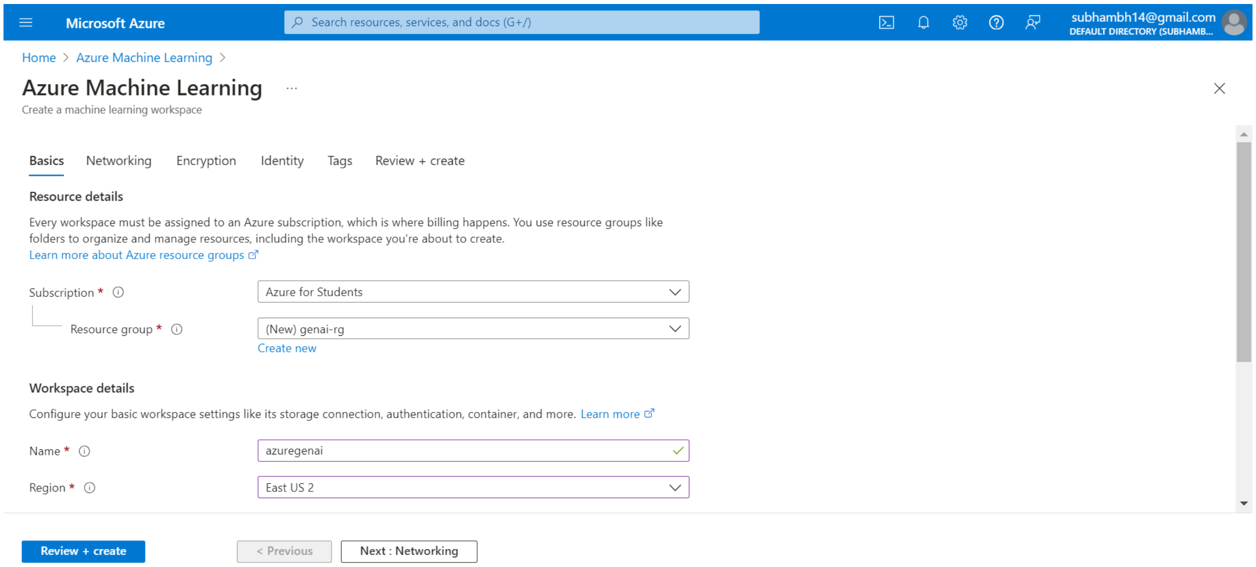
Generative AI-1



Azure Machine Learning

Create a machine learning workspace

Resource details

Every workspace must be assigned to an Azure subscription, which is where billing happens. You use resource groups like folders to organize and manage resources, including the workspace you're about to create.
Learn more about Azure resource groups

