

[Open in app](#)[Write](#)

Building your own LLM API with AWS Bedrock and Lambda

James Pog · [Follow](#)

2 min read · Nov 30, 2023



24



1



In this short and sweet tutorial, I wanted to highlight some tricky parts to the processing of using the recent AWS Bedrock release to build an LLM API.

This will be useful if you trying to learn how to build your own Large Language Model (LLM) app/API using AWS Bedrock and AWS Lambda.

Navigate to the AWS Bedrock page in the AWS Console and request access to the models you want to use. (in my experience access was granted instantly!)

Choice of foundation models from leading AI startups and Amazon

Jurassic-2 AI21labs Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch	Claude ANTHROPIC LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems	Stable Diffusion stability.ai Generation of unique, realistic, high-quality images, art, logos, and designs	Amazon Titan amazon Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings and search
--	---	---	---

Once you have access to the models, create a new Lambda in your AWS Account through the console, or however you usually provision infrastructure, and write out your code to access the bedrock service.

Now if you deploy your lambda already using Bedrock you'll notice something, the boto3 library inside the lambda doesn't actually have access to Invoke Bedrock.

Yup. That's right, AWS didn't update boto3 to know about Bedrock yet.

You need to add that dependency to your lambda, you can do it in a few ways, but one of them is through Lambda layers.

Here's a great video showing the workaround Step by step:

<https://www.youtube.com/watch?v=7PK4zdUgAt0>

Once you have the lambda working with Bedrock, all that is left to do is to link the lambda to API Gateway. There are plenty of tutorials for this online — <https://www.freecodecamp.org/news/how-to-setup-a-basic-serverless-backend-with-aws-lambda-and-api-gateway/>

You'll then have an API you can query from anywhere to access Bedrock LLMs and other models!

Note: Here's some of the code I used for my lambda, in python:

```
import os
from typing import Optional
import json

# External Dependencies:
import boto3
from botocore.config import Config

session_kwargs = {"region_name": "us-east-1"}
client_kwargs = {**session_kwargs}
retry_config = Config(
    region_name="us-east-1",
    retries={
        "max_attempts": 10,
        "mode": "standard",
    },
)
session = boto3.Session(**session_kwargs)
sts = session.client("sts")
response = sts.assume_role(RoleArn="role_arn", RoleSessionName="bedrockAssumeRole")
client_kwargs["aws_access_key_id"] = response["Credentials"]["AccessKeyId"]
client_kwargs["aws_secret_access_key"] = response["Credentials"]["SecretAccessKey"]
client_kwargs["aws_session_token"] = response["Credentials"]["SessionToken"]
client_kwargs["endpoint_url"] = "https://bedrock.us-east-1.amazonaws.com"
```

```
bedrock_client = session.client(service_name="bedrock",config=retry_config,**cli
prompt_data = "sample text"
body = json.dumps({"inputText": prompt_data})
modelId = "amazon.titan-tg1-large"
accept = "application/json"
contentType = "application/json"
response = bedrock_client.invoke_model_with_response_stream(
    body=body, modelId=modelId, accept=accept, contentType=contentType
)
response_body = json.loads(response.get('body').read())
outputText = response_body.get('results')[0].get('outputText')
```

AWS

AWS Lambda

Aws Bedrock

Software Engineering

**Written by James Pog**

Follow

1.1K Followers

Ex-Uber, Software Engineer at Amazon, I level up devs by 100x. Building AI startups. Follow me on twitter for more: <https://twitter.com/jamescodez>

More from James Pog



 James Pog

Why I will never be a Senior Software Engineer at Amazon

Probably..

🌟 · 4 min read · Jan 31, 2023



2.3K



42



 James Pog

Building the Ultimate AI Web Scraper..

building the best web scraper with GPT4 (part 1)

🌟 · 6 min read · Jan 23, 2024



12



1



 James Pog

How is Software developed at Amazon?

What the heck?!?!

🌟 · 5 min read · Jan 16, 2023



528



3



 James Pog

You will regret not learning AI

Here's why..

3 min read · May 27, 2023



392



7



[See all from James Pog](#)

Recommended from Medium



Neylson Crepalde

How to use Amazon Bedrock to write an article about Amazon...

Using Python SDK—Boto3

4 min read · Oct 10, 2023



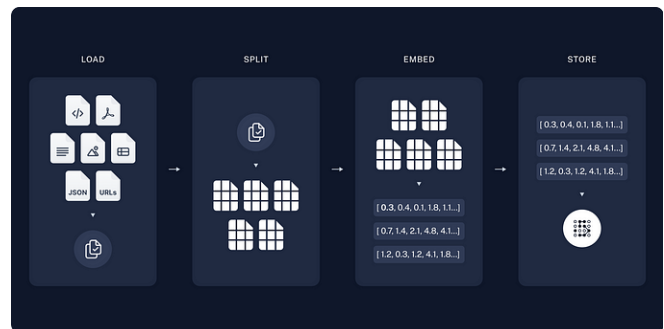
26



35



2



Heeki Park

Implementing RAG with Amazon Bedrock and AWS Lambda

Extending on my prior blog post, I wanted to continue tinkering with generative AI...

6 min read · Dec 2, 2023

Lists



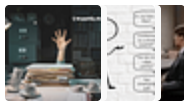
General Coding Knowledge

20 stories · 930 saves

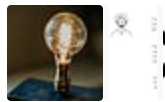


Stories to Help You Grow as a Software Developer

19 stories · 832 saves

**Leadership**

47 stories · 241 saves

**Good Product Thinking**

11 stories · 470 saves



Lakindu Hewawasam in Bits and Pieces

Should You Use Amazon API Gateway in 2024? Consider...

Did you know that you can now invoke Lambda functions using an HTTP API via...

8 min read · Feb 2, 2024



746



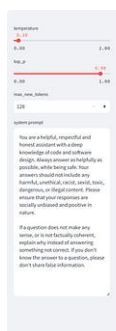
3



9.4K



122

**Code Llama Assistant**

Write a function that checks if a number is a prime number in JavaScript

To check if a number is a prime number in JavaScript, you can use the following function:

```
function isPrime(num) {  
  if (num <= 1) {  
    return false;  
  }  
  for (let i = 2; i <= Math.sqrt(num); i++) {  
    if (num % i === 0) {  
      return false;  
    }  
  }  
  return true;  
}
```

This function checks if the input number is less than 2, and if it is, it returns false. If it is greater than or equal to 2, it returns true.

How do I do Python

Sure! Here's the same function in Python:

```
def is_prime(num):  
    if num <= 1:  
        return False  
    for i in range(2, int(num**0.5) + 1):  
        if num % i == 0:  
            return False  
    return True
```

This function takes an integer as input and checks if it is a prime number by testing over the numbers from 2 to the square root of the number. If it is divisible by one of these, it is not a prime number. If it is not divisible by any of these, it is a prime number.

Dr. Leon Eversberg in Towards AI

How to Build Your Own LLM Coding Assistant With Code Llama

Creating a local LLM-chatbot with CodeLlama-7b-Instruct-hf and Streamlit

★ · 7 min read · 6 days ago



Hugo Lu in Towards Data Science

Microservices vs. Monolithic Approaches in Data

The Microservice vs. Monolith debate rages in software, but is reduced to a gentle simmer i...

★ · 8 min read · 4 days ago



606



4



218



3



See more recommendations