Lets Get Building

So you know the basics around OpenAI but where to from here?

https://www.linkedin.com/in/shanebaldacchino/

Talk is cheap.
Show me the code.

Linus Torvalds

quotefancy

Your Journey

Demystifying Machine Learning: Exploring the History and Future of AI with OpenAI

May 24th

How creative should your solution be or is there only one possible response your code should generate? In this session we will introduce concepts such as temperature, tokens, fine tuning, custom models along with the basics of the Open AI SDK's before opening the IDE and building together a simple python based application that pulls everything we have discussed together in a fun little application.

Prompt Engineering & Embeddings : Unleashing the Full Potential of GPT and LLM's

May 31st

To unlock the true potential of GPT, we will talk through the key principles of prompt engineering, and how to effectively craft prompts that yield the desired outputs.

We will explore the various approaches to prompt engineering, including the use of natural language instructions, pre- and post-processing techniques, and fine-tuning methods.

June 7th

Let's Get Building : So, you know the basics around Open AI, but where to from here?

June 14th

This session!

Azure OpenAI | Top 4 Capabilities & Use Cases

Content Generation

- Call Center Analytics: Automatically generate responses to customer inquiries
- Generate personalised UI for your website

Summarization

- Call Center Analytics: Summary of customer support conversation logs
- Subject Matter Expert Document Summarization (e.g. financial reporting, analyst articles)
- Social Media Trends Summarization

Code generation

- Convert Natural Language to SQL (or vice versa) for telemetry data
- Convert Natural Language to Query Proprietary Data Models
- Code Documentation

Semantic Search

- Search reviews for a specific product / service
- Information Discovery and Knowledge Mining

Examples of Multiple Model Use Cases

End to End Call Center Analytics: Classification, Sentiment, Entity Extraction, Summarization and Email Generation

Customer 360: Hyper-personalisation using timely Summarization of customer queries & trends, Search, and Content Generation

Business Process Automation: Search through structured & unstructured documentation, Generate Code to query data models, Content Generation

What We Will Build / Stand Up Today

- ☑ Your Environment
- Private Chat Bot – Prompt Chaining
- Q&A – Semantic Search On Your Data
- Classification and Sentiment Analysis

Let's Build

Your Guide To Driving OpenAI

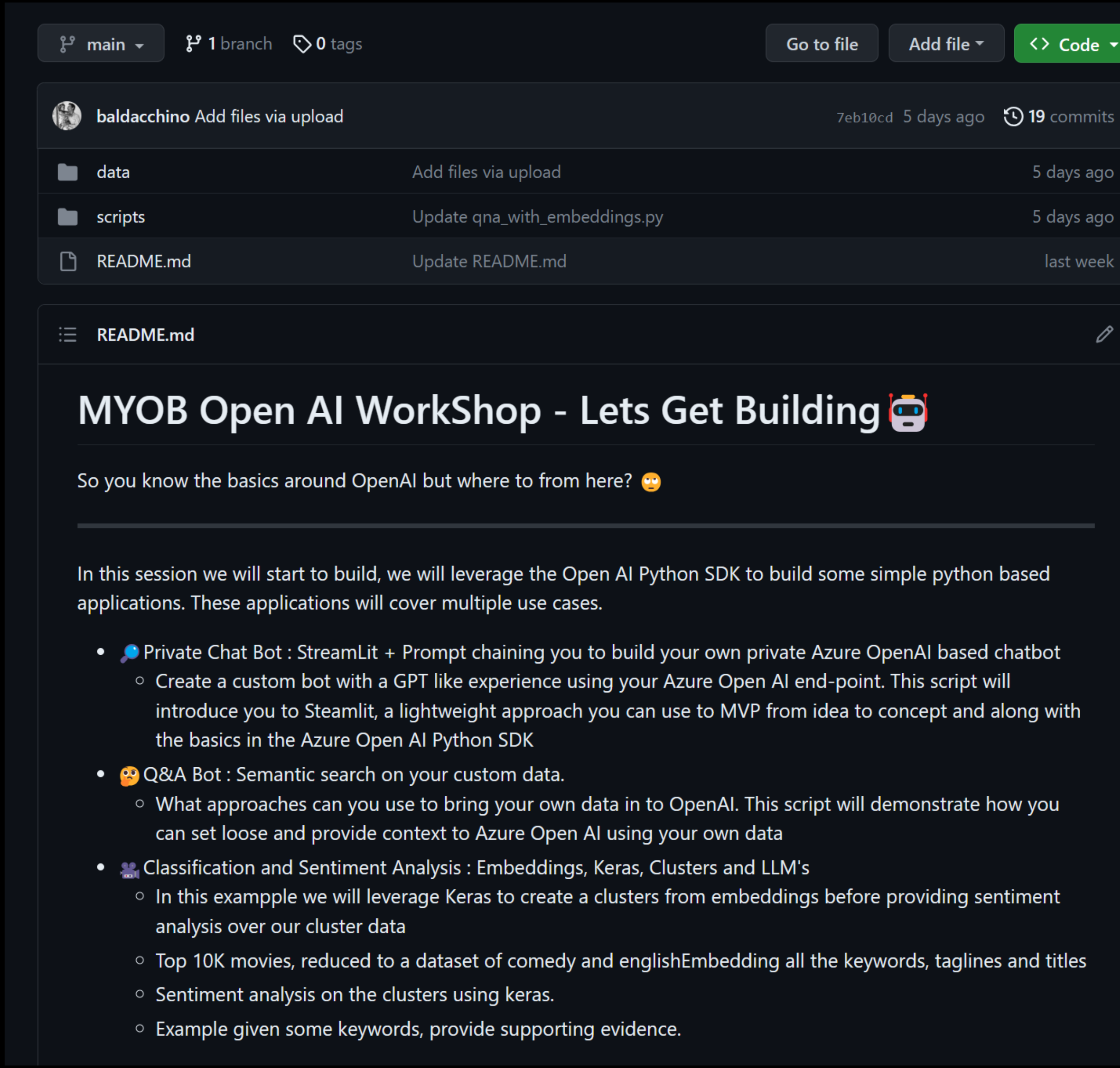
Download Repo

HTTP

https://github.com/baldacchino/AzureOpenAI-LetsBuild/

GIT

git clone https://github.com/baldacchino/AzureOpenAI-LetsBuild.git



baldacchino Add files via upload 7eb10cd 5 days ago 19 commits

data	Add files via upload	5 days ago
scripts	Update qna_with_embeddings.py	5 days ago
README.md	Update README.md	last week

