




I got access to Amazon Titan Embeddings G1 - Text and Anthropic Claude 2.1 through Bedrock. Titan Embeddings is used to convert data into vectors for embeddings. This is useful because it captures the main idea of a sentence, which is needed when answering questions and queries. Claude 2.1 handles a lot of natural language processing tasks such as providing human-like responses.

Amazon Bedrock > Model catalog > Titan Embeddings G1 - Text



Titan Embeddings G1 - Text


By: Amazon |  Access granted




Open in playground


The new Titan Embeddings G1 – Text v1.2 can intake up to 8k tokens and outputs a vector of 1536 dimensions. The model also works in 25+ different languages. The model is optimized for text retrieval tasks but can also perform additional tasks such as semantic similarity and clustering. Titan Embeddings G1 – Text v1.2 also supports long documents, however, for retrieval tasks it is recommended to segment documents into logical segments (such as paragraphs or sections).

Amazon Bedrock > Model catalog > Claude 2.1



Claude 2.1


By: Anthropic |  Access granted




Open in playground

An update to Claude 2 that features double the context window, plus improvements across reliability, hallucination rates, and evidence-based accuracy in long document and RAG contexts.

I created and downloaded PDFs containing data relating to Formula 1’s history and statistics from 2000 to the present day, key technical terms, and general information about different grand prix. I put all the files into an S3 Bucket


 Upload succeeded

For more information, see the Files and folders table.



Upload: status

Close


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

Summary


Destination

s3://f1-test-atul

Succeeded

 3 files, 6.9 MB (100.00%)


Failed

 0 files, 0 B (0%)

Files and folders

Configuration

Files and folders (3 total, 6.9 MB)






Find by name

<

1

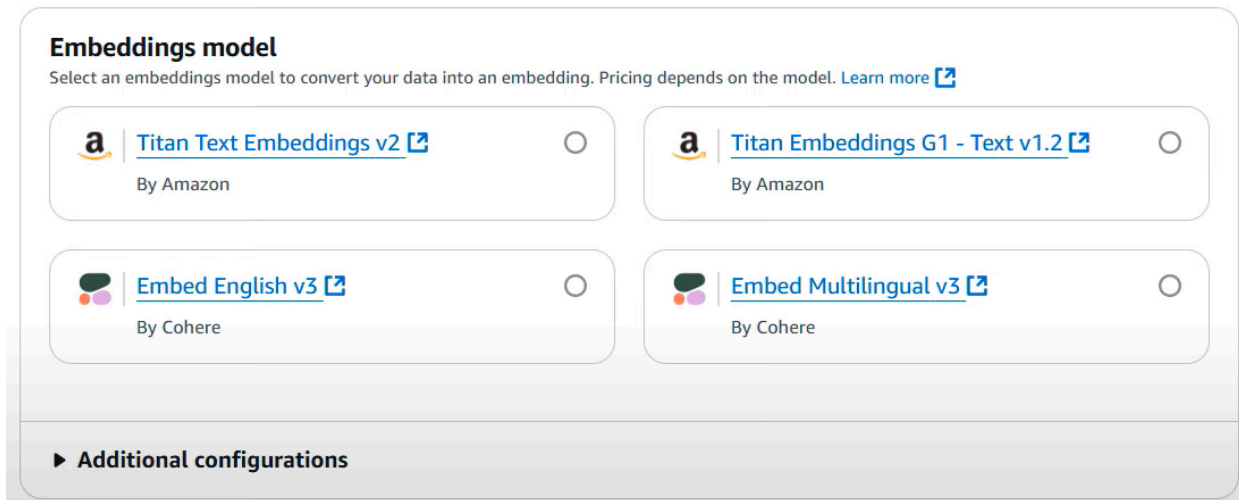
>

Name	Folder	Type	Size	Status	Error
F1-aero.pdf	-	application/pdf	1.5 MB	 Succeeded	-
Formula1-History.pdf	-	application/pdf	88.6 KB	 Succeeded	-
2018formula1.pdf	-	application/pdf	5.3 MB	 Succeeded	-

Next I created a knowledgebase in Bedrock that hooks into the S3Bucket along with Titan Embeddings to convert the data to a vector database.





