END-TO-END RETRIEVAL-AUGMENTED GENERATION WITH AMAZON BEDROCK

A Seamless Data Extraction and Response System

● ● ● Github-Link

Project Overview

Objective:

Develop a comprehensive RAG system using Amazon Bedrock, focused on efficient information retrieval and natural language processing.

INTRODUCTION TO AMAZON BEDROCK

- What is Amazon Bedrock?
- A managed service that provides access to state-of-the-art foundation models for generative AI.
- Offers pre-trained models from various AI providers, including those optimized for different use cases like text generation, summarization, and more.
- Benefits:
- Ease of Use: Simplifies the integration of advanced AI models into applications.
- Scalability: Automatically scales resources to meet demand.
- Security & Compliance: Provides robust security measures and compliance certifications.

Project Workflow

- Data Source: Start with a PDF file.
- Data Extraction: Extract and preprocess text from the PDF.
- Text Chunking: Split the extracted text into smaller chunks.
- Embedding Generation: Use Amazon Bedrock models to convert text chunks into vector embeddings.
- Vector Database Storage:
- Database Used: FAISS (Facebook AI Similarity Search)
- Purpose: Efficiently store and search for vector embeddings based on similarity.

Query Handling

- User Query Input: User submits a query.
- Similarity Check: Embedding model retrieves similar chunks from FAISS.
- Ranked Results: Retrieves and ranks similar results.

Language Model Interaction

- LLM Query: The query is sent to an LLM via Amazon Bedrock.
- Comparison & Matching: LLM compares the query with the retrieved embeddings.
- Result Selection: LLM selects the best matching response.

