



AWS GenAI Workshop

Michael Lin

Sr. Solutions Architect
Amazon Web Services

Labs

- Get hands-on with Bedrock playground
- Set up a Q&A application with knowledge retrieval
- Build a LLM application with Bedrock SDKs

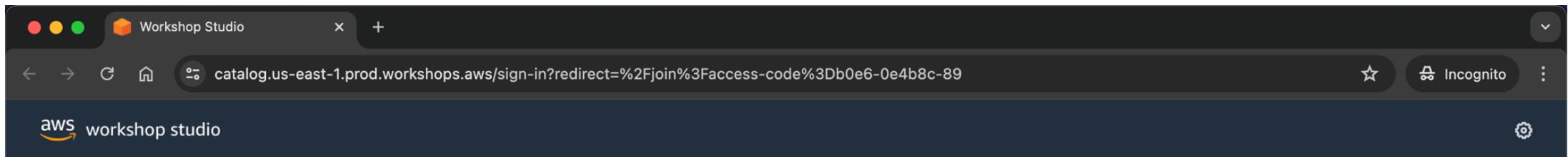


© 2024, Amazon Web Services, Inc. or its affiliates.

<https://ppt.cc/fAYV1x>



© 2024, Amazon Web Services, Inc. or its affiliates.



[Workshop Studio](#) > Sign in

Sign in

Choose a preferred sign-in method

Email one-time password (OTP)

Enter your personal or corporate email to receive a one-time password

AWS Builder ID

Login with AWS Builder ID, a new personal profile for builders

Amazon employee

Login with your Amazon Corporate account. Only for Amazon Employees.

One-time email passcode

login.us-east-1.auth.skillbuilder.aws/otp/input/response?redirect_uri=https%3A%2F%2Fgandalf-prod.auth.us-east-1.amazoncognito.com%2Foauth2%2Fidpr...

Incognito

aws training and certification

English

One-time email passcode

Send a passcode to the email below.

Email

michael.tw.lin@

Back Send passcode

[Get help signing in](#)

Verify one-time email passcode

login.us-east-1.auth.skillbuilder.aws/otp/challenge?redirect_uri=https%3A%2F%2Fgandalf-prod.auth.us-east-1.amazoncognito.com%2Foauth2%2Fidprespo... ☆ Incognito :

aws training and certification English ▾

One-time email passcode

We sent a passcode to michael.tw.lin@gmail.com. You should receive it within 5 minutes.

Passcode (9-digit) [Resend passcode](#)

Back **Sign in**

[Get help signing in](#)



- Step 1
Enter event access code
- Step 2
Review and join

Review and join

Event details

Name
TCC-Bedrock-Dryrun

Start time
7/27/2024 03:02 PM

Duration
72 hours

Level
300

Description

TCC-Bedrock-Dryrun

Terms and Conditions

Read and accept before joining the event

Read and accept before joining the event:

1. By using AWS Workshop Studio for the relevant event, You agree to [the AWS Event Terms and Conditions](#), the [AWS Responsible AI Policy](#), and the [AWS Acceptable Use Policy](#).
2. If You are under 18 years old, you may participate in the relevant event using AWS Workshop Studio: (a) if You are at least the minimum age below based on the country or region in which You reside, and (b) with the involvement of a parent, guardian, or educator.



Country or region	Minimum age
All countries or regions not listed below (including the United States, Brazil, the United Kingdom, and India)	13
Canada, China, Republic of Korea (South Korea)	14
Australia	15

* Philippines, Thailand, Turkey, and countries in Africa

3. You acknowledge and agree that You are using an AWS-owned account that You will only be able to access during the relevant event. You have no ownership rights over this AWS-owned account.
4. During the relevant event, while using this AWS-owned account, You will not use, import, input, or introduce any data, dataset, or other material that contains personal data, financial information, or any other data or materials that may be subject to laws and regulations (such as the General Data Protection Regulation or The Health Insurance Portability and Accountability Act of 1996).
5. If You find residual resources or materials in this AWS-owned account, You will notify your Event Operator immediately.
6. AWS, its affiliates, and any entities or persons acting on AWS's behalf reserves the right to terminate this AWS-owned account and to delete its contents at any time, without any notice to You.
7. During the relevant event, while using this AWS-owned account, You will not process or run any operation on any data other than test datasets or lab materials that have been approved by AWS.
8. You will not copy, import, export or otherwise create derivative works of materials provided by AWS for use outside of the relevant event.
9. AWS, its affiliates, and any entities or persons acting on AWS's behalf have no obligation to enable the transmission of Your materials through AWS Workshop Studio, and may, in their discretion, edit, block, refuse to post, or remove Your materials at any time, without notice to You.
10. If You access and use a service and/or third-party models that have their own terms during the relevant event, while in the AWS-owned account, You agree to review those terms and comply with them during the event.
11. If You are an AWS Partner using AWS Workshop Studio as part of Your participation in the AWS Partner Network Program, Your use of AWS Workshop Studio is governed by these terms, the AWS Partner Network Terms and Conditions, and the AWS Customer Agreement or other agreement with us governing your use of AWS Services.
12. Your use of AWS Workshop Studio will comply with these terms and all applicable laws. If You fail to comply with any of these terms, Your access to AWS Workshop Studio may be immediately terminated, without notice to You.

I agree with the Terms and Conditions

Cancel

Previous

Join event

catalog.us-east-1.prod.workshops.aws/join

aws workshop studio michael_tw_lin

Philippines, Thailand, Turkey, and countries in Africa

3. You acknowledge and agree that You are using an AWS-owned account that You will only be able to access during the relevant event. You have no ownership rights over this AWS-owned account.

4. During the relevant event, while using this AWS-owned account, You will not use, import, input, or introduce any data, dataset, or other material that contains personal data, financial information, or any other data or materials that may be subject to laws and regulations (such as the General Data Protection Regulation or The Health Insurance Portability and Accountability Act of 1996).

5. If You find residual resources or materials in this AWS-owned account, You will notify your Event Operator immediately.

6. AWS, its affiliates, and any entities or persons acting on AWS's behalf reserves the right to terminate this AWS-owned account and to delete its contents at any time, without any notice to You.

7. During the relevant event, while using this AWS-owned account, You will not process or run any operation on any data other than test datasets or lab materials that have been approved by AWS.

8. You will not copy, import, export or otherwise create derivative works of materials provided by AWS for use outside of the relevant event.

9. AWS, its affiliates, and any entities or persons acting on AWS's behalf have no obligation to enable the transmission of Your materials through AWS Workshop Studio, and may, in their discretion, edit, block, refuse to post, or remove Your materials at any time, without notice to You.

10. If You access and use a service and/or third-party models that have their own terms during the relevant event, while in the AWS-owned account, You agree to review those terms and comply with them during the event.

11. If You are an AWS Partner using AWS Workshop Studio as part of Your participation in the AWS Partner Network Program, Your use of AWS Workshop Studio is governed by these terms, the AWS Partner Network Terms and Conditions, and the AWS Customer Agreement or other agreement with us governing your use of AWS Services.

12. Your use of AWS Workshop Studio will comply with these terms and all applicable laws. If You fail to comply with any of these terms, Your access to AWS Workshop Studio may be immediately terminated, without notice to You.

I agree with the Terms and Conditions

Cancel Previous Join event

catalog.us-east-1.prod.workshops.aws/join

aws workshop studio

michael_tw_lin

Philippines, Thailand, Turkey, and countries in Africa

3. You acknowledge and agree that You are using an AWS-owned account that You will only be able to access during the relevant event. You have no ownership rights over this AWS-owned account.

4. During the relevant event, while using this AWS-owned account, You will not use, import, input, or introduce any data, dataset, or other material that contains personal data, financial information, or any other data or materials that may be subject to laws and regulations (such as the General Data Protection Regulation or The Health Insurance Portability and Accountability Act of 1996).

5. If You find residual resources or materials in this AWS-owned account, You will notify your Event Operator immediately.

6. AWS, its affiliates, and any entities or persons acting on AWS's behalf reserves the right to terminate this AWS-owned account and to delete its contents at any time, without any notice to You.

7. During the relevant event, while using this AWS-owned account, You will not process or run any operation on any data other than test datasets or lab materials that have been approved by AWS.

8. You will not copy, import, export or otherwise create derivative works of materials provided by AWS for use outside of the relevant event.

9. AWS, its affiliates, and any entities or persons acting on AWS's behalf have no obligation to enable the transmission of Your materials through AWS Workshop Studio, and may, in their discretion, edit, block, refuse to post, or remove Your materials at any time, without notice to You.

10. If You access and use a service and/or third-party models that have their own terms during the relevant event, while in the AWS-owned account, You agree to review those terms and comply with them during the event.

11. If You are an AWS Partner using AWS Workshop Studio as part of Your participation in the AWS Partner Network Program, Your use of AWS Workshop Studio is governed by these terms, the AWS Partner Network Terms and Conditions, and the AWS Customer Agreement or other agreement with us governing your use of AWS Services.

12. Your use of AWS Workshop Studio will comply with these terms and all applicable laws. If You fail to comply with any of these terms, Your access to AWS Workshop Studio may be immediately terminated, without notice to You.

I agree with the Terms and Conditions

Cancel Previous Join event

Amazon Bedrock Workshop (|) X TCC-Bedrock-Dryrun X +

catalog.us-east-1.prod.workshops.aws/event/dashboard/en-US

aws workshop studio michael_tw_lin

Bedrock and Claude Deep Dive Workshop

► 0. Prerequisites

- 1. Text Generation
- 2. RAG and Knowledge Bases
- 3. Image and Multimodal
- 4. Tool Use
- 5. Agents for Bedrock
- 6. Security and Governance
- 7. Summary
- 99. Resource release

▼ AWS account access

[Open AWS console \(us-west-2\)](#)  

[Get AWS CLI credentials](#)

▼ Content preferences

Language English ▾

Event dashboard > Bedrock and Claude Deep Dive Workshop

Event information

Start time 7/27/2024 03:02 PM	Duration 72 hours	Accessible regions us-west-2, us-east-1
---	-----------------------------	---

Description
TCC-Bedrock-Dryrun

Workshop

Title Bedrock and Claude Deep Dive Workshop	Complexity level 300	AWS services Amazon Bedrock	Topics Generative AI
---	--------------------------------	---------------------------------------	--------------------------------

Description
Showcase Bedrock and Claude 3 functionality with scenarios like Multimodal, Tool Use, Knowledge base, ChatBot, RAG, Agent and Guardrails.

Event Outputs (0)

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Console Home | Console Home

us-west-2.console.aws.amazon.com/console/home?region=us-west-2

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Console Home

Reset to default layout + Add widgets

Recently visited

- Support
- Service Quotas
- EC2
- IAM
- Amazon SageMaker
- Billing and Cost Management
- Amazon OpenSearch Service

View all services

Applications (0)

Create application

Region: US West (Oregon)

us-west-2 (Current Region) Find applications < 1 >

Name	Description	Region	Originating account
No applications Get started by creating an application.			

Create application

Go to myApplications

Welcome to AWS

Getting started with AWS

AWS Health

Open issues

Cost and usage

Current month costs Cost (\$)

AWS Services Search bar: bedrock

Search results for 'bedrock'

Services

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

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.
Dismiss Go to Resource Explorer

Documentation See all 2,676 results ▶

- Amazon Bedrock User Guide
- Add a data source to your app User Guide
- Document history for the Amazon Bedrock Studio User Guide User Guide
- What is Amazon Bedrock Studio? User Guide

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Services Search bar: bedrock

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Search results for 'bedrock'

Services (1)

Amazon Bedrock 

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

Resources New

Documentation (2,676)

Knowledge Articles (12)

Marketplace (386)

Blogs (232)

Events (1)

Tutorials (1)

Introducing resource search

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

Dismiss Go to Resource Explorer

Documentation See all 2,676 results ▶

Amazon Bedrock 

User Guide

Add a data source to your app 

User Guide

Document history for the Amazon Bedrock Studio User Guide 

User Guide

What is Amazon Bedrock Studio? 

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

The screenshot shows the AWS Management Console with the 'Amazon Bedrock' service selected. The top navigation bar includes tabs for 'Amazon Bedrock Workshop', 'TCC-Bedrock-Dryrun', 'Amazon Bedrock | us-west-2', and 'New Tab'. The main search bar contains the URL 'us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/'. The navigation menu on the left has a red arrow pointing to the 'Machine Learning' section, which is currently selected. The main content area features the heading 'Amazon Bedrock' and the subtext 'The easiest way to build and scale generative AI applications with foundation models (FMs)'. To the right is a call-to-action box with the heading 'Try Bedrock' and a 'Get started' button.

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Machine Learning

Amazon Bedrock

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

Try Bedrock

Get started

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits



The screenshot shows a browser window with the following details:

- Tab Bar:** Contains tabs for "Amazon Bedrock Workshop", "TCC-Bedrock-Dryrun", "Amazon Bedrock | us-west-2", and "New Tab".
- Address Bar:** Displays the URL: "us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/".
- Header Bar:** Shows the AWS logo, "Services" button, a search bar with placeholder "[Option+S]", and user information: "Oregon" and "WSParticipantRole/Participant @ 1561-5387-8293".
- Left Sidebar (Foundation Models):**
 - Base models**
 - Imported models** ([Preview](#))
 - Playgrounds**
 - Chat
 - Text
 - Image
 - Safeguards**
 - Guardrails
 - Watermark detection
 - Builder tools**
 - Knowledge bases
 - Agents
 - Prompt management ([Preview](#))
 - Prompt flows ([Preview](#))
 - Assessment & deployment**
 - Model Evaluation
 - Model access** (highlighted with a red arrow)
 - Bedrock Studio** ([Preview](#))
 - Settings**
 - User guide** (with a small icon)
- Main Content Area:**
 - Machine Learning**
 - # Amazon Bedrock

The easiest way to build and scale generative AI applications with foundation models (FMs)
 - Try Bedrock** button with "Get started" link
 - ## Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.
 - ## Benefits

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock > Model access

What is Model access?

To use Bedrock, account users with the correct IAM Permissions must enable access to available Bedrock foundation models (FMs). View all Bedrock Model Terms for Bedrock FMs.

Enable all models **Enable specific models**

Visit Amazon Bedrock Quotas for a quick guide to the default quotas and limits that apply to Amazon Bedrock.

Base models (33)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

Find model Group by provider

Models	Access status	Modality	EULA
AI21 Labs (2)	0/2 access granted		
Jurassic-2 Ultra	Available to request	Text	EULA
Jurassic-2 Mid	Available to request	Text	EULA
Amazon (6)	0/6 access granted		
Titan Embeddings G1 - Text	Available to request	Embedding	EULA

Chat Text Image

Guardrails Watermark detection

Knowledge bases Agents

Prompt management Preview





- Step 1
 Edit model access
Step 2
 Review and submit

Edit model access

Base models (33/33)

[Collapse all](#)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

Find model

Group by provider ▾

<input checked="" type="checkbox"/>	Models	Access status	Modality	EULA
<input checked="" type="checkbox"/>	▼ AI21 Labs (2)	0/2 access granted		
<input checked="" type="checkbox"/>	Jurassic-2 Ultra	Available to request	Text	EULA
<input checked="" type="checkbox"/>	Jurassic-2 Mid	Available to request	Text	EULA
<input checked="" type="checkbox"/>	▼ Amazon (6)	0/6 access granted		
<input checked="" type="checkbox"/>	Titan Embeddings G1 - Text	Available to request	Embedding	EULA
<input checked="" type="checkbox"/>	Titan Text G1 - Lite	Available to request	Text	EULA
<input checked="" type="checkbox"/>	Titan Text G1 - Express	Available to request	Text	EULA
<input checked="" type="checkbox"/>	Titan Image Generator G1	Available to request	Image	EULA
<input checked="" type="checkbox"/>	Titan Multimodal Embeddings G1	Available to request	Embedding	EULA
<input checked="" type="checkbox"/>	Titan Text Embeddings V2	Available to request	Embedding	EULA
<input checked="" type="checkbox"/>	▼ Anthropic (5)	0/5 access granted		



us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Llama 3.1 40SB Instruct Available to request Text EULA

Llama 3.1 70B Instruct Available to request Text EULA

Llama 3.1 8B Instruct Available to request Text EULA

Llama 3 8B Instruct Available to request Text EULA

Llama 3 70B Instruct Available to request Text EULA

Llama 2 Chat 13B Unavailable Text EULA

Llama 2 Chat 70B Unavailable Text EULA

Llama 2 13B Unavailable Text EULA

Llama 2 70B Unavailable Text EULA

Mistral AI (4) 0/4 access granted

Mistral Large (2407) Available to request Text EULA

Mistral 7B Instruct Available to request Text EULA

Mixtral 8x7B Instruct Available to request Text EULA

Mistral Large (2402) Available to request Text EULA

Stability AI (1) 0/1 access granted

SDXL 1.0 Available to request Image EULA

Cancel Next

Model Name	Status	Type	EULA
Llama 3.1 40SB Instruct	Available to request	Text	EULA
Llama 3.1 70B Instruct	Available to request	Text	EULA
Llama 3.1 8B Instruct	Available to request	Text	EULA
Llama 3 8B Instruct	Available to request	Text	EULA
Llama 3 70B Instruct	Available to request	Text	EULA
Llama 2 Chat 13B	Unavailable	Text	EULA
Llama 2 Chat 70B	Unavailable	Text	EULA
Llama 2 13B	Unavailable	Text	EULA
Llama 2 70B	Unavailable	Text	EULA
Mistral AI (4)	0/4 access granted		
Mistral Large (2407)	Available to request	Text	EULA
Mistral 7B Instruct	Available to request	Text	EULA
Mixtral 8x7B Instruct	Available to request	Text	EULA
Mistral Large (2402)	Available to request	Text	EULA
Stability AI (1)	0/1 access granted		
SDXL 1.0	Available to request	Image	EULA

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock > Model access > Request model access

Step 1
Edit model access

Step 2
Review and submit

Review and submit

Step 1: Edit model access

Edit

Model access modifications (33)

Models	Modifications
Mistral Large (2407)	Request access
Jurassic-2 Ultra	Request access
Jurassic-2 Mid	Request access
Claude 3 Opus	Request access
Claude 3 Sonnet	Request access
Claude 3 Haiku	Request access
Claude	Request access
Claude Instant	Request access
SDXL 1.0	Request access
Command R+	Request access



us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Model access modifications (33)

Models	Modifications
Llama 5 6B Instruct	Request access
Llama 3 70B Instruct	Request access
Llama 2 Chat 13B	Request access
Llama 2 Chat 70B	Request access
Llama 2 13B	Request access
Llama 2 70B	Request access
Mistral 7B Instruct	Request access
Mixtral 8x7B Instruct	Request access
Mistral Large (2402)	Request access

Terms
By selecting Submit, 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 agreements with AWS governing your use of such services.

Cancel Previous Submit

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Model access modifications (33)

Models	Modifications
Llama 5 6B Instruct	Request access
Llama 3 70B Instruct	Request access
Llama 2 Chat 13B	Request access
Llama 2 Chat 70B	Request access
Llama 2 13B	Request access
Llama 2 70B	Request access
Mistral 7B Instruct	Request access
Mixtral 8x7B Instruct	Request access
Mistral Large (2402)	Request access

Terms
By selecting Submit, 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 agreements with AWS governing your use of such services.

Cancel Previous Submit

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents

Prompt management [Preview](#)

Access request for 7 models failed

- Claude 3 Opus - Unauthorized to perform action due to private marketplace eligibility
- Command R+ - Unauthorized to perform action due to private marketplace eligibility
- Command R - Unauthorized to perform action due to private marketplace eligibility
- Llama 2 13B - Could not create agreement - Agreement already exists
- Llama 2 70B - Could not create agreement - Agreement already exists
- Llama 2 Chat 13B - Failed to create regional entitlement. Model not available at the moment. Try again later.
- Llama 2 Chat 70B - Failed to create regional entitlement. Model not available at the moment. Try again later.

Notifications (1) (0) (0) (1) (0) ▾

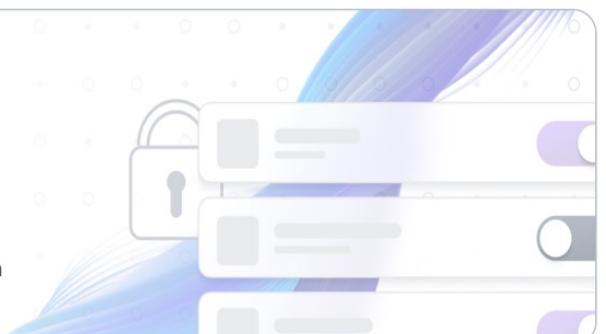
Amazon Bedrock > Model access

What is Model access?

To use Bedrock, account users with the correct [IAM Permissions](#) must enable access to available Bedrock foundation models (FMs). View all [Bedrock Model Terms](#) for Bedrock FMs.

[Modify model access](#)

Visit [Amazon Bedrock Quotas](#) for a quick guide to the default quotas and limits that apply to Amazon Bedrock.



Base models (33)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

[Collapse all](#)

[Group by provider](#)

Models Access status Modality EULA

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/modelaccess

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Model access updates submitted

It may take several minutes to receive or remove access to models. Meanwhile, you can use other Bedrock console pages. Once your access is updated, you can use the models in Bedrock. Refresh the base models table to view the updated statuses.

Amazon Bedrock > Model access

What is Model access?

To use Bedrock, account users with the correct [IAM Permissions](#) must enable access to available Bedrock foundation models (FMs). View all [Bedrock Model Terms](#) for Bedrock FMs.

[Modify model access](#)

Visit [Amazon Bedrock Quotas](#) for a quick guide to the default quotas and limits that apply to Amazon Bedrock.

Base models (33)

Not seeing a model you're interested in? Check out all supported models by region [here](#).

[Find model](#) [Collapse all](#)

[Group by provider](#)

Models	Access status	Modality	EULA
▼ AI21 Labs (2)	0/2 access granted		
Jurassic-2 Ultra	In Progress	Text	EULA
Jurassic-2 Mid	In Progress	Text	EULA

<https://ppt.cc/f23sWx>



© 2024, Amazon Web Services, Inc. or its affiliates.

Screenshot of a web browser showing the Amazon Bedrock Workshop (Chinese Version) page. The URL is catalog.us-east-1.prod.workshops.aws/workshops/17879811-bd5c-4530-8b85-f0042472f2a1/zh-TW/corefeatures/frequently/txt2txt/translation.

Amazon Bedrock Workshop (Chinese Version)

實驗環境設定

[NEW]最新特性

核心功能實驗

常見場景

文本生成

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

圖片解析

文件解析

文生圖、圖生圖

Artifacts

Content preferences

Language

Amazon Bedrock Workshop (Chinese Version) > 核心功能實驗 > 常見場景 > 文本生成 > 翻譯

翻譯

下列實驗演示在 Amazon Bedrock playground 上使用 Claude 3 進行文字翻譯：

控制台及模型選擇

- 打開 Amazon Bedrock 控制台，左側選單選擇 Playgrounds -> Text

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models

Playgrounds

- Chat
- Text
- Image

Amazon Bedrock > Text playground

Text playground

Select model

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Text playground

Text playground Info Load examples

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text** ←
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Select model

Configurations

Reset

Select model to load configs.

Try one of these examples [or view more examples](#)

Titan Text G1 - Express Action items from a meeting transcript	Claude Advanced Q&A with Citation An example	Llama 2 Chat 13B Chain of thought An example	Command Contract Entity Extraction Use generative	Jurasic-2 Ultra Earnings call summarization A prompt that	Mistral Large (2402) Finding the Difference in Payment
---	--	--	---	---	--

Run

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground

Services | Search [Option+S] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

- AI21 Labs**
- Amazon
- Anthropic
- Cohere
- Meta
- Mistral AI

2. Model

Models with access (2)

- Jurassic-2 Mid**
Text model | Context size = 8k
- Jurassic-2 Ultra**
Text model | Context size = 8k

3. Throughput

Select model to show throughput options.

Load examples

Cancel Apply

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Select model

1. Category

Model providers

- AI21 Labs
- Amazon
- Anthropic**
- Cohere
- Meta
- Mistral AI

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

Models without access (1) [Request access](#)

- Claude 3 Opus v1**
Text & vision model | Context size = up to 200k

3. Throughput

Select model to show throughput options.

Load examples

Cancel Apply

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

- AI21 Labs AI21 Labs
- Amazon
- Anthropic**
- Cohere
- Meta
- Mistral AI

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

Models without access (1) Request access

- Claude 3 Opus v1**
Text & vision model | Context size = up to 200k

3. Throughput

On-demand (ODT)

Load examples

Cancel Apply

Load configs.

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Text playground

Text playground Info

Load examples

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models Preview

Playgrounds

- Chat
- Text**
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management Preview

Claude 3 Sonnet v1 | ODT

Change

Write a prompt...

Try one of these examples [or view more examples](#)

Advanced Q&A with Citations
An example prompt for long document q&a supplemented

Craft a Design Brief
Craft a design brief for a holistic brand identity,

Meeting Notes Summarizer
Distill meetings into concise summaries including

Configurations Reset

Randomness and diversity

- Temperature: 1
- Top P: 0.999
- Top K: 250

Length

- Maximum length: 2000
- Stop sequences: Human: Add

Amazon Bedrock
Workshop (Chinese
Version)

測試翻譯情境

- 我們以一個字幕翻譯情境為例，在 playground 中輸入以下提示詞 prompt:

```
1 你是一個幫助翻譯劇本的助理。  
2 你的任務是將<text>中的英文原文翻譯成繁體中文。翻譯時，請遵守以下規則：  
3 0.不要改變原意。  
4 1.翻譯前先瞭解上下文，保持語義連貫、閱讀流暢，但不要故意誇張。  
5 2.原文大多是對話式的，因此翻譯仍應符合短影音/影音部落格/Youtube影片的上下文環境。注意避免使用平常不會在日常聊天中出現的字詞。  
6 3.適當的時候保留一些專有名詞或專業術語未翻譯，注意前後一致性。  
7 4.標點符號必須使用全形。例如，不可以使用小寫的 "，"，必須使用 "，"。  
8 5.在<result></result> 中回覆翻譯。不要包含任何額外的內容。  
9 <example>  
10 H: Welcome to the Amazon Bedrock Workshop, this workshop will help you quickly get started on your journey to generative AI application.  
11 A: 歡迎來到 Amazon Bedrock Workshop，透過此 Workshop 將協助您快速開始生成式 AI 應用程式的旅程。  
12 </example>  
13 <text>  
14 "We'll cover all of those things in a moment, but before we get started, this video doesn't have a sponsor, but it is supported by the  
15 </text>
```

Copied!



實驗環境設定

[NEW]最新特性

核心功能實驗

常見場景

文本生成

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

圖片解析

文件解析

文生圖、圖生圖

Artifacts

Content preferences

Language

中文(繁體) ▾

你是一個幫助翻譯劇本的助理。

你的任務是將<text>中的英文原文翻譯成繁體中文。翻譯時，請遵守以下規則：

0.不要改變原意。

1.翻譯前先瞭解上下文，保持語義連貫、閱讀流暢，但不要故意誇張。

2.原文大多是對話式的，因此翻譯仍應符合短影音/影音部落格/Youtube影片的上下文環境。注意避免使用平常不會在日常聊天中出現的字詞。

3.適當的時候保留一些專有名詞或專業術語未翻譯，注意前後一致性。

4.標點符號必須使用全形。例如，不可以使用小寫的 "，"，必須使用 "，"。

5.在<result></result> 中回覆翻譯。不要包含任何額外的內容。

<example>

H: Welcome to the Amazon Bedrock Workshop, this workshop will help you quickly get started on your journey to generative AI applications.

A: 歡迎來到 Amazon Bedrock Workshop，透過此 Workshop 將協助您快速開始生成式 AI 應用程式的旅程。

</example>

<text>

▼ Randomness and diversity Info

Temperature	1
Top P	0.999
Top K	250

▼ Length Info

Maximum length	2000
----------------	------

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Text playground

Text playground Info Load examples :

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

All Claude 3 Sonnet v1 | ODT Change

你是一個幫助翻譯劇本的助理。

你的任務是將<text>中的英文原文翻譯成繁體中文。翻譯時，請遵守以下規則：

0.不要改變原意。

1.翻譯前先瞭解上下文，保持語義連貫、閱讀流暢，但不要故意誇張。

2.原文大多是對話式的，因此翻譯仍應符合短影音/影音部落格/Youtube影片的上下文環境。注意避免使用平常不會在日常聊天中出現的字詞。

3.適當的時候保留一些專有名詞或專業術語未翻譯，注意前後一致性。

4.標點符號必須使用全形。例如，不可以使用小寫的 "，"，必須使用 "，"。

5.在<result></result> 中回覆翻譯。不要包含任何額外的內容。

<example>

H: Welcome to the Amazon Bedrock Workshop, this workshop will help you quickly get started on your journey to generative AI applications.

A: 歡迎來到 Amazon Bedrock Workshop，透過此 Workshop 將協助您快速開始生成式 AI 應用程式的旅程。

</example>

<text>

"We'll cover all of those things in a moment, but before we get started, this video doesn't have a sponsor, but it is supported by the thousands of you wonderful people who get value out of all of my courses, prints, presets and ebooks over at patk.com."

</text>

|

Try one of these examples [or view more examples](#)

Configurations X

Reset

Randomness and diversity Info

- Temperature 1
- Top P 0.999
- Top K 250

Length Info

- Maximum length 2000
- Stop sequences Add
- Human:

Guardrail

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

In the日常聊天中出現的字詞。

3.適當的時候保留一些專有名詞或專業術語未翻譯，注意前後一致性。

4.標點符號必須使用全形。例如，不可以使用小寫的 "，"，必須使用 "，"。

5.在<result></result>中回覆翻譯。不要包含任何額外的內容。

<example>

H: Welcome to the Amazon Bedrock Workshop, this workshop will help you quickly get started on your journey to generative AI applications.

A: 歡迎來到 Amazon Bedrock Workshop，透過此 Workshop 將協助您快速開始生成式 AI 應用程式的旅程。

</example>

<text>

"We'll cover all of those things in a moment, but before we get started, this video doesn't have a sponsor, but it is supported by the thousands of you wonderful people who get value out of all of my courses, prints, presets and ebooks over at patk.com."