README.md

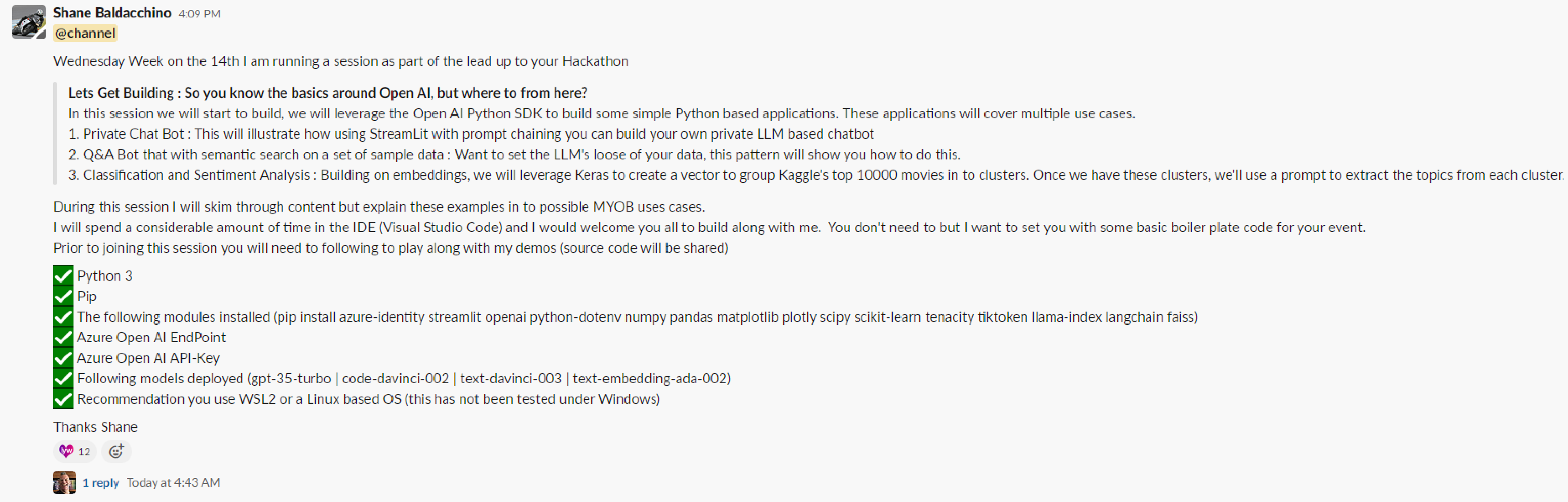
MYOB Open AI WorkShop - Lets Get Building 🤖

So you know the basics around OpenAI but where to from here? 🙄

In this session we will start to build, we will leverage the Open AI Python SDK to build some simple python based applications. These applications will cover multiple use cases.

- 🔎 Private Chat Bot : StreamLit + Prompt chaining you to build your own private Azure OpenAI based chatbot
  - Create a custom bot with a GPT like experience using your Azure Open AI end-point. This script will introduce you to Steamlit, a lightweight approach you can use to MVP from idea to concept and along with the basics in the Azure Open AI Python SDK
- 🤔 Q&A Bot : Semantic search on your custom data.
  - What approaches can you use to bring your own data in to OpenAI. This script will demonstrate how you can set loose and provide context to Azure Open AI using your own data
- 🎥 Classification and Sentiment Analysis : Embeddings, Keras, Clusters and LLM's
  - In this exampple we will leverage Keras to create a clusters from embeddings before providing sentiment analysis over our cluster data
  - Top 10K movies, reduced to a dataset of comedy and englishEmbedding all the keywords, taglines and titles
  - Sentiment analysis on the clusters using keras.
  - Example given some keywords, provide supporting evidence.

☑ Let's set up your environment



Shane Baldacchino 4:09 PM

@channel

Wednesday Week on the 14th I am running a session as part of the lead up to your Hackathon

> **Lets Get Building : So you know the basics around Open AI, but where to from here?**
> In this session we will start to build, we will leverage the Open AI Python SDK to build some simple Python based applications. These applications will cover multiple use cases.
> 1. Private Chat Bot : This will illustrate how using StreamLit with prompt chaining you can build your own private LLM based chatbot
> 2. Q&A Bot that with semantic search on a set of sample data : Want to set the LLM's loose of your data, this pattern will show you how to do this.
> 3. Classification and Sentiment Analysis : Building on embeddings, we will leverage Keras to create a vector to group Kaggle's top 10000 movies in to clusters. Once we have these clusters, we'll use a prompt to extract the topics from each cluster.

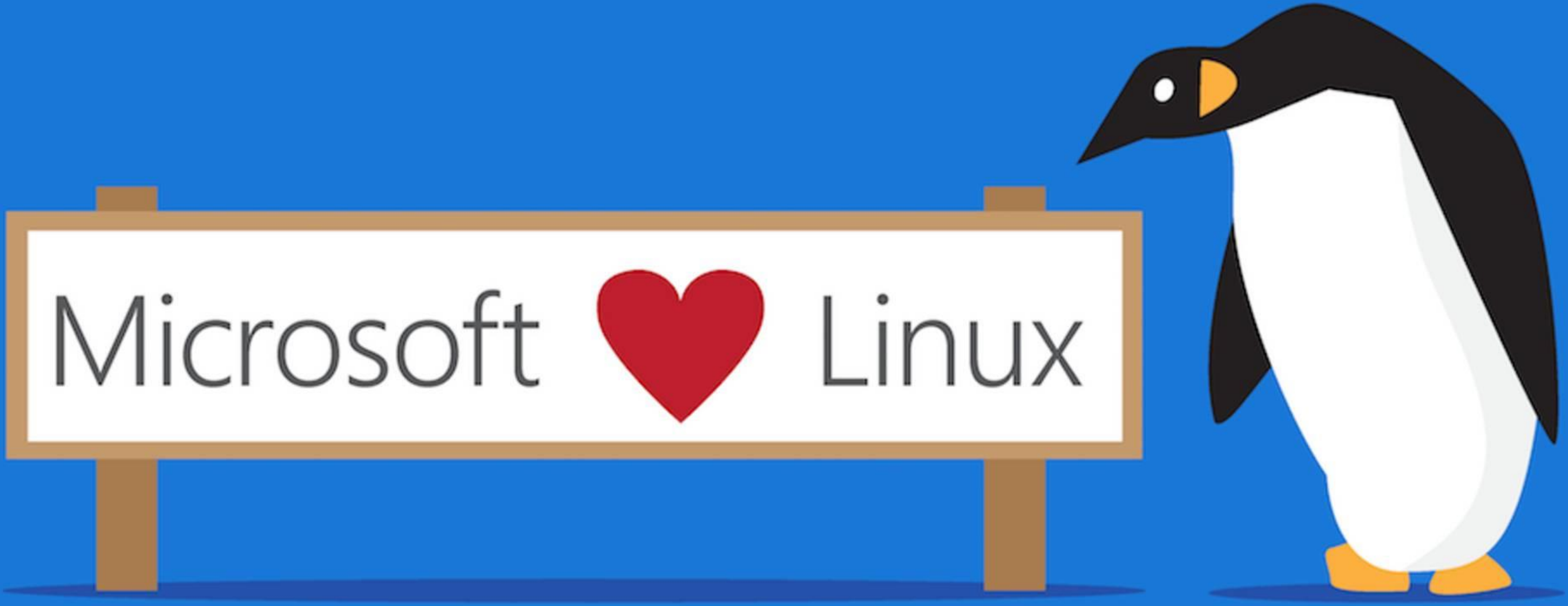
During this session I will skim through content but explain these examples in to possible MYOB uses cases.
I will spend a considerable amount of time in the IDE (Visual Studio Code) and I would welcome you all to build along with me. You don't need to but I want to set you with some basic boiler plate code for your event.
Prior to joining this session you will need to following to play along with my demos (source code will be shared)

