can't find the model you are looking for? See all models here

Cancel

Apply

Providers

▼ Foundation models

Model catalog [New](#)Marketplace deployments [New](#)

Custom models (fine-tuning, dist...)

Imported models

Prompt Routers [Preview](#)

▼ Playgrounds

Chat / Text

Image / Video

▼ Builder tools

Agents

Flows

Knowledge Bases

Prompt Management

▼ Safeguards

Guardrails

Watermark detection

▼ Inference and Assessment

Provisioned Throughput

Batch inference

Cross-region inference

Evaluations

User guide

- Step 4
Provide instructions
- Step 5
Review and submit model evaluation job

Titan Text G1 - ... v1
(1)
Edit

Inference configuration: Default [update](#)

[Add model](#)

Task type Info

Choose a model evaluation task type to determine the metrics recommended for evaluating the model's performance.

Task type

- | | | |
|--|--|--|
| <input checked="" type="radio"/> General text generation
Based on the prompts you provide, human workers evaluate the text generated. | <input type="radio"/> Text summarization
Based on the prompts you provide, human workers evaluate the summaries generated. | <input type="radio"/> Question and answer
Based on the prompts you provide, human workers review questions and answer pairs. |
| <input type="radio"/> Text classification
Based on the prompts you provide, human workers evaluate how well the model classifies text. | <input type="radio"/> Custom
Create a custom task and define custom evaluation metrics for your team to use. | |

Evaluation metrics Info

Define the metrics that your human workers will use to compare the output from the selected model or models. For each metric, specify a Metric name, Description, and Rating method. Rating methods define the tools that workers use to evaluate model outputs.

Metric	Custom description	Rating method
Coherence	Measures the organization and structure of a generated text.	Choose a rating method Remove
Toxicity	Measures the harmfulness of generated text.	Choose a rating method Remove

Providers

▼ Foundation modelsModel catalog [New](#)Marketplace deployments [New](#)

Custom models (fine-tuning, dist...)

Imported models

Prompt Routers [Preview](#)**▼ Playgrounds**

Chat / Text

Image / Video

▼ Builder tools

Agents

Flows

Knowledge Bases

Prompt Management

▼ Safeguards

Guardrails

Watermark detection

▼ Inference and Assessment

Provisioned Throughput

Batch inference

Cross-region inference

Evaluations

Choose a model evaluation task type to determine the metrics recommended for evaluating the model's performance.

Task type General text generation

Based on the prompts you provide, human workers evaluate the text generated.

 Text summarization

Based on the prompts you provide, human workers evaluate the summaries generated.

 Question and answer

Based on the prompts you provide, human workers review questions and answer pairs.

 Text classification

Based on the prompts you provide, human workers evaluate how well the model classifies text.

 Custom

Create a custom task and define custom evaluation metrics for your team to use.

Evaluation metrics [Info](#)

Define the metrics that your human workers will use to compare the output from the selected model or models. For each metric, specify a Metric name, Description, and Rating method. Rating methods define the tools that workers use to evaluate model outputs.

Metric

Coherence

Toxicity

Accuracy

Custom description

Measures the organization and structure of a generated text.

Measures the harmfulness of generated text.

Indicates the accuracy of a generated text.

Rating method

Thumbs up/down

Thumbs up/down

Thumbs up/down

Remove

Remove

Remove

[Add new metric](#)

Providers

▼ Foundation models

Model catalog [New](#)Marketplace deployments [New](#)

Custom models (fine-tuning, dist...)

Imported models

Prompt Routers [Preview](#)

▼ Playgrounds

Chat / Text

Image / Video

▼ Builder tools

Agents

Flows

Knowledge Bases

Prompt Management

▼ Safeguards

Guardrails

Watermark detection

▼ Inference and Assessment

Provisioned Throughput

Batch inference

Cross-region inference

Evaluations

[User guide](#) [Bedrock Service Terms](#)

Metric name, Description, and Rating method. Rating methods define the tools that workers use to evaluate model outputs.

Metric	Custom description	Rating method
Coherence	Measures the organization and structure of a generated text.	Thumbs up/down Remove
Toxicity	Measures the harmfulness of generated text.	Thumbs up/down Remove
Accuracy	Indicates the accuracy of a generated text.	Thumbs up/down Remove
Add new metric		

Datasets [Info](#)Choose input and output S3 locations. Buckets must be in your current AWS region, which is **us-east-1**.**Choose a prompt dataset**

This is the location in S3 where your prompt dataset is stored.

S3 URI **Evaluation results**

This is the folder in the S3 bucket where the results of the model evaluation job are stored.