</text>

Try one of these examples [or view more examples](#)

AI Claude 3 Sonnet
Advanced Q&A with Citations
An example prompt for long document q&a supplemented

AI Claude 3 Sonnet
Craft a Design Brief
Craft a design brief for a holistic brand identity,

AI Claude 3 Sonnet
Meeting Notes Summarizer
Distill meetings into concise summaries including

Run

Top K 250

Length

- Maximum length 2000
- Stop sequences Add
- Human:

Guardrail

Manage guardrails



us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Text playground

Load examples

Claude 3 Sonnet v1 | ODT

Change

在日常聊天中出現的字詞。
3.適當的時候保留一些專有名詞或專業術語未翻譯，注意前後一致性。
4.標點符號必須使用全形。例如，不可以使用小寫的 "，"，必須使用 "，"。
5.在<result></result> 中回覆翻譯。不要包含任何額外的內容。
<example>
H: Welcome to the Amazon Bedrock Workshop, this workshop will help you quickly get started on your journey to generative AI applications.
A: 歡迎來到 Amazon Bedrock Workshop，透過此 Workshop 將協助您快速開始生成式 AI 應用程式的旅程。
</example>
<text>
"We'll cover all of those things in a moment, but before we get started, this video doesn't have a sponsor, but it is supported by the thousands of you wonderful people who get value out of all of my courses, prints, presets and ebooks over at patk.com."
</text>

<result> 「我們稍後會涵蓋所有這些內容，但在開始之前，這段影片並沒有贊助商，不過它獲得來自成千上萬位從我在 patk.com 上的各種課程、美工圖案、預設和電子書中獲益良多的出色觀眾支持。」 </result>

Run

Configurations

Reset

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Length

Maximum length: 2000

Stop sequences: Human: X

Add

Guardrail



Amazon Bedrock Workshop (Chinese Version)

實驗環境設定

[NEW]最新特性

核心功能實驗

常見場景

文本生成

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

圖片解析

文件解析

文生圖、圖生圖

Artifacts

Content preferences

Language

中文(繁體) ▾

測試 RAG 問答

- 我們以通過 API 呼叫召回的 Amazon Data Firehose (前 Amazon Kinesis Data Firehose) 原始 json 作為 context 輸入，查詢 Amazon Data Firehose 在美西2的資料攝入價格
- 構造的 prompt 如下所示

```
1 您是 Amazon Web Services 的助手，在 <query> 標籤中回應用戶的問題，基於 <api_response> 中獲得的基礎資訊。
2
3 用戶的查詢是
4 <query>
5 Amazon Kinesis Firehose 在美西2的資料攝入價格是怎麼樣的？
6 </query>
7
8 <api_response>
9 { "product": { "productFamily": "Kinesis Firehose", "attributes": { "regionCode": "us-west-2", "servicecode": "AmazonKin
10 ... (其他API回應資訊)
11 </api_response>
12
13 請將您的答案放在 <response> 標籤之間，並遵循以下要求：
14 1.用問題的原始語言回答。
15 2.若消息中並沒有相關資訊，你們應該回覆用戶，你們並沒有透過原問題找到任何資訊，不要說其他的。
16 3.若提供的建議不為空，建議用戶參照建議進行詢問。如果沒有任何建議是空的，就不要再說其他的了。
17 4.不要以"API"這樣的短語開頭，跳過序言，直接進入答案。
```

- 點擊 Run，查看 RAG 回答效果

```
用戶的查詢是
<query>
Amazon Kinesis Firehose在美西2的資料攝入價格是怎麼樣的?
</query>

<api_response>
{"product": {"productFamily": "Kinesis Firehose", "attributes": {"regionCode": "us-west-2", "servicecode": "AmazonKinesisFirehose", "usageType": "USW2-VendedBilledBytes", "description": "Per GB of data ingested"}}
```

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Text playground

Load examples

Claude 3 Sonnet v1 | ODT

Change

您是 Amazon Web Services 的助手，在 <query> 標籤中回應用戶的問題，基於 <api_response> 中獲得的基礎資訊。

用戶的查詢是

```
<query>
Amazon Kinesis Firehose 在美西2的資料攝入價格是怎麼樣的？
</query>
```

<api_response>

```
{ "product": { "productFamily": "Kinesis Firehose", "attributes": { "regionCode": "us-west-2", "servicecode": "AmazonKinesisFirehose", "usagetype": "USW2-VendedLogsBilledBytes", "description": "Per GB of data ingested", "locationType": "AWS Region", "location": "US West (Oregon)", "servicename": "Amazon Kinesis Firehose", "sourcetype": "Vended Logs", "operation": "PutRecordBatch", "group": "Batch Processing" }, "sku": "4MS4CCZDG9XWYXCM", "term": { "OnDemand": { "4MS4CCZDG9XWYXCM.JRTCKXETXF": { "priceDimensions": { "4MS4CCZDG9XWYXCM.JRTCKXETXF.PERGK9UASF": { "unit": "GB", "endRange": "Inf", "description": "Tier 3 $0.11 per GB of data ingested", "appliesTo": [], "rateCode": "" } } } } }
```

Configurations

Reset

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Length

Maximum length: 2000

Stop sequences: Human: X

Guardrail

Run

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Text playground

Load examples

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

A Claude 3 Sonnet v1 | ODT

Change

</api_response>

請將您的答案放在 <response> 標籤之間，並遵循以下要求：

- 用問題的原始語言回答。
- 若消息中並沒有相關資訊，你們應該回覆用戶，你們並沒有透過原問題找到任何資訊，不要說其他的。
- 若提供的建議不為空，建議用戶參照建議進行詢問。如果沒有任何建議是空的，就不要再說其他的了。
- 不要以"API"這樣的短語開頭，跳過序言，直接進入答案。

<response>

Amazon Kinesis Firehose 在美西2區域 (奧勒岡) 的資料攝入價格採用階梯式方案，每GB數據分3個價格等級定價：

第1級 (0-512GB): \$0.16/GB
第2級 (512GB-2048GB): \$0.13/GB
第3級 (超過2048GB): \$0.11/GB

因此，數據攝入服務成本將取決於您的使用量落在哪個價格等級範圍內。

</response>

Run

Configurations

Reset

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Length

Maximum length: 2000

Stop sequences: Human: X

Add

Guardrail

Amazon Bedrock
Workshop (Chinese
Version)

測試 VOC 案例分類

- 我們以一個客服問題分類為例，測試 Claude 3 的標籤分類準確性
- 可以在這裡下載原始的客服問題和對應的分類標籤，[測試資料](#)
- 其中第一欄為客服問題，第二欄為分類標籤值
- 在 playground 中輸入以下提示詞 prompt

實驗環境設定

[NEW]最新特性

核心功能實驗

常見場景

文本生成

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

圖片解析

文件解析

文生圖、圖生圖

Artifacts

Content preferences

Language

中文(繁體)



```
1 您是一個客服代表，需要根據分類要求<instructions>對客戶的描述<description>對客服問題進行分類。  
2  
3 <description>中是問題描述：  
4 <description>  
5 某先生想諮詢一下想要購買束腹帶的問題  
6 </description>  
7  
8 <instructions>中是分類要求：  
9 <instructions>  
10 1， 將上方<description>中描述的客服問題分類到以下8個類別中  
11 備件/商務諮詢： 提到物品價格,遷移,流程等與商務有關的問題。  
12 床： 提到床無法移動,出入不便等與床操作相關的問題。  
13 掃描拱架： 提到gantry, revolution等與掃描拱架相關的關鍵詞。  
14 掃描問題： 提到無法獲得圖像,出現偽影,硬體錯誤等直接與掃描品質相關的問題。  
15 偵測器： 提到偵測器溫度,指針移動產生假影等與偵測器相關的問題。  
16 操作台： 提到開機失敗,工作站不能使用,傳不進PACS等與操作台相關的問題。  
17 球管/高壓： 提到球管故障,報錯,異響等與球管和高壓相關的問題。  
18 無法判斷： 對於一些語句資訊不足無法判斷分類的問題。  
19 2，每個問題只選擇一個最適合的類別  
20 3，如果無法判斷類別，則選擇“無法判斷”  
21 4，明確定義分類中使用的關鍵詞，例如掃描儀代表整台設備，掃描拱架特指gantry部分。  
22 5，注意問題描述的整體語意，不要過度依賴某些關鍵詞進行分類。  
23 6，進一步明確各分類類別的區分標準，避免分類混淆。  
24 7，對資訊不足無法判斷類別的問題，不要強行分類，選擇“無法判斷”。  
25 8，對於含有多个關鍵詞的問題，選擇最相關的一個類別。
```

Copied!



us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S] | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Text playground

Load examples

Claude 3 Sonnet v1 | ODT

Change

您是一個客服代表,需要根據分類要求<instructions>對客戶的描述<description>對客服問題進行分類。

<description>中是問題描述：

<description>

某先生想諮詢一下想要購買束腹帶的問題

</description>

<instructions>中是分類要求：

<instructions>

1， 將上方<description>中描述的客服問題分類到以下8個類別中

備件/商務諮詢: 提到物品價格,遷移,流程等與商務有關的問題。

床: 提到床無法移動,出入不便等與床操作相關的問題。

掃描拱架: 提到gantry, revolution等與掃描拱架相關的關鍵詞。

掃描問題: 提到無法獲得圖像,出現偽影,硬體錯誤等直接與掃描品質相關的問題。

偵測器: 提到偵測器溫度,指針移動產生假影等與偵測器相關的問題。

操作台: 提到開機失敗,工作站不能使用,傳不進PACS等與操作台相關的問題。

球管/高壓: 提到球管故障,報錯,異響等與球管和高壓相關的問題。

Run

Configurations

Reset

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Length

Maximum length: 2000

Stop sequences: Human: X

Guardrail

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/text-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Text playground

Text playground Info Load examples

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text**
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

All Claude 3 Sonnet v1 | ODT Change

4 , 明確定義分類中使用的關鍵詞,例如掃描儀代表整台設備,掃描拱架特指gantry部分。

5 , 注意問題描述的整體語意,不要過度依賴某些關鍵詞進行分類。

6 , 進一步明確各分類類別的區分標準,避免分類混淆。

7 , 對資訊不足無法判斷類別的問題,不要強行分類,選擇"無法判斷"。

8 , 對包含多個類別關鍵詞的問題描述,選擇最相關的一個類別進行分類。

9 , 必須嚴格用<example>中給出的範例格式回覆,輸出結果中必須包含:<answer></answer>和<reasoning></reasoning> :

<example>

<answer> :

操作台

</answer>

<reasoning> :

根據分類規則工作站不能使用,死機,黑畫面屬於操作台故障分類。 ←

</reasoning> :

</example>

<instructions>

Run

Configurations

Randomness and diversity

- Temperature 1
- Top P 0.999
- Top K 250

Length

- Maximum length 2000
- Stop sequences
- Human: X

Load examples

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Text playground

Load examples

Claude 3 Sonnet v1 | ODT

Reasoning

根據分類規則工作站不能使用，死機，黑畫面屬於操作台故障分類。

</reasoning> :

</example>

<instructions>

<answer>備件/商務諮詢</answer>

<reasoning>

根據問題描述"某先生想諮詢一下想要購買束腹帶的問題"，可以判斷是關於商品購買的商務諮詢問題。比較合適的分類是"備件/商務諮詢"類別。

這個類別描述為"提到物品價格,遷移,流程等與商務有關的問題"。此問題屬於諮詢購買物品的情況,符合商務諮詢的範疇。

雖然"束腹帶"是一種附件產品而不是主機的備件,但由於也涉及了物品購買的商務流程,因此分類到"備件/商務諮詢"類別是比較合適的。其他如床、球管等分類顯然與此問題描述不太相關。

</reasoning>

Run

Configurations

Reset

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Length

Maximum length: 2000

Stop sequences: Human: X

Guardrail

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

天津工廠

翻譯
事實問答
小說續寫
角色扮演
RAG 場景
程式碼輔助
客服案例分類
文字內容審核

▼ 圖片解析

IPC 圖片分析
物品辨識和計數
多模態能力
手寫內容識別問答/OCR
圖像理解
監控圖表理解
架構圖理解
時序圖片分析
截圖產生程式碼
視覺提示詞
從設計到程式碼
瑕疵偵測

▼ Content preferences

Language

中文(繁體)

演示1: 图表理解

System Prompt

- 1 你是一个分析图表的专家，特别是图表里的柱状图，折线图，或者混合折线和柱状图在一起的图表。
- 2 请首先给出趋势或概览分析，之后仔细确认每一个折线的每一个波峰和每一个波谷对应的横轴和纵轴的数字（纵轴可能有两个）并列出。使用中文回复。

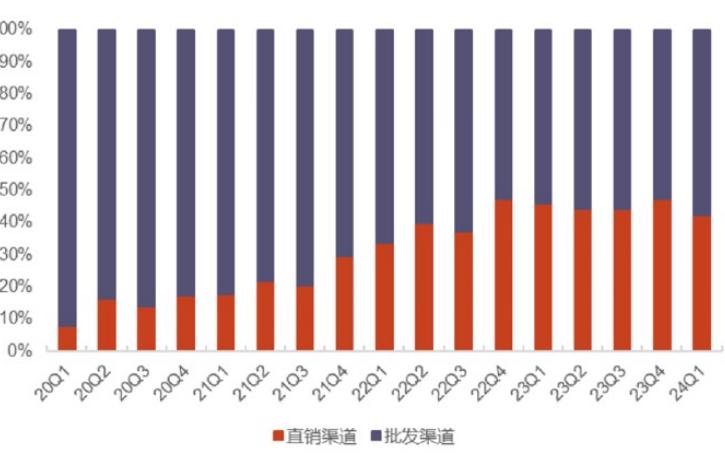


用户输入

图片1



图片2



这两张图是茅台公司酒产品20Q1-24Q1的统计图，请仔细分析并给出你的分析结果：

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground

Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models [Preview](#)

Playgrounds Chat  Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management [Preview](#)

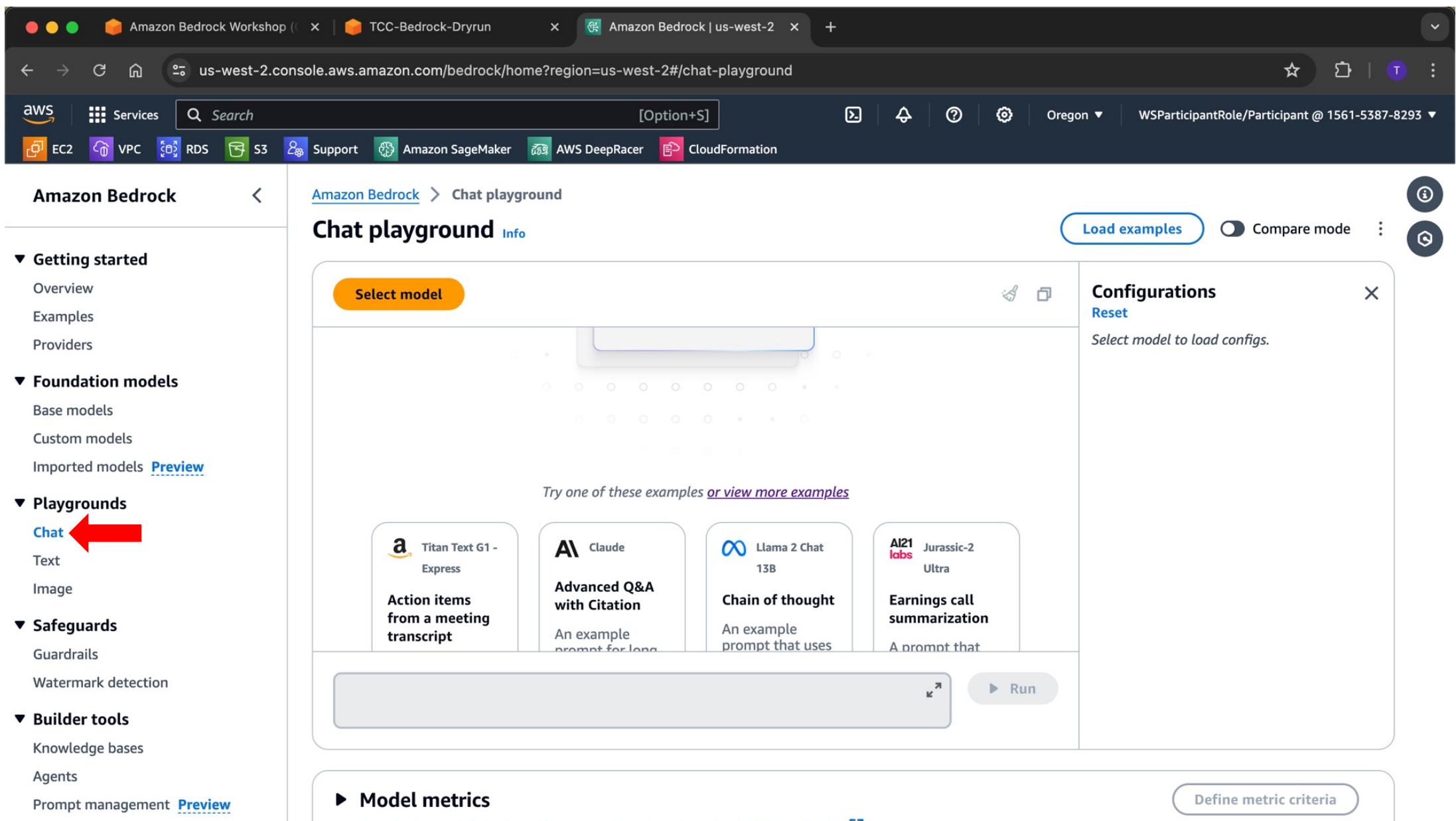
Select model

Try one of these examples [or view more examples](#)

 Titan Text G1 - Express Action items from a meeting transcript	 Claude Advanced Q&A with Citation An example prompt for long	 Llama 2 Chat 13B Chain of thought An example prompt that uses	 AI21 labs Jurassic-2 Ultra Earnings call summarization A prompt that
--	---	--	---

Run

Model metrics Define metric criteria



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

- AI21 Labs**
- Amazon
- Anthropic
- Cohere
- Meta
- Mistral AI

2. Model

Models with access (2)

- Jurassic-2 Mid**
Text model | Context size = 8k
- Jurassic-2 Ultra**
Text model | Context size = 8k

3. Throughput

Select model to show throughput options.

Compare mode

Not seeing a model you are interested in? Check out all supported models [here](#)

Cancel Apply Define metric criteria

Amazon Bedrock

Getting started

Overview
Examples
Providers

Foundation models

Base models
Custom models
Imported models [Preview](#)

Playgrounds

Chat
Text
Image

Safeguards

Guardrails
Watermark detection

Builder tools

Knowledge bases
Agents
Prompt management [Preview](#)
Prompt flows [Preview](#)

Assessment & deployment

CloudShell Feedback

Select model

1. Category

Model providers

AI21 Labs

Amazon

Anthropic

Cohere

Meta

Mistral AI

2. Model

Models with access (5)

Claude Instant 1.2 v1.2

Text model | Context size = up to 100k

Claude 2.1 v2.1

Text model | Context size = up to 200k

Claude 2 v2

Text model | Context size = up to 100k

Claude 3 Sonnet v1

Text & vision model | Context size = up to 200k

Claude 3 Haiku v1

Text & vision model | Context size = up to 200k

Models without access (1)

[Request access](#)

Claude 3 Opus v1

Text & vision model | Context size = up to 200k

3. Throughput

Select model to show throughput options.

[Cancel](#)

[Apply](#)

Defin

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS 53 Support Amazon SageMaker AWS DeepRacer CloudFormation

Select model

1. Category

Model providers

- AI21 labs AI21 Labs
- a Amazon
- AI Anthropic**
- Cohere
- Meta
- Mistral AI

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

Models without access (1) Request access ↗

- Claude 3 Opus v1**
Text & vision model | Context size = up to 200k

3. Throughput

On-demand (ODT)

Compare mode

Define metric criteria

Cancel Apply

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview Prompt flows Preview

Assessment & deployment

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview

Claude 3 Sonnet v1 | ODT

Try one of these examples or view more examples

Advanced Q&A with Citations
An example prompt for long document Q&A.

Craft a Design Brief
Craft a design brief for a holistic brand identity.

Meeting Notes Summarizer
Distill meetings into concise summaries.

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) ▶ Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Model metrics Define metric criteria

Configurations Reset

System prompts Add system prompts

Randomness and diversity Temperature 1

Top P 0.999

Top K 250

天津工廠

翻譯
事實問答
小說續寫
角色扮演
RAG 場景
程式碼輔助
客服案例分類
文字內容審核

▼ 圖片解析

IPC 圖片分析
物品辨識和計數
多模態能力
手寫內容識別問答/OCR
圖像理解
監控圖表理解
架構圖理解
時序圖片分析
截圖產生程式碼
視覺提示詞
從設計到程式碼
瑕疵偵測

▼ Content preferences

Language

中文(繁體)

演示1: 图表理解

System Prompt

- 1 你是一个分析图表的专家，特别是图表里的柱状图，折线图，或者混合折线和柱状图在一起的图表。
- 2 请首先给出趋势或概览分析，之后仔细确认每一个折线的每一个波峰和每一个波谷对应的横轴和纵轴的数字（纵轴可能有两个）并列出。使用中文回复。

Copied!

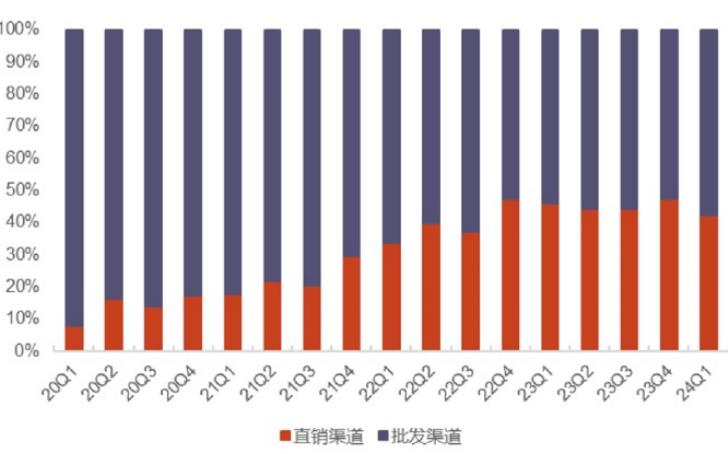


用户输入

图片1



图片2



这两张图是茅台公司酒产品20Q1-24Q1的统计图，请仔细分析并给出你的分析结果：

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Chat playground

Chat playground Info Load examples Compare mode

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

A Claude 3 Sonnet v1 | ODT Change

Configurations

System prompts

Add system prompts

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Model metrics Define metric criteria

你是分析图表的专家，特别是图表里的柱状图，折线图，或者混合折线和柱状图在一起的图表。
请首先给出趋势或概览分析，之后仔细确认每一个折线的每一个波峰和每一个波谷对应的横轴和纵轴的数字（纵轴可能有两个）并列出。使用中文回复。

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Model metrics Define metric criteria

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Chat playground

Chat playground

Load examples Compare mode

Getting started Overview Examples Providers Foundations Base models Custom imports Importers Playgrounds Chat Text Image Locations Network Media Photos

Favorites mba Applications Downloads On My Mac

iCloud iCloud Drive Desktop Shared

chart-1_1.png Today at 3:23 PM 499 KB PNG image
chart-1_2.png Today at 3:23 PM 340 KB PNG image

Show Options Cancel Open

Choose files Top K 250

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Model metrics Define metric criteria

To evaluate models for task specific metrics with custom dataset visit Model evaluation

Metrics Claude 3 Sonnet

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services: EC2, VPC, RDS, S3, Support, Amazon SageMaker, AWS DeepRacer, CloudFormation

Search: [Option+S]

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock > Chat playground

Chat playground Info

Load examples | Compare mode

Claude 3 Sonnet v1 | ODT | Change

Configurations Reset

System prompts Info

Add system prompts

Run 

Randomness and diversity Info

Temperature: 1

Top P: 0.999

Top K: 250

Model metrics Define metric criteria

Getting started:

- Overview
- Examples
- Providers

Foundation models:

- Base models
- Custom models
- Imported models Preview

Playgrounds:

- Chat
- Text
- Image

Safeguards:

- Guardrails
- Watermark detection

Builder tools:

- Knowledge bases
- Agents
- Prompt management Preview

Model description: 你是一个分析图表的专家，特别是图表里的柱状图，折线图，或者混合折线和柱状图在一起的图表。请首先给出趋势或概览分析，之后仔细确认每一个折线的每一个波峰和每一个波谷对应的横轴和纵轴的数字（纵轴可能有两个）并列出。使用中文回复。

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

chart-1_1.png 498.55 KB

chart-1_2.png 340.15 KB

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Chat playground

Chat playground Info

Load examples Compare mode ⋮

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models Preview

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management Preview

A Claude 3 Sonnet v1 | ODT Change

从图表整体趋势来看,酒收入和同比变化呈现规律性的波动。在某些季节,如2009年第一季度、2011年第一季度和2013年第一季度,酒收入和同比增长都达到峰值。而在其他季节,如2010年第三季度和2012年第三季度,酒收入和同比增长则处于低谷。此外,折线图显示了百分比格式的直接渠道和批发渠道收入占比,两条折线保持相对稳定,直接渠道收入占比略高于批发渠道。

现在我详细列出每一个波峰和波谷对应的数值:

酒收入(亿元)波峰:
2009年第1季度:158.62
2011年第1季度:264.11
2013年第1季度:392.02

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Configurations Reset

System prompts

Add system prompts

Randomness and diversity Info

- Temperature: 1
- Top P: 0.999
- Top K: 250

Model metrics Define metric criteria

天津工场

翻譯
事實問答
小說續寫
角色扮演
RAG 場景
程式碼輔助
客服案例分類
文字內容審核

▼ 圖片解析
IPC 圖片分析
物品辨識和計數
多模態能力 
手寫內容識別問答/OCR
圖像理解
監控圖表理解
架構圖理解
時序圖片分析
截圖產生程式碼
視覺提示詞
從設計到程式碼
瑕疵偵測

▼ Content preferences

Language

中文(繁體) 

演示4: 商品詳情標簽

用戶輸入

Claude 3.5 Sonnet 結果



這是一款名為"Bear小熊"的多士爐(烤面包机)的廣告圖片。以下是对該產品的詳細描述：

1. 型號：DSL-C02K8
2. 主要賣點：
 - 52秒快速烤制
 - 6档口味选择，满足不同需求
 - 一按一按，操作简单，适合快速早餐
3. 外觀：
 - 颜色：米白色
 - 设计：简约现代风格
 - 配件：包含一个面包夹
4. 功能：
 - 双槽设计，可同时烤制两片面包
 - 可调节烘烤程度
5. 广告画面：
 - 展示了正在弹出的金黄色吐司
 - 旁边放置了柠檬和其他配料，暗示多样化的早餐搭配

标签：

#小熊电器 #多士炉 #烤面包机 #快速早餐 #家用电器 #厨房用品 #简约设计 #52秒速烤 #6档口味 #DSL-C02K8

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground

Chat playground Info Load examples Compare mode

Claude 3 Sonnet v1 | ODT Change

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Configurations Reset

System prompts Add system prompts

Randomness and diversity Info

Temperature 1

Top P 0.999

Top K 250

Model metrics Define metric criteria

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Chat playground Chat playground Load examples Compare mode

Favorites: mba, Applications, Downloads, On My Mac

iCloud: iCloud Drive, Documents, Desktop, Shared

Locations: Network

Media: Photos

Show Options Cancel Open

Choose files Top K 250

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Model metrics Define metric criteria

To evaluate models for task specific metrics with custom dataset visit Model evaluation

Knowledge bases Agents Prompt management Preview Prompt flows Preview

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+S] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Chat playground

Chat playground Info

Load examples Compare mode ⋮

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models Preview

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management Preview

A Claude 3 Sonnet v1 | ODT Change

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

product1.png 659.65 KB X

Configurations Reset

System prompts

Add system prompts

Randomness and diversity Info

Temperature 1

Top P 0.999

Top K 250

Model metrics Define metric criteria



文本生成

- 翻譯
- 事實問答
- 小說續寫
- 角色扮演
- RAG 場景
- 程式碼輔助
- 客服案例分類
- 文字內容審核

▼ 圖片解析

- IPC 圖片分析
- 物品辨識和計數
- 多模態能力**
- 手寫內容識別問答/OCR
- 圖像理解
- 監控圖表理解
- 架構圖理解
- 時序圖片分析
- 截圖產生程式碼
- 視覺提示詞
- 從設計到程式碼
- 瑕疵偵測

▼ Content preferences

Language

中文(繁體) ▾



这是一个商品的宣传图，请进行商品的详细描述，并打标签

- 颜色：米白色

- 设计：简约现代风格

- 配件：包含一个面包夹

4. 功能：

- 双槽设计，可同时烤制两片面包

- 可调节烘烤程度

5. 广告画面：

- 展示了正在弹出的金黄色吐司

- 旁边放置了柠檬和其他配料，暗示多样化的早餐搭配

标签：

#小熊电器 #多士炉 #烤面包机 #快速早餐 #家用电器 #厨房用品 #简约设计 #52秒速烤 #6档口味

#DSL-C02K8



Here's a product description for this dress in the requested format:

【Fabric】

Made from soft, lightweight polyester fabric with a vibrant floral print. The material offers breathability and comfort for all-day wear.

【Occasions】

Perfect for summer parties, beach vacations, weddings, date nights, or any casual to semi-formal event where you want to make a stylish statement.

【Features】

- Stunning navy blue maxi dress with a colorful floral pattern in shades of pink, orange, and purple
- Wrap-style V-neckline
- Short sleeves

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview

Model metrics Define metric criteria

AI Claude 3 Sonnet v1 | ODT Change

Configurations Reset System prompts Add system prompts

Randomness and diversity Temperature 1 Top P 0.999 Top K 250

Model metrics Define metric criteria

这是商品的宣传图，请进行商品的详细描述，并打标签

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

product1.png 659.65 KB

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services: EC2, VPC, RDS, S3, Support, Amazon SageMaker, AWS DeepRacer, CloudFormation

Search: [Option+S]

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock > Chat playground

Chat playground Info

Load examples | Compare mode

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models Preview

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management Preview

Model metrics

Define metric criteria

Configurations

Reset

System prompts

Add system prompts

Randomness and diversity

Temperature: 1

Top P: 0.999

Top K: 250

AI Claude 3 Sonnet v1 | ODT

Change

Run

这是商品的宣传图，请进行商品的详细描述，并打标签

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

product1.png
659.65 KB

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services: EC2, VPC, RDS, S3, Support, Amazon SageMaker, AWS DeepRacer, CloudFormation

Search: [Option+5]

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Chat playground

Chat playground Info

Load examples Compare mode ⋮

AI Claude 3 Sonnet v1 | ODT Change

这是张小熊多士炉DSL-C02K8型号的产品宣传图片。图片展示了产品的外观设计和主要功能特点。

产品描述:

1. 造型简洁干净,主色为白色和咖啡色。
2. 可同时放入两片面包片烘烤。
3. 配有拔高烘烤架,可容纳较高的面包片。
4. 具有一拉一按的操作方式,方便使用。
5. 烘烤效果均匀,可做出金黄酥脆的多士。
6. 适合家用,满足早餐烘烤多士的需求。

标签:

#小家电 #多士炉 #烘焙 #厨房电器 #早餐 #面包机

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Configurations Reset

System prompts

Add system prompts

Randomness and diversity Info

Temperature: 1

Top P: 0.999

Top K: 250

Model metrics Define metric criteria

⋮

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models Preview

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management Preview

文本生成

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

▼ 圖片解析

IPC 圖片分析

物品辨識和計數

多模態能力

手寫內容識別問答/OCR

圖像理解

監控圖表理解

架構圖理解

時序圖片分析

截圖產生程式碼

視覺提示詞

從設計到程式碼

瑕疵偵測

▼ Content preferences

Language

中文(繁體)

高鐵票訊息提取

2018/10/24

車次Train 125

單程票

南港
Nangang11:20 → 左營
13:05
Zuoying5 標準廂
car 2C

NT\$1530 現金 P1

02-2-62-0-297-0006 成人

06960009 2018/10/24發行

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview

Claude 3 Sonnet v1 | ODT

Change

Try one of these examples or view more examples

Claude 3 Sonnet
Advanced Q&A with Citations
An example prompt for long document Q&A.

Claude 3 Sonnet
Craft a Design Brief
Craft a design brief for a holistic brand identity.

Claude 3 Sonnet
Meeting Notes Summarizer
Distill meetings into concise summaries.

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Configurations Reset

System prompts Add system prompts

Randomness and diversity Temperature 1

Top P 0.999

Top K 250

Model metrics Define metric criteria

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Chat playground Chat playground Load examples Compare mode

Getting started Overview Examples Providers Foundations Base models Custom imports Importers Playgrounds Chat Text Image Locations Network Media Photos Show Options Cancel Open

Favorites mba Applications Downloads On My Mac

iCloud iCloud Drive Documents Desktop Shared

2.jpeg Today at 3:27 PM 84 KB JPEG image

Choose files The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Top K 250

Model metrics Define metric criteria To evaluate models for task specific metrics with custom dataset visit Model evaluation

Knowledge bases Agents Prompt management Preview Prompt flows Preview

Instances a X | SAP Instances - X | Migrate from X | AWS DMS V X | Migrate you X | Migrate with X | TCC-Bedrock X | Amazon Bedrock X | Amazon Bedrock X | +

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Chat playground Chat playground Info Load examples Compare mode

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

AI Claude 3 Sonnet v1 | ODT Change

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

2.jpeg 83.91 KB X

Configurations

System prompts

Add system prompts

Randomness and diversity

Temperature 1

Top P 0.999

Top K 250

Model metrics Define metric criteria

This screenshot shows the Amazon Bedrock Chat playground interface. The sidebar on the left lists various sections such as Getting started, Foundation models, Playgrounds, Safeguards, and Builder tools. The main workspace displays the Claude 3 Sonnet model (v1). It includes a text input field for prompts, a file upload section for supporting files (with a single JPEG file named '2.jpeg' uploaded), and configuration sliders for Randomness and diversity parameters like Temperature, Top P, and Top K. The interface is designed to facilitate AI interaction and experimentation.

Screenshot of a web browser showing the AWS Workshop Studio interface. The tabs at the top include "Amazon Bedrock Workshop", "TCC-Bedrock-Dryrun", "Amazon Bedrock | us-west-2", and "New Tab". The main content area displays a prompt about extracting information from a high-speed rail ticket, followed by the extracted results and detailed information.

翻譯

事實問答

小說續寫

角色扮演

RAG 場景

程式碼輔助

客服案例分類

文字內容審核

▼ 圖片解析

IPC 圖片分析

物品辨識和計數

多模態能力

手寫內容識別問答/OCR

圖像理解

監控圖表理解

架構圖理解

時序圖片分析

截圖產生程式碼

視覺提示詞

從設計到程式碼

瑕疵偵測

▼ Content preferences

Language

Prompt:

這是一張台灣的高鐵車票，準確提取姓名、出發站、抵達站、出發日期和時間、車廂、座位、票價等資訊，中文輸出。



提取結果如下:



這是一張台灣的高鐵車票，準確提取出發站、抵達站、出發日期和時間、車廂、座位、票價等資訊，中文輸出。



這張車票的相關資訊如下:

出發站: 南港

抵達站: 左營

出發日期及時間: 2018/10/24 11:20

車次: Train 125

車廂: 5

座位: 2C

票價: NT\$1530



數學公式提取

Hence, for $(x, u) \in D$.

SAP Instances - | Migrate from AWS DMS v | Migrate you | Migrate with TCC-Bedrock | Amazon Bedrock | Amazon Bedrock

Instances a | us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Chat playground

Chat playground Info Load examples Compare mode

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Claude 3 Sonnet v1 | ODT

Change

Try one of these examples [or view more examples](#)

Claude 3 Sonnet

Advanced Q&A with Citations
An example prompt for long document Q&A.

Claude 3 Sonnet

Craft a Design Brief
Craft a design brief for a holistic brand identity,

Claude 3 Sonnet

Meeting Notes Summarizer
Distill meetings into concise summaries

這是一張台灣的高鐵車票，準確提取姓名、出發站、抵達站、出發日期和時間、車廂、座位、票價等資訊，中文輸出。

▶ Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

2.jpeg 83.91 KB

Configurations

Reset

System prompts

Add system prompts

Randomness and diversity

Temperature 1

Top P 0.999

Top K 250

Model metrics Define metric criteria

Instances a X | SAP Instances - X | Migrate from X | AWS DMS V X | Migrate you X | Migrate with X | TCC-Bedrock X | Amazon Bedrock X | Amazon Bedrock X | +

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+5] Services EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview

AI Claude 3 Sonnet v1 | ODT

Change Try one of these examples or view more examples

Claude 3 Sonnet Advanced Q&A with Citations An example prompt for long document Q&A.

Claude 3 Sonnet Craft a Design Brief Craft a design brief for a holistic brand identity.

Claude 3 Sonnet Meeting Notes Summarizer Distill meetings into concise summaries.

Run

Configurations Reset System prompts Add system prompts Randomness and diversity Temperature 1 Top P 0.999 Top K 250

Model metrics Define metric criteria

A red arrow points to the "Run" button.

Instances a X | SAP Instances - X | Migrate from X | AWS DMS V X | Migrate you X | Migrate with X | TCC-Bedrock X | Amazon Bedrock X | Amazon Bedrock X | +

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/chat-playground?modelId=anthropic.claude-3-sonnet-20240229-v1%3A0

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Chat playground Chat playground Info Load examples Compare mode

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models Preview

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools Knowledge bases Agents Prompt management Preview

Model metrics Define metric criteria

AI Claude 3 Sonnet v1 | ODT

Change

根據圖片所示高鐵車票,我提取出以下資訊:

姓名: 成人
出發站: 南港
抵達站: 左營
出發日期和時間: 2018/10/24 11:20
車廂: 5車廂
座位: 2C
票價: NT\$1530現金

Write a prompt... (Shift + ENTER to start a new line, and ENTER to generate a response) ▶ Run

Choose files

The supported file types are .csv, .doc, .docx, .html, .md, .pdf, .txt, .xls, .xlsx, .gif, .jpeg, .png, .webp. Maximum 5 files, 4.5 MB each.

Configurations Reset

System prompts Add system prompts

Randomness and diversity Info

Temperature 1

Top P 0.999

Top K 250

Labs

- Get hands-on with Bedrock playground
- Set up a Q&A application with knowledge retrieval
- Build a LLM application with Bedrock SDKs



© 2024, Amazon Web Services, Inc. or its affiliates.

