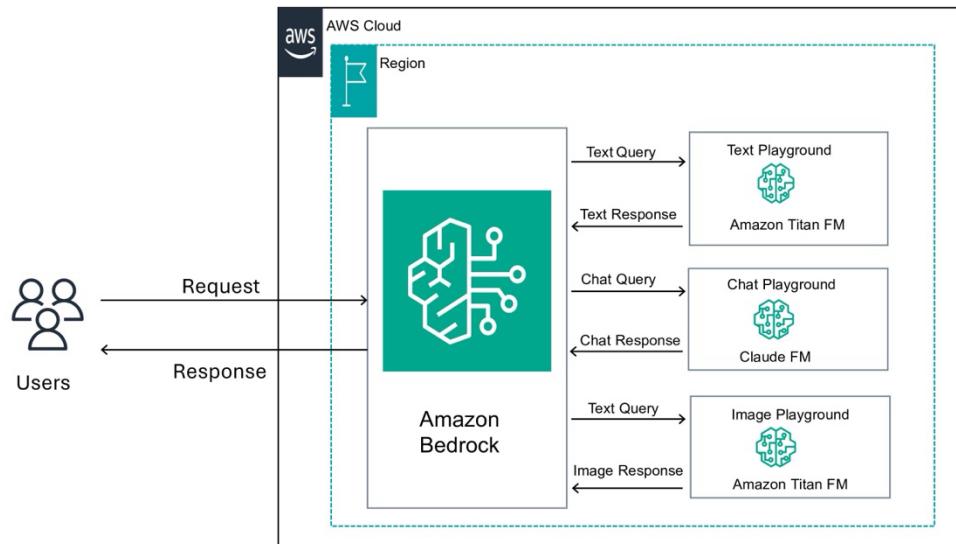


Guide to Using Amazon Bedrock on AWS: Model Access, Configuration, and Content Generation



By
Mahendran Selvakumar
<https://devopstronaut.com/>

What is GenAI?

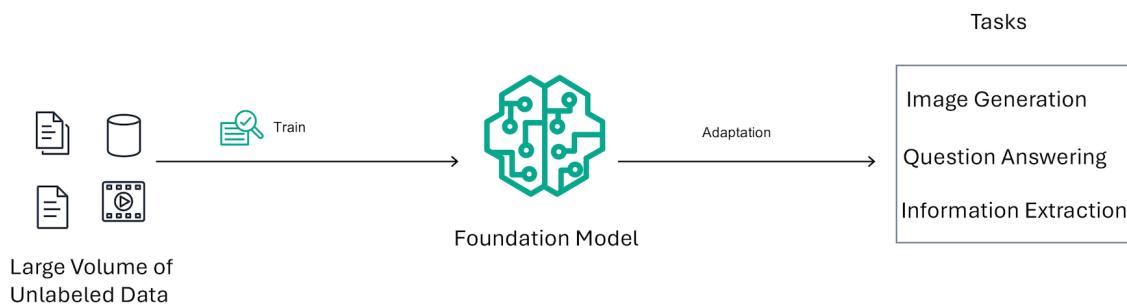
GenAI, or Generative AI, is technology that creates new content like text, images, music, or videos based on patterns learned from data. It powers tools that generate human-like writing, artwork, or even software code, enabling creative outputs across various fields, from entertainment to education.

GenAI on AWS:

GenAI on AWS refers to the use of Amazon's cloud services to create, deploy, and scale systems that generate content like text, images, and code. AWS offers tools like Amazon Bedrock and Sage Maker to help developers access and build these systems, leveraging powerful infrastructure for efficient model training, deployment, and integration into applications.

What is Foundation Model?

A foundation model is a large, pre-trained system that serves as a base for various tasks, such as text generation, translation, or image analysis. It is trained on vast datasets and can be adapted or fine-tuned for specific uses, making it versatile and efficient for different applications.



What is Amazon Bedrock?

Amazon Bedrock is a fully managed service that enables users to access and build on pre-trained foundation models from providers like Anthropic, Stability AI, and AWS's own models. It simplifies the integration of generative AI into applications without requiring specialized machine learning expertise. Key features include:

- **Fully Managed Service:** No need to manage servers.
- **Data Control:** Maintain control over your training data.
- **Pay-Per-Use Pricing:** Only pay for what you use.
- **Unified APIs:** Seamlessly integrate with a range of models.
- **Diverse Models:** Access a wide array of foundation models for various applications.