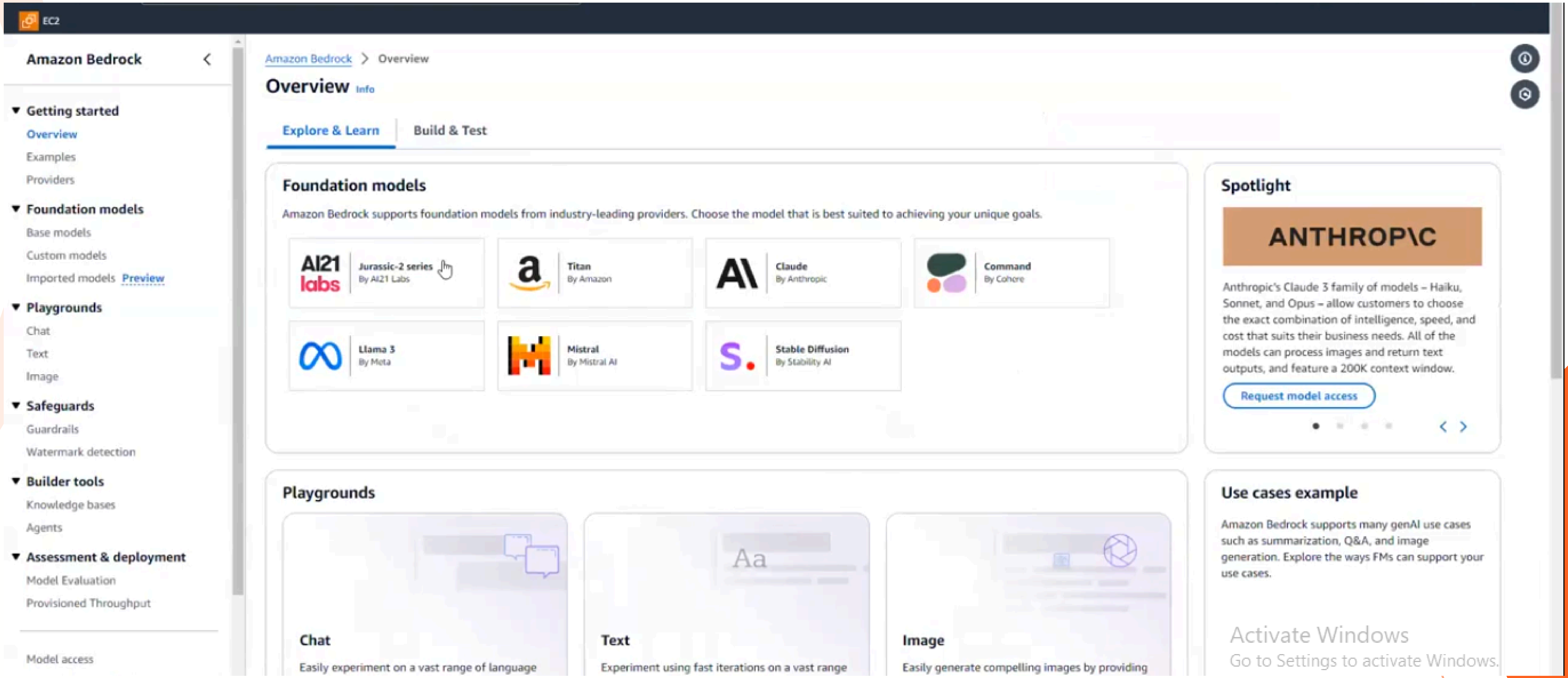
TECHNOLOGIES & TOOLS

- Amazon Bedrock: Central to utilizing pre-trained models.
- Vector Database: FAISS for storing and retrieving vector embeddings.
- Previous & Current Models:
- Previous: Sentence Transformer, Mistral-7B-Instruct (Hugging Face)
- Current: Mistral model from Amazon Bedrock
- Deployment Environment: EC2 instance, Streamlit app, and additional AWS services.

IMPLEMENTATION DETAILS

- Environment Setup: Conda environment and package installation.
- Deployment Process: EC2 instance setup, repository cloning, and app deployment.

WORKING WITH AMAZON BEDROCK



Foundation models

Amazon Bedrock supports foundation models from industry-leading providers. Choose the model that is best suited to achieving your unique goals.

Spotlight

Anthropic's Claude 3 family of models – Haiku, Sonnet, and Opus – allow customers to choose the exact combination of intelligence, speed, and cost that suits their business needs. All of the models can process images and return text outputs, and feature a 200K context window.