Subscription * : Azure for Students

Resource group * : (New) genai-rg

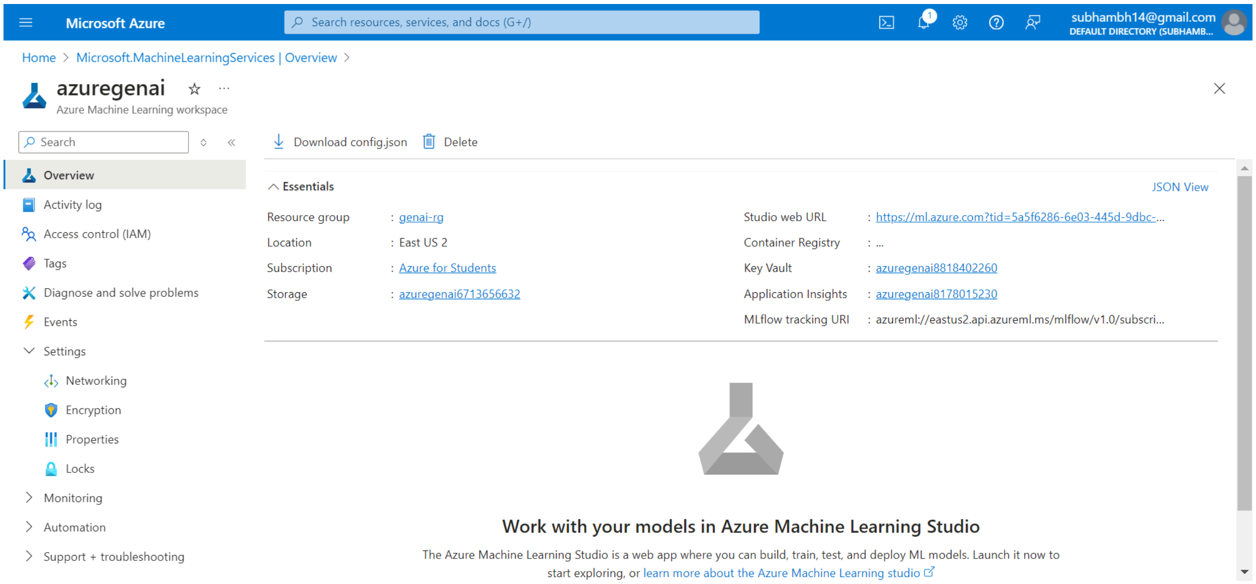
Create new

Workspace details

Configure your basic workspace settings like its storage connection, authentication, container, and more. Learn more

Name * : azuregenai

Region * : East US 2



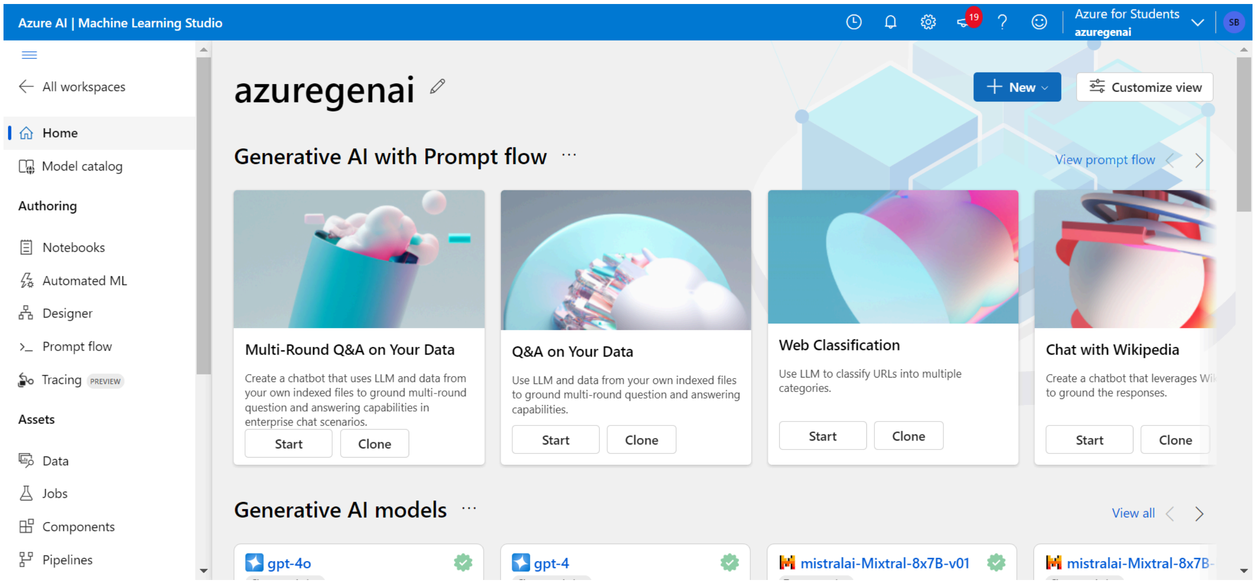
azuregenai

Azure Machine Learning workspace

Resource group	: genai-rg	Studio web URL	: https://ml.azure.com?tid=5a5f6286-6e03-445d-9dbc-...
Location	: East US 2	Container Registry	: ...
Subscription	: Azure for Students	Key Vault	: azuregenai8818402260
Storage	: azuregenai6713656632	Application Insights	: azuregenai8178015230
		MLflow tracking URI	: azureml://eastus2.api.azureml.ms/mlflow/v1.0/subscri...