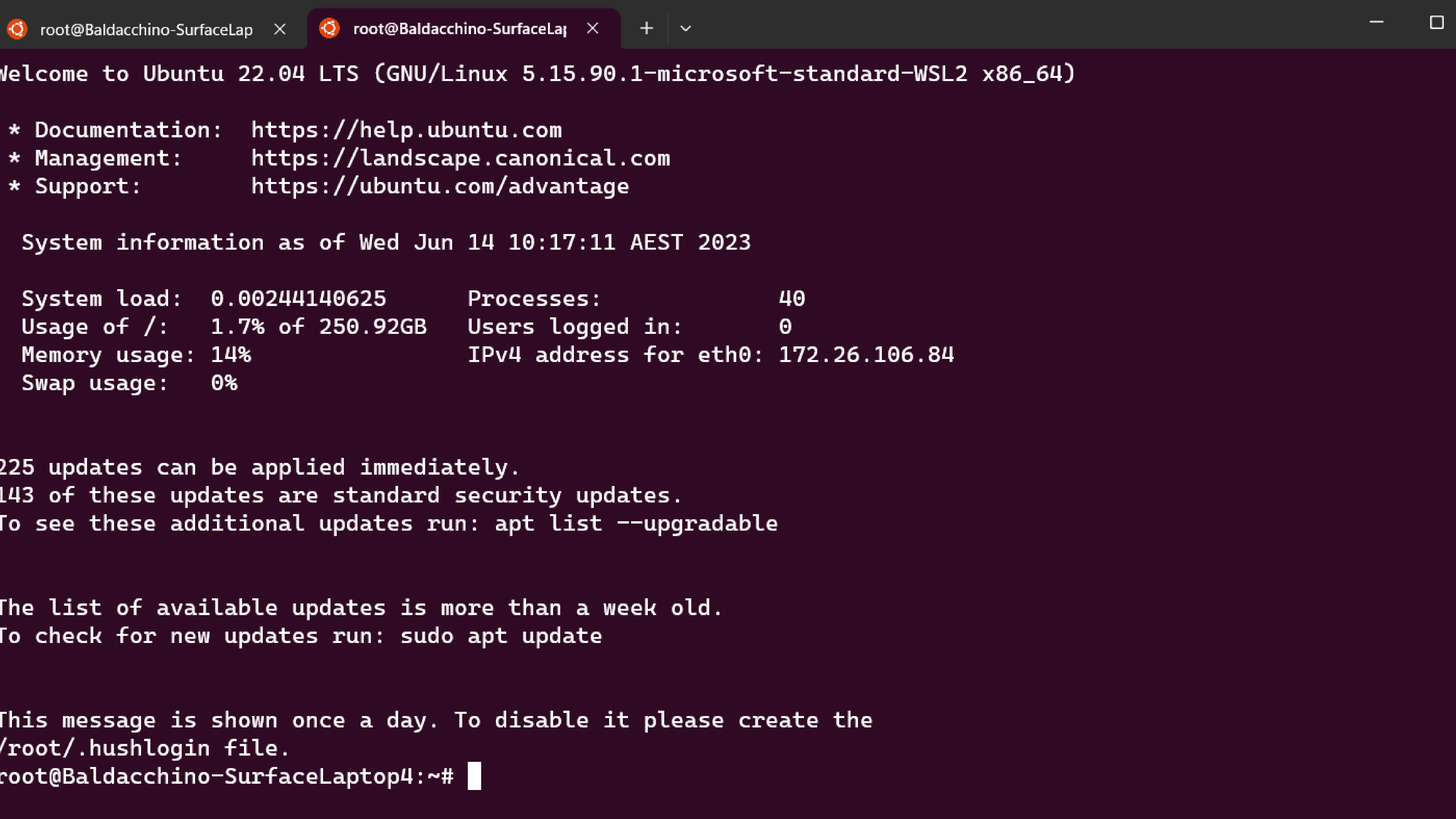
✅ Python 3
✅ Pip
✅ The following modules installed (pip install azure-identity streamlit openai python-dotenv numpy pandas matplotlib plotly scipy scikit-learn tenacity tiktoken llama-index langchain faiss)
✅ Azure Open AI EndPoint
✅ Azure Open AI API-Key
✅ Following models deployed (gpt-35-turbo | code-davinci-002 | text-davinci-003 | text-embedding-ada-002)
✅ Recommendation you use WSL2 or a Linux based OS (this has not been tested under Windows)

Thanks Shane

1 reply Today at 4:43 AM

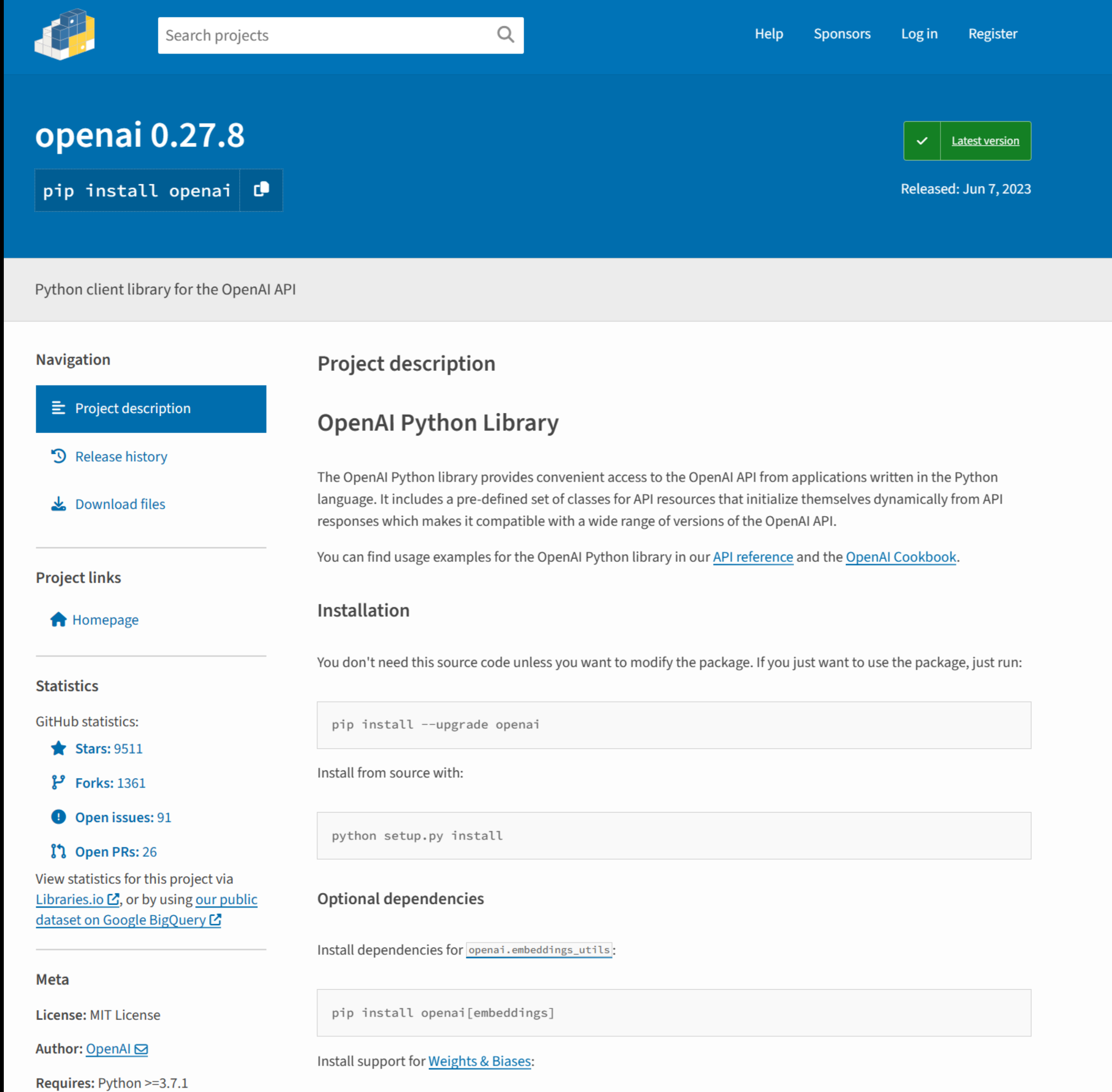
Microsoft





```
Welcome to Ubuntu 22.04 LTS (GNU/Linux 5.15.90.1-microsoft-standard-WSL2 x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  System information as of Wed Jun 14 10:17:11 AEST 2023

  System load:  0.00244140625      Processes:             40
  Usage of /:   1.7% of 250.92GB   Users logged in:       0
  Memory usage: 14%                IPv4 address for eth0: 172.26.106.84
  Swap usage:   0%


225 updates can be applied immediately.
143 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable


The list of available updates is more than a week old.
To check for new updates run: sudo apt update


This message is shown once a day. To disable it please create the
/root/.hushlogin file.
root@Baldacchino-SurfaceLaptop4:~# 
```