瑕疵偵測

影片精彩片段

文件解析

文生圖、圖生圖

Artifacts

比較模式

▼ 安全防護

圖片浮水印

模型評估

安全圍籬 Guardrail

▼ 建構工具

▼ 知識庫 (Knowledge Base)

知識庫 (Knowledge Base)

與文件聊天

► 智能代理

提示詞管理

提示詞工作流

► 定製模型

► 開發維運實踐

實驗總結

行業實驗資源

▼ Content preferences

Language

中文(繁體)

知識庫 (Knowledge Base)

使用 Knowledge Bases for Amazon Bedrock 快速構建知識庫

一、上傳文件到 S3 bucket

1. 下載要查詢的文件 [\[\]](#)，這個連結是 Amazon Bedrock 的英文操作手冊，也可以使用您自己的文件。

支援的文件格式如下:

- Plain text (.txt)
- Markdown (.md)
- HyperText Markup Language (.html)
- Microsoft Word document (.doc/.docx)
- Comma-separated values (.csv)
- Microsoft Excel spreadsheet (.xls/.xlsx)
- Portable Document (.pdf)

注意：單個文件不可以大於 50MB

2. 通過 <https://console.aws.amazon.com/s3/> 打開 Amazon S3 控制台，然後點擊 建立儲存貯體(Create bucket) 按鈕。
3. 紿定儲存貯體名稱，例如：`kbbucketxxxx`，`xxxx`請改為隨機的4位數字，點擊最下面的 建立儲存貯體(Create bucket) 按鈕。如果已存在，請修改名稱，繼續建立
4. 點擊建立好的儲存貯體，點擊建立資料夾，建立一個資料夾，點擊上傳(Upload)按鈕，點擊新增檔案(Add File)按鈕，上傳前面步驟已下載的文件或您自己的文件，上傳成功後進入對應資料夾中，點擊 複製 S3 URI 按鈕。您也可以直接上傳資料夾(Add Folder)，成功後複製文件夾的 S3 URI。

二、知識庫

瑕疵偵測
影片精彩片段

文件解析
文生圖、圖生圖

Artifacts
比較模式

▼ 安全防護
圖片浮水印
模型評估
安全圍籬 Guardrail

▼ 建構工具
▼ 知識庫 (Knowledge Base)
知識庫 (Knowledge Base)

與文件聊天
▶ 智能代理
提示詞管理
提示詞工作流

▶ 定製模型

▶ 開發維運實踐
實驗總結
行業實驗資源

▼ Content preferences

Language
中文(繁體) ▾

知識庫 (Knowledge Base)

使用 Knowledge Bases for Amazon Bedrock 快速構建知識庫

一、上傳文件到 S3 bucket

1. 下載要查詢的文件 [\[\]](#)，這個連結是 Amazon Bedrock 的英文操作手冊，也可以使用您自己的文件。

支援的文件格式如下:

- Plain text (.txt)
- Markdown (.md)
- HyperText Markup Language (.html)
- Microsoft Word document (.doc/.docx)
- Comma-separated values (.csv)
- Microsoft Excel spreadsheet (.xls/.xlsx)
- Portable Document (.pdf)

注意：單個文件不可以大於 50MB

2. 通過 <https://console.aws.amazon.com/s3/> 打開 Amazon S3 控制台，然後點擊 建立儲存貯體(Create bucket) 按鈕。

3. 紿定儲存貯體名稱，例如：kbbucketxxxx，xxxx 請改為隨機的4位數字，點擊最下面的 建立儲存貯體(Create bucket) 按鈕。如果已存在，請修改名稱，繼續建立

4. 點擊建立好的儲存貯體，點擊建立資料夾，建立一個資料夾，點擊上傳(Upload)按鈕，點擊新增檔案(Add File)按鈕，上傳前面步驟已下載的文件或您自己的文件，上傳成功後進入對應資料夾中，點擊 複製 S3 URI 按鈕。您也可以直接上傳資料夾(Add Folder)，成功後複製文件夾的 S3 URI。

二、知識庫

docs.aws.amazon.com/pdfs/bedrock/latest/userguide/bedrock-ug.pdf

Amazon Bedrock - User Guide

1 / 1535 | - 100% + ⌂ ⌂

Download Print ...

1

2

3

aws User Guide

Amazon Bedrock



The screenshot shows a PDF viewer interface displaying the "Amazon Bedrock - User Guide". The top navigation bar includes standard browser controls like back, forward, and search, along with a star icon, download, print, and more options. The page header is "Amazon Bedrock - User Guide". The main content area features the AWS logo and the word "User Guide" above the title "Amazon Bedrock". To the left of the main content, there are three smaller thumbnail images labeled 1, 2, and 3, which appear to be screenshots of the user guide's Table of Contents page. The main content area contains a large, abstract graphic composed of light gray geometric shapes, including rectangles, circles, and a double-angle bracket symbol (</>), arranged in a way that suggests a complex system or architecture.

AWS Services Search: s3

EC2 VPC

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management
- Prompt flows

Assessment & deployment

Metrics

Search results for 's3'

Services

- S3** ★ Scalable Storage in the Cloud
- S3 Glacier** ☆ Archive Storage in the Cloud
- AWS Snow Family** ☆ Large Scale Data Transport
- Storage Gateway** ☆ Hybrid Storage Integration

See all 8 results ▶

Features

- Imports from S3** DynamoDB feature
- Feature spotlight** S3 feature
- S3 Access Grants** S3 feature

See all 39 results ▶

Configurations

System prompts

Add system prompts

Randomness and diversity

- Temperature: 1
- Top P: 0.999
- Top K: 250

Define metric criteria

Claude 3 Sonnet

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates.

Privacy Terms Cookie preferences

Storage

Amazon S3

Store and retrieve any amount of data from anywhere

Amazon S3 is an object storage service that offers industry-leading scalability, data availability, security, and performance.

How it works



Create a bucket

Every object in S3 is stored in a bucket. To upload files and folders to S3, you'll need to create a bucket where the objects will be stored.

[Create bucket](#)



Pricing

With S3, there are no minimum fees. You only pay for what you use. Prices are based on the location of your S3 bucket.

Estimate your monthly bill using the [AWS Simple Monthly Calculator](#)

[View pricing details](#)

Resources

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Create S3 bucket | S3 | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/s3/bucket/create?region=us-west-2

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3 > Buckets > Create bucket

Create bucket Info

Buckets are containers for data stored in S3.

General configuration

AWS Region
US West (Oregon) us-west-2

Bucket type Info

General purpose
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

Directory - New
Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name Info

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Format: s3://bucket/prefix

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Create S3 bucket | S3 | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/s3/bucket/create?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3 > Buckets > Create bucket

Create bucket Info

Buckets are containers for data stored in S3.

General configuration

AWS Region
US West (Oregon) us-west-2

Bucket type Info

General purpose
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

Directory - New
Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name Info
kbbucket-20240727 

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

Choose bucket

Format: s3://bucket/prefix

us-west-2.console.aws.amazon.com/s3/bucket/create?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Add tag

Default encryption [Info](#)

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type [Info](#)

Server-side encryption with Amazon S3 managed keys (SSE-S3)
 Server-side encryption with AWS Key Management Service keys (SSE-KMS)
 Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)

Secure your objects with two separate layers of encryption. For details on pricing, see [DSSE-KMS pricing](#) on the Storage tab of the [Amazon S3 pricing page](#).

Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

Disable
 Enable

Advanced settings

After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

Cancel **Create bucket**

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | S3 buckets | S3 | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/s3/buckets?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Successfully created bucket "kbbucket-20240727" To upload files and folders, or to configure additional bucket settings, choose View details.

View details X

Amazon S3 > Buckets

Account snapshot - updated every 24 hours All AWS Regions

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

General purpose buckets Directory buckets

General purpose buckets (1) Info All AWS Regions

Buckets are containers for data stored in S3.

Find buckets by name < 1 > ⚙️

C Copy ARN Empty Delete Create bucket

Name	AWS Region	IAM Access Analyzer	Creation date
kbbucket-20240727	US West (Oregon) us-west-2	View analyzer for us-west-2	July 27, 2024, 15:38:19 (UTC+08:00)

A red arrow points to the bucket name "kbbucket-20240727" in the table.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | kbbucket-20240727 - S3 bu | New Tab

us-west-2.console.aws.amazon.com/s3/buckets/kbbucket-20240727?region=us-west-2&bucketType=general&tab=objects

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727

kbbucket-20240727 [Info](#)

Objects Properties Permissions Metrics Management Access Points

Objects (0) [Info](#)

C 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](#)

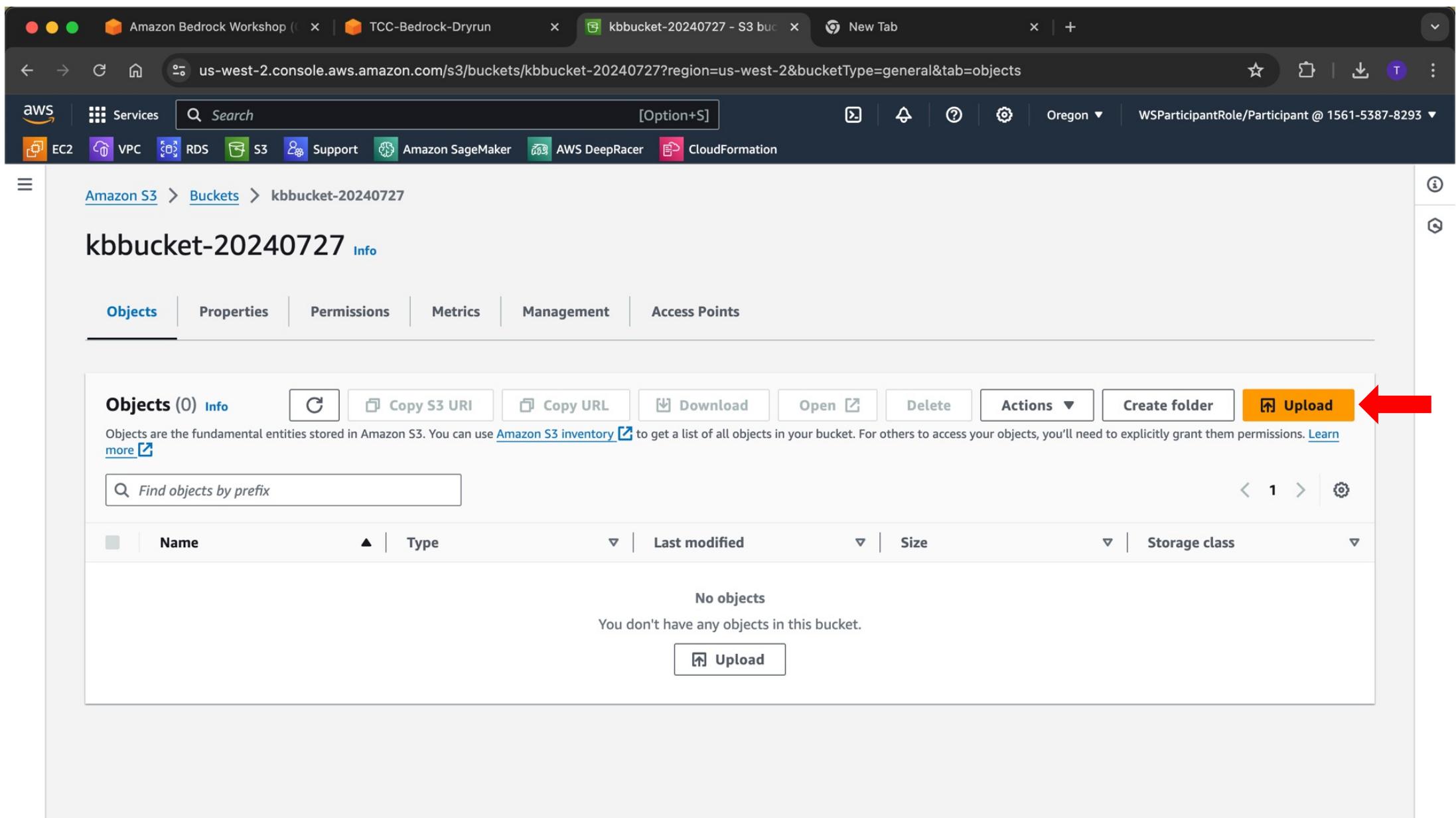
Find objects by prefix

Name Type Last modified Size Storage class

No objects

You don't have any objects in this bucket.

Upload



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket | New Tab

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (0)

All files and folders in this table will be uploaded.

Find by name < 1 >

Name	Folder	Type
No files or folders		
You have not chosen any files or folders to upload.		

Destination Info

Destination

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Favorites mba Applications Downloads On My Mac

iCloud iCloud Drive Documents Desktop Shared

Locations Network

Media Music Photos Movies

Desktop — iCloud

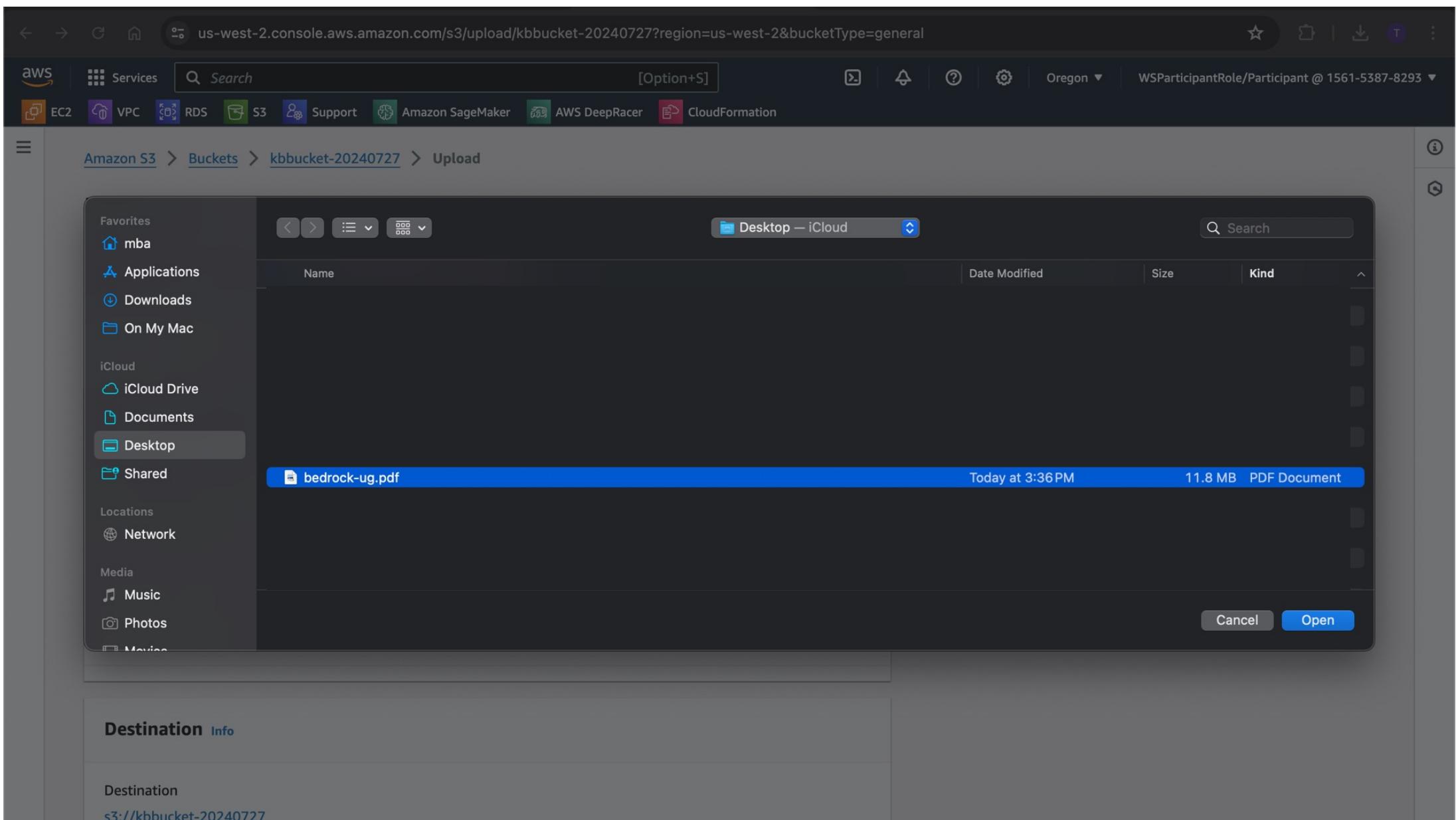
Name Date Modified Size Kind

bedrock-ug.pdf Today at 3:36 PM 11.8 MB PDF Document

Cancel Open

Destination Info

Destination s3://kbbucket-20240727



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket | New Tab

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (1 Total, 11.2 MB)		Remove	Add files	Add folder
All files and folders in this table will be uploaded.				
<input type="text"/> Find by name		<	1	>
<input type="checkbox"/>	Name	Folder	Type	
<input type="checkbox"/>	bedrock-ug.pdf	-	application/pdf	

Destination Info

Destination

s3://kbbucket-20240727

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Drag and drop files and folders you want to upload here, or choose Add files or Add folder.

Files and folders (1 Total, 11.2 MB)

All files and folders in this table will be uploaded.

<input type="checkbox"/>	Name	Folder	Type
<input type="checkbox"/>	bedrock-ug.pdf	-	application/pdf

Destination Info

Destination <s3://kbbucket-20240727>

▶ **Destination details**
Bucket settings that impact new objects stored in the specified destination.

▶ **Permissions**
Grant public access and access to other AWS accounts.

▶ **Properties**
Specify storage class, encryption settings, tags, and more.

Cancel **Upload** 

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket | New Tab

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Upload succeeded
View details below.

Upload: status Close

The information below will no longer be available after you navigate away from this page.

Summary

Destination	Succeeded	Failed
s3://kbbucket-20240727	1 file, 11.2 MB (100.00%)	0 files, 0 B (0%)

Files and folders Configuration

Files and folders (1 Total, 11.2 MB)

Name	Folder	Type	Size	Status	Error
bedrock-ua	-	application/	11.2 MB	Succeeded	-

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | kbbucket-20240727 - S3 bu | New Tab

us-west-2.console.aws.amazon.com/s3/buckets/kbbucket-20240727?region=us-west-2&bucketType=general&tab=objects

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3 > Buckets > kbbucket-20240727

kbbucket-20240727 [Info](#)

Objects Properties Permissions Metrics Management Access Points

Objects (1) [Info](#)

C 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

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	bedrock-ug.pdf	pdf	July 27, 2024, 15:39:10 (UTC+08:00)	11.2 MB	Standard

AWS Services Search bar: bedrock

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC

Amazon S3 > kbbuck

Objects

Services (1)

Resources New

Documentation (2,676)

Knowledge Articles (12)

Marketplace (386)

Blogs (232)

Events (1)

Tutorials (1)

Find objects

Names

Tags

Search results for 'bedrock'

Services

Amazon Bedrock ☆

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

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.

Dismiss Go to Resource Explorer

Documentation See all 2,676 results ▶

Amazon Bedrock ↗

User Guide

Add a data source to your app ↗

User Guide

Document history for the Amazon Bedrock Studio User Guide ↗

User Guide

What is Amazon Bedrock Studio? ↗

Create folder Upload

explicitly grant them permissions. Learn

< 1 > Settings

Storage class ▾

Standard

The screenshot shows the AWS Management Console with the 'Amazon Bedrock' service selected. The top navigation bar includes tabs for 'Amazon Bedrock Workshop' (active), 'TCC-Bedrock-Dryrun', 'Amazon Bedrock | us-west-2' (selected), and 'New Tab'. The URL in the address bar is `us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#`. The main menu bar has 'Services' selected, showing options like EC2, VPC, RDS, S3, Support, Amazon SageMaker, AWS DeepRacer, and CloudFormation. A red arrow points to the three-line menu icon on the left. The page content features the 'Machine Learning' section, the 'Amazon Bedrock' heading, and a 'Try Bedrock' call-to-action button.

Amazon Bedrock

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

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock Machine Learning

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

Try Bedrock Get started

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/

aws Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Imported models [Preview](#)

▼ Playgrounds

- Chat
- Text
- Image

▼ Safeguards

- Guardrails
- Watermark detection

▼ Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

▼ Assessment & deployment

- Model Evaluation
- Provisioned Throughput

Model access [4 new](#)

Bedrock Studio [Preview](#)

Settings

User guide

Machine Learning

Amazon Bedrock

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

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Knowledge bases

Knowledge bases Chat with your document

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- [Knowledge bases](#)
- Agents
- Prompt management [Preview](#)

Amazon Bedrock > Knowledge bases

Knowledge bases

▼ How it works

Upload and chat

Quickly query foundation models with context provided by ad-hoc dataset.

[Chat with your document](#)

Create a knowledge base

To create a knowledge base, specify the location of your data, select an embedding model, and configure a vector store for Bedrock to store and update your embeddings.

Test the knowledge base

Query your knowledge base in the test window. You can get source text chunks, or you can use the chunks to get responses from a foundation model.

Use the knowledge base

Integrate your knowledge base into your application as is or add it to agents.

Knowledge bases (0)

Edit Delete Test knowledge base Create knowledge base

Find knowledge base

Name	Status	Description	Source files	Creation time	Last sync w...	Last sync
No knowledge base						
No knowledge base to display						

A red arrow points to the "Create knowledge base" button.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1 Provide knowledge base details

Step 2 Configure data source

Step 3 Select embeddings model and configure vector store

Step 4 Review and create

Provide knowledge base details

Knowledge base details

Knowledge base name

knowledge-base-quick-start-6zlo3

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

Knowledge base description - optional

Enter description

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 200 characters.

IAM permissions

Certain permissions are necessary to access other services or perform actions in order to create this resource. For more information, see [service role](#) for Amazon Bedrock

Runtime role

Create and use a new service role

Use an existing service role

Service role name

AmazonBedrockExecutionRoleForKnowledgeBase-6zlo3

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

AWS Services Search [Option+5] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

[Knowledge bases](#)

- Agents
- Prompt management [Preview](#)

Choose data source

Select the data source that you want to configure in the next step.

 **Amazon S3** 

Object storage service that stores data as objects within buckets.

 **Web Crawler - Preview** 

Web page crawler that extracts content from public web pages you are authorized to crawl.

Third party data sources

 **Confluence - Preview** 

Collaborative work-management tool designed for project planning, software development and product management.

 **Salesforce - Preview** 

Customer relationship management (CRM) tool for managing support, sales, and marketing data.

 **Sharepoint - Preview** 

Collaborative web-based service for working on documents, web pages, web sites, lists, and more.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

No tags associated with the resource.

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Imported models [Preview](#)

▼ Playgrounds

- Chat
- Text
- Image

▼ Safeguards

- Guardrails
- Watermark detection

▼ Builder tools

- Knowledge bases**
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

▼ Assessment & deployment

- Model Evaluation
- Provisioned Throughput

Model access [4 new](#)

Bedrock Studio [Preview](#)

Settings

User guide

Bedrock Service Terms

Object storage service that stores data as objects within buckets.

Web Crawler - [Preview](#)
Web page crawler that extracts content from public web pages you are authorized to crawl.

Third party data sources

Confluence - [Preview](#)
Collaborative work-management tool designed for project planning, software development and product management.

Salesforce - [Preview](#)
Customer relationship management (CRM) tool for managing support, sales, and marketing data.

Sharepoint - [Preview](#)
Collaborative web-based service for working on documents, web pages, web sites, lists, and more.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

No tags associated with the resource.

Add new tag

You can add up to 50 more tags.

Cancel

Next

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1: Provide knowledge base details
Step 2: Configure data source (selected)
Step 3: Select embeddings model and configure vector store
Step 4: Review and create

Configure data source

Configure for the chosen data source

Amazon S3 Info

Provide details to connect Amazon Bedrock to your S3 data source.

▼ Data source: knowledge-base-quick-start-cdw4o-data-source Delete

Data source name: knowledge-base-quick-start-cdw4o-data-source

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 100 characters.

Data source location

This AWS account
 Other AWS account

S3 URI

To increase the accuracy and relevance of your responses, add a .metadata.json file containing metadata for your data source to your S3 bucket. Info

Choose a s3 location View **Browse S3**

Add customer-managed KMS key for S3 data - optional
If you encrypted your S3 data, provide the KMS key here so that Bedrock can decrypt it.

Chunking and parsing configurations Info

Choose between default or advanced customization.

Default Advanced

Servicess Search [Option+S] Services EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Amazon Bedrock > Knowledge bases > Create knowledge base

Step 1 Provide knowledge base details Step 2 Configure data source

Configure data source Configure for the chosen data source

Amazon S3 Info

Choose an archive in S3

S3 buckets Buckets (1/1)

Find S3 bucket < 1 >

Name	Creation date
kbbucket-20240727	2024-07-27T07:38:20.000Z

Cancel Choose

Choose a s3 location View Browse S3

Add customer-managed KMS key for S3 data - optional
If you encrypted your S3 data, provide the KMS key here so that Bedrock can decrypt it.

Chunking and parsing configurations Info
Choose between default or advanced customization.

Default
Uses default parsing and chunking strategy.

Custom
Customize the parsing and chunking strategy, including using advanced parsing.

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1 Provide knowledge base details
Step 2 Configure data source
Step 3 Select embeddings model and configure vector store
Step 4 Review and create

Configure data source

Configure for the chosen data source

Amazon S3 Info

Provide details to connect Amazon Bedrock to your S3 data source.

▼ Data source: knowledge-base-quick-start-cdw4o-data-source Delete

Data source name
knowledge-base-quick-start-cdw4o-data-source

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 100 characters.

Data source location

This AWS account
 Other AWS account

S3 URI
To increase the accuracy and relevance of your responses, add a .metadata.json file containing metadata for your data source to your S3 bucket. Info

s3://kbbucket-20240727 View Browse S3

Add customer-managed KMS key for S3 data - optional
If you encrypted your S3 data, provide the KMS key here so that Bedrock can decrypt it.

Chunking and parsing configurations Info

Choose between default or advanced customization.

Default Advanced

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Step 4 Review and create

configure vector store

Data source: knowledge-base-quick-start-cdw4o-data-source

Data source name
knowledge-base-quick-start-cdw4o-data-source

Valid characters are a-z, A-Z, 0-9, _ (underscore) and - (hyphen). The name can have up to 100 characters.

Data source location
 This AWS account
 Other AWS account

S3 URI
To increase the accuracy and relevance of your responses, add a .metadata.json file containing metadata for your data source to your S3 bucket. [Info](#)

s3://kbbucket-20240727 [View](#) [Browse S3](#)

Add customer-managed KMS key for S3 data - *optional*
If you encrypted your S3 data, provide the KMS key here so that Bedrock can decrypt it.

Chunking and parsing configurations [Info](#)
Choose between default or advanced customization.

Default
Uses default parsing and chunking strategy.

Custom
Customize the parsing and chunking strategy, including using advanced parsing.

Advanced settings - optional

[Add data source](#)

You can add 4 more data source(s).

Cancel Previous Next

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1: Provide knowledge base details
Step 2: Configure data source
Step 3: Select embeddings model and configure vector store
Step 4: Review and create

Select embeddings model and configure vector store

Choose an embeddings model to convert the data that you will provide in the next step, and provide details for a vector data store in which Bedrock can store, manage, and update your embeddings. The embeddings model and vector store cannot be changed after creation of knowledge base.

Embeddings model

Select an embeddings model to convert your data into an embedding. Pricing depends on the model. [Learn more](#)

-  [Titan Text Embeddings v2](#)
By Amazon
-  [Titan Embeddings G1 - Text v1.2](#)
By Amazon
-  [Embed English v3](#)
By Cohere
-  [Embed Multilingual v3](#)
By Cohere

Vector dimensions

Select the vector dimension size for your embeddings model to balance accuracy, cost, and latency. Higher dimensions improves overall accuracy and requires more vector storage. [Learn more](#)

1024

Vector database

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1: Provide knowledge base details
Step 2: Configure data source
Step 3: Select embeddings model and configure vector store
Step 4: Review and create

Select embeddings model and configure vector store

Choose an embeddings model to convert the data that you will provide in the next step, and provide details for a vector data store in which Bedrock can store, manage, and update your embeddings. The embeddings model and vector store cannot be changed after creation of knowledge base.

Embeddings model

Select an embeddings model to convert your data into an embedding. Pricing depends on the model. [Learn more](#)

Titan Text Embeddings v2 By Amazon 

Titan Embeddings G1 - Text v1.2 By Amazon

Embed English v3 By Cohere

Embed Multilingual v3 By Cohere

Vector dimensions

Select the vector dimension size for your embeddings model to balance accuracy, cost, and latency. Higher dimensions improves overall accuracy and requires more vector storage. [Learn more](#)

1024

Vector database

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Embed English v3 By Cohere

Embed Multilingual v3 By Cohere

Vector dimensions
Select the vector dimension size for your embeddings model to balance accuracy, cost, and latency. Higher dimensions improves overall accuracy and requires more vector storage. [Learn more](#)

1024

Vector database
Let Amazon create a vector store on your behalf or select a previously created store to allow Bedrock to store, update and manage embeddings. You will be billed directly from the vector store provider. [Learn more](#)

Select how you want to create your vector store.

Quick create a new vector store - Recommended
We will create an Amazon OpenSearch Serverless vector store on your behalf. This cost-efficient option is intended only for development and can't be migrated to production workload later. [Learn more](#)

Choose a vector store you have created
Select Amazon OpenSearch Serverless, Amazon Aurora, MongoDB Atlas, Pinecone or Redis Enterprise Cloud and provide field mappings.

Enable redundancy (active replicas) - optional
The default configuration has active replicas disabled, which is optimal for development workloads. Enable this option if you want to enable redundant active replicas, which may increase storage costs.

Add customer-managed KMS key for Amazon OpenSearch Serverless vector - optional
If you encrypted your OpenSearch data, provide the KMS key here so that Bedrock can decrypt it.

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

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services | Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases > Create knowledge base

Step 1: Provide knowledge base details

Step 2: Configure data source

Step 3: Select embeddings model and configure vector store

Step 4: Review and create

Review and create

Step 1: Provide details

Edit

Knowledge base name knowledge-base-quick-start-6zlo3	Knowledge base description —	Service role AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3
--	--	---

Tags (0)

Key	Value
No tags to display	

Step 2: Setup up data source

Edit

Data source: knowledge-base-quick-start-cdw4o-data-source

Data source name	Account ID	S3 URI
-------------------------	-------------------	---------------



us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Customer-managed KMS Key for S3	KMS key for transient data storage	Chunking strategy
-	-	Default
Parsing strategy	Lambda function	S3 bucket for Lambda function
Default	-	-
Data deletion policy		
Delete		

Step 3: Select embeddings model and configure vector store [Edit](#)

Embeddings model

Model	Vector dimensions
Titan Text Embeddings v2	1024

Vector store

Quick create vector store - Recommended
We will create an Amazon OpenSearch Serverless vector store in your account on your behalf.

[Cancel](#) [Previous](#) **Create knowledge base** 

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Preparing vector database in Amazon OpenSearch Serverless. This process may take several minutes to complete.

Amazon Bedrock > Knowledge bases > Create knowledge base

Review and create

Step 1: Provide details

Knowledge base details

Knowledge base name	knowledge-base-quick-start-6zlo3	Knowledge base description	—
Service role	AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3		

Tags (0)

Key	Value
No tags to display	

Step 2: Setup up data source

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/create-knowledge-base

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Notifications 0 0 1 0 2

Amazon OpenSearch Serverless vector database is ready.

Review and create

Step 1: Provide details

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base description: —

Service role: AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Tags (0)

Key	Value
No tags to display	

Step 2: Setup up data source

Edit

A red arrow points to the notification message "Amazon OpenSearch Serverless vector database is ready."

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Amazon Bedrock > **Knowledge bases** > knowledge-base-quick-start-6zlo3

knowledge-base-quick-start-6zlo3

[Test](#) [Delete](#) [Edit](#)

Knowledge base overview

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Status: Ready

Created date: July 27, 2024, 15:51 (UTC+08:00)

Service Role: AmazonBedrockExecutionRoleForKnowl edgeBase_6zlo3

Log Deliveries
Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS

Test knowledge base

Generate responses [Edit](#)

Select model

One or more data sources have not been synced. [Go to data sources](#)

Configure your retrieval and responses
To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model [Run](#)

Red arrow pointing to the success message: Knowledge base 'knowledge-base-quick-start-6zlo3' is created successfully.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases**
- Agents
- Prompt management [Preview](#)

Test knowledge base

Generate responses

Select model

One or more data sources have not been synced.

Go to data sources

Configure your retrieval and responses

To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model 

knowledge-base-quick-start-6zlo3

Knowledge base overview

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Status: Ready

Created date: July 27, 2024, 15:51 (UTC+08:00)

Service Role: AmazonBedrockExecutionRoleForKnowl edgeBase_6zlo3 

Log Deliveries

Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- [Knowledge bases](#)
- Agents
- Prompt management [Preview](#)

Test knowledge base

Generate responses

Select model

One or more data sources have not been synced. [Go to data sources](#)

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model 

Key Value

No tags No tags to display [Manage tags](#)

Data source (1) [Add](#) [Edit](#) [Delete](#) [Sync](#)

Data sources contain information returned when querying a Knowledge base.

Find data source

knowledge... Available S3 15615387... s3://k

Embeddings model

Model Titan Text Embeddings v2 Vector dimensions 1024



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

No tags
No tags to display
Manage tags

Data source (1)
Data sources contain information returned when querying a Knowledge base.