Playgrounds

Chat

Easily experiment on a vast range of language

Text

Experiment using fast iterations on a vast range

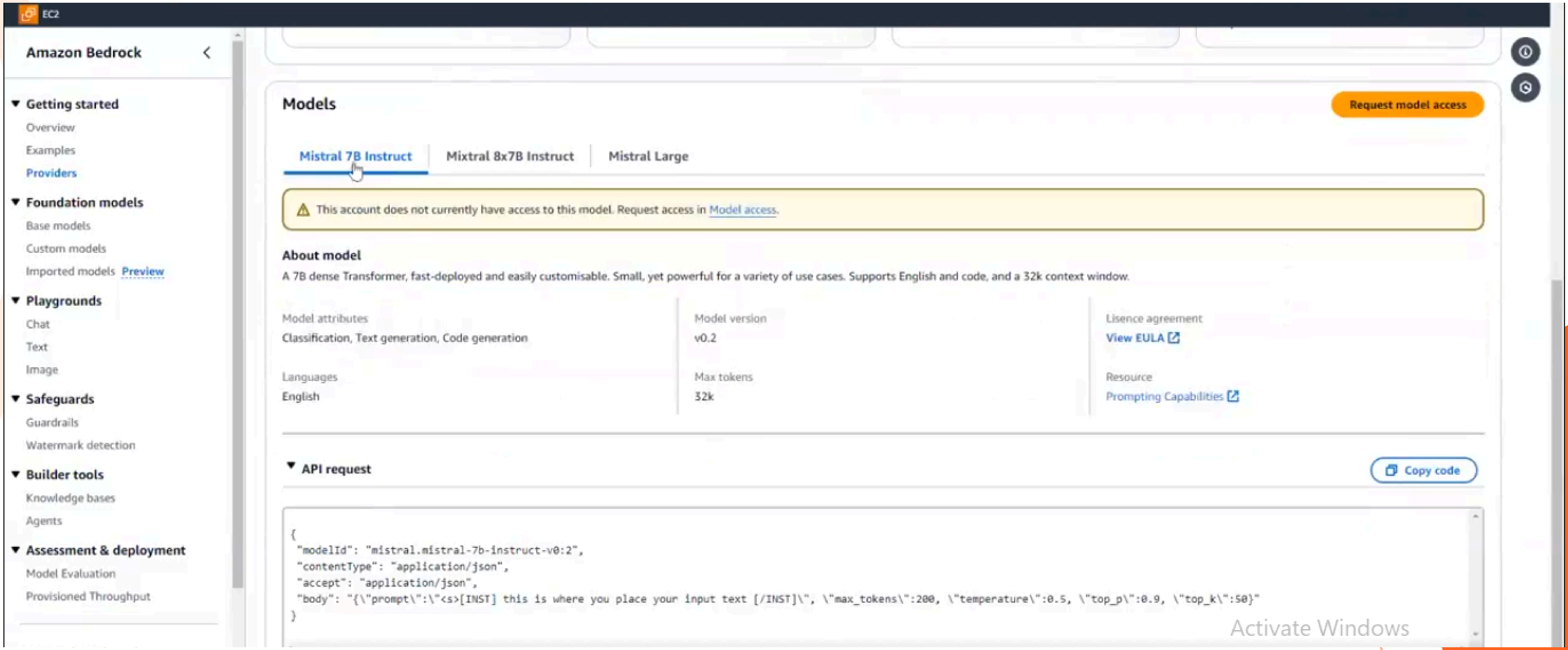
Image

Easily generate compelling images by providing

Use cases example

Amazon Bedrock supports many genAI use cases such as summarization, Q&A, and image generation. Explore the ways FMs can support your use cases.

WORKING WITH AMAZON BEDROCK



Models

This account does not currently have access to this model. Request access in Model access.