1. Create Amazon Bedrock and Enable Model Access:

Go to the AWS Console and select the US East region to access all Generative AI services

The screenshot shows the AWS Console homepage. In the top right corner, there is a dropdown menu for selecting a region. The 'us-east-1' region is highlighted as the current selection. Other regions listed include 'US East (Ohio)', 'US West (N. California)', 'US West (Oregon)', and 'Asia Pacific (Malaysia)'. The 'us-east-1' entry also includes 'us-east-1' in parentheses.

Search “Amazon Bedrock”

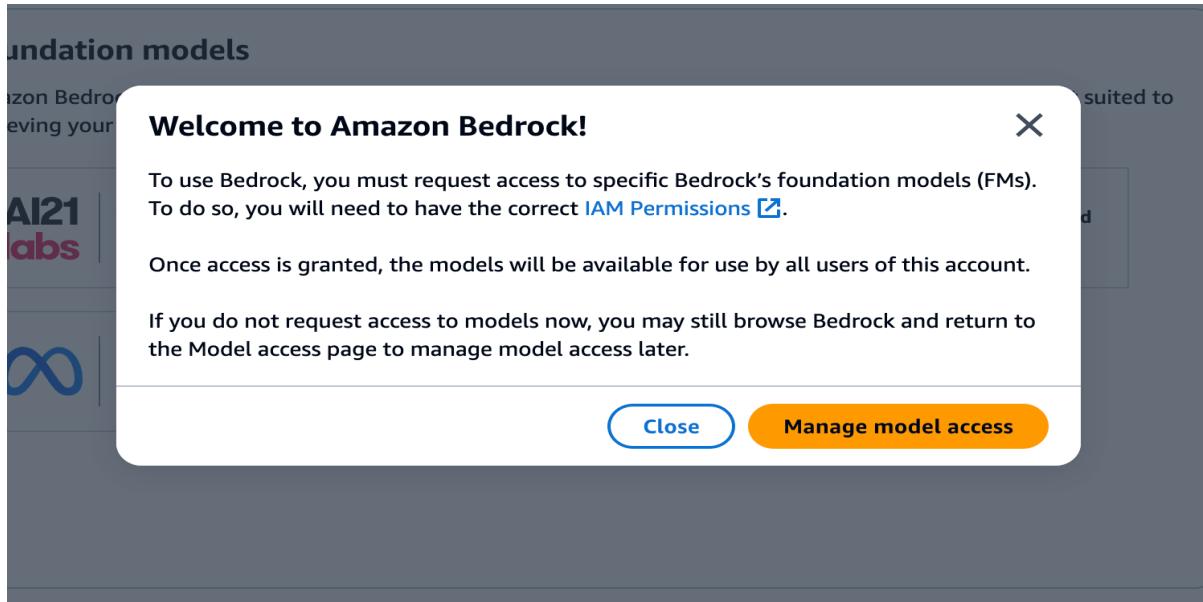
The screenshot shows the AWS search results for 'Amazon Bedrock'. The search bar at the top contains the query 'Amazon Bedrock'. Below the search bar, the results are categorized under 'Services'. The first result is 'Amazon Bedrock', which is described as 'The easiest way to build and scale generative AI applications with foundation models (FMs)'. Other results listed include 'AWS Private Certificate Authority', 'Amazon Redshift', and 'Amazon EventBridge'.

Click “Get started”

The screenshot shows the Amazon Bedrock landing page. At the top, there is a 'Machine Learning' navigation item. The main heading is 'Amazon Bedrock', followed by the subtext 'The easiest way to build and scale generative AI applications with foundation models (FMs)'. To the right of the text, there is a white call-to-action box containing a 'Try Bedrock' button and a 'Get started' button.



Click “Manage model access”



Now we can see the access status as 'Available to request' so we need to submit a request for model access

Models	Access status	Modality	EULA
▼ AI21 Labs (5)	0/5 access granted		
Jamba 1.5 Mini	⚠️ Unavailable	Text	EULA
Jamba 1.5 Large	⚠️ Unavailable	Text	EULA
Jamba-Instruct	⌚ Available to request	Text	EULA
Jurassic-2 Ultra	⌚ Available to request	Text	EULA
Jurassic-2 Mid	⌚ Available to request	Text	EULA
▼ Amazon (8)	0/8 access granted		
Titan Embeddings G1 - Text	⌚ Available to request	Embedding	EULA
Titan Text G1 - Lite	⌚ Available to request	Text	EULA
Titan Text G1 - Express	⌚ Available to request	Text	EULA
Titan Image Generator G1 v2	⌚ Available to request	Image	EULA
Titan Image Generator G1	⌚ Available to request	Image	EULA
Titan Multimodal Embeddings G1	⌚ Available to request	Embedding	EULA
Titan Text G1 - Premier	⌚ Available to request	Text	EULA
Titan Text Embeddings V2	⌚ Available to request	Embedding	EULA