Work with your models in Azure Machine Learning Studio

The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning studio



Azure AI | Machine Learning Studio

Azure for Students
azuregenai

All workspaces

Home
Model catalog

Authoring

Notebooks
Automated ML
Designer
Prompt flow
Tracing PREVIEW

Assets

Data
Jobs
Components
Pipelines

azuregenai

New | Customize view

Generative AI with Prompt flow

View prompt flow

Multi-Round Q&A on Your Data

Create a chatbot that uses LLM and data from your own indexed files to ground multi-round question and answering capabilities in enterprise chat scenarios.

Start | Clone

Q&A on Your Data

Use LLM and data from your own indexed files to ground multi-round question and answering capabilities.

Start | Clone

Web Classification

Use LLM to classify URLs into multiple categories.

Start | Clone

Chat with Wikipedia

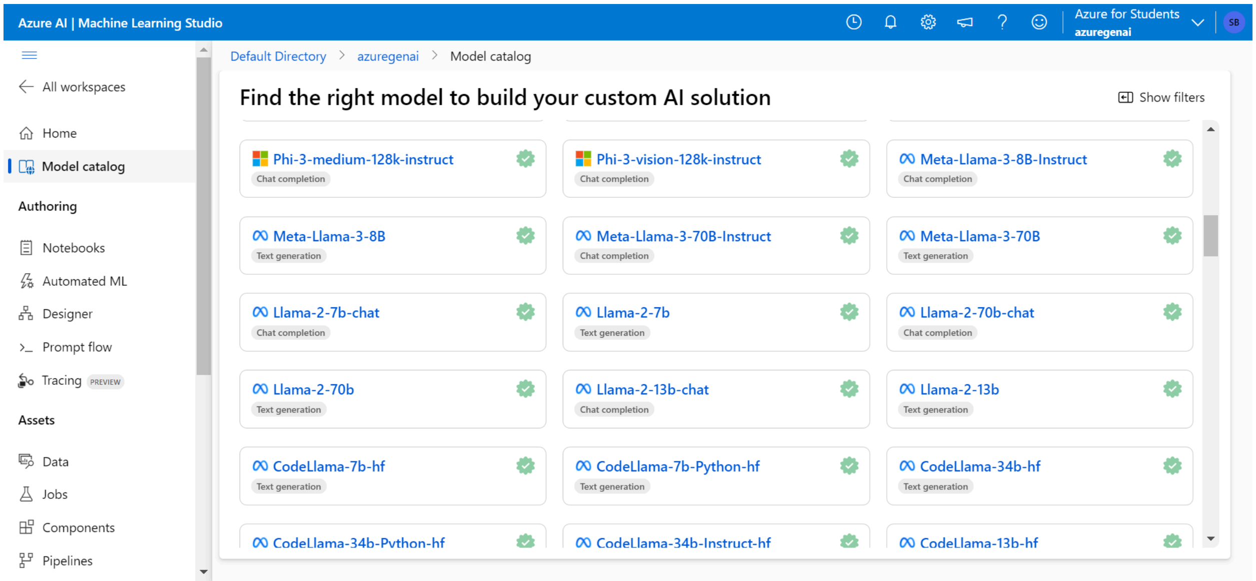
Create a chatbot that leverages Wi to ground the responses.