About model

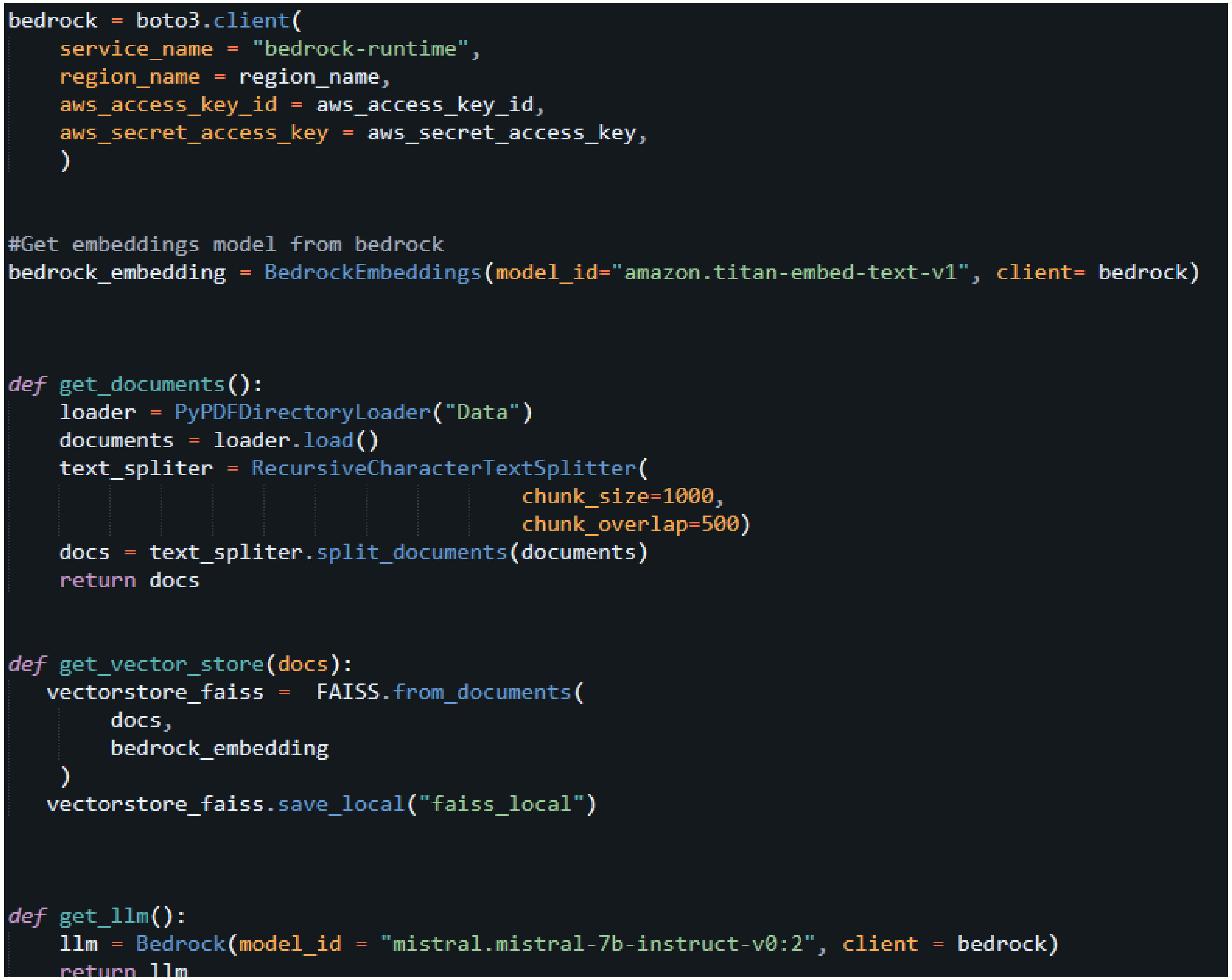
A 7B dense Transformer, fast-deployed and easily customisable. Small, yet powerful for a variety of use cases. Supports English and code, and a 32k context window.

Model attributes Classification, Text generation, Code generation	Model version v0.2	Lisence agreement View EULA
Languages English	Max tokens 32k	Resource Prompting Capabilities