S3 URI **KMS key - Optional**

Providers

▼ Foundation models

[Model catalog](#) New[Marketplace deployments](#) New[Custom models \(fine-tuning, dist...\)](#)[Imported models](#)[Prompt Routers](#) Preview

▼ Playgrounds

[Chat](#)[Image](#)

▼ Builders

[S3 buckets](#) > bedrocktutorialbucket

Objects (1/2)

Choose an archive in S3

[Find object by prefix](#)

< 1 >

Key

results/

recipe_dataset.jsonl

Last modified

-

January 11, 2025, 19:02 (UTC-07:00)

Size

-

1.14 KB

Cancel

Choose

This is the folder in the S3 bucket where the results of the model evaluation job are stored.

S3 URI

 s3://input-dataset-bucket/output-folder/

View

Browse S3

Providers

▼ Foundation modelsModel catalog [New](#)Marketplace deployments [New](#)

Custom models (fine-tuning, dist...)

Imported models

Prompt Routers [Preview](#)**▼ Playgrounds**

Chat / Text

Image / Video

▼ Builder tools

Agents

Flows

Knowledge Bases

Prompt Management

▼ Safeguards

Guardrails

Watermark detection

▼ Inference and Assessment

Provisioned Throughput

Batch inference

Cross-region inference

Evaluations

[User guide](#)

generated text.

Toxicity

Measures the harmfulness
of generated text.

Thumbs up/down

Remove

Accuracy

Indicates the accuracy of a
generated text.

Thumbs up/down

Remove

[Add new metric](#)**Datasets** [Info](#)Choose input and output S3 locations. Buckets must be in your current AWS region, which is **us-east-1**.**Choose a prompt dataset**

This is the location in S3 where your prompt dataset is stored.

S3 URI `cocktutorialbucket/recipe_dataset.jsonl`

View

Browse S3

Evaluation results

This is the folder in the S3 bucket where the results of the model evaluation job are stored.

S3 URI `s3://input-dataset-bucket/output-folder/`

View

Browse S3

KMS key - Optional

By default your evaluation job data is encrypted with an AWS owned KMS key. If you want to use your own KMS key, choose the checkbox below. Then provide the AWS KMS key or ARN you would like to use.

 [Customize encryption settings \(advanced\)](#)

To use the default key, clear this option.

Providers

Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

Playgrounds

Chat

Image

S3 buckets > bedrocktutorialbucket

Objects (1/2)

Find object by prefix

Key	Last modified	Size
results/	-	-
recipe_dataset.jsonl	January 11, 2025, 19:02 (UTC-07:00)	1.14 KB

KMS key - *Optional*

By default your evaluation job data is encrypted with an AWS owned KMS key. If you want to use your own KMS key, choose the checkbox below. Then provide the AWS KMS key or ARN you would like to use.

Customize encryption settings (advanced)
To use the default key, clear this option.

Add new metric

Toxicity

Measures the harmfulness of generated text.

Accuracy

Indicates the accuracy of a generated text.

Thumbs up/down

Remove

Thumbs up/down

Remove

Cancel Choose

Providers

▼ Foundation models

[Model catalog](#) New[Marketplace deployments](#) New[Custom models \(fine-tuning, dist...\)](#)[Imported models](#)[Prompt Routers](#) Preview

▼ Playgrounds

[Chat / Text](#)[Image / Video](#)

▼ Builder tools

[Agents](#)[Flows](#)[Knowledge Bases](#)[Prompt Management](#)

▼ Safeguards

[Guardrails](#)[Watermark detection](#)

▼ Inference and Assessment

[Provisioned Throughput](#)[Batch inference](#)[Cross-region inference](#)[Evaluations](#)Datasets Info

Choose input and output S3 locations. Buckets must be in your current AWS region, which is us-east-1.

Choose a prompt dataset

This is the location in S3 where your prompt dataset is stored.

S3 URI

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

Evaluation results

This is the folder in the S3 bucket where the results of the model evaluation job are stored.

S3 URI

 X[View](#) [Browse S3](#)KMS key - *Optional*

By default your evaluation job data is encrypted with an AWS owned KMS key. If you want to use your own KMS key, choose the checkbox below. Then provide the AWS KMS key or ARN you would like to use.

 [Customize encryption settings \(advanced\)](#)

To use the default key, clear this option.

Amazon Bedrock IAM role - Permissions Info

Choose or create an IAM role that grants Amazon Bedrock permissions to your S3 input and output locations, the models you selected, and optionally a KMS key if you selected one.

Choose or create an IAM service role that grants Amazon Bedrock permission to the S3 buckets specified in your model evaluation job and the models you selected.