Add Edit Delete Sync

Find data source

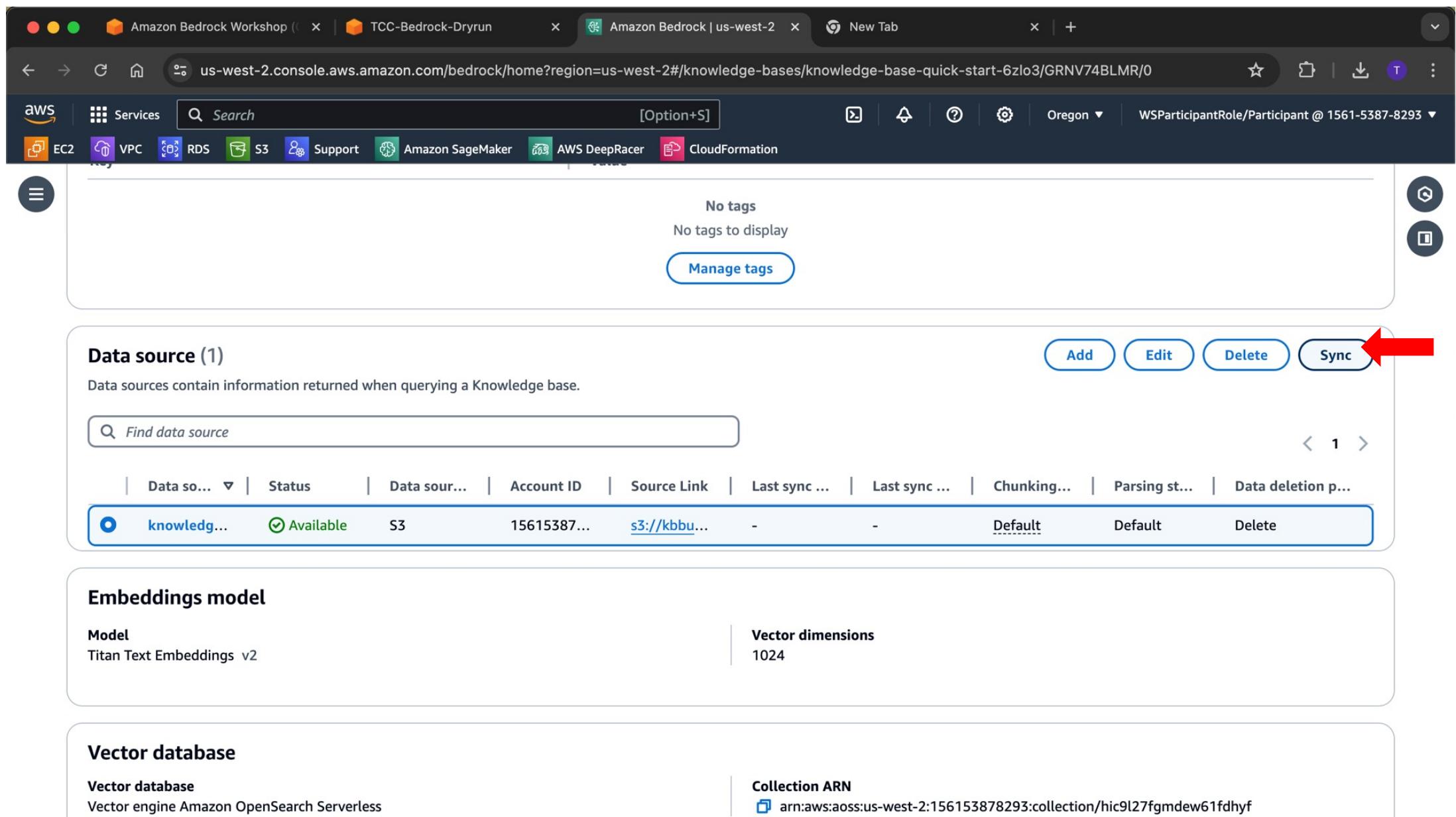
Data so...	Status	Data sour...	Account ID	Source Link	Last sync ...	Last sync ...	Chunking...	Parsing st...	Data deletion p...
knowledge... knowledge...	Available	S3	15615387...	s3://kbbu...	-	-	Default	Default	Delete

Embeddings model

Model: Titan Text Embeddings v2 | **Vector dimensions**: 1024

Vector database

Vector database: Vector engine Amazon OpenSearch Serverless | **Collection ARN**: arn:aws:aoss:us-west-2:156153878293:collection/hic9l27fgmdew61fdhyf



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

No tags
No tags to display

Manage tags

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so...	Status	Data sour...	Account ID	Source Link	Last sync ...	Last sync ...	Chunking...	Parsing st...	Data deletion p...
knowledge... knowledge...	Syncing	S3	15615387...	s3://kb... s3://kb...	July 27, 2...	-	Default	Default	Delete

Embeddings model

Model: Titan Text Embeddings v2

Vector dimensions: 1024

Vector database

Vector database: Vector engine Amazon OpenSearch Serverless

Collection ARN: arn:aws:aoss:us-west-2:156153878293:collection/hic9l27fgmdew61fdhyf

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | New Tab

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases**
- Agents
- Prompt management [Preview](#)

Test knowledge base

Generate responses

Select model

The system is syncing your data source. Wait for the sync to complete before starting next sync job. [Go to data sources](#)

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model 

Key Value

No tags No tags to display [Manage tags](#)

Data source (1) [Add](#) [Edit](#) [Delete](#) [Sync](#)

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so... Status Data sour... Account ID Sour...

knowledge... Syncing 15615387... s3://k...

Model Vector dimensions

Titan Text Embeddings v2 1024

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases**
- Agents
- Prompt management [Preview](#)

Test knowledge base

Generate responses

Select model

Configure your retrieval and responses

To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model 

Key	Value
No tags	No tags to display

Manage tags

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so... Status Data sour... Account ID Sour...

knowledgebase Available 15615387... s3://knowledgebase

Embeddings model

Model: Titan Text Embeddings v2 Vector dimensions: 1024

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services Search [Option+5]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- [Knowledge bases](#)
- Agents
- Prompt management [Preview](#)

Key Value

No tags
No tags to display

Manage tags

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so... Status Data sour... Account ID Sour...

knowledgebase	Available	S3	15615387...	s3://k...
---------------	-----------	----	-------------	-----------

Embeddings model

Model Vector dimensions

Titan Text Embeddings v2 1024

Test knowledge base

Generate responses

Select model

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model Run

The screenshot shows the Amazon Bedrock console interface. On the left, there's a sidebar with various navigation options like 'Getting started', 'Foundation models', 'Playgrounds', 'Safeguards', and 'Builder tools'. The main area displays a 'Data source (1)' table with one entry: 'knowledgebase' (Status: Available, Source: S3). Below it is an 'Embeddings model' section showing 'Titan Text Embeddings v2' with 'Vector dimensions 1024'. On the right, there's a 'Test knowledge base' panel with a 'Generate responses' toggle (which is turned on) and a 'Select model' button (which is highlighted with a red arrow). A tooltip for 'Select model' explains that it allows customizing the search strategy for the knowledge base. At the bottom of the 'Test knowledge base' panel, there's a note about running queries and generating responses, and a 'Please select a model' message with a 'Run' button.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

Anthropic

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

Not seeing a model you are interested in? Check out all supported models [here](#)

3. Throughput

Provisioned throughput is not supported for knowledge bases.

base by running a query to disable response generation information stored from your Generate responses above.

Cancel | Apply | Run

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

Anthropic

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

3. Throughput

On-demand (ODT)

base by running a query to disable response generation information stored from your Generate responses above.

Cancel | Apply | Run

Getting started
Overview
Examples
Providers

Foundation models
Base models
Custom models
Imported models [Preview](#)

Playgrounds
Chat
Text
Image

Safeguards
Guardrails
Watermark detection

Builder tools
Knowledge bases
Agents
Prompt management [Preview](#)

retrieval and search strategy base, select the on .

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

Anthropic

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2** Text model | Context size = up to 100k
- Claude 2.1 v2.1** Text model | Context size = up to 200k
- Claude 2 v2** Text model | Context size = up to 100k
- Claude 3 Sonnet v1** Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1** Text & vision model | Context size = up to 200k

Not seeing a model you are interested in? Check out all supported models [here](#)

3. Throughput

On-demand (ODT)

base by running a query to disable response generation information stored from your Generate responses above.

Cancel **Apply** Run

The screenshot shows the 'Select model' interface in the Amazon Bedrock console. On the left, a sidebar lists categories like 'Getting started', 'Foundation models', 'Playgrounds', 'Safeguards', and 'Builder tools'. The main area is divided into three sections: 'Category' (Model providers), 'Model' (Models with access), and 'Throughput' (On-demand). In the 'Category' section, 'Anthropic' is selected. In the 'Model' section, 'Claude 3 Sonnet v1' is selected. At the bottom right, there are 'Cancel', 'Apply', and 'Run' buttons. A red arrow points to the 'Apply' button.



瑕疵偵測
影片精彩片段
文件解析
文生圖、圖生圖
Artifacts
比較模式

▼ 安全防護

圖片浮水印
模型評估
安全圍籬 Guardrail

▼ 建構工具

▼ 知識庫 (Knowledge Base)
[知識庫 \(Knowledge Base\)](#)

與文件聊天

► 智能代理
提示詞管理
提示詞工作流

► 定製模型

► 開發維運實踐
實驗總結
行業實驗資源

▼ Content preferences

Language

中文(繁體) ▾

Cancel

Apply

然後輸入要查詢的問題,等待輸出。

1. 如何配置 Amazon Bedrock 知識庫? 

輸出如下圖:

 Generate responses



Claude 3 Sonnet v1 | ODT

[Change](#)

 **Configure your retrieval and responses**



To customize the search strategy for your knowledge base, select the configurations icon .

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

[Knowledge bases](#)

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Amazon Bedrock > Knowledge bases > knowledge-base-quick-start-6zlo3

knowledge-base-quick-start-6zlo3

Test Delete Edit

Knowledge base overview

Knowledge base name knowledge-base-quick-start-6zlo3

Knowledge base ID GRNV74BLMR

Status Ready

Created date July 27, 2024, 15:51 (UTC+08:00)

Service Role AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Log Deliveries
Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags No tags to display	

Manage tags

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

如何配置 Amazon Bedrock 知識庫? Run

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

[Knowledge bases](#)

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

Amazon Bedrock > Knowledge bases > knowledge-base-quick-start-6zlo3

knowledge-base-quick-start-6zlo3

Test Delete Edit

Knowledge base overview

Knowledge base name knowledge-base-quick-start-6zlo3

Knowledge base ID GRNV74BLMR

Knowledge base description —

Status Ready

Service Role AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Created date July 27, 2024, 15:51 (UTC+08:00)

Log Deliveries
Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags No tags to display	

Manage tags

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

如何配置 Amazon Bedrock 知識庫? Run

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/0?modelId=a...

Services Search [Option+5] | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases**
- Agents
- Prompt management [Preview](#)

Knowledge base overview

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Status: Ready

Created date: July 27, 2024, 15:51 (UTC+08:00)

Service Role: [AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3](#)

Log Deliveries: Configure log deliveries and event logs in the [Edit](#) page.

Tags: A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags	

No tags to display

[Manage tags](#)

configurations icon .

如何配置 Amazon Bedrock 知識庫?

要配置 Amazon Bedrock 知識庫,您可以按照以下步驟: 1. 收集源文件並將它們存儲在支持的數據源中,例如 Amazon S3。2. 在 Amazon Bedrock 控制台中創建一個新的知識庫,提供知識庫名稱、描述和 IAM 角色。3. 選擇您的數據源並配置連接信息。4. 選擇一個向量存儲來存儲數據的嵌入向量表示。您可以快速創建一個新的 Amazon OpenSearch Serverless 向量數據庫,或使用現有的向量存儲。5. 選擇一個模型將您的數據轉換為嵌入向量。6. 檢查知識庫配置並創建知識庫。創建完成後,知識庫的狀態將變為"就緒"。[\[1\]](#)[\[2\]](#)[\[3\]](#)[\[4\]](#)[\[5\]](#) 一旦知識庫就緒,您就可以同步數據源以保持內容最新。在控制台中選擇您的知識庫,然後在數據源概覽部分選擇"同步"。[\[6\]](#)

Show source details

Enter your message here ▶ Run

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Test Knowledge base

Generate responses Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon.

如何配置 Amazon Bedrock 知識庫?

要配置 Amazon Bedrock 知識庫,您可以按照以下步驟: 1. 收集源文件並將它們存儲在支持的數據源中,例如 Amazon S3。 2. 在 Amazon Bedrock 控制台中創建一個新的知識庫,提供知識庫名稱、描述和 IAM 角色。 3. 選擇您的數據源並配置連接信息。 4. 選擇一個向量存儲來存儲數據的嵌入向量表示。 您可以快速創建一個新的 Amazon OpenSearch Serverless 向量數據庫,或使用現有的向量存儲。 5. 選擇一個模型將您的數據轉換為嵌入向量。 6. 檢查知識庫配置並創建知識庫。創建完成後,知識庫的狀態將變為“就緒”。[\[1\]](#)[\[2\]](#)[\[3\]](#)[\[4\]](#)[\[5\]](#) 一旦知識庫就緒,您就可以同步數據源以保持內容最新。在控制台中選擇您的知識庫,然後在數據源概覽部分選擇“同步”。[\[6\]](#)

Source details (6)

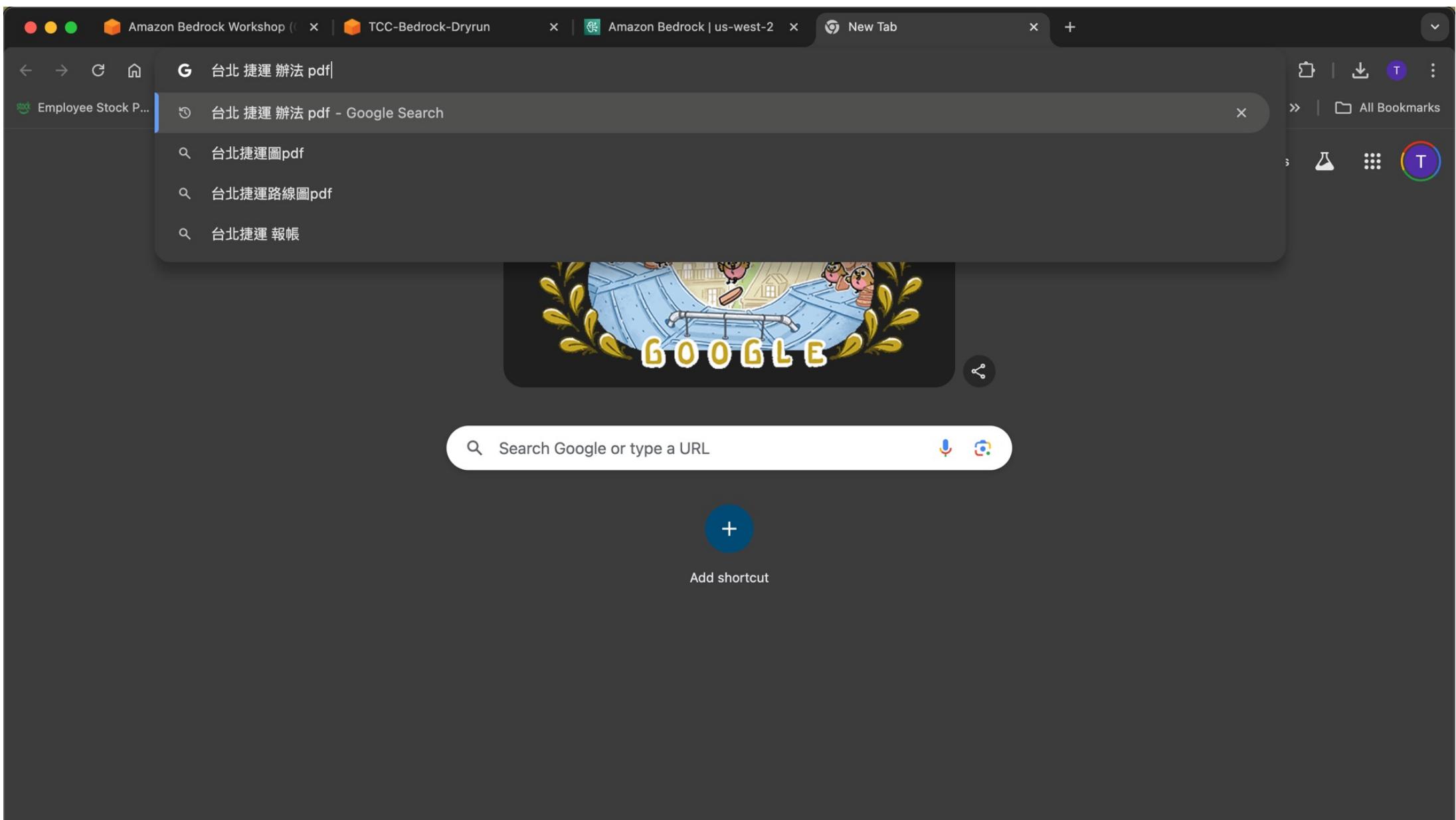
Each container below shows a chunk that comes from a data source and that relates to your query.

Query configurations (1)

Source chunk 1

Gather source documents to add to your knowledge base. 2. Store your source documents in a supported data source and configure the connection information to connect to and crawl your data. 3. (Optional if using Amazon S3 to store your source documents) Create a metadata file for each source document to allow for filtering of results during knowledge base query. 4. (Optional) Set up a vector index in a supported vector store to index your data. You can use the Amazon Bedrock console to create an Amazon OpenSearch Serverless vector database for you. 5. Create and configure your knowledge base. 6. Ingest your data by letting knowledge bases generate embeddings with an embeddings model and storing them in a supported vector store. 7. Set up your application or agent to query the knowledge base and return augmented responses. 492
<https://docs.aws.amazon.com/bedrock/latest/userguide/kb-how-it-works.html>
https://docs.aws.amazon.com/bedrock/latest/APIReference/API_agent-runtime_RetrieveAndGenerate.html
https://docs.aws.amazon.com/bedrock/latest/APIReference/API_agent-runtime_Retrieve.html
<https://docs.aws.amazon.com/bedrock/latest/userguide/data-source-integration.html>

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences





台北 捷運 辦法 pdf



All Images News Videos Shopping Books Web More

Tools



臺北大眾捷運股份有限公司

<https://web.metro.taipei> QRCode PDF

⋮

臺北捷運系統旅客須知

Jun 19, 2024 — 同警察人員強制或護送其離開站、車或大眾捷運系統區域：. (一) 違反法令、
公共秩序、善良風俗或本公司旅客運送章則等各項 規定。 (二) 有明顯傷害他人 ...



臺北大眾捷運股份有限公司

<https://web.metro.taipei> TaipeiMetroGuides T... PDF

⋮

台北捷運營運資訊簡介

攜帶自行車 單程票 開放時段:假日全天,平日10至16時、22時至營運結束。除文湖線各車站、淡水
站、**台北**車站、忠孝新生站、忠孝 復興站、南京復興站、大安站以外,餘85個車站 ...

2 pages



臺北大眾捷運股份有限公司

<https://web.metro.taipei> QRCode PDF

⋮

臺北捷運系統營運服務規約

一、營運時間：06:00~24:00. 二、服務內容. (一) 平均班距. 1.主線：. (1) 尖峰時段：. A.淡水
站－象山站、松山站－新店站、南港展覽館站－.



臺北市政府全球資訊網

<https://www.laws.taipei.gov.tw> caseatt, 臺北市... PDF

⋮

臺北市大眾捷運系統與地下街設施移設及連通申請自治條例

第四條本自治條例用詞定義如下：. 一移設：將原計畫(指於都市計畫變更書圖註明者)或已. 興建於
都市計畫道路上之**捷運**或地下街設施改設於. 申請人或都市計畫規定之建築基地內 ...

7 pages

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2 | 台北大眾捷運股份有限公司票證

web.metro.taipei/QRCode/Regulations%20for%20Use%20of%20the%20Taipei%20Metro%20System-Chinese.pdf?t=20240619

1 / 10 | - 100% + | ☰ ⚡

臺北大眾捷運股份有限公司票證作業管理要點修訂意見表

1

2

3

臺北捷運系統旅客須知

中華民國八十五年三月二十八日公告修正
中華民國八十七年十二月二十四日公告修正
中華民國八十八年十一月十日公告修正
中華民國九十一一年八月二十日公告修正
中華民國九十三年十二月二十四日公告修正
中華民國九十七年三月二十八日公告修正
中華民國九十七年六月三日公告修正
中華民國九十八年三月十二日公告修正
中華民國一〇三年五月二十七日公告修正
中華民國一〇四年三月十八日公告修正
中華民國一〇四年十二月十一日公告修正
中華民國一〇六年五月十二日公告修正
中華民國一〇六年十月五日公告修正
中華民國一〇七年十月二十六日公告修正
中華民國一一一年三月二十九日公告修正
中華民國一一二年二月七日公告修正
中華民國一一二年十一月八日公告修正
中華民國一一三年六月十三日公告修正

壹、一般規定

一、臺北大眾捷運股份有限公司(以下簡稱本公司)為提供捷運系統旅客安全、可靠、便捷、舒適之服務，特依「臺北市大眾捷運系統旅客運送自治條例」規定訂定本須知，並於車站公告，變更或調整時亦同。

二、本須知用語定義如下：

(一) 捷運範圍：為本公司所經營之大眾捷運系統路網範圍內所有路線、場、站與列車等區域。

(二) 旅客：指搭乘本公司列車，或持有有效車票並進出乘車處所車站大廳之人。

AWS Services Search: s3

Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started: Overview, Examples, Providers

Foundation models: Base models, Custom models, Imported models (Preview)

Playgrounds: Chat, Text, Image

Safeguards: Guardrails, Watermark detection

Builder tools: Knowledge bases, Agents, Prompt management, Prompt flows (Preview)

Assessment & deployment

Services (8): Features (39), Resources (New), Documentation (26,830), Knowledge Articles (288), Marketplace (1,850), Blogs (1,415), Events (26), Tutorials (12)

Search results for 's3'

Services

- S3** ★ Scalable Storage in the Cloud
- S3 Glacier** ☆ Archive Storage in the Cloud
- AWS Snow Family** ☆ Large Scale Data Transport
- Storage Gateway** ☆ Hybrid Storage Integration

See all 8 results ▶

Features

- Imports from S3** DynamoDB feature
- Feature spotlight** S3 feature
- S3 Access Grants** S3 feature

See all 39 results ▶

from a data source and that relates to

your knowledge base. 2. Store your source and configure the and crawl your data. 3. (Optional documents) Create a metadata for filtering of results during up a vector index in a data. You can use the Amazon OpenSearch Serverless vector your knowledge base. 6. bases generate embeddings them in a supported vector to query the knowledge base

/latest/userguide/kb-how-it-

/latest/APIReference/API_agent

/latest/APIReference/API_agent

/latest/userguide/data-source-

chunk

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | S3 buckets | S3 | us-west-2

us-west-2.console.aws.amazon.com/s3/home?region=us-west-2

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3

Buckets Access Grants Access Points Object Lambda Access Points Multi-Region Access Points Batch Operations IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens Dashboards Storage Lens groups AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3

▶ Account snapshot - updated every 24 hours All AWS Regions

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

[View Storage Lens dashboard](#)

General purpose buckets Directory buckets

General purpose buckets (1) [Info](#) All AWS Regions

Buckets are containers for data stored in S3.

Find buckets by name

Name	AWS Region	IAM Access Analyzer	Creation date
kbbucket-20240727	US West (Oregon) us-west-2	View analyzer for us-west-2	July 27, 2024, 15:38:19 (UTC+08:00)

[C](#) [Copy ARN](#) [Empty](#) [Delete](#) [Create bucket](#)

< 1 > [⚙️](#)

A red arrow points to the bucket name "kbbucket-20240727" in the table.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | kbbucket-20240727 - S3 bucket

us-west-2.console.aws.amazon.com/s3/buckets/kbbucket-20240727?region=us-west-2&bucketType=general&tab=objects

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon S3

Buckets

- Access Grants
- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

- Dashboards
- Storage Lens groups
- AWS Organizations settings

Feature spotlight 7

AWS Marketplace for S3

Amazon S3 > Buckets > kbbucket-20240727

kbbucket-20240727 [Info](#)

Objects Properties Permissions Metrics Management Access Points

Objects (1) [Info](#)

C 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

Name	Type	Last modified	Size	Storage class
bedrock-ug.pdf	pdf	July 27, 2024, 15:39:10 (UTC+08:00)	11.2 MB	Standard

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket k... | +

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (0)

All files and folders in this table will be uploaded.

Find by name < 1 >

Name	Folder	Type
No files or folders		
You have not chosen any files or folders to upload.		

Destination Info

Destination

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Favorites mba Applications Downloads On My Mac

iCloud iCloud Drive Shared

Locations Network

Media Music Photos

Workshop_KB

Name Size Kind Date Added

Regulations for Use of the Taipei Metro System-Chinese.pdf 226 KB PDF Document Today at 3:58 PM

Cancel Open

Destination Info

Destination

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket k... | +

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon S3 > Buckets > kbbucket-20240727 > Upload

Upload Info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (1 Total, 220.4 KB)		Remove	Add files	Add folder
All files and folders in this table will be uploaded.				
<input type="text"/> Find by name		<	1	>
<input type="checkbox"/>	Name	Folder	Type	
<input type="checkbox"/>	Regulations for Use of the Taipei Metro ...	-	application/pdf	

Destination Info

Destination

s3://kbbucket-20240727

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Drag and drop files and folders you want to upload here, or choose Add files or Add folder.

Files and folders (1 Total, 220.4 KB)

All files and folders in this table will be uploaded.

<input type="checkbox"/>	Name	Folder	Type
<input type="checkbox"/>	Regulations for Use of the Taipei Metro ...	-	application/pdf

Destination Info

Destination
s3://kbbucket-20240727

▶ **Destination details**
Bucket settings that impact new objects stored in the specified destination.

▶ **Permissions**
Grant public access and access to other AWS accounts.

▶ **Properties**
Specify storage class, encryption settings, tags, and more.

Cancel Upload



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Upload objects - S3 bucket k... | +

us-west-2.console.aws.amazon.com/s3/upload/kbbucket-20240727?region=us-west-2&bucketType=general

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Upload succeeded
View details below.

Upload: status Close

The information below will no longer be available after you navigate away from this page.

Summary

Destination	Succeeded	Failed
s3://kbbucket-20240727	✓ 1 file, 220.4 KB (100.00%)	... 0 files, 0 B (0%)

Files and folders Configuration

Files and folders (1 Total, 220.4 KB)

Name	Folder	Type	Size	Status	Error
Regulations	-	application/	220.4 KB	✓ Succeeded	-

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock < Knowledge bases Chat with your document

Getting started Overview Examples Providers

Foundation models Base models Custom models Imported models [Preview](#)

Playgrounds Chat Text Image

Safeguards Guardrails Watermark detection

Builder tools [Knowledge bases](#) Agents Prompt management [Preview](#)

Knowledge bases

How it works

Upload and chat

Create a knowledge base

Test the knowledge base

Use the knowledge base

Quickly query foundation models with context provided by ad-hoc dataset.

To create a knowledge base, specify the location of your data, select an embedding model, and configure a vector store for Bedrock to store and update your embeddings.

Query your knowledge base in the test window. You can get source text chunks, or you can use the chunks to get responses from a foundation model.

Integrate your knowledge base into your application as is or add it to agents.

Knowledge bases (1)

Edit Delete Test knowledge base Create knowledge base

Name	Status	Description	Source files	Creation time	Last sync w...	Last sync
knowledge-b...	Ready	-	1	July 27, 2024, ...	-	July 27, 2...

Find knowledge base < 1 > ⚙️

A red arrow points to the "knowledge-b..." entry in the table, highlighting the newly created knowledge base.

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

knowledge-base-quick-start-6zlo3

Test Delete

Knowledge base overview

Edit

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Status: Ready

Created date: July 27, 2024, 15:51 (UTC+08:00)

Service Role: AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Log Deliveries

Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags	

Please select a model Run

Test knowledge base

Generate responses

Select model

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so...	Status	Data sour...	Account ID	Source
knowledge...	Available	S3	15615387...	s3://k...

A red arrow points to the "knowledge..." entry in the table.

Embeddings model

Model: Titan Text Embeddings v2 | Vector dimensions: 1024

Vector database

Vector database: Vector engine Amazon OpenSearch Serverless | Collection ARN: arn:aws:ao:ss:us-west-2:156153878293:collection/hic9l27fgmdew61fdhyf

Vector index name: bedrock-knowledge-base-default-index | Vector field name: bedrock-knowledge-base-default-vector

Test knowledge base

Generate responses

Select model

Configure your retrieval and responses: To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model 

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- Knowledge bases
- Agents
- Prompt management [Preview](#)

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so... Status Data sour... Account ID Sour...

knowledge... Syncing	S3	15615387...	s3://k...
----------------------	----	-------------	-----------

Embeddings model

Model Titan Text Embeddings v2 Vector dimensions 1024

Vector database

Vector database Vector engine Amazon OpenSearch Serverless

Collection ARN arn:aws:aoss:us-west-2:156153878293:collection/hic9l27fgm dew61fdhyf

Vector index name bedrock-knowledge-base-default-index

Vector field name bedrock-knowledge-base-default-vector

Test knowledge base

Generate responses

Select model

The system is syncing your data source. Wait for the sync to complete before starting next sync job.

Go to data sources

Configure your retrieval and responses

To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Please select a model Run



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

AWS Services Search [Option+S]

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Oregon WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

Data source (1)

Add Edit Delete Sync

Data sources contain information returned when querying a Knowledge base.

Find data source

Data so...	Status	Data sour...	Account ID	Sourc...
knowledge...	Available	S3	15615387...	s3://k...

Embeddings model

Model: Titan Text Embeddings v2 Vector dimensions: 1024

Vector database

Vector database: Vector engine Amazon OpenSearch Serverless Collection ARN: arn:aws:ao:ss:us-west-2:156153878293:collection/hic9l27fgm dew61fdhyf

Vector index name: bedrock-knowledge-base-default-index Vector field name: bedrock-knowledge-base-default-vector

Test knowledge base

Generate responses

Select model

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Please select a model Run



Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

Anthropic

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet v1**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

Not seeing a model you are interested in? Check out all supported models [here](#)

3. Throughput

Provisioned throughput is not supported for knowledge bases.

base by running a query to disable response generation information stored from your Generate responses above.

Cancel | Apply | Run

Amazon Bedrock

Getting started
Overview
Examples
Providers

Foundation models
Base models
Custom models
Imported models [Preview](#)

Playgrounds
Chat
Text
Image

Safeguards
Guardrails
Watermark detection

Builder tools
Knowledge bases
Agents
Prompt management [Preview](#)

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1

Services | Search [Option+5] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Select model

1. Category

Model providers

Anthropic

2. Model

Models with access (5)

- Claude Instant 1.2 v1.2**
Text model | Context size = up to 100k
- Claude 2.1 v2.1**
Text model | Context size = up to 200k
- Claude 2 v2**
Text model | Context size = up to 100k
- Claude 3 Sonnet**
Text & vision model | Context size = up to 200k
- Claude 3 Haiku v1**
Text & vision model | Context size = up to 200k

3. Throughput

On-demand (ODT)

base by running a query to disable response generation information stored from your Generate responses above.

Cancel | Apply | Run

Getting started
Overview
Examples
Providers

Foundation models
Base models
Custom models
Imported models [Preview](#)

Playgrounds
Chat
Text
Image

Safeguards
Guardrails
Watermark detection

Builder tools
Knowledge bases
Agents
Prompt management [Preview](#)

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1?modelId=a...

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases [Agents](#) [Prompt management](#) [Preview](#)

knowledge-base-quick-start-6zlo3

Test Delete

Knowledge base overview

Edit

Knowledge base name	knowledge-base-quick-start-6zlo3
Knowledge base ID	GRNV74BLMR
Knowledge base description	—
Service Role	AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3
Log Deliveries	Configure log deliveries and event logs in the Edit page.

Status Ready

Created date July 27, 2024, 15:51 (UTC+08:00)

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags	

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

Enter your message here ▶ Run

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1?modelId=a... ☆

Services Search [Option+S] EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock < Amazon Bedrock > Knowledge bases knowledge-base-quick-start-6zlo3

knowledge-base-quick-start-6zlo3

Test Delete Edit

Knowledge base overview

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Knowledge base description: —

Status: Ready

Service Role: AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Created date: July 27, 2024, 15:51 (UTC+08:00)

Log Deliveries

Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags No tags to display	

Manage tags

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

在台北捷運內可不可以溜直排輪？ 

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1?modelId=a... ☆

Services Search [Option+S] | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

EC2 VPC RDS S3 Support Amazon SageMaker AWS DeepRacer CloudFormation

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

knowledge-base-quick-start-6zlo3

Test Delete Edit

Knowledge base overview

Knowledge base name: knowledge-base-quick-start-6zlo3

Knowledge base ID: GRNV74BLMR

Knowledge base description: —

Status: Ready

Service Role: AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3

Created date: July 27, 2024, 15:51 (UTC+08:00)

Log Deliveries

Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags	

No tags to display

Manage tags

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT Change

Configure your retrieval and responses To customize the search strategy for your knowledge base, select the configurations icon .

Test your knowledge base by running a query to generate responses. To disable response generation and only see retrieved information stored from your vector store, turn off Generate responses above.

在台北捷運內可不可以溜直排輪?

Run

Amazon Bedrock Workshop | TCC-Bedrock-Dryrun | Amazon Bedrock | us-west-2

us-west-2.console.aws.amazon.com/bedrock/home?region=us-west-2#/knowledge-bases/knowledge-base-quick-start-6zlo3/GRNV74BLMR/1?modelId=a...

Services | Search [Option+S] | AWS | EC2 | VPC | RDS | S3 | Support | Amazon SageMaker | AWS DeepRacer | CloudFormation | Oregon | WSParticipantRole/Participant @ 1561-5387-8293

Amazon Bedrock

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

Knowledge bases

- Agents
- Prompt management [Preview](#)

knowledge-base-quick-start-6zlo3

Knowledge base overview

[Edit](#)

Knowledge base name knowledge-base-quick-start-6zlo3	Knowledge base ID GRNV74BLMR
Knowledge base description —	Status Ready
Service Role AmazonBedrockExecutionRoleForKnowledgeBase_6zlo3	Created date July 27, 2024, 15:51 (UTC+08:00)

Log Deliveries

Configure log deliveries and event logs in the [Edit](#) page.

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
No tags No tags to display.	

Test knowledge base

Generate responses

Claude 3 Sonnet v1 | ODT [Change](#)

Configure your retrieval and responses
To customize the search strategy for your knowledge base, select the configurations icon .