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

► **Additional configurations**

I created an Amazon Lex bot that uses Bedrock and the knowledgebase I previously created and synced.

Then I built out two intents in Lex. The first was a general welcome intent that greets the user and prompts them to ask questions about Formula when the user greets it. The second intent is an AI QnA intent that uses Claude 2.1 to answer questions using the knowledgebase.

Amazon Lex



Draft version ▼

English (US) ▼

Not built

[← Back to intents list \(2\)](#)

Sort by last updated ▼

WelcomeIntent

FallbackIntent

Hey

Hello

Maximum 500 characters.

Add utterance

Initial response [Info](#)

You can provide messages to acknowledge the user's initial request. You can also configure next step in the conversation and branch based on conditions.

▼ Response to acknowledge the user's request

Message: Hi, what questions do you have about Formula 1 from 2000 to the present?

▼ Message group [Info](#)

You can define a text message group to respond using plain text.

Message - *optional*► Variations - *optional*[Advanced options](#)

Configure user request acknowledgement response, dialog code hook and conditional branches.

Editor

Visual builder

[New](#)

QnA configuration

Select model

The Knowledge base for Amazon Bedrock used to answer user queries.



Anthropic ▼

Claude V2:1 ▼

Choose a knowledge store

☒ **Knowledge base for Amazon bedrock**

Knowledge base for Amazon Bedrock – For more information, see [Building a knowledge base](#).



☐ **Opensearch**

Amazon OpenSearch Service database – For more information, see [Creating and managing Amazon OpenSearch Service domains](#).



☐ **Amazon Kendra**

Amazon Kendra index – For more information, see [Creating an index](#).



Knowledge base for Amazon Bedrock Id

The Knowledge base for Amazon Bedrock used to answer user queries.

NLMDCHOCJP

Maximum 10 characters. Valid characters: A-Z, a-z, 0-9

Associated ARN

arn:aws:bedrock:us-east-1:954332720039:knowledge-base/NLMDCHOCJP

Copy

Exact Response

Specifies whether to return an exact response or to return an answer generated by the model using the fields you specify from the database.

☐ Yes ☒ No

I didn't end up deploying it because of potential costs, but below are some tests that show the intents working and the chatbot functioning as intended and answering questions.

Inspect

Summary

JSON input and output

Intent

WelcomeIntent

Slots

Elicitation

Active contexts

Number of turns or seconds

Test Draft version

Last build submitted: Now

Inspect

Hi

Hi, what questions do you have about Formula 1 from 2000 to the present?

Ready for complete testing

Type a message

Inspect

Summary

JSON input and output

Intent

FallbackIntent

Slots

Elicitation

Active contexts

Number of turns or seconds

Test Draft version

Last build submitted: 1 minute ago

Inspect

Hi

Hi, what questions do you have about Formula 1 from 2000 to the present?

Who won the Driver's Championship in 2011?

Intent FallbackIntent is fulfilled

Ready for complete testing

Type a message

Test Draft version

Last build submitted: 2 minutes ago



Inspect

Hi

Hi, what questions do you have about Formula 1 from 2000 to the present?

Who won the World Driver's Championship in 2011?

Sebastian Vettel won the World Driver's Championship in 2011 driving for Red Bull Racing.

What races did he win that season?

In 2011, Sebastian Vettel won 11 races driving for Red Bull Racing.

List the races he won /

✓ Ready for complete testing



Type a message

Test Draft version

Last build submitted:

18 minutes ago



Inspect

What is DRS and how does it help cars go faster?

The Drag Reduction System (DRS) is a system that allows cars to reduce aerodynamic drag on straights, giving them a temporary speed boost. It can be activated if a car is within 1 second of the car in front when entering a DRS zone. By opening a flap in the rear wing, it reduces downforce and drag.

Who won the championship in 2008

Lewis Hamilton won the championship in 2008.

✓ Ready for complete testing



Type a message