Click “Enable all models”

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 (36) C **Collapse all**

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

Click “Next” to enable access

Edit model access

Base models (34/36)

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

Model	Access status	Modality	EULA
AWSCloudLabs	6/6 access granted	Text	EULA
Amazon T.5 Micro	▲ Available	Text	EULA
Amazon T.5 Large	▲ Available	Text	EULA
Amazon T.5 Instruct	▢ Available to request	Text	EULA
Amazon T.0 Ultra	▢ Available to request	Text	EULA
Amazon T.0 Mid	▢ Available to request	Text	EULA
Amazon T.0 Small	▢ Available to request	Text	EULA
Amazon ImageEmbedding G1 - Text	▢ Available to request	Embedding	EULA
Amazon Text G1 - Text	▢ Available to request	Text	EULA
Amazon Text G1 - Images	▢ Available to request	Text	EULA
Amazon ImageGenerator G1	▢ Available to request	Image	EULA
Amazon ImageGenerator G1	▢ Available to request	Image	EULA
Amazon MultimodalEmbedding G1	▢ Available to request	Embedding	EULA
Amazon Text G1 - Poetry	▢ Available to request	Text	EULA
Amazon Text Embeddings V2	▢ Available to request	Embedding	EULA
Amazon (S)	6/6 access granted	Text & Vision	EULA
Amazon S.5 Small	▲ Available	Text & Vision	EULA
Amazon S.5 Medium	▲ Available	Text & Vision	EULA
Amazon S.5 Large	▢ Available to request	Text & Vision	EULA
Amazon S.5 Instruct	▢ Available to request	Text & Vision	EULA
Amazon S.0 Ultra	▢ Available to request	Text	EULA
Amazon S.0 Mid	▢ Available to request	Text	EULA
Amazon S.0 Small	▢ Available to request	Text	EULA
Amazon S.0 Large	▢ Available to request	Text	EULA
Amazon S.0 Instruct	▢ Available to request	Text	EULA
Amazon T.0 Chat 12B	▲ Available	Text	EULA
Amazon T.0 Chat 7B	▲ Available	Text	EULA
Amazon T.0 Chat 1B	▲ Available	Text	EULA
Amazon T.0 Chat 0B	▲ Available	Text	EULA
Amazon T.0 Text	6/6 access granted	Text	EULA
Amazon T.0 Image	▢ Available to request	Text	EULA
Amazon T.0 Video	▢ Available to request	Text	EULA
Amazon T.0 Object	▢ Available to request	Text	EULA
Amazon T.0 MultiModal	▢ Available to request	Text	EULA
Amazon Large (4X4)	▢ Available to request	Text	EULA
Amazon Small (2X2)	▢ Available to request	Text	EULA
Amazon X-Tiny	6/6 access granted	Image	EULA

Cancel **Next**

Provide the company name, URL, and select 'Internal Employees' as the intended users and the relevant industry type. Then, describe the use case and click 'Next'



Use case details

You must provide the following information to access the Anthropic models. While you will instantly get access to the models, the information you provide below will be shared with Anthropic.

Company name

Devopstronaut

Company website URL

<https://devopstronaut.com>

What industry do you operate in?

Education

Who are your intended users?

- Internal employees
- External users

Describe your use cases (do not share any PII or IP information).

Tip: Avoid mentioning the term "Claude", as that will be identified as PII and access will automatically be denied.

Learn Amazon Bedrock

Maximum 500 characters

i Aggregated metrics (not including Your Content) arising from your use of Anthropic models on Amazon Bedrock will be shared with Anthropic for purposes of trust and safety. [Learn more](#)

[Cancel](#)

[Previous](#)

[Next](#)

Click 'Submit' and it take some time to enable the model access

Step 2: Add use case details for Anthropic

[Edit](#)

Use case details

Company name
Devopstronaut

Company website URL
<https://devopstronaut.com>

What industry do you operate in?
Education

Who are your intended users?
Internal employees

Describe your use cases (do not share any PII or IP information).
Learn Amazon Bedrock

i **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](#)

We might see a 'failed' error, but you can ignore it because the model is not available at the moment



- ⌚ Access request for 8 models failed
- Jamba 1.5 Mini - Your account is not authorized to invoke this API operation.
 - Jamba 1.5 Large - Your account is not authorized to invoke this API operation.
 - Llama 2 13B - Could not create agreement - Agreement already exists
 - Llama 2 70B - Could not create agreement - Agreement already exists
 - Claude 3.5 Sonnet - Failed to create regional entitlement. Account is not authorized for the model.
 - Claude 3 Opus - Failed to create regional entitlement. Account is not authorized for the model.
 - 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](#)