 在台北捷運內可不可以溜直排輪？

 根據台北捷運系統旅客須知的規定,在捷運範圍內是不允許使用直排輪的。具體來說,第8條明確規定"在捷運範圍內,不得有下列行為:(一)使用直排輪、溜冰鞋、滑板、滑板車或其他類似器材。"因此,在台北捷運內是禁止溜直排輪的。[\[1\]](#)

[Show source details >](#)

Enter your message here 

Amazon Bedrock



Amazon Bedrock > Knowledge bases > knowledge-base-quick-start-6zlo3



Test Knowledge base

Getting started

- Overview
- Examples
- Providers

Foundation models

- Base models
- Custom models
- Imported models [Preview](#)

Playgrounds

- Chat
- Text
- Image

Safeguards

- Guardrails
- Watermark detection

Builder tools

- [Knowledge bases](#)
- Agents
- Prompt management [Preview](#)
- Prompt flows [Preview](#)

Assessment & deployment

[CloudShell](#) [Feedback](#)



Generate responses



Claude 3 Sonnet v1 | ODT

[Change](#)

Configure your retrieval and responses

To customize the search strategy for your knowledge base, select the configurations icon



在台北捷運內可不可以溜直排輪？



根據台北捷運系統旅客須知的規定,在捷運範圍內是不允許使用直排輪的。具體來說,第8條明確規定"在捷運範圍內,不得有下列行為:(一)使用直排輪、溜冰鞋、滑板、滑板車或其他類似器材。"因此,在台北捷運內是禁止溜直排輪的。[\[1\]](#)

[< Hide source details](#)

Source details (1)

Each container below shows a chunk that comes from a data source and that relates to your query.

Query configurations (1)

Source chunk 1

(七) 其他經本公司公告禁止之事項。八、在捷運範圍內，不得有下列行為：（一）使用直排輪、溜冰鞋、滑板、滑板車或其他類似器材。（二）推擠或影響自己或他人安全。（三）坐、臥於車廂、車站地板、設備或設施上。北捷024184臺北捷運系統旅客須知 3 （四）吸菸。（五）其他行為有造成大眾捷運系統設備損壞、運轉障礙或構成危險之虞。九、在捷運範圍內為下列行為，應向本公司申請許可後，始得为之：（一）聚眾講演、播放音響、演奏樂器或其他干擾之行為。（二）張貼、塗抹、刻畫任何文字、圖畫或其他類似東西於各項設施及建築物上。（三）於車站或車廂內，照相、拍攝或攝影，而妨礙他人或系統安全之虞者。（四）非營運時間內，於車站或車廂內逗留。（五）向他人為傳教、市場調查或其他類似行為。（六）散發報紙、傳單、廣告物或宣傳品。（七）使用車站、車廂內未開放使用之電源插座。十、在捷運範圍內，電扶梯之搭乘規定及注意事項如下：（一）年長及行動不便者宜改搭電梯。（二）須遵循電扶梯方向搭乘，握好扶手、站穩踏階，勿倚靠側板。（三）禁止於電扶梯上奔跑、嬉戲、跳躍、跨越兩側護欄或其他危險行為。（四）禁止攜帶大型物品搭乘電扶梯。貳、車票使用規定十一、車票種類：（一）單程票：提供旅客單次使用之車票。（二）團體票：提供旅客 2人以上，全程同行且起訖站相同使用之車票。（三）定期票：提供旅客於一定期間內使用之車票。（四）回數票：可供旅客於一定區間或不限區間搭乘一定次數之車票。（五）儲值卡：各發行機構合法發行，經本公司同意使用之車票。（六）其他種類車票。

Labs

- Get hands-on with Bedrock playground
- Set up a Q&A application with knowledge retrieval
- Build a LLM application with Bedrock SDKs



© 2024, Amazon Web Services, Inc. or its affiliates.

catalog.us-east-1.prod.workshops.aws/event/dashboard/en-US/workshop/60-chatbot/61-chatbot-with-models

aws workshop studio michael_tw_lin

moda-workshop-dryrun-20241125

Amazon Bedrock Workshop - Moderated

- ▶ Prerequisites
- ▶ Prompt Engineering
- ▶ Text Generation
- ▶ Text Summarization
- ▶ Question Answering
- ▼ Conversational Interfaces - Chatbots
 - Chatbot using Bedrock Models**
 - ▶ Image Generation

▶ AWS account access

- [Open AWS console \(us-west-2\)](#)
- [Get AWS CLI credentials](#)

Exit event

Event dashboard > [Conversational Interfaces - Chatbots](#) > Chatbot using Bedrock Models

Chatbot using Bedrock Models

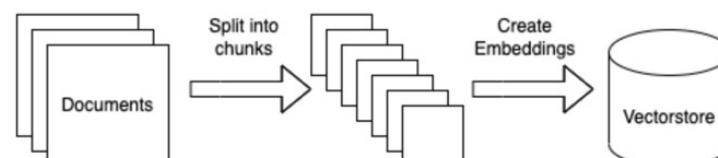
Overview

In this module, you will work with three distinct models on Bedrock to power your chatbot application.

- [Claude](#) is a LLM for conversations, question answering, and workflow automation that is accessible through an API. Claude is capable of a wide variety of conversational and text processing tasks while maintaining a high degree of reliability and predictability.
- [Amazon Titan Text](#) is pre-trained on large datasets, making it a powerful, general-purpose model. Use it as is or privately customize it with your own data for a particular task without annotating large volumes of data. Amazon Titan Text is a generative large language model (LLM) for tasks such as summarization, text generation (for example, creating a blog post), classification, open-ended Q&A, and information extraction.

Architecture

The first process in building a contextual-aware chatbot is to generate embeddings for the context. Typically, you will have an ingestion process which will run through your embedding model and generate the embeddings which will be stored in a sort of vector store. In this example we are using the Amazon Titan Embeddings model:



AWS Services Search: sagemaker|Canvas Oregon WSParticipantRole/Participant @ 5407-0146-3772

Support Builder tools Services Features Resources New Documentation Amazon SageMaker Build, Train, and Deploy Machine Learning Models

Safeguards Guardrails Marketplace Show more ▶

Inference Provisioned Throughput Events Tutorials

Assessment Model Evaluation

User guide Bedrock Service Terms

Bedrock configuration Model access Preview Settings

Introducing resource search To search for resources, Resource Explorer must be active in at least one AWS Region and you must have permission to use the default view in the account. Learn more Access granted

Titan Multimodal Embeddings G1 Embedding EULA

CloudShell Feedback © 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

moda-workshop-dryrun-2021

Amazon SageMaker | us-west-2

us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/landing

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 5407-0146-3772

Support

Amazon SageMaker

MACHINE LEARNING

Getting started

Applications and IDEs

- Studio **←**
- Canvas
- RStudio
- TensorBoard
- Profiler
- Notebooks

Admin configurations

- Domains
- Role manager
- Images
- Lifecycle configurations

SageMaker dashboard

Search

Amazon SageMaker

Build, train, and deploy machine learning models at scale

The quickest and easiest way to get ML models from idea to production.

How it works

What is Amazon SageMaker?

New to SageMaker?

Quick setup for a single user **i**
This is perfect for first time users to try capabilities in just a few clicks.

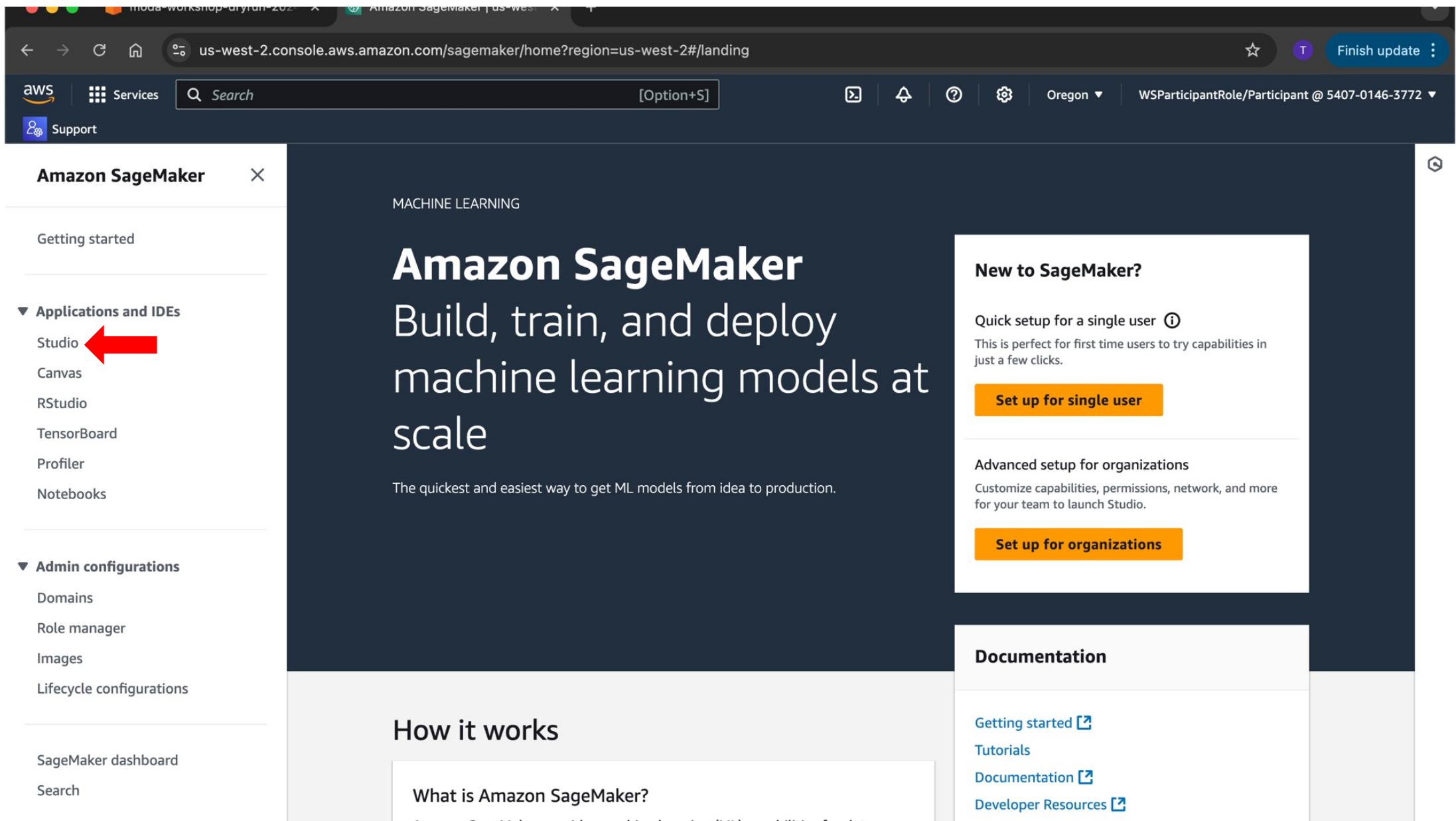
Set up for single user

Advanced setup for organizations
Customize capabilities, permissions, network, and more for your team to launch Studio.

Set up for organizations

Documentation

Getting started **↗**
Tutorials
Documentation **↗**
Developer Resources **↗**



us-west-2.console.aws.amazon.com/sagemaker/home?region=us-west-2#/studio-landing

Finish update :

AWS Services Search [Option+S] Oregon WSParticipantRole/Participant @ 5407-0146-3772

Support

Amazon SageMaker X

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 16 hours, 46 minutes) [Learn more](#)

We will automatically migrate all accounts to the new Studio experience, featuring JupyterLab 4-based notebooks, in phased rollouts before this date. You can manually migrate to the new Studio before this date and take advantage of the updated features. Your EFS data will automatically be available in the new Studio.

Getting started

▼ Applications and IDEs

- Studio
- Canvas
- RStudio
- TensorBoard
- Profiler
- Notebooks

▼ Admin configurations

- Domains
- Role manager
- Images
- Lifecycle configurations

SageMaker dashboard

Search

▼ JumpStart

Amazon SageMaker

SageMaker Studio

The first fully integrated development environment (IDE) for machine learning.

Get Started

Select user profile SageMakerUser

Open Studio ←

How it works

What is Studio?

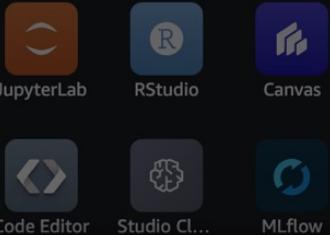
Amazon SageMaker Studio provides a single, web-based visual interface where you can perform all ML development steps, improving data science

Pricing (US)

View All SageMaker Studio Pricing



Applications (6)



Home

Running instances

Data

Auto ML

Experiments

Jobs

Pipelines

Models

JumpStart

Collapse Menu

Home

Launch new workflows, open getting started materials, and view the latest feature updates

Onboarding

To get the most ou



Welcome to the new

SageMaker Studio

We've built a new experience to empower you and your work.

Want to take a quick tour?

[Skip Tour for now](#)

[Take a quick tour](#)



Are you an existing Studio Classic user and looking to migrate
your data and notebooks? Click here to learn how.



← → G ⌂ studio-d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/home?tab=Overview ⭐ T Finish update :

SageMaker Studio

Applications (6)

- JupyterLab
- RStudio
- Canvas
- Code Editor
- Studio Cl...
- MLflow

Studio Classic

Home

Launch new workflows, open getting started materials, and view the latest feature updates

Onboarding plan

To get the most out of the new Studio experience, explore the onboarding steps below.

Take the tour

Quick tour highlights where you can find key features and how to navigate the new experience. See what's new and where to locate the tools you need to be productive.

[Take the tour >](#)

Access your EFS data in JupyterLab and CodeEditor

Automatically available in private spaces.

[Try JupyterLab](#)

Access your Studio Classic apps

Pickup where you left off and access your Studio Classic apps from within the updated Studio experience.

[View Studio Classic](#)

Not ready to use the new experience? Revert to Studio Classic experience in domain settings. [Learn more ↗](#)

Overview Getting started What's new

Overview

Start a new ML workflow or jump back into your workflow



SageMaker Studio | Amazon Sagemaker Studio Classic - SageMaker S +

studio-d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/studio-classic

Finish update

SageMaker Studio > Studio Classic

Provide feedback

Applications (6)

- JupyterLab
- RStudio
- Canvas
- Code Editor
- Studio Cl...
- MLflow

⚠ Studio Classic will reach end of maintenance on December 31, 2024
We recommend using JupyterLab 4, which offers a faster and more reliable experience. Your EFS data will automatically be available for use.

Try JupyterLab

SageMaker Studio Classic

+ Create Studio Classic space

▼ About

Studio Classic is a legacy IDE that allows you to access the previous iteration of SageMaker Studio from within the new Studio experience.

See features | Learn more about Studio Classic

Search...

Filter spaces: Running

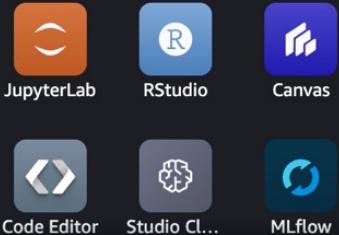
Name	Application	Status	Type	Last modified	Action
SageMakerUser	Studio Classic	Stopped	Private	0 seconds ago	Run

1 results Results are cached Refresh Go to page 1 Page 1 of 1 < >

Introducing spaces New



Applications (6)



Home

Running instances

Data

Auto ML

Experiments

Jobs

Pipelines

Models

JumpStart

[Collapse Menu](#)**⚠ Studio Classic will reach end of maintenance on December 31, 2024**

We recommend using JupyterLab 4, which offers a faster and more reliable experience. Your EFS data will automatically be available for use.

[Try JupyterLab](#)

SageMaker Studio Classic

[+ Create Studio Classic space](#)

About

Studio Classic is a legacy IDE that allows you to access the previous iteration of SageMaker Studio from within the new Studio experience.

[See features](#) | [Learn more about Studio Classic](#) Search...Filter spaces:

Name	Application	Status	Type	Last modified	Action
SageMakerUser	Studio Classic	✓ Running	🔒 Private	3 minutes ago	Stop Open

1 results Results are cached ⟳ Refresh Go to page Page 1 of 1 < >

Introducing spaces New

JupyterLab and Code Editor now have their own separate spaces, with improved options, and configurable storage.

✓ Starting SageMakerUser in progress [Learn more](#)

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 33 minutes)

[Learn More ↗](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic

File

Run Kernel Git Tabs Settings Help



SageMakerUser / Personal Studio



Home



Data



AutoML



Geospatial ML New



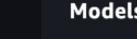
Experiments



Notebook jobs



Pipelines



Models

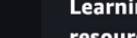


Deployments



SageMaker

JumpStart



Learning

resources



Home



Home

[Customize layout](#)

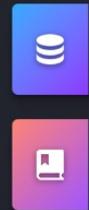


Quick actions



Open Launcher

Create notebooks and other resources



Import & prepare data visually



Open the Getting Started notebook



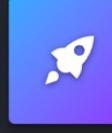
Read documentation



View guided tutorials

Prebuilt and automated solutions

Deploy built-in algorithms, pre-built solutions, example notebooks, and build models from visual interface.



JumpStart

Pretrained models, notebooks, and prebuilt solutions



AutoML

Automatically build, train, and tune the best ML models

Workflows and tasks

Kick off a new step in the machine learning workflow.

Prepare data

Build, train, tune model

Deploy model

← → G ⌂ d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab? ⌂ Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 33 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic

File Edit View Run Kernel Git Tabs Settings Help

Home X

New

- Console
- Notebook
- Data Wrangler Flow
- Example Notebook
- Terminal 
- Text File
- Markdown File
- Python File
- Model group
- Create AutoML experiment

Customize layout

Home

Quick actions

- + Open Lab Create notebook
- Close Tab ⌘ W
- Close and Shutdown ⌘ ⌘ Q
- Close All Tabs
- Save ⌘ S
- Save As... ⌘ ⌘ S
- Save All
- Reload from Disk
- Revert to Checkpoint
- Rename...
- Download
- Save and Export Notebook As... ↗
- Save Current Workspace As...
- Save Current Workspace
- Print... ⌘ P
- Log Out
- Shut Down

Open the Getting Started notebook

View guided tutorials

JumpStart Pretrained model

AutoML Automatically build, train, and tune the best ML models

Workflows and tasks Kick off a new step in the machine learning workflow.

Prepare data Build, train, tune model Deploy model

The screenshot shows a web browser window with four tabs open:

- moda-workshop-dryrun-2024
- SageMaker Studio | Amazon
- Studio Classic - SageMaker
- lab - JupyterLab

The address bar shows the URL: `d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab`.

A prominent yellow banner at the top of the page reads:

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 33 minutes) [Learn More](#)

The main content area is a terminal window titled "Terminal 1". It contains the following text:

```
!!!!!! Welcome to SageMaker Studio System Terminal !!!!!!!

Below are some useful tips:

* Activate studio conda environment using "conda activate studio" to install extensions, list extension etc. For more information, see https://docs.aws.amazon.com/sagemaker/latest/dg/studio-jl.html#studio-jl-install
* Post JupyterServer extension installation, if needed, restart just the server(not app) using "restart-jupyter-server"

sagemaker-user@studio$
```

SageMaker Studio | Amazon | Studio Classic - SageMaker | lab - JupyterLab

catalog.us-east-1.prod.workshops.aws/event/dashboard/en-US/workshop/20-intro

aws workshop studio

michael_tw_lin

moda-workshop-dryrun-20241125

Amazon Bedrock Workshop -
Moderated

▼ Prerequisites

- ▶ Environment Setup
- ▶ Prompt Engineering
- ▶ Text Generation
- ▶ Text Summarization
- ▶ Question Answering
- ▶ Conversational Interfaces - Chatbots
- ▶ Image Generation

▼ AWS account access

[Open AWS console
\(us-west-2\)](#)

[Get AWS CLI credentials](#)

[Exit event](#)

Repository Download (code samples)

Clone the following repository, which includes examples showcased in the workshop:

```
1 git clone https://github.com/aws-samples/amazon-bedrock-workshop
```

Copied!



Navigate the workshop artifacts

1. In the left hand side navigation, click the Folder icon. You should see a folder called `amazon-bedrock-workshop`. This is where you will run your notebooks.

IAM Policy for Amazon Bedrock

The following IAM policy is a sample that should be created to grant access to Bedrock APIs:

```
1 {
2     "Version": "2012-10-17",
3     "Statement": [
4         {
5             "Sid": "Statement1",
6             "Effect": "Allow",
7             "Action": "bedrock:*",
8             "Resource": "*"
9         }
10    ]
11 }
```



The screenshot shows a browser window with four tabs open:

- moda-workshop-dryrun-2024
- SageMaker Studio | Amazon
- Studio Classic - SageMaker
- lab - JupyterLab

The address bar shows the URL: `d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab`.

A prominent yellow warning banner at the top states: "JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 33 minutes)" with a "Learn More" link.

The main content area is a terminal window titled "Terminal 1". It displays the following text:

```
!!!!!! Welcome to SageMaker Studio System Terminal !!!!!!!

Below are some useful tips:

* Activate studio conda environment using "conda activate studio" to install extensions, list extension etc. For more information, see https://docs.aws.amazon.com/sagemaker/latest/dg/studio-jl.html#studio-jl-install
* Post JupyterServer extension installation, if needed, restart just the server(not app) using "restart-jupyter-server"

sagemaker-user@studio$ git clone https://github.com/aws-samples/amazon-bedrock-workshop
Cloning into 'amazon-bedrock-workshop'...
remote: Enumerating objects: 3210, done.
remote: Counting objects: 100% (165/165), done.
remote: Compressing objects: 100% (110/110), done.
remote: Total 3210 (delta 90), reused 109 (delta 55), pack-reused 3045 (from 1)
Receiving objects: 100% (3210/3210), 40.93 MiB | 34.02 MiB/s, done.
Resolving deltas: 100% (2118/2118), done.
Updating files: 100% (184/184), done.
sagemaker-user@studio$
```

A large red arrow points to the first line of the terminal output: "sagemaker-user@studio\$ git clone https://github.com/aws-samples/amazon-bedrock-workshop".

The screenshot shows a web browser window with the address bar pointing to `d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab`. The page title is "SageMaker Studio Classic - SageMaker". A prominent yellow banner at the top states: "JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 32 minutes)" with a link to "Learn More". Below the banner, a message says: "We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in the new Studio." The main content area is titled "Terminal 1" and displays the following text:

```
!!!!!! Welcome to SageMaker Studio System Terminal !!!!!!!  
Below are some useful tips:  
* Activate studio conda environment using "conda activate studio" to install extensions, list extension etc. For more information, see https://docs.aws.amazon.com/sagemaker/latest/dg/studio-jl.html#studio-jl-install  
* Post JupyterServer extension installation, if needed, restart just the server(not app) using "restart-jupyter-server"  
  
sagemaker-user@studio$ git clone https://github.com/aws-samples/amazon-bedrock-workshop  
Cloning into 'amazon-bedrock-workshop'...  
remote: Enumerating objects: 3210, done.  
remote: Counting objects: 100% (165/165), done.  
remote: Compressing objects: 100% (110/110), done.  
remote: Total 3210 (delta 90), reused 109 (delta 55), pack-reused 3045 (from 1)  
Receiving objects: 100% (3210/3210), 40.93 MiB | 34.02 MiB/s, done.  
Resolving deltas: 100% (2118/2118), done.  
Updating files: 100% (184/184), done.  
sagemaker-user@studio$
```

The left sidebar contains icons for Home, Create New, Upload, Filter files by name, and a folder icon with a red arrow pointing to it. The "Name" column in the file browser lists "amazon-bedrock-wor". The top navigation bar includes File, Edit, View, Run, Kernel, Git, Tabs, Settings, Help, and a user profile icon.

The screenshot shows the Amazon SageMaker Studio Classic interface. At the top, there are four tabs: "moda-workshop-dryrun-2024" (closed), "SageMaker Studio | Amazon S" (closed), "Studio Classic - SageMaker S" (closed), and "lab - JupyterLab" (active). Below the tabs is a browser-style address bar with the URL "d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab". To the right of the address bar are icons for "Finish update" and a menu.

A prominent yellow banner at the top states: "JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 32 minutes)" with a link to "Learn More". Below the banner, a message says: "We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in the new Studio." A "Learn More" button is also present here.

The main area is a JupyterLab interface. On the left is a sidebar with icons for Home, Filter files by name, and a file tree showing a single folder named "amazon-bedrock-wor...". A red arrow points from the text below to this folder icon. The central area has a "Home" tab and a "Terminal 1" tab. The "Terminal 1" tab is active, displaying the following text:

```
!!!!!! Welcome to SageMaker Studio System Terminal !!!!!!!

Below are some useful tips:
* Activate studio conda environment using "conda activate studio" to install extensions, list extension etc. For more information, see https://docs.amazonaws.com/sagemaker/latest/dg/studio-jl.html#studio-jl-install

* Post JupyterServer extension installation, if needed, restart just the server(not app) using "restart-jupyter-server"

sagemaker-user@studio$ git clone https://github.com/aws-samples/amazon-bedrock-workshop
Cloning into 'amazon-bedrock-workshop'...
remote: Enumerating objects: 3210, done.
remote: Counting objects: 100% (165/165), done.
remote: Compressing objects: 100% (110/110), done.
remote: Total 3210 (delta 90), reused 109 (delta 55), pack-reused 3045 (from 1)
Receiving objects: 100% (3210/3210), 40.93 MiB | 34.02 MiB/s, done.
Resolving deltas: 100% (2118/2118), done.
Updating files: 100% (184/184), done.
sagemaker-user@studio$
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon Sagemaker Studio Classic - SageMaker S 00_Prerequisites - JupyterLab +

← → C ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites ⌂ Finish update ⌂

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 32 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1

!!!!!! Welcome to SageMaker Studio System Terminal !!!!!!!

Below are some useful tips:

- * Activate studio conda environment using "conda activate studio" to install extensions, list extension etc. For more information, see <https://docs.aws.amazon.com/sagemaker/latest/dg/studio-jl.html#studio-jl-install>
- * Post JupyterServer extension installation, if needed, restart just the server(not app) using "restart-jupyter-server"

```
sagemaker-user@studio$ git clone https://github.com/aws-samples/amazon-bedrock-workshop
Cloning into 'amazon-bedrock-workshop'...
remote: Enumerating objects: 3210, done.
remote: Counting objects: 100% (165/165), done.
remote: Compressing objects: 100% (110/110), done.
remote: Total 3210 (delta 90), reused 109 (delta 55), pack-reused 3045 (from 1)
Receiving objects: 100% (3210/3210), 40.93 MiB | 34.02 MiB/s, done.
Resolving deltas: 100% (2118/2118), done.
Updating files: 100% (184/184), done.
sagemaker-user@studio$
```

Filter files by name / amazon-bedrock-workshop / 00_Prerequisites /

Name	Last Modified
bedrock_basics.ipynb	a minute ago
Getting_started_with_Convers...	a minute ago
README.md	a minute ago

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 31 minutes) ⓘ

Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic

File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio



Filter files by name

/ amazon-bedrock-workshop / 00_Prerequisites /

Name	Last Modified
bedrock_basics.ipynb	a minute ago
Getting_started_with_Convers...	a minute ago
README.md	a minute ago

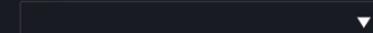
Amazon Bedrock boto3 Prerequisites

This notebook should work well with the Data Science 3.0 kernel in SageMaker Studio

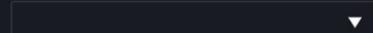
Set up notebook environment

Set up environment for "bedrock_basics.ipynb".

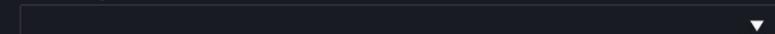
Image



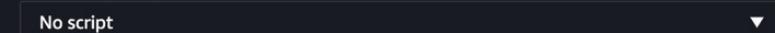
Kernel



Instance type



Start-up script ⓘ



Cancel

Select

running this workshop at an AWS event, it is likely you have limited access to third party models.

Prerequisites

Run the cells in this section to install the packages needed by the notebooks in this workshop. ⚠ You will see pip dependency errors, you can safely ignore these errors. ⚠

IGNORE ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 31 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb

Filter files by name / amazon-bedrock-workshop / 00_Prerequisites /

Name Last Modified

- bedrock_basics.ipynb a minute ago
- Getting_started_with_Convers... a minute ago
- README.md a minute ago

Amazon Bedrock boto3 Prerequisites

This notebook should work well with the Data Science 3.0 kernel in SageMaker Studio

Set up notebook environment

Set up environment for "bedrock_basics.ipynb".

Image: Data Science 3.0 Data Science 3.0

Kernel: Python 3

Instance type: ml.t3.medium

Start-up script: No script

Cancel Select

You have manually enabled access to the following models

running this workshop at an AWS event, it is likely you have limited access to third party models.

Prerequisites

Run the cells in this section to install the packages needed by the notebooks in this workshop. ⚠️ You will see pip dependency errors, you can safely ignore these errors. ⚠️

IGNORE ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.

Starting notebook kernel...

Amazon Bedrock boto3 Prerequisites

This notebook should work well with the **Data Science 3.0** kernel in SageMaker Studio

In this demo notebook, we demonstrate how to use the `boto3` Python SDK to work with Amazon Bedrock Foundation Models.

⚠ Before you get started with this workshop... make sure you have access you have manually enabled access to the following models in the "Model access" tab:

- All Titan Models
- Claude 3 Haiku
- Claude 3 Sonnet
- Stable Diffusion
- Llama 2 13B
- Llama 2 70B

Remember that you can modify model access at any point of time, but if you are running this workshop at an AWS event, it is likely you have limited access to third party models.

Prerequisites

Run the cells in this section to install the packages needed by the notebooks in this workshop. **⚠ You will see pip dependency errors, you**

Simple (1) \$ 0 main No Kernel | Initializing Kernel: Starting... | Instance MEM

Cookie Preferences Mode: Command Ln 1, Col 1 bedrock_basics.ipynb 0

Home Terminal 1 bedrock_basics.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Filter files by name

Name Last Modified

- bedrock_basics.ipynb 6 minutes ago
- Getting_started_with_Convers... 14 minutes ago
- README.md 14 minutes ago

Amazon Bedrock boto3 Prerequisites

*This notebook should work well with the **Data Science 3.0** kernel in SageMaker Studio*

In this demo notebook, we demonstrate how to use the `boto3` Python SDK to work with Amazon Bedrock Foundation Models.

⚠ Before you get started with this workshop... make sure you have access you have manually enabled access to the following models in the "Model access" tab:

- All Titan Models
- Claude 3 Haiku
- Claude 3 Sonnet
- Stable Diffusion
- Llama 2 13B
- Llama 2 70B

Remember that you can modify model access at any point of time, but if you are running this workshop at an AWS event, it is likely you have limited access to third party models.

Prerequisites

Run the cells in this section to install the packages needed by the notebooks in this workshop. **⚠ You will see pip dependency errors, you can safely ignore these errors. ⚠**

IGNORE ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflicts.

Simple 1 \$ 1 main Python 3 (Data Science 3.0) | Idle Kernel: Idle | Instance MEM

Cookie Preferences Mode: Command Ln 1, Col 1 bedrock_basics.ipynb 0

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab | +

← → C ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb ⌂ Finish update ⌂

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 19 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

Filter files by name / amazon-bedrock-workshop / 00_Prerequisites /

Name Last Modified

- bedrock_basics.ipynb 7 minutes ago
- Getting_started_with_Convers... 14 minutes ago
- README.md 14 minutes ago

Terminal 1

```
[*]: %pip install --no-build-isolation --force-reinstall \
    "boto3>=1.28.57" \
    "awscli>=1.29.57" \
    "botocore>=1.31.57"

Collecting boto3>=1.28.57
  Downloading boto3-1.35.68-py3-none-any.whl.metadata (6.7 kB)
Collecting awscli>=1.29.57
  Downloading awscli-1.36.9-py3-none-any.whl.metadata (11 kB)
Collecting botocore>=1.31.57
  Downloading botocore-1.35.68-py3-none-any.whl.metadata (5.7 kB)
Collecting jmespath<2.0.0,>=0.7.1 (from boto3>=1.28.57)
  Downloading jmespath-1.0.1-py3-none-any.whl.metadata (7.6 kB)
Collecting s3transfer<0.11.0,>=0.10.0 (from boto3>=1.28.57)
  Downloading s3transfer-0.10.4-py3-none-any.whl.metadata (1.7 kB)
Collecting docutils<0.17,>=0.10 (from awscli>=1.29.57)
  Downloading docutils-0.16-py2.py3-none-any.whl.metadata (2.7 kB)
Collecting PyYAML<6.1,>=3.10 (from awscli>=1.29.57)
  Downloading PyYAML-6.0.2-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (2.1 kB)
Collecting colorama<0.4.7,>=0.2.5 (from awscli>=1.29.57)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Collecting rsa<4.8,>=3.1.2 (from awscli>=1.29.57)
  Downloading rsa-4.7.2-py3-none-any.whl.metadata (3.6 kB)
Collecting python-dateutil<3.0.0,>=2.1 (from botocore>=1.31.57)
  Downloading python_dateutil-2.3.0-py3-none-any.whl.metadata (2.4 kB)
```

Create the boto3 client

Interaction with the Bedrock API is done via the AWS SDK for Python: [boto3](#).

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab +

← → C ⌘ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb ⌘ T Finish update ⌋

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 18 minutes) ⓘ [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

AmazonUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

Create the boto3 client

Interaction with the Bedrock API is done via the AWS SDK for Python: [boto3](#).

Use different clients

The `boto3` provides different clients for Amazon Bedrock to perform different actions. The actions for `InvokeModel` and `InvokeModelWithResponseStream` are supported by Amazon Bedrock Runtime where as other operations, such as `ListFoundationModels`, are handled via Amazon Bedrock client.

Use the default credential chain

If you are running this notebook from [Amazon Sagemaker Studio](#) and your Sagemaker Studio execution role has permissions to access Bedrock you can just run the cells below as-is. This is also the case if you are running these notebooks from a computer whose default AWS credentials have access to Bedrock.