API request

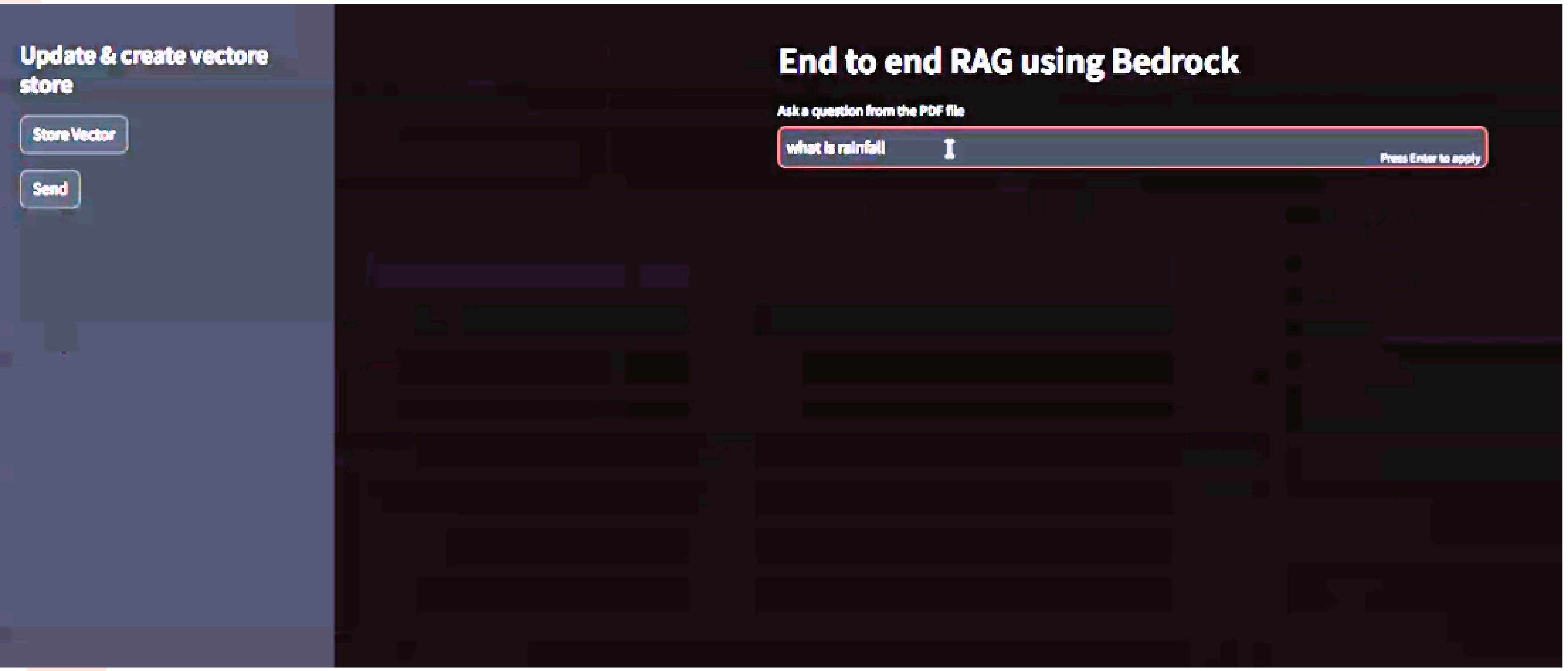
```
{
 "modelId": "mistral.mistral-7b-instruct-v0:2",
 "contentType": "application/json",
 "accept": "application/json",
 "body": "{\"prompt\":\"<s>[INST] this is where you place your input text [/INST]\", \"max_tokens\":200, \"temperature\":0.5, \"top_p\":0.9, \"top_k\":50}"
}
```