Python Modules



openai 0.27.8

```
pip install openai
```

Released: Jun 7, 2023

Python client library for the OpenAI API

Statistics

GitHub statistics:

- **Stars:** 9511
- **Forks:** 1361
- **Open issues:** 91
- **Open PRs:** 26

View statistics for this project via Libraries.io, or by using our public dataset on Google BigQuery

Meta

License: MIT License

Author: OpenAI

Requires: Python >=3.7.1

Project description

OpenAI Python Library

The OpenAI Python library provides convenient access to the OpenAI API from applications written in the Python language. It includes a pre-defined set of classes for API resources that initialize themselves dynamically from API responses which makes it compatible with a wide range of versions of the OpenAI API.

You can find usage examples for the OpenAI Python library in our API reference and the OpenAI Cookbook.

Installation

You don't need this source code unless you want to modify the package. If you just want to use the package, just run:

```
pip install --upgrade openai
```

Install from source with:

```
python setup.py install
```

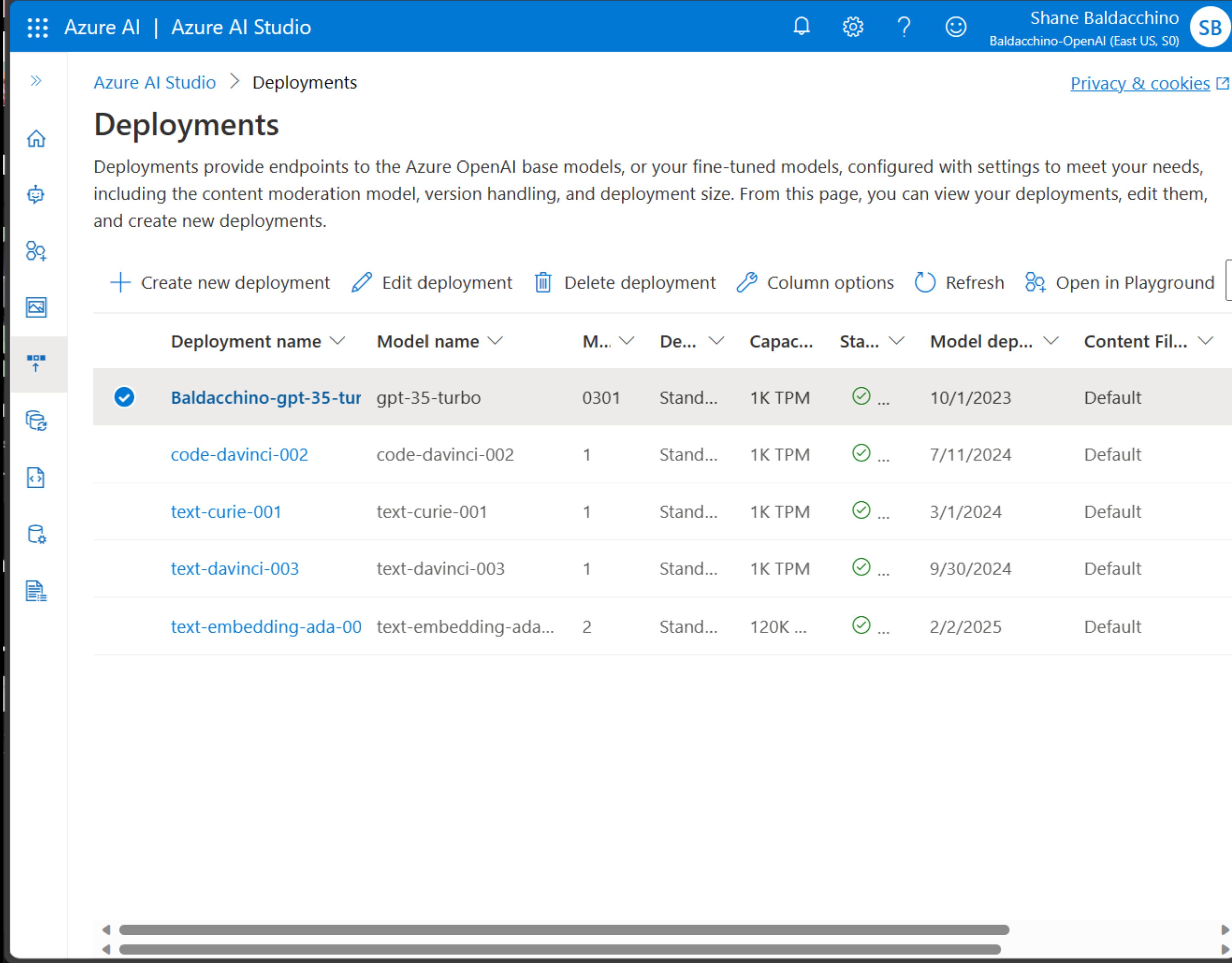
Optional dependencies

Install dependencies for `openai.embeddings_utils`:

```
pip install openai[embeddings]
```

Install support for Weights & Biases:

Models

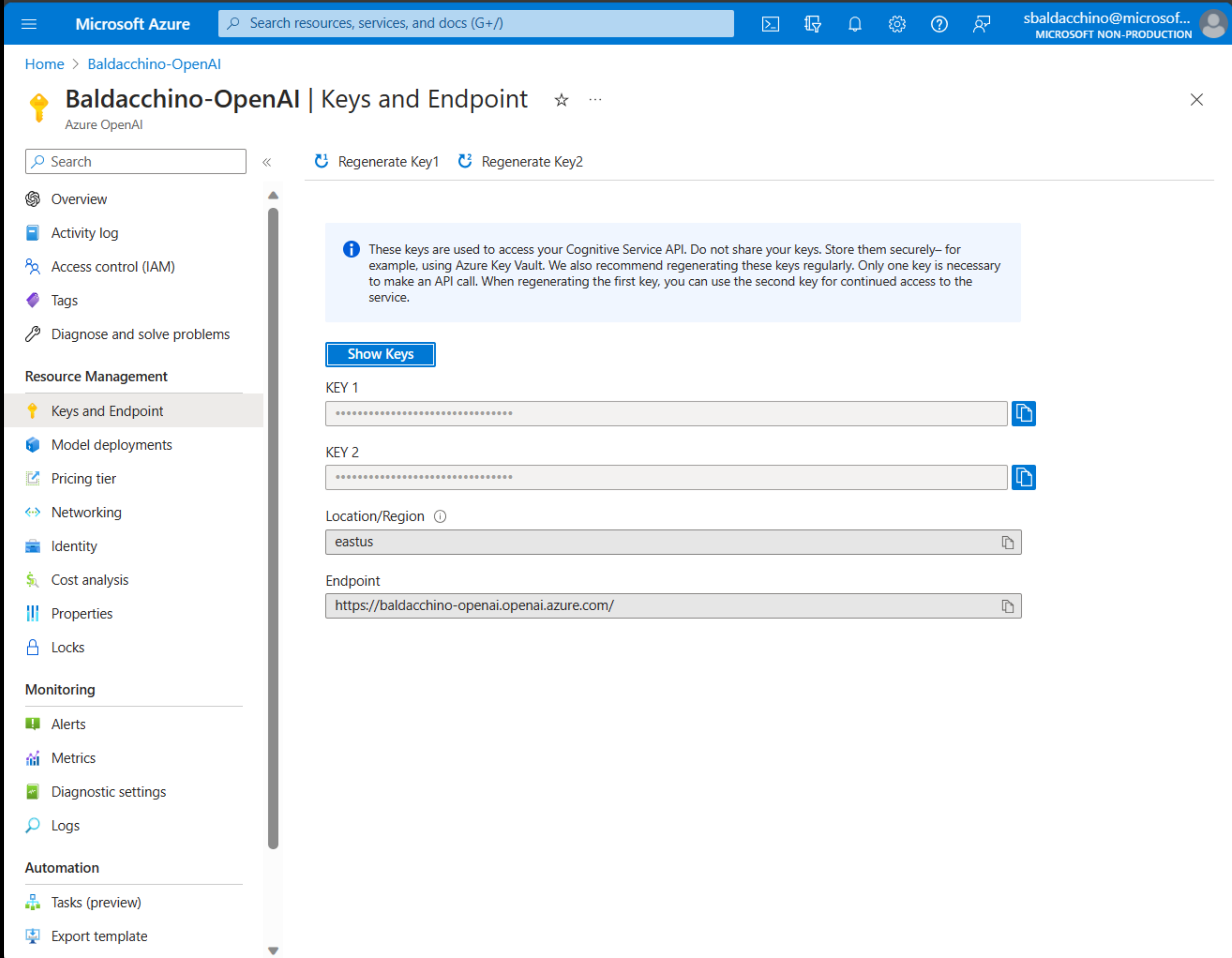


Deployments

Deployments provide endpoints to the Azure OpenAI base models, or your fine-tuned models, configured with settings to meet your needs, including the content moderation model, version handling, and deployment size. From this page, you can view your deployments, edit them, and create new deployments.

	Deployment name	Model name	M...	De...	Capac...	Sta...	Model dep...	Content Fil...
✓	Baldacchino-gpt-35-tur	gpt-35-turbo	0301	Stand...	1K TPM	✓ ...	10/1/2023	Default
	code-davinci-002	code-davinci-002	1	Stand...	1K TPM	✓ ...	7/11/2024	Default
	text-curie-001	text-curie-001	1	Stand...	1K TPM	✓ ...	3/1/2024	Default
	text-davinci-003	text-davinci-003	1	Stand...	1K TPM	✓ ...	9/30/2024	Default
	text-embedding-ada-00	text-embedding-ada...	2	Stand...	120K ...	✓ ...	2/2/2025	Default

Access Keys



Baldacchino-OpenAI | Keys and Endpoint

Azure OpenAI

These keys are used to access your Cognitive Service API. Do not share your keys. Store them securely– for example, using Azure Key Vault. We also recommend regenerating these keys regularly. Only one key is necessary to make an API call. When regenerating the first key, you can use the second key for continued access to the service.