Now we can see that the model access status is 'Access granted'

AI21 Labs (5)		3/5 access granted		
Jamba 1.5 Mini		Unavailable	Text	EULA
Jamba 1.5 Large		Unavailable	Text	EULA
Jamba-Instruct		Access granted	Text	EULA
Jurassic-2 Ultra		Access granted	Text	EULA
Jurassic-2 Mid		Access granted	Text	EULA
Amazon (8)		8/8 access granted		
Titan Embeddings G1 - Text		Access granted	Embedding	EULA
Titan Text G1 - Lite		Access granted	Text	EULA
Titan Text G1 - Express		Access granted	Text	EULA
Titan Image Generator G1 v2		Access granted	Image	EULA
Titan Image Generator G1		Access granted	Image	EULA
Titan Multimodal Embeddings G1		Access granted	Embedding	EULA
Titan Text G1 - Premier		Access granted	Text	EULA
Titan Text Embeddings V2		Access granted	Embedding	EULA
Anthropic (6)		4/6 access granted		

2. Using Text Playground:

Go to the Playground, choose 'Text' and click 'Select model'

Amazon Bedrock > Text playground

Text playground [Info](#)

[Select model](#)



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

Titan

Claude

Llama

AI21 Jamba-

Command

Mistral

<https://www.linkedin.com/in/mahendran-selvakumar- 36444a77/>



Select the model category as 'Amazon' and choose 'Titan Text G1-Express' ,Click 'Apply' with throughput set to 'OnDemand'

Select model

1. Category

Model providers

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

2. Model

Models with access (3)

- Titan Text G1 - Premier v1**
Text model | Context size = 32k
- Titan Text G1 - Lite v1**
Text model | Context size = 4k
- Titan Text G1 - Express v1**
Text model | Context size = 8k

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

3. Throughput

On-demand

No throughput options available. [Purchase Provisioned Throughput](#) to continue.

Cancel **Apply**

Enter the prompt and click “Run”

[Amazon Bedrock](#) > Text playground

Text playground [Info](#) [Load examples](#) [⋮](#)

Titan Text G1 - Express v1 | On-demand throughput [Change](#)

what is GenAI?

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

Titan Text G1 - Express

Action items from a meeting transcript
This prompt summarizes a meeting transcript into the action items assigned to each person

Titan Text G1 - Express

Creating a table of product descriptions
This is a prompt for generating product descriptions that incorporate keywords

Titan Text G1 - Express

JSON creation
This prompt creates synthetic data for a product in a JSON format.

Run ⏪ ⏴



We will receive the details based on the prompt

Amazon Bedrock > Text playground

Text playground Info

Titan Text G1 - Express v1 | On-demand throughput Change

what is GenAI?

GenAI is a new generation of artificial intelligence that is designed to be more human-like and capable of learning and adapting to new situations more quickly than previous generations of AI. It is based on a new type of neural network that is more efficient and scalable than previous networks. GenAI is being developed by a team of researchers at the University of California, San Francisco, and is being funded by a number of major technology companies.

One of the key features of GenAI is its ability to learn from large amounts of data. This data can come from a variety of sources, including text, images, and videos, and can be used to train the neural network to perform a wide range of tasks. GenAI can also learn from its own experiences, which can help it to improve its performance over time.

Another key feature of GenAI is its ability to adapt to new situations. This is done by using a technique called reinforcement learning, which involves training the neural network to make decisions based on feedback from its environment. GenAI can learn to perform tasks that it has never been trained on before, and can even learn to perform tasks that are impossible for humans to perform.

GenAI has a number of potential applications, including in healthcare, finance, and transportation. In healthcare, GenAI can be used to develop new treatments and therapies, and can be used to analyze medical data to identify patterns and trends. In finance, GenAI can be used to develop new trading strategies and to analyze financial data to identify risks and opportunities. In transportation, GenAI can be used to develop new autonomous vehicles and to improve traffic flow.