SOURCE CODE



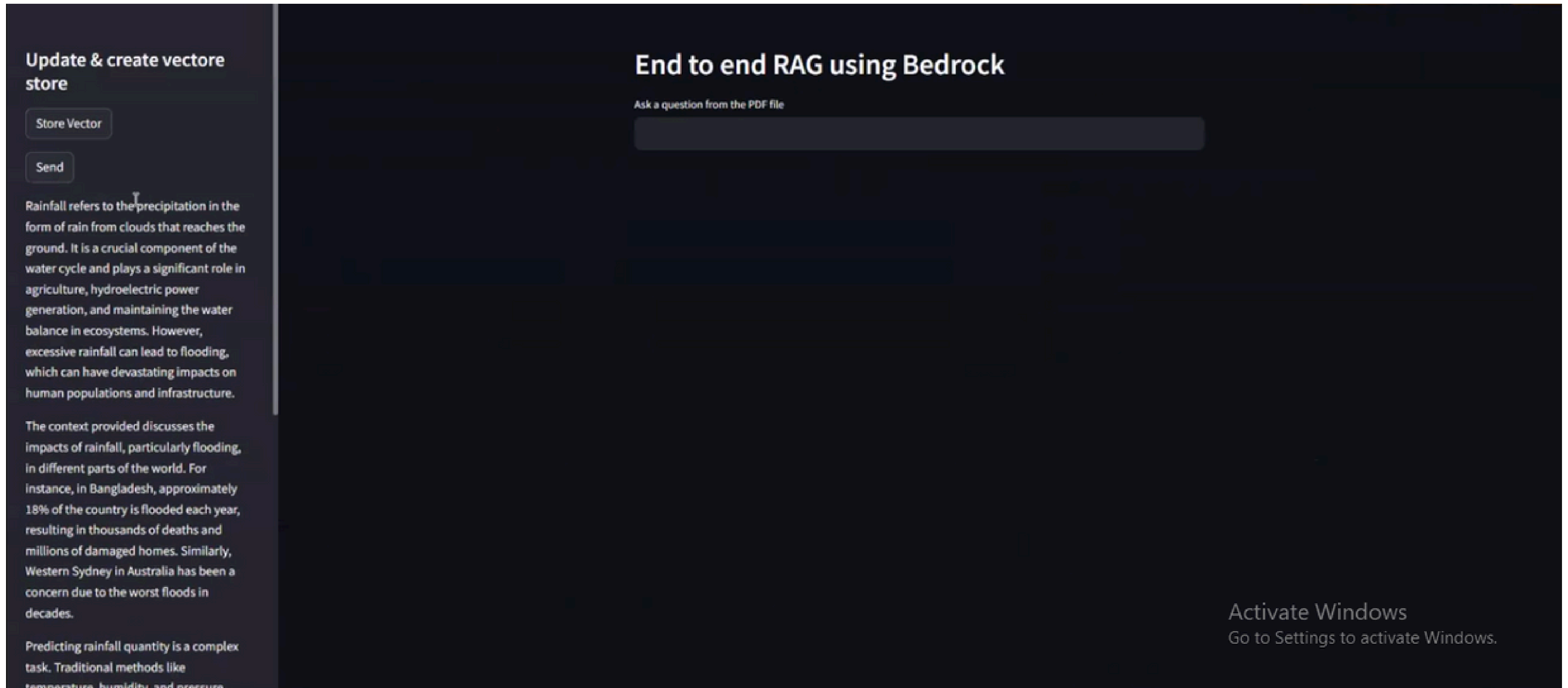
```python
bedrock = boto3.client(
    service_name = "bedrock-runtime",
    region_name = region_name,
    aws_access_key_id = aws_access_key_id,
    aws_secret_access_key = aws_secret_access_key,
    )


#Get embeddings model from bedrock
bedrock_embedding = BedrockEmbeddings(model_id="amazon.titan-embed-text-v1", client= bedrock)



def get_documents():
    loader = PyPDFDirectoryLoader("Data")
    documents = loader.load()
    text_spliter = RecursiveCharacterTextSplitter(
                                        chunk_size=1000,
                                        chunk_overlap=500)
    docs = text_spliter.split_documents(documents)
    return docs


def get_vector_store(docs):
   vectorstore_faiss =  FAISS.from_documents(
        docs,
        bedrock_embedding
    )
   vectorstore_faiss.save_local("faiss_local")



def get_llm():
    llm = Bedrock(model_id = "mistral.mistral-7b-instruct-v0:2", client = bedrock)
    return llm
```

RESULTS


Update & create vectore store
Store Vector
Send
End to end RAG using Bedrock
Ask a question from the PDF file
what is rainfall
Press Enter to apply

i asked my model ..what is rainfall?

this is what my model replied....!


Update & create vectore store
Store Vector
Send
End to end RAG using Bedrock
Ask a question from the PDF file
Activate Windows
Go to Settings to activate Windows.

Rainfall refers to the precipitation in the form of rain from clouds that reaches the ground. It is a crucial component of the water cycle and plays a significant role in agriculture, hydroelectric power generation, and maintaining the water balance in ecosystems. However, excessive rainfall can lead to flooding, which can have devastating impacts on human populations and infrastructure.

The context provided discusses the impacts of rainfall, particularly flooding, in different parts of the world. For instance, in Bangladesh, approximately 18% of the country is flooded each year, resulting in thousands of deaths and millions of damaged homes. Similarly, Western Sydney in Australia has been a concern due to the worst floods in decades.

Predicting rainfall quantity is a complex task. Traditional methods like temperature, humidity, and pressure