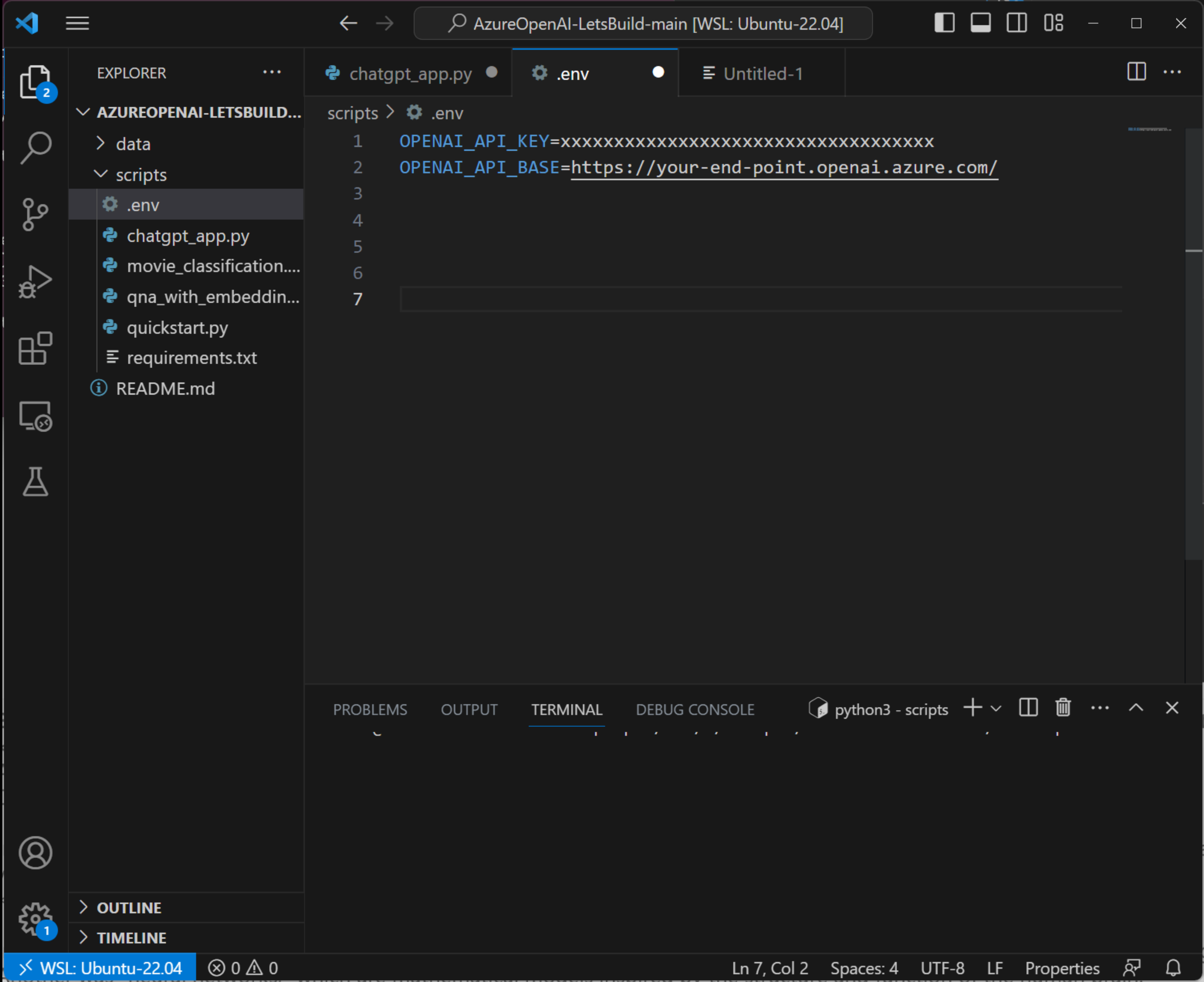
KEY 1: ••••••••••••••••••••••••••••••••

KEY 2: ••••••••••••••••••••••••••••••••

Location/Region: eastus

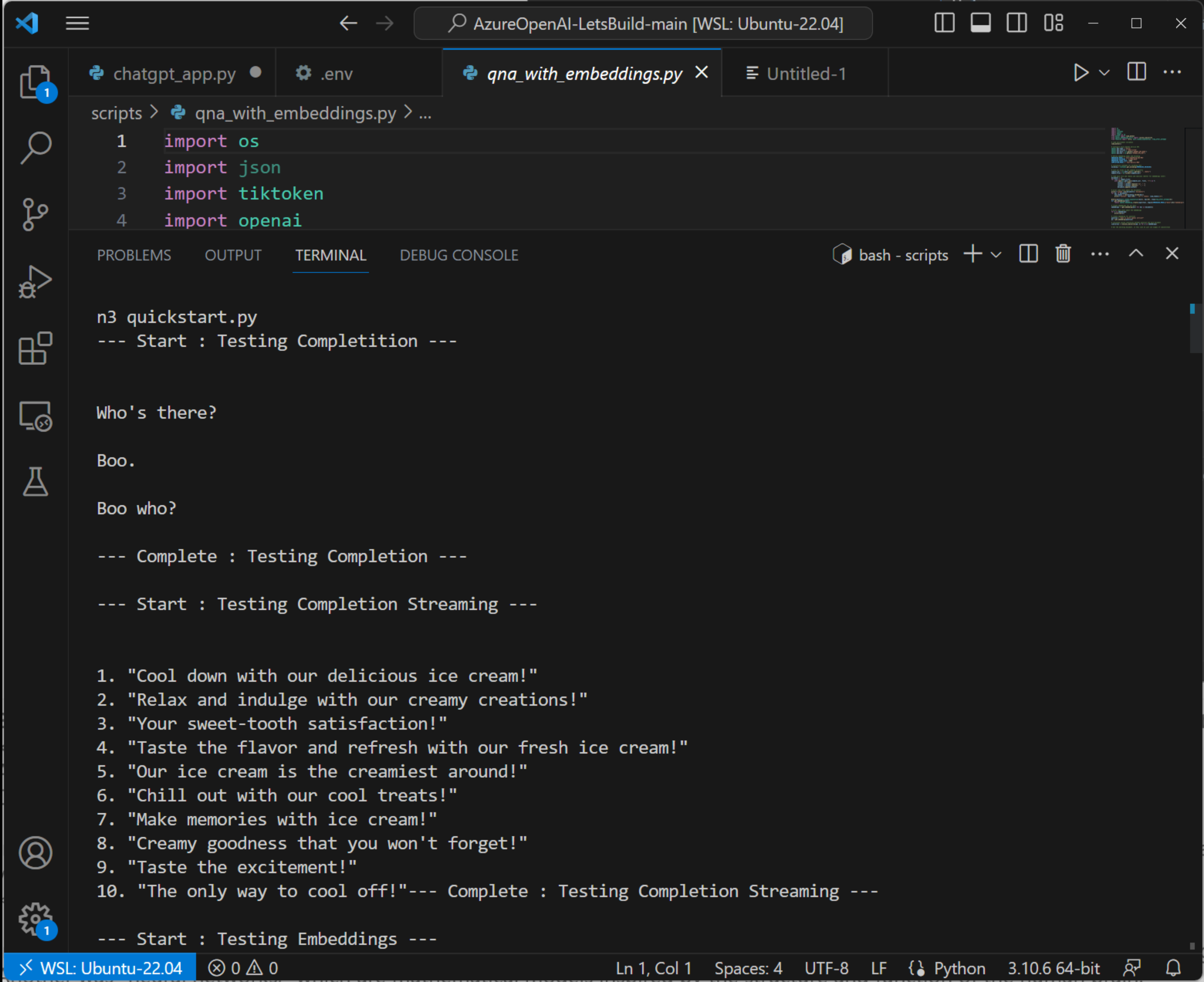
Endpoint: https://baldacchino-openai.openai.azure.com/

Environment File



```
OPENAI_API_KEY=xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
OPENAI_API_BASE=https://your-end-point.openai.azure.com/
```

Validate Your Environment

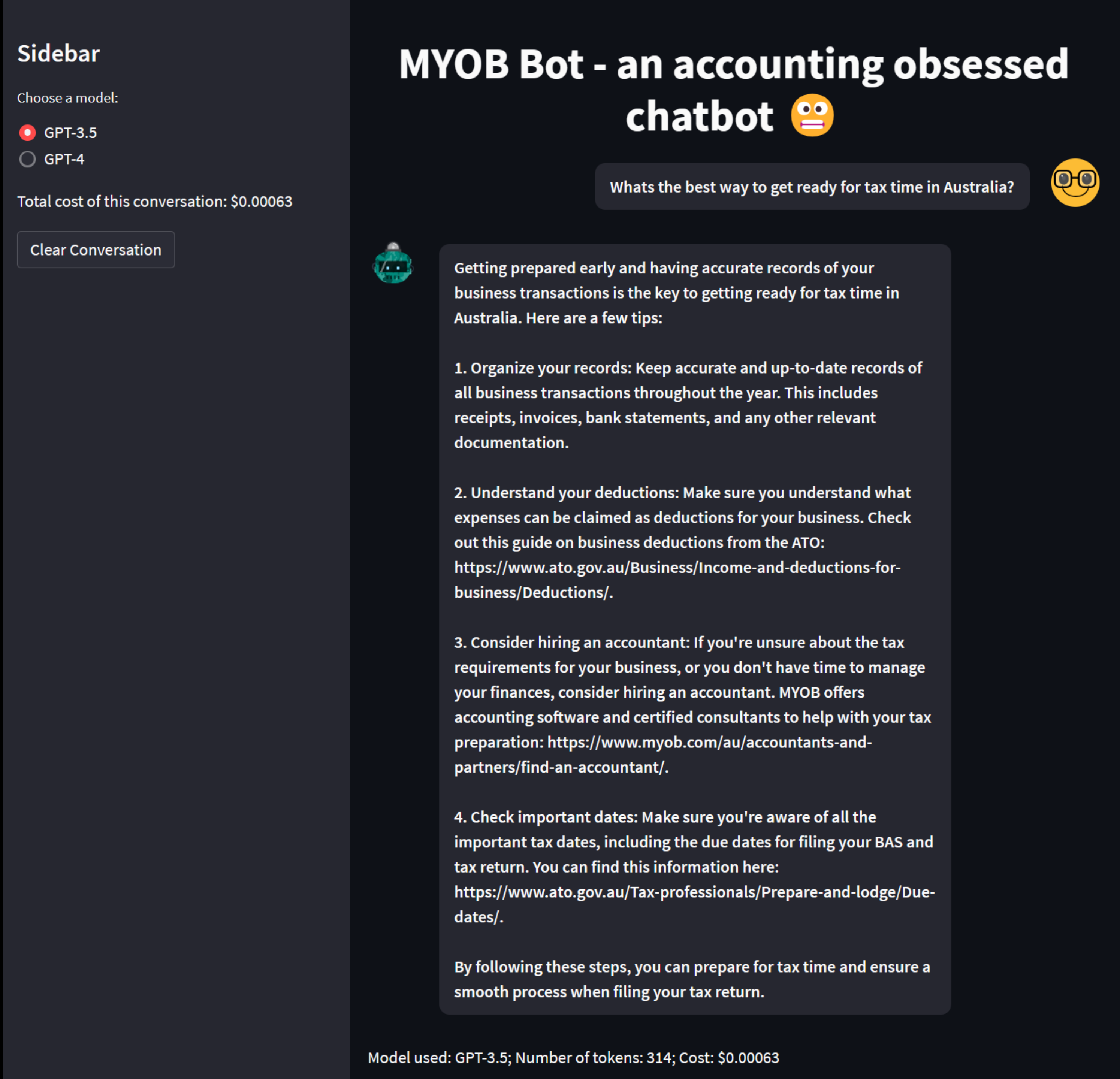


```python
import os
import json
import tiktoken
import openai
```

```
n3 quickstart.py
--- Start : Testing Completition ---


Who's there?

Boo.

Boo who?

--- Complete : Testing Completion ---

--- Start : Testing Completion Streaming ---


1. "Cool down with our delicious ice cream!"
2. "Relax and indulge with our creamy creations!"
3. "Your sweet-tooth satisfaction!"
4. "Taste the flavor and refresh with our fresh ice cream!"
5. "Our ice cream is the creamiest around!"
6. "Chill out with our cool treats!"
7. "Make memories with ice cream!"
8. "Creamy goodness that you won't forget!"
9. "Taste the excitement!"
10. "The only way to cool off!"--- Complete : Testing Completion Streaming ---

--- Start : Testing Embeddings ---
```