Start | Clone

Generative AI models

View all

gpt-4o
gpt-4
mistralai-Mixtral-8x7B-v01
mistralai-Mixtral-8x7B-



Azure AI | Machine Learning Studio

Azure for Students
azuregenai

All workspaces

Home
Model catalog

Authoring

Notebooks
Automated ML
Designer
Prompt flow
Tracing PREVIEW

Assets

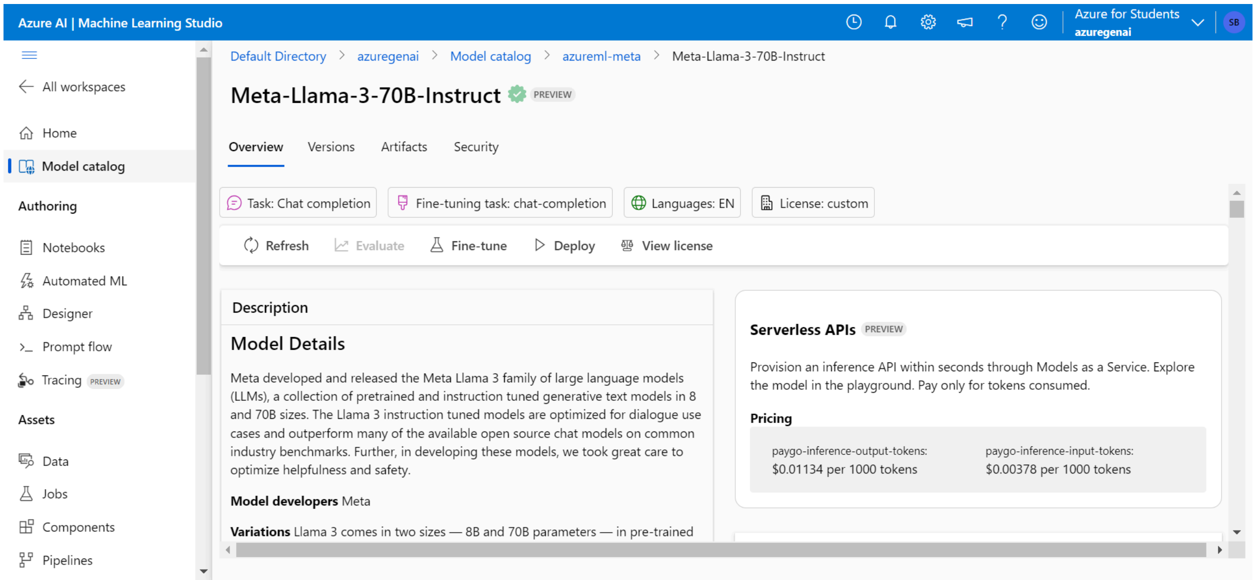
Data
Jobs
Components
Pipelines

Default Directory > azuregenai > Model catalog

Find the right model to build your custom AI solution

Show filters

Phi-3-medium-128k-instruct Chat completion	Phi-3-vision-128k-instruct Chat completion	Meta-Llama-3-8B-Instruct Chat completion
Meta-Llama-3-8B Text generation	Meta-Llama-3-70B-Instruct Chat completion	Meta-Llama-3-70B Text generation
Llama-2-7b-chat Chat completion	Llama-2-7b Text generation	Llama-2-70b-chat Chat completion
Llama-2-70b Text generation	Llama-2-13b-chat Chat completion	Llama-2-13b Text generation
CodeLlama-7b-hf Text generation	CodeLlama-7b-Python-hf Text generation	CodeLlama-34b-hf Text generation
CodeLlama-34b-Python-hf	CodeLlama-34b-Instruct-hf	CodeLlama-13b-hf



Description

Model Details

Meta developed and released the Meta Llama 3 family of large language models (LLMs), a collection of pretrained and instruction tuned generative text models in 8 and 70B sizes. The Llama 3 instruction tuned models are optimized for dialogue use cases and outperform many of the available open source chat models on common industry benchmarks. Further, in developing these models, we took great care to optimize helpfulness and safety.