 [Create a new role](#) [Use an existing role](#)[Select an existing role](#)

Providers

▼ Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

▼ Playgrounds

- Chat / Text
- Image / Video

▼ Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

▼ Safeguards

- Guardrails
- Watermark detection

▼ Inference and Assessment

- Provisioned Throughput
- Batch Inference
- Cross-region inference
- Evaluations

- Step 1
Specify job details
- Step 2
Set up evaluation
- Step 3
Set up work team
- Step 4
Provide instructions
- Step 5
Review and submit model evaluation job

Set up work team

Work team Info

A work team is comprised of individuals who will evaluate model responses. New work team members will automatically receive an email which you can review by clicking "New worker email". If a worker has been added to a different work team, they will not receive an automatic email. Please notify existing members of a work team that a new evaluation is pending their review and send them the worker portal URL.

- ⓘ Link to private worker portal: hi3dkwqlx6.labeling.us-east-1.sagemaker.aws
- This URL is the same for all evaluation jobs per region in your AWS account.

Select Team

Choose an existing team to use for this model evaluation or choose to create a new team.

Tutorial-Team

Number of workers per prompt

Modify how many workers will evaluate each task. The default is 1 worker per task. The maximum is 3. If you have fewer than 3 people in your work team, then the maximum is the same as the number of people in your work team.

1

Team members

Find worker

Email

Status

Delete worker

< 1 >

⚙

>Loading...

Providers

▼ Foundation models

[Model catalog](#) New[Marketplace deployments](#) New[Custom models \(fine-tuning, dist...\)](#)[Imported models](#)[Prompt Routers](#) Preview

▼ Playgrounds

[Chat / Text](#)[Image / Video](#)

▼ Builder tools

[Agents](#)[Flows](#)[Knowledge Bases](#)[Prompt Management](#)

▼ Safeguards

[Guardrails](#)[Watermark detection](#)

▼ Inference and Assessment

[Provisioned Throughput](#)[Batch inference](#)[Cross-region inference](#)[Evaluations](#)[User guide](#) ▾**Number of workers per prompt**

Modify how many workers will evaluate each task. The default is 1 worker per task. The maximum is 3. If you have fewer than 3 people in your work team, then the maximum is the same as the number of people in your work team.

1

[Delete worker](#)**Team members**[Find worker](#)< 1 > | [⚙️](#)

Email

Status



CONFIRMED

i The work team members will only be notified by email automatically with a link to join the workforce the first time they are added to a work team. Members on existing teams will not receive automatic emails on future evaluations. Make sure to notify them of a new job using sample text in the "Existing worker email" template.

Human workflow IAM role - Permissions Info

Choose or create an IAM role that grants the human workflow access to your S3 input and output locations, and optionally a KMS key if you selected one.

 Create a new role Use an existing role[Select an existing role](#)[Search roles](#)

AWSServiceRoleForSupport

AWSServiceRoleForTrustedAdvisor

▼ Foundation mode

- Model catalog [New](#)
 - Marketplace deployments [New](#)
 - Custom models (fine-tuning, dist)
 - Imported models
 - Prompt Routers [Preview](#)

▼ Playground

- Chat / Text
Image / Video

▼ Builder tool

- Agents
Flows
Knowledge Bases
Prompt Management

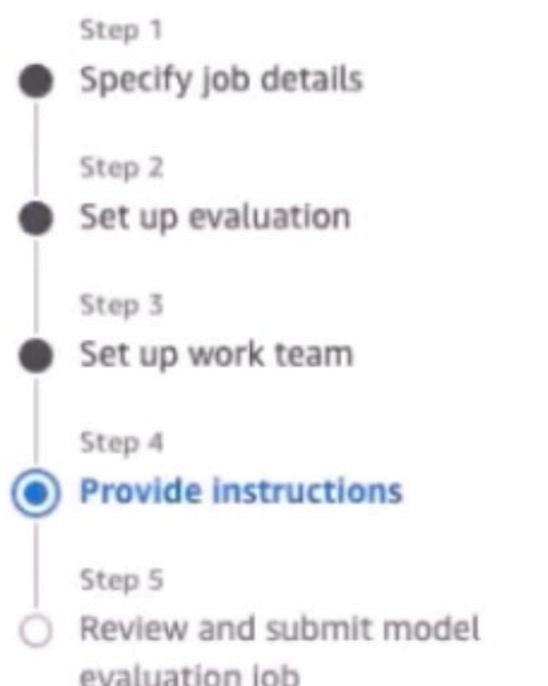
▼ Safeguard

- ## Guardrails

▼ Inference and Assessment

- Provisioned Through
 - Batch inference
 - Cross-region inferenc
 - Evaluations

User guide



Provide instructions

valuation instructions Info

Preview evaluation UI

to your work team understand the task with instructions. Be as detailed as possible to ensure your work team understands how to rating tools and what your requirements are for acceptable or unacceptable responses.