Private ChatGPT

- ☑ Prompt Chaining
- ☑ Streamlit
- ☑ Tiktoken
- ☑ System Message



Sidebar

Choose a model:

● GPT-3.5
○ GPT-4

Total cost of this conversation: $0.00063

MYOB Bot - an accounting obsessed chatbot 😬

Whats the best way to get ready for tax time in Australia?

Getting prepared early and having accurate records of your business transactions is the key to getting ready for tax time in Australia. Here are a few tips:

1. Organize your records: Keep accurate and up-to-date records of all business transactions throughout the year. This includes receipts, invoices, bank statements, and any other relevant documentation.

2. Understand your deductions: Make sure you understand what expenses can be claimed as deductions for your business. Check out this guide on business deductions from the ATO: https://www.ato.gov.au/Business/Income-and-deductions-for-business/Deductions/.

3. Consider hiring an accountant: If you're unsure about the tax requirements for your business, or you don't have time to manage your finances, consider hiring an accountant. MYOB offers accounting software and certified consultants to help with your tax preparation: https://www.myob.com/au/accountants-and-partners/find-an-accountant/.

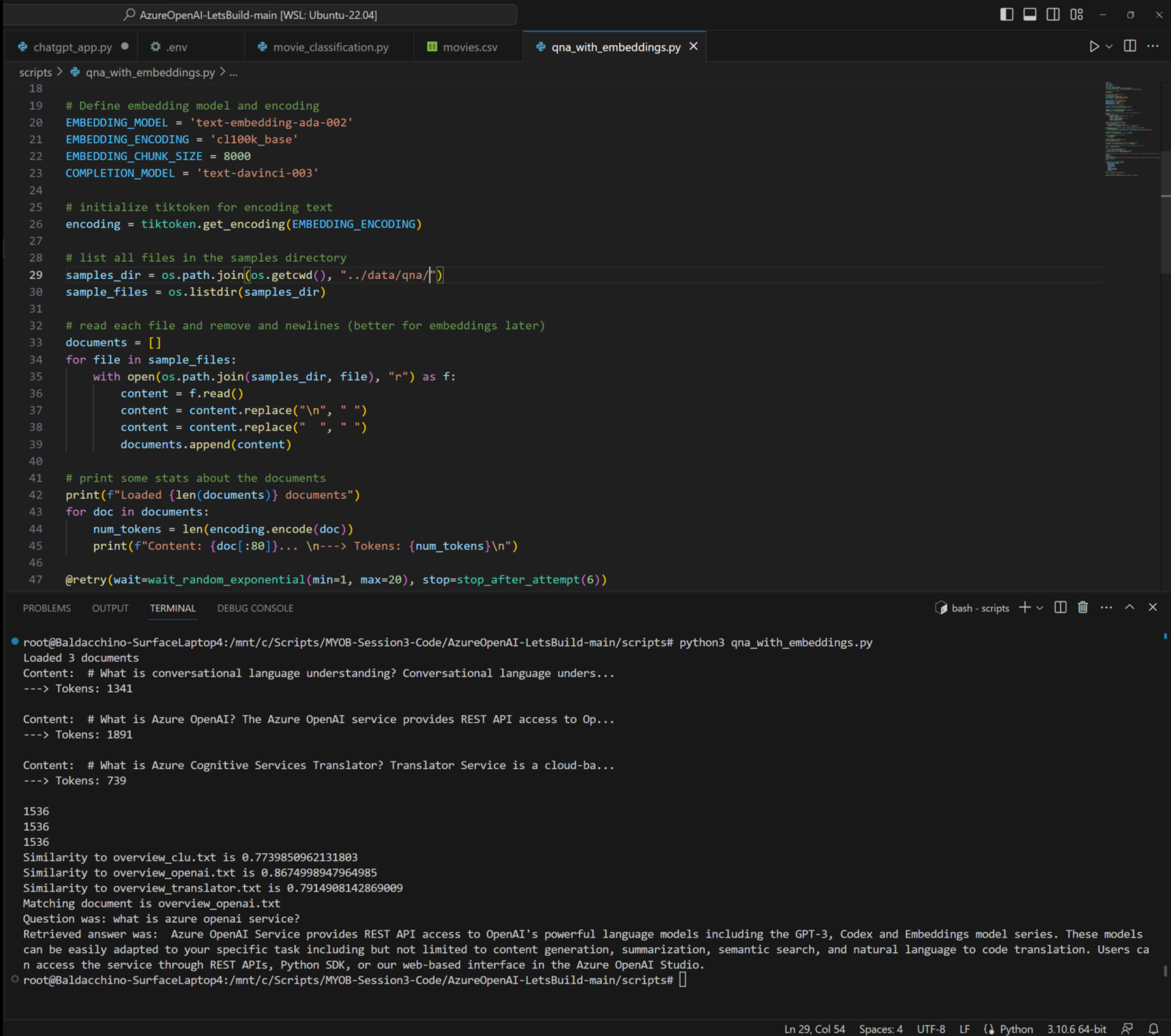
4. Check important dates: Make sure you're aware of all the important tax dates, including the due dates for filing your BAS and tax return. You can find this information here: https://www.ato.gov.au/Tax-professionals/Prepare-and-lodge/Due-dates/.

By following these steps, you can prepare for tax time and ensure a smooth process when filing your tax return.