```
[3]: import json
import os
import sys

import boto3

boto3_bedrock = boto3.client('bedrock')
```

Validate the connection

We can check the client works by trying out the `list.foundation_models()` method, which will tell us all the models available for us to use

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | bedrock_basi... - JupyterLab +

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb

Finish update

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 18 minutes)

Learn More

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

Validate the connection

We can check the client works by trying out the `list.foundation_models()` method, which will tell us all the models available for us to use

```
[4]: [models['modelId'] for models in boto3_bedrock.list.foundation_models()['modelSummaries']]
```

```
[4]: ['amazon.titan-tg1-large',
      'amazon.titan-embed-g1-text-02',
      'amazon.titan-text-lite-v1:0:4k',
      'amazon.titan-text-lite-v1',
      'amazon.titan-text-express-v1:0:8k',
      'amazon.titan-text-express-v1',
      'amazon.titan-embed-text-v1:2:8k',
      'amazon.titan-embed-text-v1',
      'amazon.titan-embed-text-v2:0',
      'amazon.titan-embed-image-v1:0',
      'amazon.titan-embed-image-v1',
      'amazon.titan-image-generator-v1:0',
      'amazon.titan-image-generator-v1',
      'amazon.titan-image-generator-v2:0',
      'stability.stable-diffusion-xl-v1:0',
      'stability_stable-diffusion-xl-v1']
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab | +

← → C ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb ⌂ Finish update ⌂

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 18 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

+ ☰ ⌂ ⌂ Filter files by name

/ amazon-bedrock-workshop / 00_Prerequisites /

Name Last Modified

- bedrock_basics.ipynb 7 minutes ago
- Getting_started_with_Convers... 15 minutes ago
- README.md 15 minutes ago

Home Terminal 1 bedrock_basics.ipynb

InvokeModel body and output

The `invoke_model()` method of the Amazon Bedrock runtime client (`InvokeModel` API) will be the primary method we use for most of our Text Generation and Processing tasks - whichever model we're using.

Although the method is shared, the format of input and output varies depending on the foundation model used - as described below:

▶ **Amazon Titan Large and Premier**

Input

```
{  
    "inputText": "<prompt>",  
    "textGenerationConfig" : {  
        "maxTokenCount": 512,  
        "stopSequences": [],  
        "temperature": 0.1,  
        "topP": 0.9  
    }  
}
```

Output

```
{  
    "inputTextTokenCount": 613,  
    "results": [  
        {"tokenCount": 219,  
         "outputText": "<output>"  
    ]  
}
```

Feedback icon

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab +

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 18 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Help

SageMakerUser / Personal Studio

Try out the models

With some theory out of the way, let's see the models in action! Run the cells below to see basic, synchronous example invocations for each model:

```
[5]: import boto3
import botocore
import json

bedrock_runtime = boto3.client('bedrock-runtime')
```

Amazon Titan Text Premier

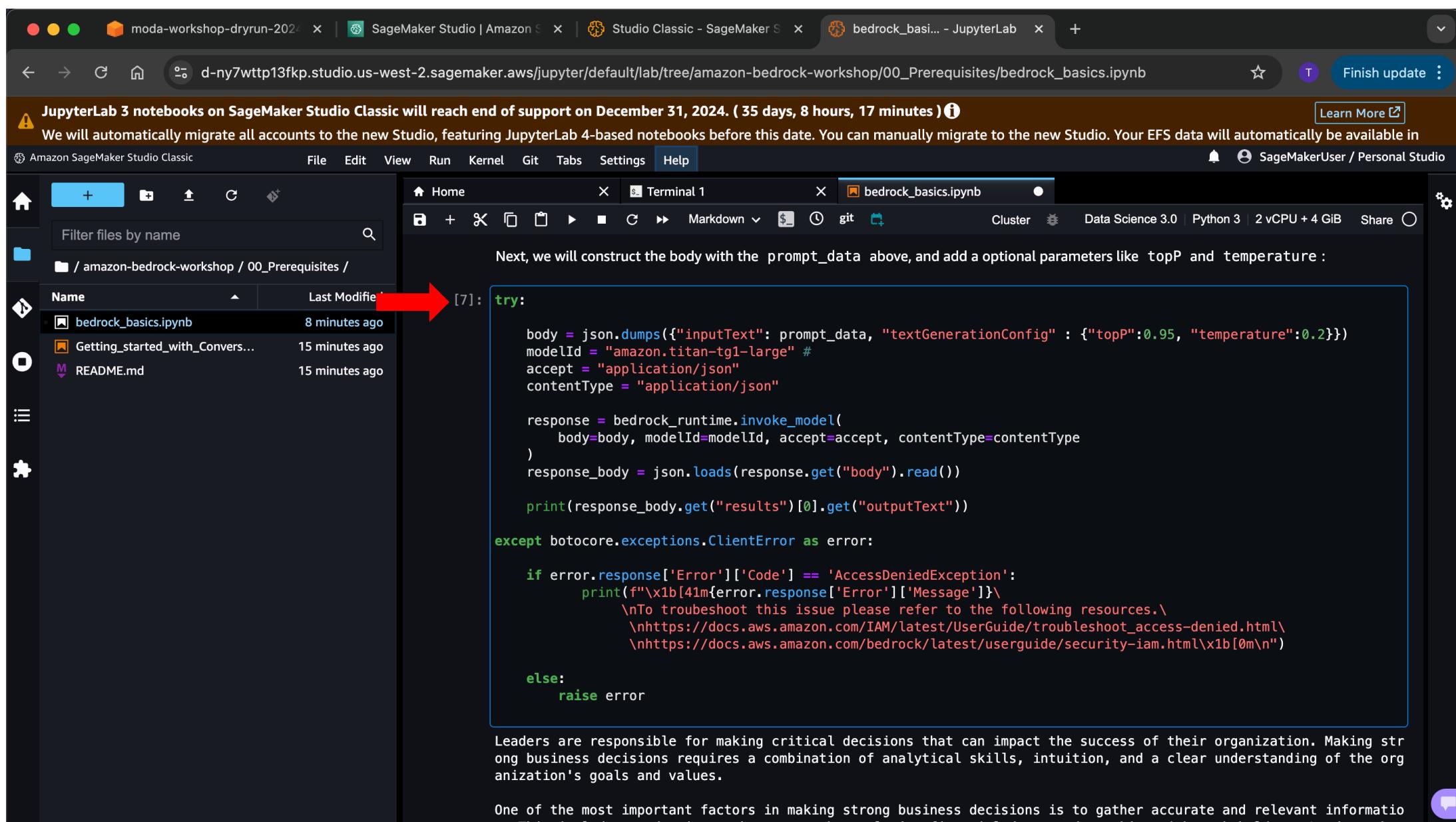
```
[6]: # If you'd like to try your own prompt, edit this parameter!
prompt_data = """Command: Write me a blog about making strong business decisions as a leader.

Blog:
"""

Next, we will construct the body with the prompt_data above, and add a optional parameters like topP and temperature :
```

```
[7]: try:
```





SageMaker Studio | Amazon | Studio Classic - SageMaker | bedrock_basi... - JupyterLab +

moda-workshop-dryrun-2024 | d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 17 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb

Anthropic Claude (messages API)

[8]:

```
# If you'd like to try your own prompt, edit this parameter!
prompt_data = """Human: Write me 500 word paragraph about making strong business decisions as a leader.

Assistant:
"""
```

[9]:

```
messages_API_body = {
    "anthropic_version": "bedrock-2023-05-31",
    "max_tokens": int(500/0.75),
    "messages": [
        {
            "role": "user",
            "content": [
                {
                    "type": "text",
                    "text": prompt_data
                }
            ]
        }
    ]
}
```

[10]:

```
from IPython.display import clear_output, display, display_markdown, Markdown

body = json.dumps(messages_API_body)
modelId = "anthropic.claude-3-haiku-20240307-v1:0" # (Change this to try different model versions)
accept = "application/json"
contentType = "application/json"
```

+

Filter files by name / amazon-bedrock-workshop / 00_Prerequisites /

Name Last Modified

- bedrock_basics.ipynb seconds ago
- Getting_started_with_Convers... 15 minutes ago
- README.md 15 minutes ago

File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb

Anthropic Claude (messages API)

[8]:

```
# If you'd like to try your own prompt, edit this parameter!
prompt_data = """Human: Write me 500 word paragraph about making strong business decisions as a leader.

Assistant:
"""
```

[9]:

```
messages_API_body = {
    "anthropic_version": "bedrock-2023-05-31",
    "max_tokens": int(500/0.75),
    "messages": [
        {
            "role": "user",
            "content": [
                {
                    "type": "text",
                    "text": prompt_data
                }
            ]
        }
    ]
}
```

[10]:

```
from IPython.display import clear_output, display, display_markdown, Markdown

body = json.dumps(messages_API_body)
modelId = "anthropic.claude-3-haiku-20240307-v1:0" # (Change this to try different model versions)
accept = "application/json"
contentType = "application/json"
```

SageMaker Studio | Amazon | Studio Classic - SageMaker | bedrock_basi... - JupyterLab +

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 17 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb

Anthropic Claude (messages API)

[8]: *# If you'd like to try your own prompt, edit this parameter!*
prompt_data = """Human: Write me 500 word paragraph about making strong business decisions as a leader.

Assistant:
"""

[9]: messages_API_body = {
 "anthropic_version": "bedrock-2023-05-31",
 "max_tokens": int(500/0.75),
 "messages": [
 {
 "role": "user",
 "content": [
 {
 "type": "text",
 "text": prompt_data
 }
]
 }
]
}

[10]: from IPython.display import clear_output, display, display_markdown, Markdown

body = json.dumps(messages_API_body)
modelId = "anthropic.claude-3-haiku-20240307-v1:0" # (Change this to try different model versions)
accept = "application/json"
contentType = "application/json"

+

Filter files by name

/ amazon-bedrock-workshop / 00_Prerequisites /

Name Last Modified

bedrock_basics.ipynb seconds ago

Getting_started_with_Convers... 15 minutes ago

README.md 15 minutes ago

Red arrow pointing to the code block at line [9]

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 17 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

[10]:

```
from IPython.display import clear_output, display, display_markdown, Markdown

body = json.dumps(messages_API_body)
modelId = "anthropic.claude-3-haiku-20240307-v1:0" # (Change this to try different model versions)
accept = "application/json"
contentType = "application/json"

try:
    response = bedrock_runtime.invoke_model_with_response_stream(
        body=body, modelId=modelId, accept=accept, contentType=contentType
    )

    stream = response.get('body')

    output = []

    if stream:
        for event in stream:
            chunk = event.get('chunk')
            if chunk:
                chunk_obj = json.loads(chunk.get('bytes').decode())
                if 'delta' in chunk_obj:
                    delta_obj = chunk_obj.get('delta', None)
                    if delta_obj:
                        text = delta_obj.get('text', None)
                        print(text,end='')

if not text :
```



moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab | +

← → C ⌘ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb ⌘ Finish update ⌋

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 16 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

display_markdown(Markdown(text))

```
except botocore.exceptions.ClientError as error:
    if error.response['Error']['Code'] == 'AccessDeniedException':
        print(f"\x1b[41m{error.response['Error']['Message']}\\nTo troubleshoot this issue please refer to the following resources.\\nhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.html\\nhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m\\n")
    else:
        raise error
```

Here is a 525 word paragraph about making strong business decisions as a leader:

As a business leader, the decisions you make can have a profound impact on the success and trajectory of your organization. It is a responsibility that should not be taken lightly. Strong leadership and decision-making are essential to navigating the challenges and complexities that arise in the modern business world.

First and foremost, effective business decision-making requires a clear vision and strategic long-term thinking. Leaders must have a firm understanding of their company's mission, values, and objectives, and be able to chart a course that aligns with those guiding principles. This means looking beyond short-term gains or quick fixes, and considering the broader implications and sustainability of any decision. What may seem like the expedient choice in the moment could ultimately undermine the company's long-term competitive position or brand reputation.

Closely related to this is the ability to think critically and see the big picture. Successful leaders don't get bogged down in minutiae or knee-jerk reactions. They step back, analyze the available information objectively, and consider multiple perspectives before coming to a conclusion. This requires strong emotional intelligence and the capacity to override personal biases. It's about making dispassionate, rational decisions that serve the greater good of the organization, not one's own ego or agenda.

Of course, business decisions are rarely black and white. Ambiguity, conflicting priorities, and uncertain outcomes

Feedback

Extra Credits



© 2024, Amazon Web Services, Inc. or its affiliates.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | bedrock_basi... - JupyterLab | +

← → C ⌘ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/00_Prerequisites/bedrock_basics.ipynb ⌘ Finish update ⌋

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 16 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

display_markdown(Markdown(text))

```
except botocore.exceptions.ClientError as error:
    if error.response['Error']['Code'] == 'AccessDeniedException':
        print(f"\x1b[41m{error.response['Error']['Message']}\\nTo troubleshoot this issue please refer to the following resources.\\nhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.html\\nhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m\\n")
    else:
        raise error
```

Here is a 525 word paragraph about making strong business decisions as a leader:

As a business leader, the decisions you make can have a profound impact on the success and trajectory of your organization. It is a responsibility that should not be taken lightly. Strong leadership and decision-making are essential to navigating the challenges and complexities that arise in the modern business world.

First and foremost, effective business decision-making requires a clear vision and strategic long-term thinking. Leaders must have a firm understanding of their company's mission, values, and objectives, and be able to chart a course that aligns with those guiding principles. This means looking beyond short-term gains or quick fixes, and considering the broader implications and sustainability of any decision. What may seem like the expedient choice in the moment could ultimately undermine the company's long-term competitive position or brand reputation.

Closely related to this is the ability to think critically and see the big picture. Successful leaders don't get bogged down in minutiae or knee-jerk reactions. They step back, analyze the available information objectively, and consider multiple perspectives before coming to a conclusion. This requires strong emotional intelligence and the capacity to override personal biases. It's about making dispassionate, rational decisions that serve the greater good of the organization, not one's own ego or agenda.

Of course, business decisions are rarely black and white. Ambiguity, conflicting priorities, and uncertain outcomes

File Edit View Run Kernel Git Tabs Settings Help

Amazon SageMaker Studio Classic

Filter files by name

/ amazon-bedrock-workshop / Prerequisites /

Name Last Modified

- bedrock_basics.ipynb a minute ago
- Getting_started_with_Convers... 16 minutes ago
- README.md 16 minutes ago

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 16 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in the new Studio.

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

+ Filter files by name / amazon-bedrock-workshop /

Name	Last Modified
00_Prerequisites	2 minutes ago
01_Text_generation	17 minutes ago
02_KnowledgeBases_and_RAG	17 minutes ago
03_Model_customization	17 minutes ago
04_Image_and_Multimodal	17 minutes ago
05_Agents	17 minutes ago
06_OpenSource_examples	17 minutes ago
imgs	17 minutes ago
CODE_OF_CONDUCT.md	17 minutes ago
CONTRIBUTING.md	17 minutes ago
LICENSE	17 minutes ago
README.md	17 minutes ago
RELEASE_NOTES.md	17 minutes ago

```
except botocore.exceptions.ClientError as error:  
  
    if error.response['Error']['Code'] == 'AccessDeniedException':  
        print(f"\x1b[41m{error.response['Error']['Message']}")  
        \nTo troubleshoot this issue please refer to the following resources.\n        \nhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.html\n        \nhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m\n")  
  
    else:  
        raise error
```

Here is a 525 word paragraph about making strong business decisions as a leader:

As a business leader, the decisions you make can have a profound impact on the success and trajectory of your organization. It is a responsibility that should not be taken lightly. Strong leadership and decision-making are essential to navigating the challenges and complexities that arise in the modern business world.

First and foremost, effective business decision-making requires a clear vision and strategic long-term thinking. Leaders must have a firm understanding of their company's mission, values, and objectives, and be able to chart a course that aligns with those guiding principles. This means looking beyond short-term gains or quick fixes, and considering the broader implications and sustainability of any decision. What may seem like the expedient choice in the moment could ultimately undermine the company's long-term competitive position or brand reputation.

Closely related to this is the ability to think critically and see the big picture. Successful leaders don't get bogged down in minutiae or knee-jerk reactions. They step back, analyze the available information objectively, and consider multiple perspectives before coming to a conclusion. This requires strong emotional intelligence and the capacity to override personal biases. It's about making dispassionate, rational decisions that serve the greater good of the organization, not one's own ego or agenda.

Of course, business decisions are rarely black and white. Ambiguity, conflicting priorities, and uncertain outcomes are the norm. Effective leaders are comfortable with this level of complexity and are able to make decisions with a

Generate embeddings

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 15 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb

```
text = delta_obj.get('text', None)
print(text,end='')
if not text :
    break
# output.append(text[0]) if type(text) is list and len(text)>0 else output.
# display_markdown(Markdown(text))

except botocore.exceptions.ClientError as error:

    if error.response['Error']['Code'] == 'AccessDeniedException':
        print(f"\x1b[41m{error.response['Error']['Message']}\x1b[0m")
        \nTo troubleshoot this issue please refer to the following resources.\n\nhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.\n\nhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m

else:
    raise error
```

Here is a 525 word paragraph about making strong business decisions as a leader:

As a business leader, the decisions you make can have a profound impact on the success and trajectory of your organization. It is a responsibility that should not be taken lightly. Strong leadership and decision-making are essential to navigating the challenges and complexities that arise in the modern business world.

First and foremost, effective business decision-making requires a clear vision and strategic long-term thinking. Leaders must have a firm understanding of their company's mission, values, and objectives, and be able to chart a course that aligns with those guiding principles. This means looking beyond short-term gains or quick fixes, and considering the broader implications and sustainability of any decision. What may seem like the expedient choice in the moment could ultimately undermine the company's long-term competitive position or brand reputation.

Closely related to this is the ability to think critically and see the big picture. Successful leaders don't get bogged down in minutiae or knee-jerk reactions. They step back, analyze the available information objectively, and consider multiple perspectives before coming to a conclusion. This requires strong analytical skills and the ability to prioritize effectively.

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 15 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Filter files by name

Name Last Modified

18 minutes ago

00_Chatbot_Claude.ipynb (highlighted with a red arrow)

00_Chatbot_Llama2.ipynb

00_Chatbot_Titan.ipynb

00_Chatbot_V3_Claude.ipynb

00_medibot_3V_agents.ipynb

00_medibot_V3_memory_rag.ipynb

00_medibot_V3_prompts.ipynb

00_medibot_V3.ipynb

00_MutliModal_Claude_search.ipynb

03_chatbot_mapreduce_claude_v3.ipynb

invoke_end_point.py

README.md

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

No Kernel Share

Conversational Interface - Chatbot with Claude LLM

This notebook should work well with the Data Science 3.0 kernel in SageMaker Studio

Foundation Models (FMs) in Amazon Bedrock. For our use-case

Set up environment for "00_Chatbot_Claude.ipynb".

Image Kernel

Instance type

Start-up script

No script

Cancel Select

1

2

Chat History

3

Store chat history to a database

Question + chat history

1

2

3

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 15 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb No Kernel Share

Conversational Interface - Chatbot with Claude LLM

This notebook should work well with the Data Science 3.0 kernel in SageMaker Studio

Set up environment for "00_Chatbot_Claude.ipynb".

Image: Data Science 3.0 (highlighted with a red arrow)

Kernel: Python 3

Instance type: ml.t3.medium

Start-up script: No script

Cancel Select

Chat History

Store chat history to a database

Question + chat history

The screenshot shows the Amazon SageMaker Studio Classic interface. On the left, there is a file browser with a sidebar containing various files and folders related to a workshop. On the right, a modal dialog is open for setting up a notebook environment. The dialog includes fields for 'Image' (set to 'Data Science 3.0'), 'Kernel' (set to 'Python 3'), 'Instance type' (set to 'ml.t3.medium'), and 'Start-up script' (set to 'No script'). A large red arrow points to the 'Image' dropdown. The background shows a preview of the notebook content with the title 'Conversational Interface - Chatbot with Claude LLM'.

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 14 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Filter files by name / ...

Name Last Modified

- images 18 minutes ago
- rag_data 18 minutes ago
- 00_Chatbot_Claude.ipynb 18 minutes ago
- 00_Chatbot_Llama2.ipynb 18 minutes ago
- 00_Chatbot_Titan.ipynb 18 minutes ago
- 00_Chatbot_V3_Claude... 18 minutes ago
- 00_medibot_3V_agents... 18 minutes ago
- 00_medibot_V3_memo... 18 minutes ago
- 00_medibot_V3_promp... 18 minutes ago
- 00_medibot_V3.ipynb 18 minutes ago
- 00_MultiModal_Claude... 18 minutes ago
- 03_chatbot_mapreduce... 18 minutes ago
- invoke_end_point.py 18 minutes ago
- README.md 18 minutes ago

Conversational Interface - Chatbot with Claude LLM

This notebook should work well with the **Data Science 3.0** kernel in SageMaker Studio

In this notebook, we will build a chatbot using the Foundation Models (FMs) in Amazon Bedrock. For our use-case we use Claude as our FM for building the chatbot.

Overview

Conversational interfaces such as chatbots and virtual assistants can be used to enhance the user experience for your customers. Chatbots use natural language processing (NLP) and machine learning algorithms to understand and respond to user queries. Chatbots can be used in a variety of applications, such as customer service, sales, and e-commerce, to provide quick and efficient responses to users. They can be accessed through various channels such as websites, social media platforms, and messaging apps.

Chatbot using Amazon Bedrock

1 Store chat history to

2 Chat History

3 Question + chat

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

Finish update

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 14 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Conversational Interface - Chatbot with Claude LLM

This notebook should work well with the **Data Science 3.0** kernel in SageMaker Studio

In this notebook, we will build a chatbot using the Foundation Models (FMs) in Amazon Bedrock. For our use-case we use Claude as our FM for building the chatbot.

Overview

Conversational interfaces such as chatbots and virtual assistants can be used to enhance the user experience for your customers. Chatbots use natural language processing (NLP) and machine learning algorithms to understand and respond to user queries. Chatbots can be used in a variety of applications, such as customer service, sales, and e-commerce, to provide quick and efficient responses to users. They can be accessed through various channels such as websites, social media platforms, and messaging apps.

Chatbot using Amazon Bedrock

1 Store chat 2 Chat History 3 Question

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 14 minutes)

[Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic

File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio



Home X

Terminal 1 X

bedrock_basics.ipynb X

00_Chatbot_Claude.ipynb ●

Cluster



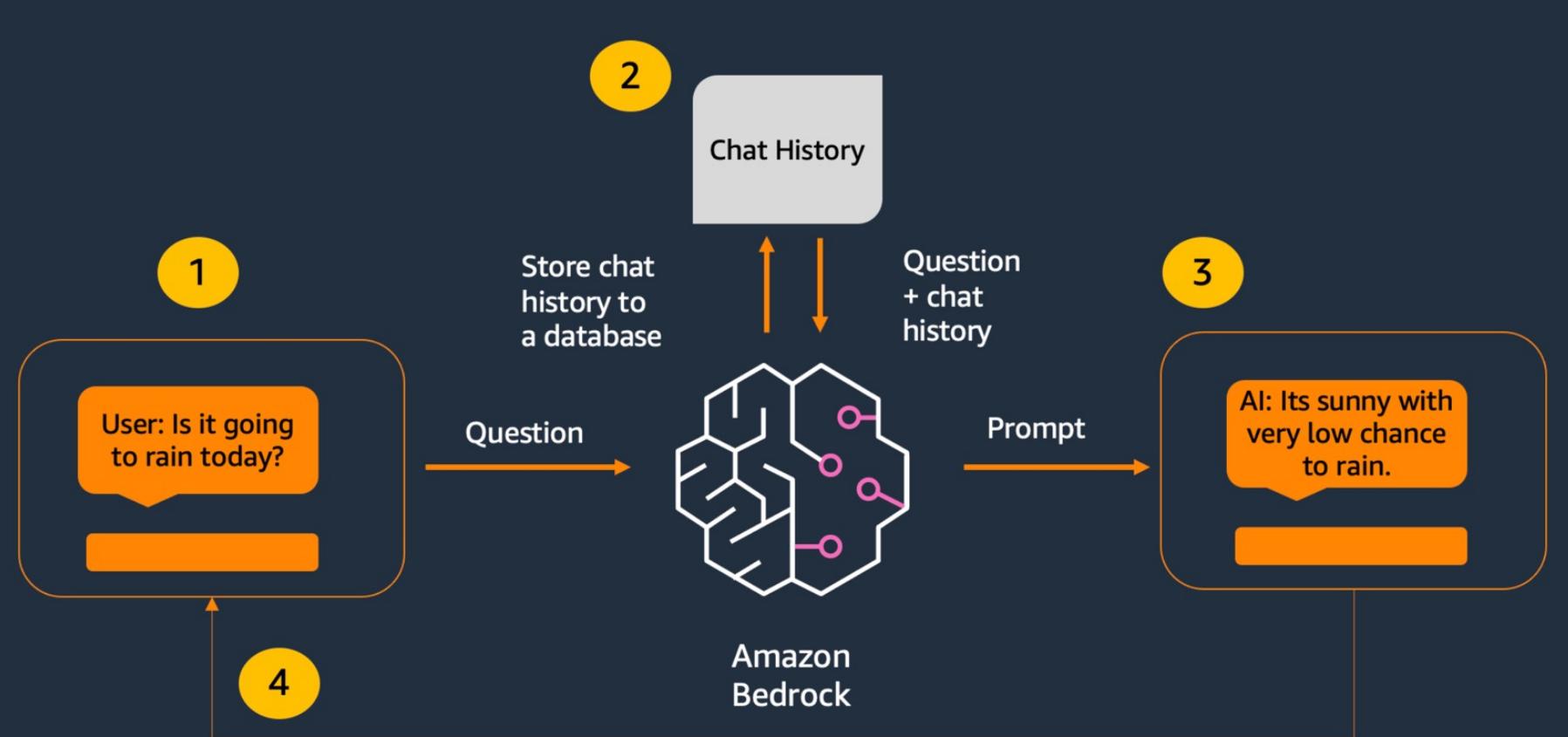
Data Science 3.0

Python 3 | 2 vCPU + 4 GiB

Share



Chatbot using Amazon Bedrock



Simple

1 \$ 2



main Python 3 (Data Science 3.0) | Idle

Kernel: Idle | Instance MEM

Cookie Preferences

Mode: Command



Ln 1, Col 1 00_Chatbot_Claude.ipynb

0



Home X Terminal 1 X bedrock_basics.ipynb X 00_Chatbot_Claude.ipynb ●

+ X C ► Markdown \$ git Share ⚙

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB

Setup

⚠️⚠️⚠️ Before running this notebook, ensure you've run the Bedrock boto3 setup notebook notebook. ⚠️⚠️⚠️

```
[*]: %pip install -U --no-cache-dir boto3
%pip install -U --no-cache-dir \
    "langchain>=0.1.11" \
    sqlalchemy -U \
    "faiss-cpu>=1.7,<2" \
    "pypdf>=3.8,<4" \
    pinecone-client==2.2.4 \
    apache-beam==2.52. \
    tiktoken==0.5.2 \
    "ipywidgets>=7,<8" \
    matplotlib==3.8.2 \
    anthropic==0.9.0
%pip install -U --no-cache-dir transformers
```

Requirement already satisfied: boto3 in /opt/conda/lib/python3.10/site-packages (1.35.68)
Requirement already satisfied: botocore<1.36.0,>=1.35.68 in /opt/conda/lib/python3.10/site-packages (from boto3) (1.35.68)
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /opt/conda/lib/python3.10/site-packages (from boto3) (1.0.1)
Requirement already satisfied: s3transfer<0.11.0,>=0.10.0 in /opt/conda/lib/python3.10/site-packages (from boto3) (0.10.4)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /opt/conda/lib/python3.10/site-packages (from botocore<1.36.0,>=1.35.68->boto3) (2.9.0.post0)
Requirement already satisfied: urllib3!=2.2.0,<3,>=1.25.4 in /opt/conda/lib/python3.10/site-packages (from botocore<1.36.0,>=1.35.68->boto3) (2.2.3)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.10/site-packages (from python-dateutil<3.0.0,>=2.1->botocore<1.36.0,>=1.35.68->boto3) (1.16.0)
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>
Note: you may need to restart the kernel to use updated packages.
Collecting langchain>=0.1.11
 Downloading langchain-0.3.8-py3-none-any.whl.metadata (7.1 kB)
Requirement already satisfied: sqlalchemy in /opt/conda/lib/python3.10/site-packages (2.0.29)
Collecting sqlalchemy
 Downloading SQLAlchemy-2.0.36-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.7 kB)

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter | +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 12 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

```
[3]: import warnings
from io import StringIO
import sys
import textwrap
import os
from typing import Optional

# External Dependencies:
import boto3

warnings.filterwarnings('ignore')
def print_ww(*args, width: int = 100, **kwargs):
    """Like print(), but wraps output to `width` characters (default 100)"""
    buffer = StringIO()
    try:
        _stdout = sys.stdout
        sys.stdout = buffer
        print(*args, **kwargs)
        output = buffer.getvalue()
    finally:
        sys.stdout = _stdout
    for line in output.splitlines():
        print("\n".join(textwrap.wrap(line, width=width)))
```

```
[4]: import json
import os
import sys

import boto3
import botocore

boto3_bedrock = boto3.client('bedrock-runtime')
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 12 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Chatbot (Basic - without context)

We use `CoversationChain` from LangChain to start the conversation. We also use the `ConversationBufferMemory` for storing the messages. We can also get the history as a list of messages (this is very useful in a chat model).

Chatbots needs to remember the previous interactions. Conversational memory allows us to do that. There are several ways that we can implement conversational memory. In the context of LangChain, they are all built on top of the `ConversationChain`.

Note: The model outputs are non-deterministic

```
[ ]: from langchain.chains import ConversationChain
from langchain.llms.bedrock import Bedrock
from langchain.memory import ConversationBufferMemory
modelId = "anthropic.claude-v2"
cl_llm = Bedrock(
    model_id=modelId,
    client=boto3_bedrock,
    model_kwargs={"max_tokens_to_sample": 1000},
)
memory = ConversationBufferMemory()
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)

try:
    print_ww(conversation.predict(input="Hi there!"))

except ValueError as error:
    if "AccessDeniedException" in str(error):
        print(f"\x1b[41m{error}\nTo troubleshoot this issue please refer to the following resources.\x1b[0m")
```

moda-workshop-dryrun-2024 X | SageMaker Studio | Amazon Sagemaker Studio Classic - SageMaker S 00_Chatbot_C... (2) - Jupyter X +

← → C ⌂ d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 12 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

ModuleNotFoundError: No module named 'langchain_community'

The above exception was the direct cause of the following exception:

```
ModuleNotFoundError Traceback (most recent call last)
Cell In[5], line 2
      1 from langchain.chains import ConversationChain
----> 2 from langchain.llms.bedrock import Bedrock
      3 from langchain.memory import ConversationBufferMemory
      4 modelId = "anthropic.claude-v2"

File /opt/conda/lib/python3.10/site-packages/langchain,llms/bedrock.py:22, in __getattr__(name)
   20 def __getattr__(name: str) -> Any:
   21     """Look up attributes dynamically."""
-> 22     return _import_attribute(name)

File /opt/conda/lib/python3.10/site-packages/langchain/_api/module_import.py:72, in create_importer.<locals>.import_by_name(name)
   70 except ModuleNotFoundError as e:
   71     if new_module.startswith("langchain_community"):
-> 72         raise ModuleNotFoundError(
   73             f"Module {new_module} not found."
   74             "Please install langchain-community to access this module."
   75             "You can install it using `pip install -U langchain-community`"
   76         ) from e
   77     raise
   79 try:

ModuleNotFoundError: Module langchain_community.llms not found. Please install langchain-community to access this module. You can install it using `pip install -U langchain-community`
```

As you can see above, we used a basic prompt ("Hi there!") with no special formatting. In the next sections we will see how to enrich it with more detail.

To learn more about how to write prompts for Claude, check Anthropic documentation.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 11 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

+ Chatbot (Basic - without context) Insert a cell below (B)

We use `ConversationChain` from `LangChain` to start the conversation. We also use the `ConversationBufferMemory` for storing the messages. We can also get the history as a list of messages (this is very useful in a chat model).

Chatbots needs to remember the previous interactions. Conversational memory allows us to do that. There are several ways that we can implement conversational memory. In the context of `LangChain`, they are all built on top of the `ConversationChain`.

Note: The model outputs are non-deterministic

```
[5]: from langchain.chains import ConversationChain
from langchain.llms.bedrock import Bedrock
from langchain.memory import ConversationBufferMemory
modelId = "anthropic.claude-v2"
cl_llm = Bedrock(
    model_id=modelId,
    client=boto3_bedrock,
    model_kwargs={"max_tokens_to_sample": 1000},
)
memory = ConversationBufferMemory()
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)

try:

    print_ww(conversation.predict(input="Hi there!"))

except ValueError as error:
    if "AccessDeniedException" in str(error):
        print(f"\x1b[41m{error}\x1b[0m")
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → ⌛ ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 11 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Chatbot (Basic - without context)

We use `ConversationChain` from `LangChain` to start the conversation. We also use the `ConversationBufferMemory` for storing the messages. We can also get the history as a list of messages (this is very useful in a chat model).

Chatbots needs to remember the previous interactions. Conversational memory allows us to do that. There are several ways that we can implement conversational memory. In the context of `LangChain`, they are all built on top of the `ConversationChain`.