Model developers Meta

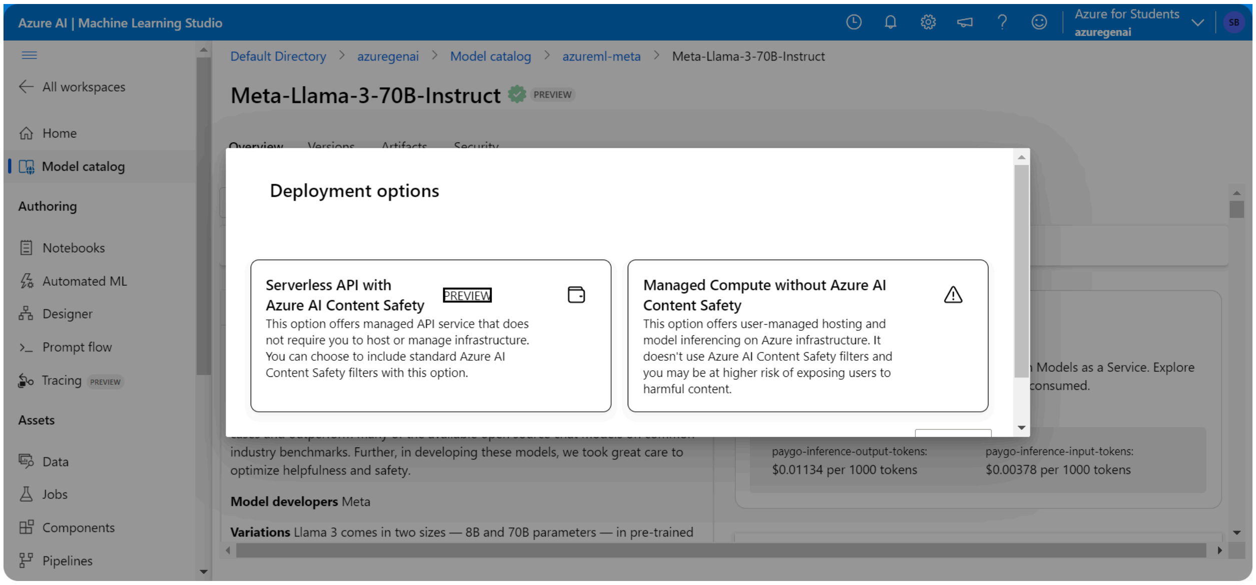
Variations Llama 3 comes in two sizes — 8B and 70B parameters — in pre-trained

Serverless APIs PREVIEW

Provision an inference API within seconds through Models as a Service. Explore the model in the playground. Pay only for tokens consumed.

Pricing

paygo-inference-output-tokens:	paygo-inference-input-tokens:
$0.01134 per 1000 tokens	$0.00378 per 1000 tokens


Azure AI | Machine Learning Studio
Azure for Students
azuregenai
SB
Default Directory > azuregenai > Model catalog > azureml-meta > Meta-Llama-3-70B-Instruct
Meta-Llama-3-70B-Instruct PREVIEW