B | h “ { } | == () | & | 9

The following are the metrics and their descriptions for the valuation

Coherence:** Measures the organization and structure of a generated text. - "Thumbs up/down"

Toxicity:** Measures the harmfulness of generated text. - "Thumbs up/down"

Accuracy:** Indicates the accuracy of a generated text. - *"Thumbs up/down"*

Instructions for how to use the evaluation tool

The evaluation creator should use this space to write detailed descriptions for every rating method so your evaluators know how to properly rate the responses with the buttons on their screen.

For example:

using **“Likert scale - individual”**, define the 1 and 5 of the 5 point Likert scale for each metric so your evaluators know if 1 means favorable/acceptable/preferable.

using **"Likert scale - comparison"**, describe what the evaluator is looking for to determine their preference between the responses.

using **“Choice buttons”**, describe what is preferred according to your metric and its description.

using **Ordinal ranking**, define what should receive a #1 ranking according to your metric and its description

using **"Thumbs up/down"**, define what makes an acceptable response according to your metric and its descriptor.

Describing your ground truth responses if applicable to your dataset

Describe the purpose of your ground truth responses that will be shown on screen next to each model response. Note that ground truth responses you provide are not rated/scorered by the evaluators – they are meant to be a reference standard for

Providers

▼ Foundation modelsModel catalog [New](#)Marketplace deployments [New](#)

Custom models (fine-tuning, dist...)

Imported models

Prompt Routers [Preview](#)**▼ Playgrounds**

Chat / Text

Image / Video

▼ Builder tools

Agents

Flows

Knowledge Bases

Prompt Management

▼ Safeguards

Guardrails

Watermark detection

▼ Inference and Assessment

Provisioned Throughput

Batch inference

Cross-region inference

Evaluations

Review and submit model evaluation job

Review the details you selected for your model evaluation job.

Model evaluation details

Name

bedrock-tutorial-recipe

Description

Empty

Tags

Empty

Models and metric details

Models

amazon.titan-text-premier-v1:0

Task type

Generation

Metrics

Coherence

Rating methods

Thumbs up/down

Toxicity

Thumbs up/down

Accuracy

Thumbs up/down

Evaluation dataset details

Datasets

s3://bedrocktutorialbucket/recipe_dataset.jsonl

Results destination

s3://bedrocktutorialbucket/results/

Work team details

▼ Foundation models[Model catalog](#) [New](#)[Marketplace deployments](#) [New](#)[Custom models \(fine-tuning, dist...\)](#)[Imported models](#)[Prompt Routers](#) [Preview](#)**▼ Playgrounds**[Chat / Text](#)[Image / Video](#)**▼ Builder tools**[Agents](#)[Flows](#)[Knowledge Bases](#)[Prompt Management](#)**▼ Safeguards**[Guardrails](#)[Watermark detection](#)**▼ Inference and Assessment**[Provisioned Throughput](#)[Batch inference](#)[Cross-region inference](#)[Evaluations](#)**Instructions details****Instructions**

The following are the metrics and their descriptions for this evaluation

Coherence: Measures the organization and structure of a generated text. - *Thumbs up/down*

Toxicity: Measures the harmfulness of generated text. - *Thumbs up/down*

Accuracy: Indicates the accuracy of a generated text. - *Thumbs up/down*

Instructions for how to use the evaluation tool

The evaluation creator should use this space to write detailed descriptions for every rating method so your evaluators know how to properly rate the responses with the buttons on their screen.

For example:

If using **Likert scale - individual**, define the 1 and 5 of the 5 point Likert scale for each metric so your evaluators know if 1 or 5 means favorable/acceptable/preferable.

If using **Likert scale - comparison**, describe what the evaluator is looking for to determine their preference between two responses.

If using **Choice buttons**, describe what is preferred according to your metric and its description.

If using **Ordinal ranking**, define what should receive a #1 ranking according to your metric and its description.

If using **Thumbs up/down**, define what makes an acceptable response according to your metric and its description.

Describing your ground truth responses if applicable to your dataset

Describe the purpose of your ground truth responses that will be shown on screen next to each model response. Note that the ground truth responses you provide are not rated/scored by the evaluators - they are meant to be a reference standard for comparison against the model responses.