Note: The model outputs are non-deterministic

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

```
[5]: from langchain.chains import ConversationChain
from langchain.llms.bedrock import Bedrock
from langchain.memory import ConversationBufferMemory
modelId = "anthropic.claude-v2"
cl_llm = Bedrock(
    model_id=modelId,
    client=boto3_bedrock,
    model_kwargs={"max_tokens_to_sample": 1000},
)
memory = ConversationBufferMemory()
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 11 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

```
File /opt/conda/lib/python3.10/site-packages/langchain/llms/bedrock.py:22, in __getattr__(name)
 20 def __getattr__(name: str) -> Any:
 21     """Look up attributes dynamically."""
--> 22     return _import_attribute(name)

File /opt/conda/lib/python3.10/site-packages/langchain/_api/module_import.py:72, in create_importer.<locals>.import_by_name(name)
 70 except ModuleNotFoundError as e:
 71     if new_module.startswith("langchain_community"):
--> 72         raise ModuleNotFoundError(
 73             f"Module {new_module} not found. "
 74             "Please install langchain-community to access this module. "
 75             "You can install it using `pip install -U langchain-community`"
 76         ) from e
 77     raise
 79 try:

ModuleNotFoundError: Module langchain_community.llms not found. Please install langchain-community to access this module. You can install it using `pip install -U langchain-community`
```

[]:

As you can see above, we used a basic prompt ("Hi there!") with no special formatting. In the next sections we will see how to enrich it with more detail.

To learn more about how to write prompts for Claude, check Anthropic documentation.

Chatbot using prompt template (Langchain)

LangChain provides several classes and functions to make constructing and working with prompts easy. We are going to use the `PromptTemplate` class to construct the prompt from a f-string template.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 11 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

```
import os
import sys

import boto3
import botocore

boto3_bedrock = boto3.client('bedrock-runtime')
```

Chatbot (Basic - without context)

We use `CoversationChain` from `LangChain` to start the conversation. We also use the `ConversationBufferMemory` for storing the messages. We can also get the history as a list of messages (this is very useful in a chat model).

Chatbots needs to remember the previous interactions. Conversational memory allows us to do that. There are several ways that we can implement conversational memory. In the context of `LangChain`, they are all built on top of the `ConversationChain`.

Note: The model outputs are non-deterministic

```
[ ]:
```

```
[ ]:
```

```
pip install -U langchain-community
```

5: `from langchain.chains import ConversationChain
from langchain.llms.bedrock import Bedrock
from langchain.memory import ConversationBufferMemory
modelId = "anthropic.claude-v2"
cl_llm = Bedrock(
 model_id=modelId,
 client=boto3_bedrock.`

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end-of-support on December 31, 2024. (35 days, 0 hours, 25 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in the new Studio.

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Chatbot (Basic - without context)

We use `ConversationChain` from LangChain to start the conversation. We also use the `ConversationBufferMemory` for storing the messages. We can also get the history as a list of messages (this is very useful in a chat model).

Chatbots need to remember the previous interactions. Conversational memory allows us to do that. There are several ways that we can implement conversational memory. In the context of LangChain, they are all built on top of the `ConversationChain`.

Note: The model outputs are non-deterministic

```
[ ]:
```

```
[ ]:
```

```
[6]: !pip install -U langchain-community
```

```
Collecting langchain-community
  Downloading langchain_community-0.3.8-py3-none-any.whl.metadata (2.9 kB)
Requirement already satisfied: PyYAML>=5.3 in /opt/conda/lib/python3.10/site-packages (from langchain-community) (6.0.2)
Collecting SQLAlchemy<2.0.36,>=1.4 (from langchain-community)
  Downloading SQLAlchemy-2.0.35-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (9.6 kB)
Requirement already satisfied: aiohttp<4.0.0,>=3.8.3 in /opt/conda/lib/python3.10/site-packages (from langchain-community) (3.11.7)
Collecting dataclasses-json<0.7,>=0.5.7 (from langchain-community)
  Downloading dataclasses_json-0.6.7-py3-none-any.whl.metadata (25 kB)
WARNING: Retrying (Retry(total=4, connect=None, read=None, redirect=None, status=None)) after connection broken by 'ProtocolError('Connection aborted.', RemoteDisconnected('Remote end closed connection without response'))': /simple/httpx-sse/
Collecting httpx-sse<0.5.0,>=0.4.0 (from langchain-community)
  Downloading httpx_sse-0.4.0-py3-none-any.whl.metadata (9.0 kB)
```

Home X Terminal 1 X bedrock_basics.ipynb X 00_Chatbot_Claude.ipynb ● Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share ○

[7]:

```
from langchain.chains import ConversationChain
from langchain.llms.bedrock import Bedrock
from langchain.memory import ConversationBufferMemory
modelId = "anthropic.claude-v2"
cl_llm = Bedrock(
    model_id=modelId,
    client=boto3_bedrock,
    model_kwarg={"max_tokens_to_sample": 1000},
)
memory = ConversationBufferMemory()
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)

try:
    print_ww(conversation.predict(input="Hi there!"))

except ValueError as error:
    if "AccessDeniedException" in str(error):
        print(f"\x1b[41m{error}\x1b[0m\nTo troubleshoot this issue please refer to the following resources.\n\x1b[42mhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.html\x1b[0m\n\x1b[42mhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m\n")
    class StopExecution(ValueError):
        def _render_traceback_(self):
            pass
    raise StopExecution
else:
    raise error
```

/tmp/ipykernel_125/863247160.py:5: LangChainDeprecationWarning: The class `Bedrock` was deprecated in LangChain 0.0.34 and will be removed in 1.0. An updated version of the class exists in the :class:`~langchain-aws` package and should be used instead. To use it run `pip install -U :class:`~langchain-aws` and import as `from :class:`~langchain_aws import BedrockLLM`.
 cl_llm = Bedrock(
/tmp/ipykernel_125/863247160.py:10: LangChainDeprecationWarning: Please see the migration guide at: https://python.langchain.com/docs/versions/migrating_memory/
 memory = ConversationBufferMemory()

Simple (●) 1 \$ 2 ⚙ main Python 3 (Data Science 3.0) | Idle Kernel: Idle | Instance MEM (Cookie Preferences Mode: Command ✅ Ln 1, Col 1 00_Chatbot_Claude.ipynb 0 🔍)

The screenshot shows a JupyterLab interface within the SageMaker Studio Classic environment. The top navigation bar includes tabs for 'moda-workshop-dryrun-2024', 'SageMaker Studio | Amazon S3', 'Studio Classic - SageMaker S3', '00_Chatbot_C... (2) - Jupyter...', and a '+' icon. A prominent yellow banner at the top states: 'JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 10 minutes)' with a 'Learn More' link. Below the banner, a message encourages migration to the new Studio. The main workspace shows two open notebooks: 'bedrock_basics.ipynb' and '00_Chatbot_Claude.ipynb'. The '00_Chatbot_Claude.ipynb' tab is active, displaying a conversation chain:

```
> Entering new ConversationChain chain...
Prompt after formatting:
The following is a friendly conversation between a human and an AI. The AI is talkative and provides lots of specific details from its context. If the AI does not know the answer to a question, it truthfully says it does not know.

Current conversation:

Human: Hi there!
AI:

> Finished chain.
Hello! Nice to meet you. How's your day going so far? I'm Claude, an AI assistant created by Anthropic to be helpful, harmless, and honest.
```

As you can see above, we used a basic prompt ("Hi there!") with no special formatting. In the next sections we will see how to enrich it with more detail.

To learn more about how to write prompts for Claude, check [Anthropic documentation](#).

Chatbot using prompt template (Langchain)

LangChain provides several classes and functions to make constructing and working with prompts easy. We are going to use the `PromptTemplate` class to construct the prompt from a f-string template.

```
[ ]: from langchain.memory import ConversationBufferMemory
from langchain.prompts import PromptTemplate

# turn verbose to true to see the full logs and documents
conversation = ConversationChain(
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

Finish update

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 7 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Chatbot using prompt template (Langchain)

LangChain provides several classes and functions to make constructing and working with prompts easy. We are going to use the `PromptTemplate` class to construct the prompt from a f-string template.

```
[8]: from langchain.memory import ConversationBufferMemory
from langchain.prompts import PromptTemplate

# turn verbose to true to see the full logs and documents
conversation= ConversationChain(
    llm=cl_llm, verbose=False, memory=ConversationBufferMemory() #memory_chain
)

# langchain prompts do not always work with all the models. This prompt is tuned for Claude
claude_prompt = PromptTemplate.from_template("""
```

Human: The following is a friendly conversation between a human and an AI.
The AI is talkative and provides lots of specific details from its context. If the AI does not know the answer to a question, it truthfully says it does not know.

Current conversation:
<conversation_history>
{history}
</conversation_history>

Here is the human's next reply:
<human_reply>
{input}
</human_reply>

Assistant:
"""



moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share ⚙️

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Code git

```
<conversation_history>
{history}
</conversation_history>

Here is the human's next reply:
<human_reply>
{input}
</human_reply>

Assistant:
""")

conversation.prompt = claude_prompt

print_ww(conversation.predict(input="Hi there!"))
```

Hello! Nice to meet you. I'm Claude, an AI assistant created by Anthropic to be helpful, harmless, and honest. How can I assist you today?

New Questions

Model has responded with initial message, let's ask few questions

[9]: print_ww(conversation.predict(input="Give me a few tips on how to start a new garden.))

Here are some tips for starting a new garden:

- Decide what you want to grow – Vegetables, flowers, herbs? Make a list of what you'd like to plant so you can prepare accordingly. Popular beginner vegetables include tomatoes, zucchini, beans, lettuce, etc.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

New Questions

Model has responded with intial message, let's ask few questions

[9]: `print_ww(conversation.predict(input="Give me a few tips on how to start a new garden. "))`

Here are some tips for starting a new garden:

- Decide what you want to grow – Vegetables, flowers, herbs? Make a list of what you'd like to plant so you can prepare accordingly. Popular beginner vegetables include tomatoes, zucchini, beans, lettuce, etc.
- Prepare the soil – Remove any weeds or grass and mix in compost or fertilizer to enrich the soil. You want nice loose, nutrient-rich soil for planting. Adding mulch helps retain moisture.
- Determine the amount of sun – Most vegetables and herbs need at least 6 hours of direct sunlight per day. Be mindful of shady spots in the yard and plant sun-loving plants elsewhere.
- Make a planting schedule – Look up ideal planting times in your region for what you want to grow. Many seeds are best started indoors first before transplanting outside based on frost dates.
- Arrange your garden layout – Consider spacing, trellising needs, and companion planting as you map out your garden beds or rows. Group plants with similar care needs together.
- Prepare your seeds – Read seed packet instructions. Some seeds need pre-sowing indoors first before transplanting outside at the right time.
- Water regularly – New gardens need about 1 inch of water per week from rain or watering. Consider setting up a drip irrigation system on a timer if you'll be away.

I hope these tips help you get started planning and planting your new garden! Let me know if you have any other questions.

Build and Run Cell

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Build on the questions

Let's ask a question without mentioning the word garden to see if model can understand previous conversation

[10]: `print_ww(conversation.predict(input="Cool. Will that work with tomatoes?"))`

Yes, the tips I provided will work well for growing tomatoes in a new garden. Here are some more specific tomato gardening tips:

- Choose disease-resistant tomato varieties – Look for labels like VFN, which indicate resistance to verticillium, fusarium, and nematodes. This will help avoid common tomato diseases.
- Plant tomatoes in full sun – Tomatoes need at least 6–8 hours of direct sunlight per day. Make sure to pick a sunny spot in your garden.
- Prepare the soil – Tomatoes thrive in rich, well-draining soil. Mix in compost or other organic matter before planting. The soil pH should be between 6.0–6.8.
- Space plants adequately – Give tomatoes plenty of room to grow. Plant them 18–24 inches apart in rows 3 feet apart. Cage or stake plants to support growth.
- Water regularly – Tomatoes need consistent moisture, about 1–2 inches per week. Water at the base to keep foliage dry and prevent disease.
- Mulch around plants – Use 2–3 inches of mulch like straw or bark to retain moisture and prevent weeds.
- Fertilize occasionally – Use a balanced vegetable fertilizer once every 2–3 weeks when plants start flowering and producing fruit.

Let me know if you need any other tomato growing tips! I'm happy to provide more details on how to successfully grow them in your new garden.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Finishing this conversation

[11]: `print_ww(conversation.predict(input="That's all, thank you!"))`

You're very welcome! I'm glad I could provide some helpful tips on starting a new garden, especially for growing tomatoes. It was a pleasure chatting with you. Please let me know if you have any other gardening questions in the future. I'm happy to help out with more specific advice and details whenever needed. Have a great rest of your day!

Claude is still really talkative. Try changing the prompt to make Claude provide shorter answers.

Interactive session using ipywidgets

The following utility class allows us to interact with Claude in a more natural way. We write out the question in an input box, and get Claude's answer. We can then continue our conversation.

[12]:

```
import ipywidgets as ipw
from IPython.display import display, clear_output

class ChatUX:
    """ A chat UX using IPWidgets
    """
    def __init__(self, qa, retrievalChain = False):
        self.qa = qa
        self.name = None
        self.b=None
        self.retrievalChain = retrievalChain
        self.out = ipw.Output()

    def start_chat(self):
        print("Starting chat bot")
        display(self.out)
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share ⚙

Interactive session using ipywidgets

The following utility class allows us to interact with Claude in a more natural way. We write out the question in an input box, and get Claude's answer. We can then continue our conversation.

```
[12]: import ipywidgets as ipw
from IPython.display import display, clear_output

class ChatUX:
    """ A chat UX using IPWidgets
    """
    def __init__(self, qa, retrievalChain = False):
        self.qa = qa
        self.name = None
        self.b=None
        self.retrievalChain = retrievalChain
        self.out = ipw.Output()

    def start_chat(self):
        print("Starting chat bot")
        display(self.out)
        self.chat(None)

    def chat(self, _):
        if self.name is None:
            prompt = ""
        else:
```

← → ⌂ d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 6 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

```
def start_chat(self):
    print("Starting chat bot")
    display(self.out)
    self.chat(None)

def chat(self, _):
    if self.name is None:
        prompt = ""
    else:
        prompt = self.name.value
    if 'q' == prompt or 'quit' == prompt or 'Q' == prompt:
        print("Thank you , that was a nice chat!!")
        return
    elif len(prompt) > 0:
        with self.out:
            thinking = ipw.Label(value="Thinking...")
            display(thinking)
            try:
                if self.retrievalChain:
                    result = self.qa.run({'question': prompt })
                else:
                    result = self.qa.run({'input': prompt }) #, 'history':chat_history})
            except:
                result = "No answer"
            thinking.value=""
            print_wv(f"AI:{result}")
            self.name.disabled = True
            self.b.disabled = True
            self.name = None

    if self.name is None:
        with self.out:
            self.name = ipw.Text(description="You:", placeholder='q to quit')
            self.b = ipw.Button(description="Send")
            self.b.on_click(self.chat)
            display(ipw.Box(children=(self.name, self.b)))
```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 5 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Code git Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Let's start a chat. You can also test the following questions:

1. tell me a joke
2. tell me another joke
3. what was the first joke about
4. can you make another joke on the same topic of the first joke

[13]: `chat = ChatUX(conversation)
chat.start_chat()`

Starting chat bot

You: tell me a joke Send

AI: Here is a joke for you:
Why don't eggs tell jokes? They'd crack each other up!

You: q to quit Send

Chatbot with persona

AI assistant will play the role of a career coach. Role Play Dialogue requires user message to be set in before starting the chat. ConversationBufferMemory is used to pre-populate the dialog

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | 00_Chatbot_C... (2) - Jupyter | +

Finish update

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 5 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Let's start a chat. You can also test the following questions:

1. tell me a joke
2. tell me another joke
3. what was the first joke about
4. can you make another joke on the same topic of the first joke

[13]: `chat = ChatUX(conversation)
chat.start_chat()`

Starting chat bot

You: tell me a joke Send

AI: Here is a joke for you:

Why don't eggs tell jokes? They'd crack each other up!

You: tell me another joke Send

AI: Here is another joke for you:

What do you call an alligator in a vest? An investigator!

You: q to quit Send

Chatbot with persona

moda-workshop-dryrun-2024 | Amazon Bedrock Workshop | Amazon Bedrock | us-west-2 | Studio Classic - SageMaker S | 00_Chatbot_C... (auto-r : 4) +

← → C ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/worksheets/auto-r/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (33 days, 5 hours, 45 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Console 1 bedrock_basics.ipynb 02_rag_claude_titan_pinecone.ipynb 00_Chatbot_Claude.ipynb 00_Chatbot_V3_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

[13]: chat = ChatUX(conversation)
chat.start_chat()

Starting chat bot

You: tell me a joke Send

AI: Here is a joke for you:

Why don't eggs tell jokes? They'd crack each other up!

You: tell me another joke Send

AI: Here is another joke for you:

What do you call an alligator in a vest? An investigator!

You: what was the first joke about Send

AI: The first joke I told was:

Why don't eggs tell jokes? They'd crack each other up!

The joke plays on the double meaning of "crack up" – eggs cracking their shells, and also laughing very hard (cracking up). It suggests eggs don't tell each other jokes because they would literally crack their shells from laughing so hard.

You: q Send

Chatbot with persona

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 3 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Chatbot with persona

AI assistant will play the role of a career coach. Role Play Dialogue requires user message to be set in before starting the chat. ConversationBufferMemory is used to pre-populate the dialog

```
[ ]: # store previous interactions using ConversationBufferMemory and add custom prompts to the chat.
memory = ConversationBufferMemory()
memory.chat_memory.add_user_message("You will be acting as a career coach. Your goal is to give career advice to users")
memory.chat_memory.add_ai_message("I am a career coach and give career advice")
cl_llm = Bedrock(model_id="anthropic.claude-v2",client=boto3_bedrock)
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)

conversation.prompt = claude_prompt

print_ww(conversation.predict(input="What are the career options in AI?"))

[ ]: print_ww(conversation.predict(input="What these people really do? Is it fun?"))
```

Let's ask a question that is not specialty of this Persona and the model shouldn't answer that question and give a reason for that

```
[ ]: conversation.verbose = False
print_ww(conversation.predict(input="How to fix my car?"))
```

▼ Chatbot with Context

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

Finish update

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 2 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Chatbot with persona

AI assistant will play the role of a career coach. Role Play Dialogue requires user message to be set in before starting the chat. ConversationBufferMemory is used to pre-populate the dialog

```
[14]: # store previous interactions using ConversationBufferMemory and add custom prompts to the chat.
memory = ConversationBufferMemory()
memory.chat_memory.add_user_message("You will be acting as a career coach. Your goal is to give career advice to users")
memory.chat_memory.add_ai_message("I am a career coach and give career advice")
cl_llm = Bedrock(model_id="anthropic.claude-v2",client=boto3_bedrock)
conversation = ConversationChain(
    llm=cl_llm, verbose=True, memory=memory
)

conversation.prompt = claude_prompt

print_ww(conversation.predict(input="What are the career options in AI?"))
```

> Entering new ConversationChain chain...

Prompt after formatting:

Human: The following is a friendly conversation between a human and an AI.
The AI is talkative and provides lots of specific details from its context. If the AI does not know the answer to a question, it truthfully says it does not know.

Current conversation:
<conversation_history>
Human: You will be acting as a career coach. Your goal is to give career advice to users
AI: I am a career coach and give career advice
</conversation_history>

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 2 minutes) ⓘ [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

> Finished chain.

Here are some of the top career options in AI:

- Machine Learning Engineer – These engineers develop and implement various machine learning algorithms and models to enable systems to improve and learn. Skills needed include programming languages like Python and familiarity with ML frameworks like TensorFlow.
- Data Scientist – Data scientists analyze and interpret complex data sets to identify trends and patterns. They utilize statistical and machine learning techniques to build predictive models and optimize data. Skills include SQL, Python, R, and knowledge of ML algorithms.
- AI Researcher – AI researchers advance the field of artificial intelligence through developing new theories, algorithms, and applications. A PhD in computer science, math or a related field is generally required.
- AI/ML Software Developer – These developers code and engineer AI/machine learning software and applications. They build platforms and frameworks to deploy AI models and systems. Skills include languages like Python and Java and ML frameworks.
- Robotics Engineer – Robotics engineers design, develop, and test robotic systems and components. This involves integrating AI decision-making capabilities into robots and autonomous systems. Skills needed are C++, Python, mechanical engineering.
- Computer Vision Engineer – Computer vision engineers develop algorithms and systems that can process, analyze, and understand visual data like images and videos. Skills include OpenCV, TensorFlow, neural networks.
- Natural Language Processing Engineer – NLP engineers build systems that can understand, interpret, and generate human languages. This involves machine learning algorithms like sentiment analysis. Skills include Python, TensorFlow, linguistics knowledge.

Let me know if you would like me to expand on any of these career options! I'm happy to provide more details on the roles, responsibilities, and qualifications needed for each field.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | 00_Chatbot_C... (2) - Jupyter | +

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 2 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

[15]: `print_ww(conversation.predict(input="What these people really do? Is it fun?"))`

> Entering new ConversationChain chain...
Prompt after formatting:

Human: The following is a friendly conversation between a human and an AI.
The AI is talkative and provides lots of specific details from its context. If the AI does not know
the answer to a question, it truthfully says it does not know.

Current conversation:
<conversation_history>
Human: You will be acting as a career coach. Your goal is to give career advice to users
AI: I am a career coach and give career advice
Human: What are the career options in AI?
AI: Here are some of the top career options in AI:

- Machine Learning Engineer – These engineers develop and implement various machine learning algorithms and models to enable systems to improve and learn. Skills needed include programming languages like Python and familiarity with ML frameworks like TensorFlow.
- Data Scientist – Data scientists analyze and interpret complex data sets to identify trends and patterns. They utilize statistical and machine learning techniques to build predictive models and optimize data. Skills include SQL, Python, R, and knowledge of ML algorithms.
- AI Researcher – AI researchers advance the field of artificial intelligence through developing new theories, algorithms, and applications. A PhD in computer science, math or a related field is generally required.
- AI/ML Software Developer – These developers code and engineer AI/machine learning software and applications. They build platforms and frameworks to deploy AI models and systems. Skills include languages like Python and Java and ML frameworks.
- Robotics Engineer – Robotics engineers design, develop, and test robotic systems and components. This involves integrating AI decision-making capabilities into robots and autonomous systems. Skills needed are C++, Python, mechanical engineering.
- Computer Vision Engineer – Computer vision engineers develop algorithms and systems that can process, analyze, and understand visual data like images and video

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 2 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

> Finished chain.

Here are some more details on what these AI careers entail and whether they are enjoyable:

Machine Learning Engineers – These engineers get to apply cutting-edge ML algorithms to real world problems. They analyze data, train models, and optimize performance. It involves a lot of coding, testing, and troubleshooting. Can be fun to see your models make predictions and improve over time.

Data Scientists – Data scientists get to uncover impactful insights from data. They conceptualize analysis techniques and visualize results. It involves statistics, storytelling, and translating analysis into action. Fun creative problem-solving with data.

AI Researchers – Researchers enjoy exploring new theories and innovations in AI. It involves deep intellectual curiosity and passion for advancing science. Formulating hypotheses and designing experiments is exciting. Can be fun to push boundaries of knowledge.

AI/ML Developers – Developers have fun building and shipping AI products. They get to see ideas come to life and work on a variety of applications. Fast-paced environment collaborating cross-functionally. Satisfying to deliver solutions.

Robotics Engineers – Designing autonomous robots with integrated AI is exciting. Engineers get to program decision-making while solving real world mobility and mechanics problems. Cutting edge work with visible robotic creations.

Computer Vision Engineers – Developing algorithms that can process images and videos can be highly rewarding. Fun to work on recognition, scene analysis and image processing problems. Exciting to bring visual sense to machines.

NLP Engineers – NLP projects involve high-impact applications like chatbots, text analytics and voice recognition. Dynamic work at intersection of AI and human languages. Fun to enable natural interactivity.

In general, AI careers allow for a lot of creativity, innovation and visible impact. The rapid pace of progress and learning keeps things exciting. There is a lot of job satisfaction in building systems that are intelligent, autonomous and helpful.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 1 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

[16]: `conversation.verbose = False
print_ww(conversation.predict(input="How to fix my car?"))`

I apologize, but as an AI assistant I do not have enough information or expertise to provide specific advice on how to fix a car. Cars are complex machines and diagnosing and repairing issues requires hands-on inspection, mechanical skills and knowledge that I do not possess.

Some general tips I can suggest if your car has an issue:

- Try to accurately identify the symptoms and precisely pinpoint the problem. Issues like engine troubles, electrical problems, leaks etc have different solutions.
- Refer to your owner's manual and follow the troubleshooting guidance if available. Manuals often contain directions for common problems.
- Search online forums and videos relevant to the year, make and model of your car. Other owners commonly document DIY fixes and advice.
- Consider taking your car to a trusted mechanic or dealership service center. They have the expertise and resources to accurately diagnose and fix all types of car problems.
- If a simple issue like a flat tire or dead battery, see if you are able to safely repair it yourself with instructions from online or a knowledgeable friend.
- Make sure to never attempt complex repairs that require specialized tools or put your safety at risk.

I hope these general tips give you some direction, but please seek qualified automotive assistance for any vehicle repairs or issues beyond basic maintenance. Let me know if you have any other questions!

Feedback icon

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 1 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in the new Studio.

Chatbot with Context

In this use case we will ask the Chatbot to answer question from some external corpus it has likely never seen before. To do this we apply a pattern called RAG (Retrieval Augmented Generation): the idea is to index the corpus in chunks, then look up which sections of the corpus might be relevant to provide an answer by using semantic similarity between the chunks and the question. Finally the most relevant chunks are aggregated and passed as context to the ConversationChain, similar to providing a history.

We will take a csv file and use **Titan Embeddings Model** to create vectors for each line of the csv. This vector is then stored in FAISS, an open source library providing an in-memory vector datastore. When the chatbot is asked a question, we query FAISS with the question and retrieve the text which is semantically closest. This will be our answer.

Titan embeddings Model

Embeddings are a way to represent words, phrases or any other discrete items as vectors in a continuous vector space. This allows machine learning models to perform mathematical operations on these representations and capture semantic relationships between them.

Embeddings are for example used for the RAG document search capability

```
[17]: from langchain.embeddings import BedrockEmbeddings  
br_embeddings = BedrockEmbeddings(model_id="amazon.titan-embed-text-v1", client=boto3_bedrock)
```

/tmp/ipykernel_125/2319863231.py:3: LangChainDeprecationWarning: The class `BedrockEmbeddings` was deprecated in LangChain 0.2.11 and will be removed in 1.0. An updated version of the class exists in the :class:`~langchain-aws` package and should be used instead. To use it run `pip install -U :class:`~langchain-aws` and import as `from :class:`~langchain_aws import BedrockEmbeddings``.
br_embeddings = BedrockEmbeddings(model_id="amazon.titan-embed-text-v1", client=boto3_bedrock)

FAISS as VectorStore

In order to be able to use embeddings for search, we need a store that can efficiently perform vector similarity searches. In this notebook we use FAISS, which is an in memory store. For permanently store vectors, one can use pgVector, Pinecone or Chroma.

The langchain VectorStore API's are available [here](#).

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 1 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

FAISS as VectorStore

In order to be able to use embeddings for search, we need a store that can efficiently perform vector similarity searches. In this notebook we use FAISS, which is an in memory store. For permanently store vectors, one can use pgVector, Pinecone or Chroma.

The langchain VectorStore API's are available [here](#)

To know more about the FAISS vector store please refer to this [document](#).

```
[18]: from langchain.document_loaders import CSVLoader
from langchain.text_splitter import CharacterTextSplitter
from langchain.indexes.vectorstore import VectorStoreIndexWrapper
from langchain.vectorstores import FAISS

s3_path = "s3://jumpstart-cache-prod-us-east-2/training-datasets/Amazon_SageMaker_FAQs/Amazon_SageMaker_FAQs.csv"
!aws s3 cp $s3_path ./rag_data/Amazon_SageMaker_FAQs.csv

loader = CSVLoader("./rag_data/Amazon_SageMaker_FAQs.csv") # --- > 219 docs with 400 chars, each row consists in a question column and an answer column
documents_aws = loader.load() #
print(f"Number of documents={len(documents_aws)}")

docs = CharacterTextSplitter(chunk_size=2000, chunk_overlap=400, separator=",").split_documents(documents_aws)

print(f"Number of documents after split and chunking={len(docs)}")
vectorstore_faiss_aws = None
try:

    vectorstore_faiss_aws = FAISS.from_documents(
        documents=docs,
        embedding = br_embeddings
    )

    print(f"vectorstore_faiss_aws: number of elements in the index={vectorstore_faiss_aws.index.ntotal}:::")


```



moda-workshop-dryrun-2024 | Amazon Bedrock Workshop | Amazon Bedrock | us-west-2 | Studio Classic - SageMaker S | 00_Chatbot_C... (auto-r : 4) +

← → C ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/worksheets/auto-r/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (33 days, 5 hours, 28 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home X Console 1 X bedrock_basics.ipynb X 02_rag_claude_titan_pinecone X 00_Chatbot_Claude.ipynb X 00_Chatbot_V3_Claude.ipynb X

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

```
docs = CharacterTextSplitter(chunk_size=2000, chunk_overlap=400, separator="").split_documents(documents_aws)

print(f"Number of documents after split and chunking={len(docs)}")
vectorstore_faiss_aws = None
try:
    vectorstore_faiss_aws = FAISS.from_documents(
        documents=docs,
        embedding = br_embeddings
    )

    print(f"vectorstore_faiss_aws: number of elements in the index={vectorstore_faiss_aws.index.ntotal}::")

except ValueError as error:
    if "AccessDeniedException" in str(error):
        print(f"\x1b[41m{error}\x1b[0m")
        \nTo troubleshoot this issue please refer to the following resources.\n  \nhttps://docs.aws.amazon.com/IAM/latest/UserGuide/troubleshoot_access-denied.html\n  \nhttps://docs.aws.amazon.com/bedrock/latest/userguide/security-iam.html\x1b[0m\n    class StopExecution(ValueError):
        def _render_traceback_(self):
            pass
    raise StopExecution
else:
    raise error

download: s3://jumpstart-cache-prod-us-east-2/training-datasets/Amazon_SageMaker_FAQs/Amazon_SageMaker_FAQs.csv to rag_data/Amazon_SageMaker_FAQs.csv
Number of documents=153
Number of documents after split and chunking=154
vectorstore_faiss_aws: number of elements in the index=154::
```

Semantic search

We can use a Wrapper class provided by LangChain to query the vector data base store and return to us the relevant documents. Behind the scenes this is only going to run a RetrievalQA chain.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 1 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Semantic search

We can use a Wrapper class provided by LangChain to query the vector data base store and return to us the relevant documents. Behind the scenes this is only going to run a RetrievalQA chain.

[19]: `wrapper_store_faiss = VectorStoreIndexWrapper(vectorstore=vectorstore_faiss_aws)
print(wrapper_store_faiss.query("R in SageMaker", llm=cl_llm))`

Based on the provided context, here are some key points about using R with Amazon SageMaker:

- R is supported in SageMaker through R kernels in notebook instances and the reticulate R interface to the SageMaker Python SDK. This allows you to build, train, tune, and deploy R models using SageMaker.
- RStudio on SageMaker provides a fully managed IDE for R that integrates with SageMaker Studio notebooks. You can seamlessly switch between RStudio and SageMaker Studio, with automatic syncing of your work.
- You can develop and experiment with R models in SageMaker Studio's visual interface, leveraging features like experiment tracking, model registry, automatic model creation, debugging, etc.
- Overall, Amazon SageMaker provides a fully managed environment to develop, train, tune and deploy R machine learning models at scale. The integration of RStudio IDE and reticulate R interface makes SageMaker very well suited for R developers.

Let's see how the semantic search works:

1. First we calculate the embeddings vector for the query, and
2. then we use this vector to do a similarity search on the store

[20]: `v = br_embeddings.embed_query("R in SageMaker")
print(v[0:10])
results = vectorstore_faiss_aws.similarity_search_by_vector(v, k=4)`

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 1 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic

File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio



Home



Terminal 1



bedrock_basics.ipynb



00_Chatbot_Claude.ipynb



Cluster



Data Science 3.0

Python 3

2 vCPU + 4 GB

Share



```
v = br_embeddings.embed_query("R in SageMaker")
print(v[0:10])
results = vectorstore_faiss_aws.similarity_search_by_vector(v, k=4)
for r in results:
    print_ww(r.page_content)
    print('----')
```

[1.171875, 0.333984375, 0.3125, -0.2431640625, 0.60546875, 0.419921875, -0.361328125, -6.580352783203125e-05, 0.3203125, -0.66796875]

What is Amazon SageMaker?: Is R supported with Amazon SageMaker?

Amazon SageMaker is a fully managed service to prepare data and build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows.: Yes, R is supported with Amazon SageMaker. You can use R within SageMaker notebook instances, which include a preinstalled R kernel and the reticulate library. Reticulate offers an R interface for the Amazon SageMaker Python SDK, enabling ML practitioners to build, train, tune, and deploy R models.

What is Amazon SageMaker?: What is RStudio on Amazon SageMaker?

Amazon SageMaker is a fully managed service to prepare data and build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows.: RStudio on Amazon SageMaker is the first fully managed RStudio Workbench in the cloud. You can quickly launch the familiar RStudio integrated development environment (IDE) and dial up and down the underlying compute resources without interrupting your work, making it easy to build machine learning (ML) and analytics solutions in R at scale. You can seamlessly switch between the RStudio IDE and Amazon SageMaker Studio notebooks for R and Python development. All your work, including code, datasets, repositories, and other artifacts, is automatically synchronized between the two environments to reduce context switch and boost productivity.

What is Amazon SageMaker?: What is Amazon SageMaker Studio?

Amazon SageMaker is a fully managed service to prepare data and build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows.: Amazon SageMaker Studio provides a single, web-based visual interface where you can perform all ML development steps. SageMaker Studio gives you complete access, control, and visibility into each step required to prepare data and build, train, and deploy models. You can quickly upload data, create new notebooks, train and tune models, move back and forth between steps to adjust experiments, compare results, and deploy models to production all in one place, making you much more productive. All ML development activities including notebooks, experiment management, automatic model creation, debugging and profiling, and model drift detection can be performed within the unified SageMaker Studio visual interface.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter | +

← → ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 0 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Memory

In any chatbot we will need a QA Chain with various options which are customized by the use case. But in a chatbot we will always need to keep the history of the conversation so the model can take it into consideration to provide the answer. In this example we use the `ConversationalRetrievalChain` from LangChain, together with a `ConversationBufferMemory` to keep the history of the conversation.

Source: https://python.langchain.com/docs/modules/chains/popular/chat_vector_db

Set `verbose` to `True` to see all the what is going on behind the scenes.

from langchain.chains.conversational_retrieval.prompts import CONDENSE_QUESTION_PROMPT

print_ww(CONDENSE_QUESTION_PROMPT.template)

Given the following conversation and a follow up question, rephrase the follow up question to be a standalone question, in its original language.

Chat History:
{chat_history}

Follow Up Input: {question}

Standalone question:

Parameters used for `ConversationRetrievalChain`

- retriever:** We used `VectorStoreRetriever`, which is backed by a `VectorStore`. To retrieve text, there are two search types you can choose: "similarity" or "mmr". `search_type="similarity"` uses similarity search in the retriever object where it selects text chunk vectors that are most similar to the question vector.
- memory:** Memory Chain to store the history
- condense_question_prompt:** Given a question from the user, we use the previous conversation and that question to make up a standalone question
- chain_type:** If the `chat_history` is long and doesn't fit the context you use this parameter and the options are `stuff` `refine` `map_reduce` `map_rerank`

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 0 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Parameters used for ConversationRetrievalChain

- **retriever**: We used `VectorStoreRetriever`, which is backed by a `VectorStore`. To retrieve text, there are two search types you can choose: "similarity" or "mmr". `search_type="similarity"` uses similarity search in the retriever object where it selects text chunk vectors that are most similar to the question vector.
- **memory**: Memory Chain to store the history
- **condense_question_prompt**: Given a question from the user, we use the previous conversation and that question to make up a standalone question
- **chain_type**: If the chat history is long and doesn't fit the context you use this parameter and the options are `stuff`, `refine`, `map_reduce`, `map_rerank`

If the question asked is outside the scope of context, then the model will reply it doesn't know the answer

Note: if you are curious how the chain works, uncomment the `verbose=True` line.