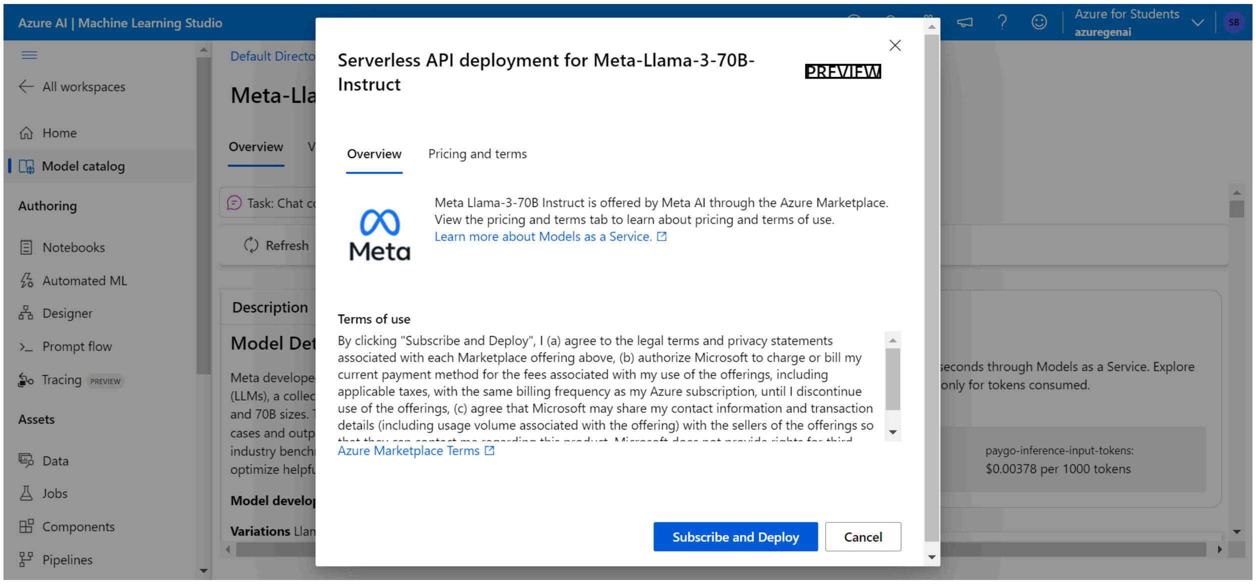
Deployment options

Serverless API with Azure AI Content Safety PREVIEW

This option offers managed API service that does not require you to host or manage infrastructure. You can choose to include standard Azure AI Content Safety filters with this option.

Managed Compute without Azure AI Content Safety

This option offers user-managed hosting and model inferencing on Azure infrastructure. It doesn't use Azure AI Content Safety filters and you may be at higher risk of exposing users to harmful content.



Serverless API deployment for Meta-Llama-3-70B-Instruct PREVIEW

Overview | Pricing and terms

Meta Llama-3-70B Instruct is offered by Meta AI through the Azure Marketplace. View the pricing and terms tab to learn about pricing and terms of use. Learn more about Models as a Service.

Terms of use

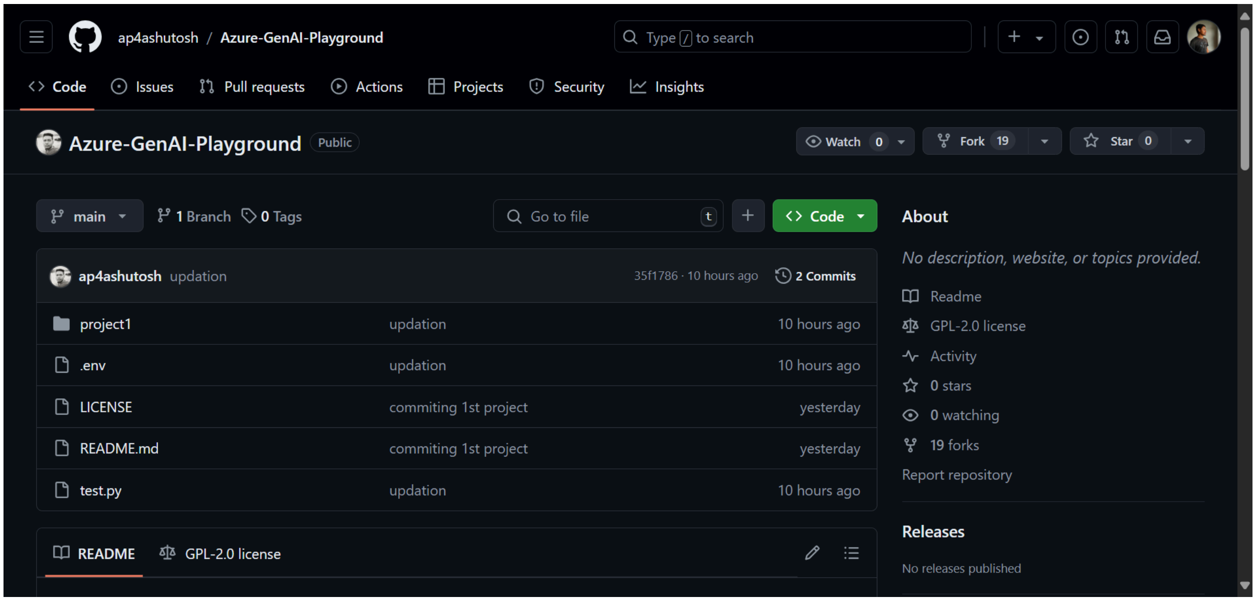
By clicking "Subscribe and Deploy", I (a) agree to the legal terms and privacy statements associated with each Marketplace offering above, (b) authorize Microsoft to charge or bill my current payment method for the fees associated with my use of the offerings, including applicable taxes, with the same billing frequency as my Azure subscription, until I discontinue use of the offerings, (c) agree that Microsoft may share my contact information and transaction details (including usage volume associated with the offering) with the sellers of the offerings so