3. Using Chat Playground:

Go to the Playground, choose 'Text' and click 'Select model'

Amazon Bedrock <

Chat playground Info

Select model

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

a Titan Text G1 - Express Action items from a meeting transcript	A Claude Advanced Q&A with Citation	∞ Llama 2 Chat 13B Chain of thought An example prompt that	AI21 labs Jamba-Instruct Clinical documentation Q&A
--	---	---	---

Run



Choose a different model provider and model, then click 'Apply' with throughput set to 'OnDemand'

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 (2)

Request access

Claude 3 Opus v1

Cross-region inference



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

3. Throughput

On-demand

Cancel

Apply

Enter the same prompt and click 'Run'

Amazon Bedrock > Chat playground

Chat playground Info

Load examples

Compare mode



Anthropic Claude 3 Haiku v1 | On-demand throughput



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

Anthropic Claude 3 Haiku

Career Coach Role-play

An example prompt to set up in-

Anthropic Claude 3 Haiku

Data Extraction with JSON Outputs

An example prompt to turn unstructured

Anthropic Claude 3 Haiku

Removing PII

Automatically detect and remove

what is GenAI?

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.



It will provide a different output based on the selected model

Amazon Bedrock > Chat playground

Chat playground Info

AI Claude 3 Haiku v1 | On-demand throughput

Change

GenAI refers to generative artificial intelligence - a type of AI that can generate new content like text, images, audio, and other media. Some key things to know about GenAI:

- It uses large language models and other deep learning techniques to generate original content, rather than just retrieving or recombining existing information.
- Popular GenAI models include GPT-3, DALL-E, Midjourney, and Stable Diffusion, which can create human-like text, images, and other media based on textual prompts.

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

Run

We can also get additional details here, such as model metrics like latency and input token count

▼ Model metrics

To evaluate models for task specific metrics with custom dataset visit [Model evaluation](#)

Metrics | **Claude 3 Haiku**

Overall summary	Define metric criteria
Latency	4653 ms
Input token count	13
Output token count	282
Cost	-

4. Using Image Playground:

Note:

Generating images and text can incur additional costs, so avoid doing it if you don't need it

Go to the Playground, choose 'Image' and click 'Select model'



Amazon Bedrock

Getting started

Overview
Examples
Providers

Foundation models

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

Playgrounds

Chat
Text
Image

Builder tools

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

[Amazon Bedrock](#) > Image playground

Image playground [Info](#)

Select model



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

SDXL 1.0

Create an image

This prompt creates an image based on a

Titan Image Generator G1

Generate images from a text prompt

This prompt generates realistic images from

Choose the model type as 'Amazon Titan Image Generator G1' to generate images from a text prompt and click 'Run.' It may take some time to generate the image

[Amazon Bedrock](#) > Image playground

Image playground [Info](#)



Titan Image Generator G1 v1 | On-demand throughput



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

SDXL 1.0

Create an image

This prompt creates an image based on a

Titan Image Generator G1

Generate images from a text prompt

This prompt generates realistic images from

blue backpack on a table

▶ Run



we will receive the image output based on the text prompt

[Amazon Bedrock](#) > [Image playground](#)

Image playground [Info](#)

Titan Image Generator G1 v1 | On-demand throughput
[Change](#)



blue backpack on a table

Run

We can adjust the image orientation, size, and number of images in the configuration, then click 'Run'

[Amazon Bedrock](#) > [Image playground](#)

Image playground [Info](#)

[Load examples](#)

Titan Image Generator G1 v1 | On-demand throughput
[Change](#)



blue backpack on a table

Run

Configurations [Reset](#)

Negative prompt

[Info](#)

Response image

[Info](#)

Orientation

Landscape

Portrait

Size

1024 x 1024

Number of images

2

Advanced configurations [Info](#)

Prompt strength

8

Seed

0



After changing the configuration (orientation to Portrait and the number of images to 2), we will see the updated output based on the text prompt

The screenshot shows the Amazon Bedrock Image playground interface. At the top left is the AWS logo and the text "Amazon Bedrock > Image playground". On the right are buttons for "Load examples" and a three-dot menu. Below this is a card for "Titan Image Generator G1 v1 | On-demand throughput". The card displays two generated images of a blue backpack on a table. A text input field at the bottom contains the prompt "blue backpack on a table". To the right of the prompt is a yellow "Run" button. On the far right is a sidebar titled "Configurations" with sections for "Negative prompt" (Info), "Response image" (Info), "Orientation" (Portrait selected), "Size" (1024 x 1024), "Number of images" (2), and "Advanced configurations" (Prompt strength 8, Seed 0).

Conclusion:

Amazon Bedrock on AWS provides a powerful platform for accessing and utilizing advanced generative models. By following the steps outlined, you can efficiently manage model access, configure settings, and generate various types of content. This flexibility allows you to leverage Amazon Bedrock's capabilities for diverse applications, enhancing productivity and creativity in your projects

Keep Learning, Keep Implementing!!!

Feel free to reach out to me, if you have any other queries or suggestions

Stay connected on LinkedIn: <https://www.linkedin.com/in/mahendran-selvakumar-36444a77/>