```
[22]: # turn verbose to true to see the full logs and documents
from langchain.chains import ConversationalRetrievalChain
from langchain.memory import ConversationBufferMemory
```

memory_chain = ConversationBufferMemory(memory_key="chat_history", return_messages=True)
qa = ConversationalRetrievalChain.from_llm(
 llm=cl_llm,
 retriever=vectorstore_faiss_aws.as_retriever(),
 memory=memory_chain,
 condense_question_prompt=CONDENSE_QUESTION_PROMPT,
 #verbose=True,
 chain_type='stuff', # 'refine',
 #max_tokens_limit=300
)

Let's chat! ask the chatbot some questions about SageMaker, like:

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

d-ny7wtp13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 8 hours, 0 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

```
[22]: # turn verbose to true to see the full logs and documents
from langchain.chains import ConversationalRetrievalChain
from langchain.memory import ConversationBufferMemory

memory_chain = ConversationBufferMemory(memory_key="chat_history", return_messages=True)
qa = ConversationalRetrievalChain.from_llm(
    llm=c1_llm,
    retriever=vectorstore_faiss_aws.as_retriever(),
    memory=memory_chain,
    condense_question_prompt=CONDENSE_QUESTION_PROMPT,
    #verbose=True,
    chain_type='stuff', # 'refine',
    #max_tokens_limit=300
)
```

Let's chat! ask the chatbot some questions about SageMaker, like:

1. What is SageMaker?
2. What is canvas?

chat = ChatUX(qa, retrievalChain=True)
chat.start_chat()

Starting chat bot

You: q to quit Send

Your mileage might vary, but after 2 or 3 questions you will start to get some weird answers. In some cases, even in other languages. This is happening for the same reasons outlined at the

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | 00_Chatbot_C... (2) - Jupyter | +

d-ny7wtt13fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot...

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 7 hours, 59 minutes) [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

Let's chat! ask the chatbot some questions about SageMaker, like:

1. What is SageMaker?
2. What is canvas?

[23]: `chat = ChatUX(qa, retrievalChain=True)`
`chat.start_chat()`

Starting chat bot

You: What is SageMaker? Send

AI: Amazon SageMaker is a fully managed service to prepare data and build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows. Key capabilities include:

- SageMaker Studio – A web-based IDE for ML development and end-to-end ML workflows
- Managed training and tuning – Train ML models using optimized algorithm containers and auto-tuning
- Model deployment – Deploy models for real-time predictions and integration into applications
- Managed inference – Host models and get predictions via real-time and batch transforms
- Data preparation – Clean, label, and preprocess data using tools like Data Wrangler
- MLOps – Model monitoring, explainability, governance, and CI/CD integration

So in summary, SageMaker provides a broad set of capabilities to support the full ML workflow from data prep to training to deployment, with fully managed infrastructure and tools optimized for ML workloads.

You: q to quit Send

Your mileage might vary, but after 2 or 3 questions you will start to get some weird answers. In some cases, even in other languages. This is happening for the same reasons outlined at the beginning of this notebook: the default langchain prompts are not optimal for Claude. In the following cell we are going to set two new prompts: one for the question rephrasing, and one to get the answer from that rephrased question.

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → ⌛ 🏷 d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 7 hours, 58 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share ⚙

data, train machine learning models, and deploy them for predictions, all within a managed AWS environment. Key features of Amazon SageMaker include:

- Fully managed infrastructure and workflows for machine learning
- Jupyter notebooks for data preparation, model training, deployment, and monitoring
- Managed training for popular ML frameworks like TensorFlow, PyTorch, XGBoost, etc.
- Managed model hosting for production deployments
- Integration with other AWS services like S3, ECR, CloudWatch, etc.
- Security measures like encryption, IAM roles, VPC support, etc.
- Pay-as-you-go pricing based on usage of compute, storage, and other resources
- Maintains privacy and prevents unauthorized access to models, data, and algorithms

You: What is canvas? Send

AI: Based on the context provided, Amazon SageMaker Canvas is a visual, no-code machine learning service that allows business analysts to build ML models and generate predictions without writing any code or requiring ML expertise. Key features of Amazon SageMaker Canvas include:

- A drag-and-drop interface to access, prepare, and combine data from multiple sources.
- Automatic data cleaning and preparation to minimize manual cleanup.
- Applying a variety of ML algorithms to find highly accurate predictive models with a single click.
- Easily publishing results, explaining and interpreting models, and sharing models with others.
- Usage-based pricing based on session time and dataset size for training models.
- Support for discovering and importing data from AWS sources like S3 and Redshift as well as third party sources.

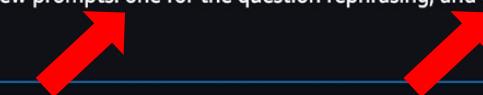
So in summary, Amazon SageMaker Canvas is a no-code, visual ML service to easily build and deploy ML models without needing coding or data science expertise.

You: q to quit Send

Your mileage might vary, but after 2 or 3 questions you will start to get some weird answers. In some cases, even in other languages. This is happening for the same reasons outlined at the beginning of this notebook: the default langchain prompts are not optimal for Claude. In the following cell we are going to set two new prompts: one for the question rephrasing, and one to get the answer from that rephrased question.

```
[ ]: # turn verbose to true to see the full logs and documents
from langchain.chains import ConversationalRetrievalChain
from langchain.schema import BaseMessage

# We are also providing a different chat history retriever which outputs the history as a Claude chat (ie including the \n\n)
_ROLE_MAP = {"human": "\n\nHuman: ", "ai": "\n\nAssistant: "}
def _get_chat_history(chat_history):
    buffer = ""
    for dialogue_turn in chat_history:
        if isinstance(dialogue_turn, BaseMessage):
            role_prefix = _ROLE_MAP.get(dialogue_turn.type, f"{dialogue_turn.type}: ")
            buffer += f"\n{role_prefix}{dialogue_turn.content}"
        elif isinstance(dialogue_turn, tuple):
            human = "\n\nHuman: " + dialogue_turn[0]
            ai = "\n\nAssistant: " + dialogue_turn[1]
            buffer += "\n" + "\n".join([human, ai])
        else:
            raise ValueError(
                f"Unsupported chat history format: {type(dialogue_turn)}."
                f" Full chat history: {chat_history} "
            )
    return buffer
```



moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter | 2月就知道謝宜容作為挨轟吃素 | +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 4 hours, 52 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

SageMakerUser / Personal Studio

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 | Python 3 | 2 vCPU + 4 GiB Share

the condense prompt for Claude
condense_prompt_claude = PromptTemplate.from_template("""{chat_history}

Answer only with the new question.

Human: How would you ask the question considering the previous conversation: {question}

Assistant: Question:""")



The screenshot shows a Jupyter Notebook interface with several tabs at the top: Home, Terminal 1, bedrock_basics.ipynb, and 00_Chatbot_Claude.ipynb. The main area displays Python code for setting up a conversational retrieval chain:

```
# recreate the Claude LLM with more tokens to sample - this provides longer responses but introduces some latency
cl_llm = Bedrock(model_id="anthropic.claude-v2", client=boto3_bedrock, model_kwarg={"max_tokens_to_sample": 500})
memory_chain = ConversationBufferMemory(memory_key="chat_history", return_messages=True)
qa = ConversationalRetrievalChain.from_llm(
    llm=cl_llm,
    retriever=vectorstore_faiss_aws.as_retriever(),
    #retriever=vectorstore_faiss_aws.as_retriever(search_type='similarity', search_kwargs={"k": 8}),
    memory=memory_chain,
    get_chat_history=_get_chat_history,
    verbose=True,
    condense_question_prompt=condense_prompt_claude,
    chain_type='stuff', # 'refine',
    #max_tokens_limit=300
)

# the LLMChain prompt to get the answer. the ConversationalRetrievalChain does not expose this parameter in the constructor
qa.combine_docs_chain.llm_chain.prompt = PromptTemplate.from_template("""
{context}

Human: Use at maximum 3 sentences to answer the question inside the <q></q> XML tags.

<q>{question}</q>

Do not use any XML tags in the answer. If the answer is not in the context say "Sorry, I don't know as the answer was not found in the context"

Assistant:""")


```

A red arrow points to the line `qa = ConversationalRetrievalChain.from_llm(`. The bottom right corner of the code cell contains a blue speech bubble icon.

At the bottom of the interface, there are various status indicators and buttons: Simple (radio button), 1 \$ 2 (radio button), main (radio button), 2 (radio button), Python 3 (Data Science 3.0) | Idle, Kernel: Idle | Instance MEM, Cookie Preferences, Mode: Command, Ln 1, Col 1, 00_Chatbot_Claude.ipynb, and a bell icon.

```
# recreate the Claude LLM with more tokens to sample - this provides longer responses but introduces some latency
cl_llm = Bedrock(model_id="anthropic.claude-v2", client=boto3_bedrock, model_kwarg={"max_tokens_to_sample": 500})
memory_chain = ConversationBufferMemory(memory_key="chat_history", return_messages=True)
qa = ConversationalRetrievalChain.from_llm(
    llm=cl_llm,
    retriever=vectorstore_faiss_aws.as_retriever(),
    #retriever=vectorstore_faiss_aws.as_retriever(search_type='similarity', search_kwargs={"k": 8}),
    memory=memory_chain,
    get_chat_history=_get_chat_history,
    verbose=True,
    condense_question_prompt=condense_prompt_claude,
    chain_type='stuff', # 'refine',
    #max_tokens_limit=300
)
# the LLMChain prompt to get the answer. the ConversationalRetrievalChain does not expose this parameter in the constructor
qa.combine_docs_chain.llm_chain.prompt = PromptTemplate.from_template("""
{context}

Human: Use at maximum 3 sentences to answer the question inside the <q></q> XML tags.

<q>{question}</q>

Do not use any XML tags in the answer. If the answer is not in the context say "Sorry, I don't know as the answer was not found in the context"

Assistant:""")


```

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | 00_Chatbot_C... (2) - Jupyter | +

← → ⌛ ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update ⋮

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 7 hours, 57 minutes) ⓘ [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Let's start another chat. Feel free to ask the following questions:

1. What is SageMaker?
2. what is canvas?

chat = ChatUX(qa, retrievalChain=True)
chat.start_chat()

Do some prompt engineering

You can "tune" your prompt to get more or less verbose answers. For example, try to change the number of sentences, or remove that instruction all-together. You might also need to change the number of max_tokens_to_sample (eg 1000 or 2000) to get the full answer.

▼ In this demo we used Claude LLM to create conversational interface with following patterns:

1. Chatbot (Basic - without context)
2. Chatbot using prompt template(Langchain)
3. Chatbot with personas
4. Chatbot with context

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon | Studio Classic - SageMaker | 00_Chatbot_C... (2) - Jupyter | +

← → ⌛ ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 7 hours, 56 minutes) ⓘ [Learn More](#)

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb

Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share

Let's start another chat. Feel free to ask the following questions:

1. What is SageMaker?
2. what is canvas?

[24]: `chat = ChatUX(qa, retrievalChain=True)`
`chat.start_chat()`

Starting chat bot

You: What is SageMaker? Send

AI: Based on the provided context, Amazon SageMaker is a fully managed service from Amazon Web Services (AWS) that helps data scientists and developers prepare, build, train, and deploy machine learning models at scale. Some key features of Amazon SageMaker include:

- Fully managed ML infrastructure and workflows to quickly build, train, and deploy models.
- Ability to bring your own notebooks, algorithms, and data and train using SageMaker's managed infrastructure.
- Tools to label and prepare data, build models, tune and optimize them for deployment.
- Secure – encrypts data in transit and at rest, integrates with IAM, VPC, etc.
- Pay only for what you use – no minimum fees or upfront commitments.

So in summary, Amazon SageMaker provides the tools and managed infrastructure to build, train, and deploy ML models in a scalable, secure, and flexible manner.

You: q to quit Send

Feedback icon

moda-workshop-dryrun-2024 | SageMaker Studio | Amazon S | Studio Classic - SageMaker S | 00_Chatbot_C... (2) - Jupyter +

← → G ⌂ d-ny7wttpl3fkp.studio.us-west-2.sagemaker.aws/jupyter/default/lab/tree/amazon-bedrock-workshop/06_OpenSource_examples/02_Langchain_Chatbot... ☆ T Finish update :

JupyterLab 3 notebooks on SageMaker Studio Classic will reach end of support on December 31, 2024. (35 days, 7 hours, 56 minutes) ⓘ Learn More ⓘ

We will automatically migrate all accounts to the new Studio, featuring JupyterLab 4-based notebooks before this date. You can manually migrate to the new Studio. Your EFS data will automatically be available in

Amazon SageMaker Studio Classic File Edit View Run Kernel Git Tabs Settings Help

Home Terminal 1 bedrock_basics.ipynb 00_Chatbot_Claude.ipynb Cluster Data Science 3.0 Python 3 2 vCPU + 4 GiB Share ⚙️

You: what is canvas? Send

AI: Amazon SageMaker Canvas is a visual, no-code machine learning service that allows you to build, train, and deploy ML models without writing any code. Some key things to know about SageMaker Canvas:

- It provides a drag-and-drop interface to visualize your data, clean and prepare it, and train ML models. No coding or ML expertise is required.
- It supports connecting to data sources like Amazon S3, Amazon Redshift, Snowflake, and uploading local files. You can combine and prepare data from multiple sources.
- It automatically cleans your data, handles missing values, and suggests data transformations to improve model accuracy.
- It trains high-quality ML models with a single click using your prepared data. A variety of algorithms are supported.
- You can interpret, explain, and share the ML models visually with others in your organization.
- It is fully managed, so you don't need to provision any infrastructure.
- You pay only for what you use based on session time and training dataset size.

In summary, SageMaker Canvas makes ML accessible to business analysts by providing a no-code, visual interface to prepare data, train models, and generate predictions.

You: q to quit Send

Do some prompt engineering

You can "tune" your prompt to get more or less verbose answers. For example, try to change the number of sentences, or remove that instruction all-together. You might also need to change the number of max_tokens_to_sample (eg 1000 or 2000) to get the full answer.

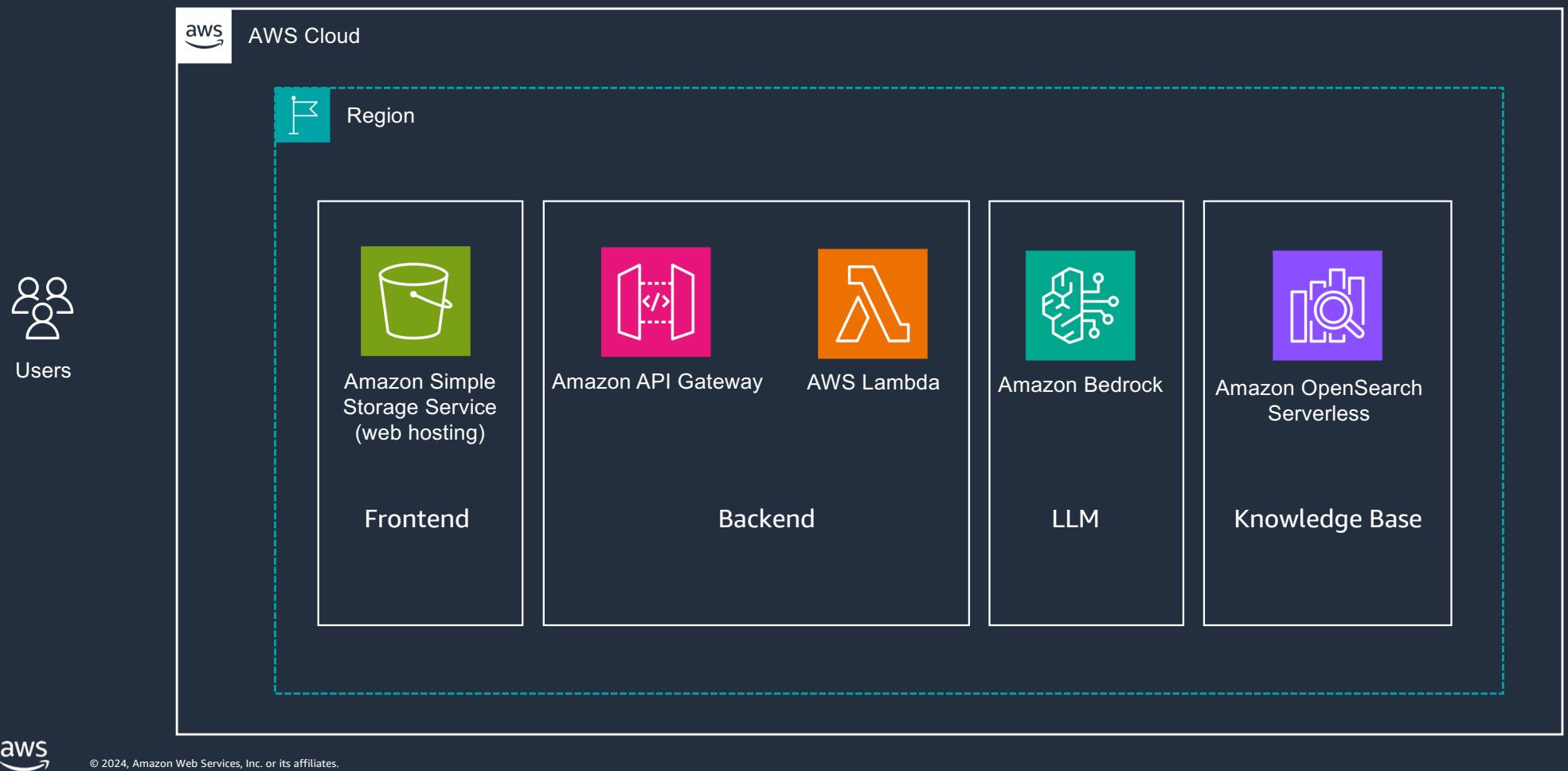
Options for Solution Adoption

- Develop prototype using AWS components, with full control on customization
- Leverage open-source project and accelerate time-to-market
- Evaluate and buy from vendor's pre-assembled solution



© 2024, Amazon Web Services, Inc. or its affiliates.

A minimalist architecture



```
[37]: # adjust modelId for the workshop demo  
modelId = 'anthropic.claude-3-5-sonnet-20240620-v1:0'
```

```
•[38]: # adjust knowledge case id for the workshop demo  
kb_id = "CJOMCDOKNC"
```

```
[ ]:
```

```
[26]: # from Langchain.Llms.bedrock import Bedrock  
import langchain  
from langchain_aws import ChatBedrock  
from langchain.retrievers.bedrock import AmazonKnowledgeBasesRetriever  
→ llm = ChatBedrock(model_id=modelId,  
                     client=bedrock_client)
```



© 2024, Amazon Web Services, Inc. or its affiliates.

```
[56]: from langchain.prompts import PromptTemplate

PROMPT_TEMPLATE = """
Human: You are an intelligent and friendly assistant, and provides answers to questions based on facts wehre possible.
Use the following pieces of information to provide a concise answer to the question enclosed in <question> tags.
Do take yuor time to validate your answer before replying.
If you don't know the answer, just say that you don't know, don't try to make up an answer.
Do not answer any political or religious questions. Just reply "我只針對氣象局內建檔資料，進行問題的回覆。如有疑問，請聯繫管理員 Michael。
Be sure to answer in traditioanl chineese, using phrases most frequently used in Taiwan.

<context>
{context}
</context>

<question>
{question}
</question>

The response should be specific and use statistics or numbers when possible.

Assistant:"""
claude_prompt = PromptTemplate(template=PROMPT_TEMPLATE,
                                input_variables=["context", "question"])
```



© 2024, Amazon Web Services, Inc. or its affiliates.

```
[39]: query = "113年度新聞媒體從業人員氣象資訊推廣說明會，總共有幾場?"  
# query = "北區場次的日期與地點?"  
# query = "111年度預算執行情形?"
```



```
retriever = AmazonKnowledgeBasesRetriever(  
    knowledge_base_id=kb_id,  
    retrieval_config={"vectorSearchConfiguration":  
        {"numberOfResults": 4,  
         'overrideSearchType': "SEMANTIC", # optional  
        },  
    },  
    # endpoint_url=endpoint_url,  
    # region_name=region,  
    # credentials_profile_name=<profile_name>,  
)  
docs = retriever.get_relevant_documents(  
    query=query  
)  
pp pprint(docs)
```

[Document(metadata={'location': {'s3Location': {'uri': 's3://cwa-rag-demo-bucket-20240220/CWA_2024_MediaBrief.pdf'}, 'type': 'S3'}, 'score': 0.7734306}, page_content='1 交通部中央氣象署 「113年度新聞媒體從業人員氣象資訊推廣說明會」實施計畫 壹、目的 交通部中央氣象署為增進國內新聞媒體從業人員對本署氣象服務之認識，正確解讀氣象資訊及妥適應用，特以氣象防災及地震防災為主題，分別 於北、東、中、南4區各舉辦1場本說明會，以深化氣象資訊之推廣與運用。貳、主(協)辦機關 一、主辦機關：中央氣象署 二、共同主辦：臺東縣政府、國立自然科學博物館 參、舉辦方式 一、參加對象：國內各媒體業者(報紙、廣播、電視等) 二、說明會共4場，日期時間及地點如下表： 場次 日期 地點 北區 2月29日 (星期四) 中央氣象署617會議室 (臺北市中正區公園路64號6樓) 南區 3月7日 (星期四) 中央氣象署臺灣南區氣象中心1樓視聽室 (臺南市中西區公園路21號) 東區 3月20日 (星期三) 中央氣象署臺東氣象站 (臺東市大同路106號) 中區 4月3日 (星期三) 國立自然科學博物館第二科學教室 (臺中市北區館前路1號) 三、研討會議程 時間 議程 講座 09:00~09:20 報到 09:20~09:30 致詞 09:30~11:00 氣候變遷下之氣象防災精進作為 氣象預報中心 11:00~11:10 休息 11:10~12:10 地震速報資訊的通報與應用 地震測報中



© 2024, Amazon Web Services, Inc. or its affiliates.

```
qa = RetrievalQA.from_chain_type(
    llm=llm,
    chain_type="stuff",
    retriever=retriever,
    return_source_documents=True,
    chain_type_kwargs={"prompt": claude_prompt}
)

[42]: answer = qa.invoke(query)
pp.pprint(answer)

{'query': '113年度新聞媒體從業人員氣象資訊推廣說明會，總共有幾場？',
 'result': '根據提供的資訊，113年度新聞媒體從業人員氣象資訊推廣說明會總共有4場。具體場次如下：\n\n'
 '1. 北區場次：2月29日(星期四)在中央氣象署617會議室舉行\n'
 '2. 南區場次：3月7日(星期四)在中央氣象署臺灣南區氣象中心1樓視聽室舉行\n'
 '3. 東區場次：3月20日(星期三)在中央氣象署臺東氣象站舉行\n'
 '4. 中區場次：4月3日(星期三)在國立自然科學博物館第二科學教室舉行\n\n'
 '這4場說明會分別在北、南、東、中四個區域舉辦，目的是為了增進國內新聞媒體從業人員對氣象實務的認識，並正確解讀和應用氣象資訊。',
 'source_documents': [ Document(metadata={'location': {'s3Location': {'uri': 's3://cwa-rag-demo-bucket-20240220/CWA_2024_MediaBrief.pdf'}}, 'type': 'PDF'}, page_content='1 交通部中央氣象署 「113年度新聞媒體從業人員氣象資訊推廣說明會」實施計畫 壹、目的 交通部中央氣象署為增進國內新聞媒體從業人員對本署氣象實務之認識，正確解讀氣象資訊及妥適應用，特以氣象防災及地震防災為主題，分別於北、東、中、南4區各舉辦1場本說明會，以深化氣象資訊之推廣與運用。貳、主(協)辦機關 一、主辦機關：中央氣象署 二、共同主辦：臺東縣政府、國立自然科學博物館 參、舉辦方式 一、參加對象：國內各媒體業者(報紙、廣播、電視等) 二、說明會共4場，日期時間及地點如下表： 場次 日期 地點 北區 2月29日 (星期四) 中央氣象署617會議室 (臺北市中正區公園路64號6樓) 南區 3月7日 (星期四) 中央氣象署臺灣南區氣象中心1樓視聽室 (臺南市中西區公園路21號) 東區 3月20日 (星期三) 中央氣象署臺東氣象站 (臺東市大同路106號) 中區 4月3日 (星期三) 國立自然科學博物館第二科學教室 (臺中市北區館前路1號) 三、研討會議程 時間 講座 09:00~09:20 報到 09:20~09:30 致詞 09:30~11:00 氣候變遷下之氣象防災精進作為 氣象預報中心 11:00~11:10 休息 11:10~12:10 地震速報資訊的通報與應用 地震測報中心 12:10~13:30 互動交流及用餐 2 四、欲參加之媒體人員可依其便利性，擇一場次參加。肆、報名方式及研習證明一、各場次報名截止日為活動舉辦日前2週，請逕至說明會資訊平台(網址：https://reurl.cc/WRX7ly)下報名頁面填寫報名表。北區場次：自即日起至113年2月14日(星期三)止。南區場次：自即日起至113年2月28日(星期三)止。東區場次：自即日起至113年3月5日(星期二)止。中區場次：自即日起至113年3月12日(星期二)止。'}
```

← → C ⌂ 2fd1ccf319ad533bc0.gradio.live ☆ | ⌂ T Error :

Knowledge Retrieval for CWA

請輸入問題

113年度新聞媒體從業人員氣象資訊推廣說明會，總共有幾場？直接給答案。

[Clear](#) [Submit](#)

這裏是答案

根據提供的資訊，113年度新聞媒體從業人員氣象資訊推廣說明會總共舉辦4場。

[Flag](#)



© 2024, Amazon Web Services, Inc. or its affiliates.

Illustrative Open Source Project

github.com/aws-samples/bedrock-claude-chat

README Code of conduct MIT-0 license Security

Bedrock Claude Chat

CDK tests passing

⚠ Warning

A major update to v2 is planned soon. v2 will not be backward compatible with v1, and **existing RAG bots will no longer be usable**. For more details, please refer to the [migration guide](#).

⚠ Warning

If you are using old version (e.g. v0.4.x) and wish to use the latest version, refer [migration guide](#). Without any care, **ALL DATA IN Aurora cluster WILL BE DESTROYED, and NO LONGER USERS CANNOT USE EXISTING BOTS WITH KNOWLEDGE AND CREATE NEW BOTS.**

This repository is a sample chatbot using the Anthropic company's LLM [Claude](#), one of the foundational models provided by [Amazon Bedrock](#) for generative AI.

Packages
No packages published

Contributors 56

+ 42 contributors

Languages

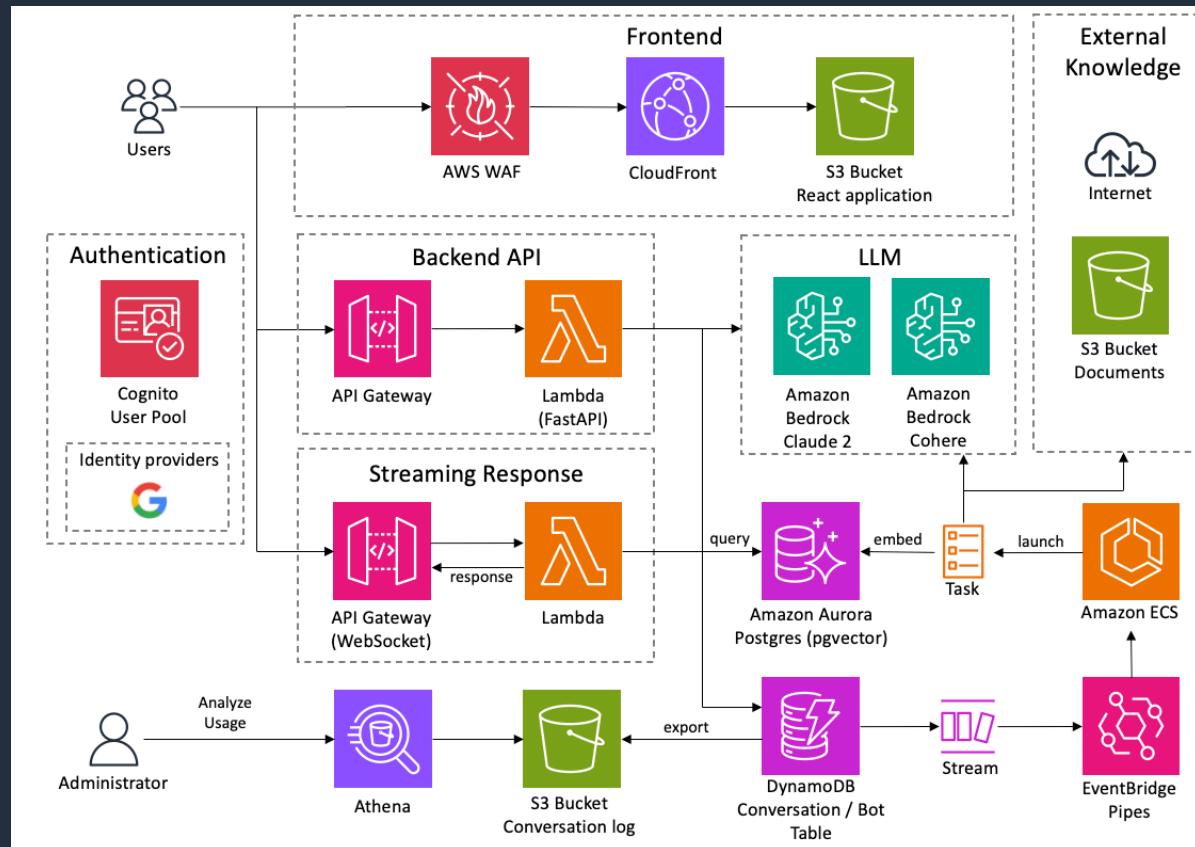
Language	Percentage
TypeScript	61.9%
Python	36.2%
JavaScript	1.2%
Shell	0.4%
Dockerfile	0.2%
HTML	0.1%

<https://github.com/aws-samples/bedrock-claude-chat>



© 2024, Amazon Web Services, Inc. or its affiliates.

With Enterprise Considerations in Mind



<https://github.com/aws-samples/bedrock-claude-chat/>



© 2024, Amazon Web Services, Inc. or its affiliates.

- 新的聊天
- Bot Console
- 我的最愛 Bots
- 最近用過的 Bots
- demo-bot-mirdc-20241014
- 交談紀錄
- 台北捷運內禁止使用

demo-bot-mirdc-20241014

沒有描述

同步完成 

...

model: claude-v3-haiku



在台北捷運內可不可以溜直排輪



knowledge_base_tool

▼



<result> 根據搜索結果,在台北捷運範圍內不得使用直排輪、溜冰鞋、滑板、滑板車或其他類似器材。

這是捷運系統的規定,旅客不得在捷運範圍內使用這些器材。

主要原因包括:

1. 為了維護捷運系統的安全運作,避免對其他旅客造成危害或干擾。
2. 直排輪等器材可能會對自己或他人的安全造成影響。
3. 在車站或月台上使用這些器材可能會阻礙其他人的通行。
4. 老、幼、重病等需要協助而無人協助的旅客,不得使用這些器材。
5. 攜帶物品可能會造成他人不便。
6. 參與選舉活動、政黨黨務活動、抗議集會或其他政治性議題活動。
7. 其他經本公司公告禁止之事項。

因此,在台北捷運範圍內不得使用直排輪、溜冰鞋、滑板、滑板車或其他類似器材。旅客如有違反,除應補繳票價外,並應支付違約金50倍。

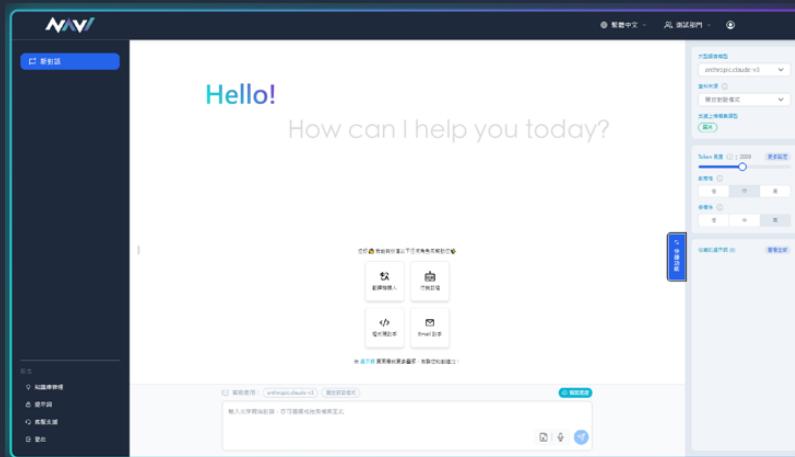
</result>

⟳ 重新生成

請輸入訊息



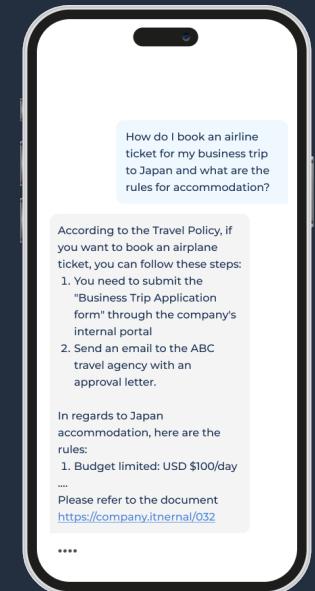
Illustrative Vendor Solutions



<https://www.netron.net/tw/brand/navi>



<https://www.no8.io/>



<https://www.going.cloud/>



© 2024, Amazon Web Services, Inc. or its affiliates.



Thank you!

Michael Lin

linmicht@amazon.com