Azure Marketplace Terms

Meta develope (LLMs), a collec and 70B sizes. cases and outp industry bench optimize helpfu

seconds through Models as a Service. Explore only for tokens consumed.

paygo-inference-input-tokens: $0.00378 per 1000 tokens



Azure-GenAI-Playground Public

ap4ashutosh updation 35f1786 · 10 hours ago 2 Commits

project1	updation	10 hours ago
.env	updation	10 hours ago
LICENSE	commiting 1st project	yesterday
README.md	commiting 1st project	yesterday
test.py	updation	10 hours ago

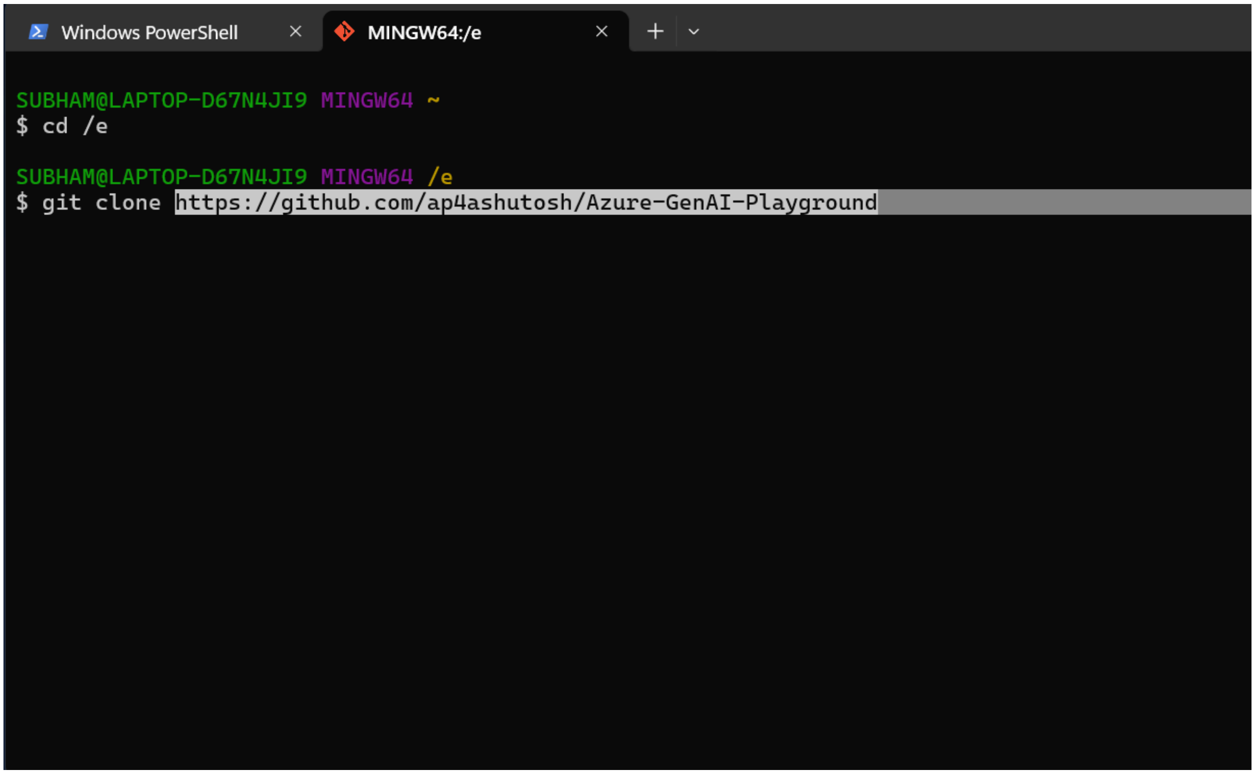
About

No description, website, or topics provided.