Model used: GPT-3.5; Number of tokens: 314; Cost: $0.00063

Q&A Using Your Data

- ☑ Embedding
- ☑ Your Local Data

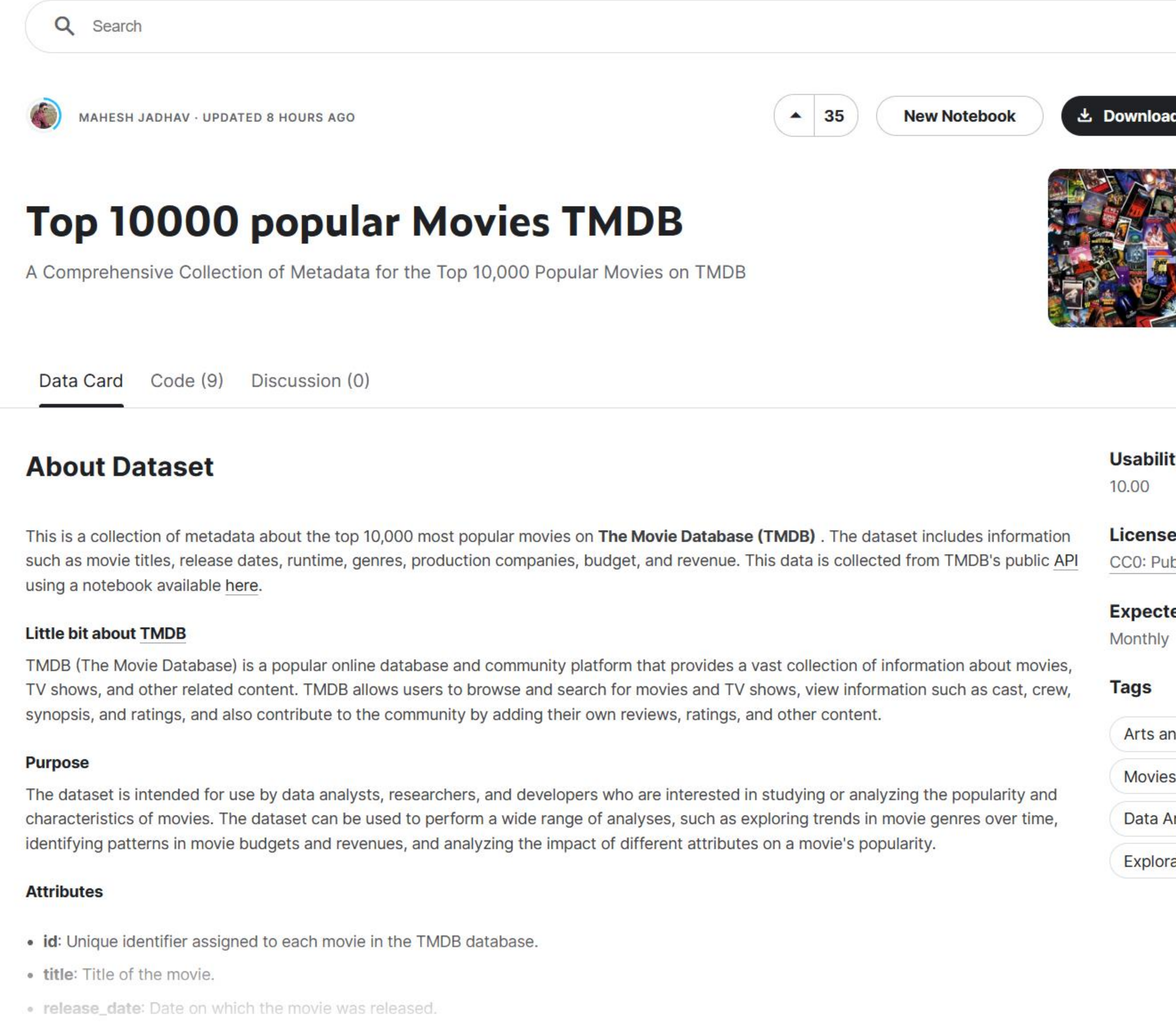


```python

# Define embedding model and encoding
EMBEDDING_MODEL = 'text-embedding-ada-002'
EMBEDDING_ENCODING = 'cl100k_base'
EMBEDDING_CHUNK_SIZE = 8000
COMPLETION_MODEL = 'text-davinci-003'

# initialize tiktoken for encoding text
encoding = tiktoken.get_encoding(EMBEDDING_ENCODING)

# list all files in the samples directory
samples_dir = os.path.join(os.getcwd(), "../data/qna/")
sample_files = os.listdir(samples_dir)

# read each file and remove and newlines (better for embeddings later)
documents = []
for file in sample_files:
    with open(os.path.join(samples_dir, file), "r") as f:
        content = f.read()
        content = content.replace("\n", " ")
        content = content.replace("  ", " ")
        documents.append(content)

# print some stats about the documents
print(f"Loaded {len(documents)} documents")
for doc in documents:
    num_tokens = len(encoding.encode(doc))
    print(f"Content: {doc[:80]}... \n---> Tokens: {num_tokens}\n")

@retry(wait=wait_random_exponential(min=1, max=20), stop=stop_after_attempt(6))
```

```
root@Baldacchino-SurfaceLaptop4:/mnt/c/Scripts/MYOB-Session3-Code/AzureOpenAI-LetsBuild-main/scripts# python3 qna_with_embeddings.py
Loaded 3 documents
Content:  # What is conversational language understanding? Conversational language unders...
---> Tokens: 1341

Content:  # What is Azure OpenAI? The Azure OpenAI service provides REST API access to Op...
---> Tokens: 1891

Content:  # What is Azure Cognitive Services Translator? Translator Service is a cloud-ba...
---> Tokens: 739

1536
1536
1536
Similarity to overview_clu.txt is 0.7739850962131803
Similarity to overview_openai.txt is 0.8674998947964985
Similarity to overview_translator.txt is 0.7914908142869009
Matching document is overview_openai.txt
Question was: what is azure openai service?
Retrieved answer was:  Azure OpenAI Service provides REST API access to OpenAI's powerful language models including the GPT-3, Codex and Embeddings model series. These models
can be easily adapted to your specific task including but not limited to content generation, summarization, semantic search, and natural language to code translation. Users ca
n access the service through REST APIs, Python SDK, or our web-based interface in the Azure OpenAI Studio.
root@Baldacchino-SurfaceLaptop4:/mnt/c/Scripts/MYOB-Session3-Code/AzureOpenAI-LetsBuild-main/scripts# 
```

Classification & Sentiment Analysis

- ☑ Keras
- ☑ Clustering
- ☑ Your Data
- ☑ Embedding
- ☑ Needle In Haystack



Top 10000 popular Movies TMDB

A Comprehensive Collection of Metadata for the Top 10,000 Popular Movies on TMDB

About Dataset

This is a collection of metadata about the top 10,000 most popular movies on **The Movie Database (TMDB)** . The dataset includes information such as movie titles, release dates, runtime, genres, production companies, budget, and revenue. This data is collected from TMDB's public API using a notebook available here.

Little bit about TMDB

TMDB (The Movie Database) is a popular online database and community platform that provides a vast collection of information about movies, TV shows, and other related content. TMDB allows users to browse and search for movies and TV shows, view information such as cast, crew, synopsis, and ratings, and also contribute to the community by adding their own reviews, ratings, and other content.

Purpose

The dataset is intended for use by data analysts, researchers, and developers who are interested in studying or analyzing the popularity and characteristics of movies. The dataset can be used to perform a wide range of analyses, such as exploring trends in movie genres over time, identifying patterns in movie budgets and revenues, and analyzing the impact of different attributes on a movie's popularity.

Attributes

- **id**: Unique identifier assigned to each movie in the TMDB database.
- **title**: Title of the movie.
- **release_date**: Date on which the movie was released.

Usabilit
10.00

License
CC0: Pub

Expecte
Monthly

Tags

- Arts an
- Movies
- Data A
- Explora

We want to hear from you!

- **Thank you for attending this session as part of our multi-part learning sessions towards MYOB's AI Hackathon**

Please take the time to provide feedback!

What worked, what didn't – help us improve

Link: https://forms.office.com/r/HXenDu8F4m

Let's Be Friends…..

LinkedIn : https://www.linkedin.com/in/shanebaldacchino/
Email : sbaldacchino@microsoft.com
Web : https://automation.baldacchino.net

SHANE BALDACCHINO | CHIEF ARCHITECT MICROSOFT ANZ

@Shane Baldacchino
#sig-azure