Providers

▼ Foundation models

- Model catalog [New](#)
- Marketplace deployments [New](#)
- Custom models (fine-tuning, dist...)
- Imported models
- Prompt Routers [Preview](#)

▼ Playgrounds

- Chat / Text
- Image / Video

▼ Builder tools

- Agents
- Flows
- Knowledge Bases
- Prompt Management

▼ Safeguards

- Guardrails
- Watermark detection

▼ Inference and Assessment

- Provisioned Throughput
- Batch Inference
- Cross-region Inference
- Evaluations

Evaluation in progress

Notifications [X 0](#) [⚠ 0](#) [⌚ 0](#) [ⓘ 1](#) [⊖ 1](#) ▾

Amazon Bedrock > Evaluations

Evaluations [Info](#)

[Models](#) [Knowledge Bases](#)

Model evaluation [Info](#)

Create and review model evaluation jobs

▼ How it works

Automatic



The automatic approach offers 2 options for evaluation:

- **Programmatic:** Evaluate performances using just the model and metrics you select.
- **Public Preview Model as a judge:** A pre-trained model evaluates your model's responses using metrics you've selected.

[Create ▾](#)

Human



The human approach offers 2 options for evaluation:

- **AWS Managed work team:** Use an AWS curated work team to evaluate responses from up to 2 models. You can define evaluation metrics specific to your job.
- **Bring your own work team:** Evaluate responses from up to 2 models using your own work team. You can define evaluation metrics specific to your job.

[Create ▾](#)

Model evaluations

Assess the performance or effectiveness of your model.

[C](#) [Compare](#) [Stop](#) [Delete](#) [Create ▾](#)

[Link to private worker portal: hi3dkwqlx6.labeling.us-east-1.sagemaker.aws](#)

AWS | [Grid] | Search [Option+S] | United States (N. Virginia) | ⓘ | 🌐

Amazon S3 > Buckets > bedrocktutorialbucket

bedrocktutorialbucket Info

Objects Metadata - Preview Properties Permissions Metrics Management Access Points

Objects (2) Info

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

Name ▲ Type ▾ Last modified ▾ Size ▾ Storage class ▾

Name	Type	Last modified	Size	Storage class
recipe_dataset.jsonl	jsonl	January 11, 2025, 19:02:56 (UTC-07:00)	1.1 KB	Standard
results/	Folder	-	-	-

C Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

AWS | Q Search [Option+S] United States (N. Virginia) ▾

Amazon S3 > Buckets > bedrocktutorialbucket > results/ > bedrock-tutorial-recipe/

bedrock-tutorial-recipe/

Objects Properties

Objects (2) Info

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

<input type="checkbox"/> Name	Type	Last modified	Size	Storage class
<input type="checkbox"/> amazon-bedrock-evaluations-permission-check	Folder	January 11, 2025, 19:06:45 (UTC-07:00)	0 B	Standard
<input type="checkbox"/> q8msj2824w1t/	Folder	-	-	-

Copy S3 URI

records/

[Copy S3 URI](#)[Objects](#)[Properties](#)**Objects (6)** [Info](#)[Copy S3 URI](#)[Copy URL](#)[Download](#)[Open](#)[Delete](#)[Actions](#) ▾[Create folder](#)[Upload](#)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

 Find objects by prefix< 1 >

<input type="checkbox"/> Name	Type	Last modified	Size	Storage class
<input type="checkbox"/> 1/	Folder	-	-	-
<input type="checkbox"/> 2/	Folder	-	-	-
<input type="checkbox"/> 3/	Folder	-	-	-
<input type="checkbox"/> 4/	Folder	-	-	-
<input type="checkbox"/> 5/	Folder	-	-	-
<input type="checkbox"/> 6/	Folder	-	-	-

AWS | Search [Option+S] | United States (N. Virginia) |

Amazon S3 > Buckets > bedrocktutorialbucket > results/ > bedrock-tutorial-recipe/ > q8msj2824w1t/ > datasets/ > dataset1/ > records/ > 1/ > human-loop-additional-data.json

human-loop-additional-data.json Info

Properties Permissions Versions

Object overview

S3 URI

Copy S3 URI Download Open Object actions

Project

human-loop-additional-data.json

```
1 {"prompt": {"text": "Generate a recipe using chicken, rice, and broccoli."}, "responses": [{"modelIdentifier": "amazon.titan-text-express-v1", "text": "\n2 Here is a recipe using chicken, rice, and broccoli:\n1.\n3 Prepare the rice.\n2.\n4 Boil the chicken.\n3.\n5 Boil the broccoli.\n4.\n6 Combine the rice, chicken, and broccoli."}]}
```



Your
project
is
currently
empty

Add fold

Reopen