Readme
GPL-2.0 license
Activity
0 stars
0 watching
19 forks
Report repository

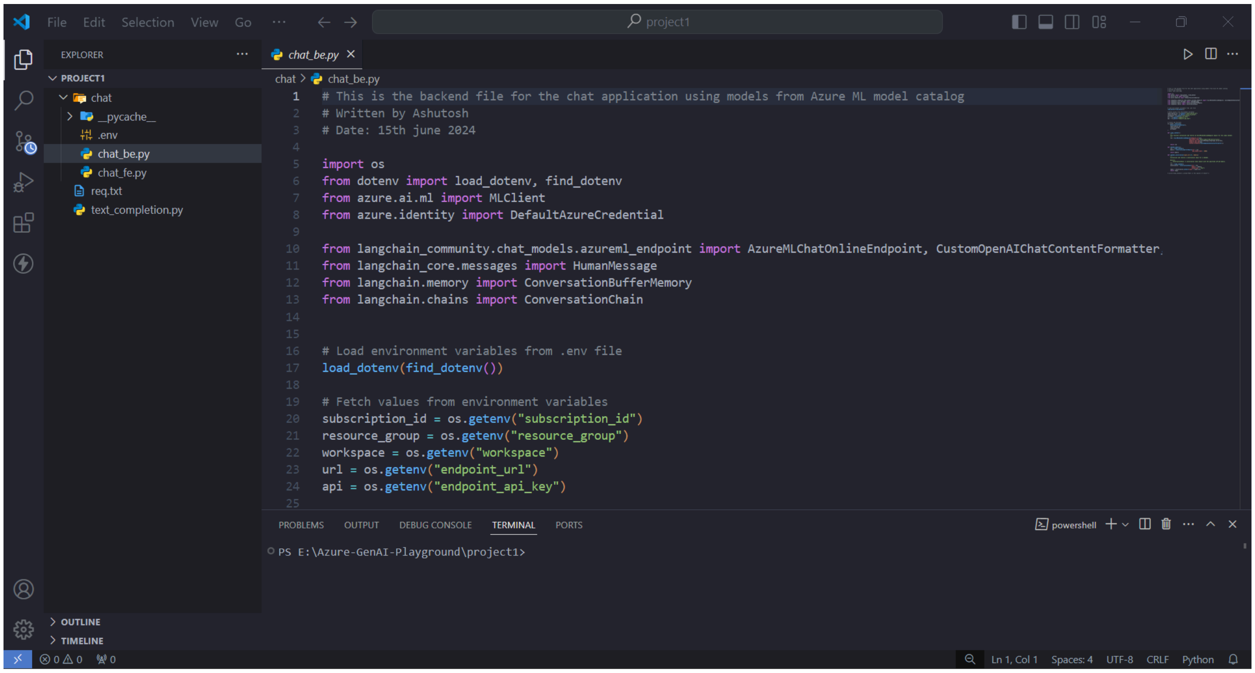
Releases

No releases published



```
SUBHAM@LAPTOP-D67N4JI9 MINGW64 ~
$ cd /e

SUBHAM@LAPTOP-D67N4JI9 MINGW64 /e
$ git clone https://github.com/ap4ashutosh/Azure-GenAI-Playground
```



```python
# This is the backend file for the chat application using models from Azure ML model catalog
# Written by Ashutosh
# Date: 15th june 2024

import os
from dotenv import load_dotenv, find_dotenv
from azure.ai.ml import MLClient
from azure.identity import DefaultAzureCredential

from langchain_community.chat_models.azureml_endpoint import AzureMLChatOnlineEndpoint, CustomOpenAIChatContentFormatter,
from langchain_core.messages import HumanMessage
from langchain.memory import ConversationBufferMemory
from langchain.chains import ConversationChain


# Load environment variables from .env file
load_dotenv(find_dotenv())

# Fetch values from environment variables
subscription_id = os.getenv("subscription_id")
resource_group = os.getenv("resource_group")
workspace = os.getenv("workspace")
url = os.getenv("endpoint_url")
api = os.getenv("endpoint_api_key")

```

```
PS E:\Azure-GenAI-Playground\project1>
```