PharmaSUG 2024 - Paper SD-429

Build up Your Own ChatGPT Environment with Azure OpenAI Platform

Bill Zhang and Jun Yang, ClinChoice Inc.

ABSTRACT

ChatGPT from OpenAI is a state-of-the-art chatbot technology that uses natural language processing to understand and respond to user queries. With ChatGPT, you can do your daily work more efficiently and effectively. This article provides a step-by-step guide on how to configure and extend a customized, secure ChatGPT environment with Azure OpenAI platform for your organization.

The article starts with an introduction to the topic and explains why it is important to configure and deploy a customized ChatGPT environment for pharmaceutical industry. It then provides some architecture information on internal ChatGPT environment with Azure OpenAI platform & Azure AI search services, explaining what they are, how they work together. It then describes the methodology used to configure and deploy it, explaining the steps involved in the process and providing details on how to implement each step, including OpenAI model deployment, Azure AI search configuration, internal business document ingestion as ChatGPT extension, and web server integration with OpenAI API etc.

The article also gives simple user guides and skills while utilizing AI features like general chat and completion, SAS/R/Python programming copilot and internal document query. At last, it also provides some recommendations for future research and development.

INTRODUCTION

In November 2022, OpenAI (openai.com) released ChatGPT which is an AI chatbot that quickly captured the interest and imagination of people across the world. ChatGPT has super-fast adoption among difference industries, including pharmaceutical industry. According to many technology surveys, companies are increasingly investing more on ChatGPT and related AI technologies and products since the lunch of ChatGPT.

ChatGPT, as a most popular AI product, is based on huge AI neural networks - Large Language Models (LLMs) that generate human-like text based on the user input it receives. People can utilize ChatGPT for all kinds of creative activities, including writing articles and communication emails, answering general questions, providing explanations, querying documents, and writing software code like R/SAS/Python.

ChatGPT is with new generation of AI technology which can be adopted for our daily life and daily work. This AI technology is much closer to us than before. Many companies are planning to leverage ChatGPT to improve their business.

There is one major concern - data security and privacy when ChatGPT is utilized. OpenAI could use user's questions and associated answers for future AI model training. This causes security concerns while ChatGPT is used for daily work. Thanks to OpenAI platform of Microsoft Azure cloud, the trained ChatGPT model can be deployed into Azure cloud environment managed by a company. So, all the pre-trained data, user input and output can be kept inside the boundary of company Azure cloud.

ChatGPT can be combined with other AI features, including Azure AI search, to ingest more internal documents into company dedicated ChatGPT environment. The company users can do more semantic search against company internal documents with nature languages.

The article will specify the setup and configuration steps to assist you to build up your own ChatGPT environment for your business usage.

ARCHTECTURE OF YOUR OWN CHATGPT ENEVIRONMENT

To implement your own ChatGPT environment inside Azure Cloud, the following components are utilized to support customized ChatGPT features.

- Azure Cloud Storage
- Azure AI Search Service
- Azure OpenAI Platform
- Azure Web App Service or others
- Microsoft Entra ID (formerly Azure Active Directory)

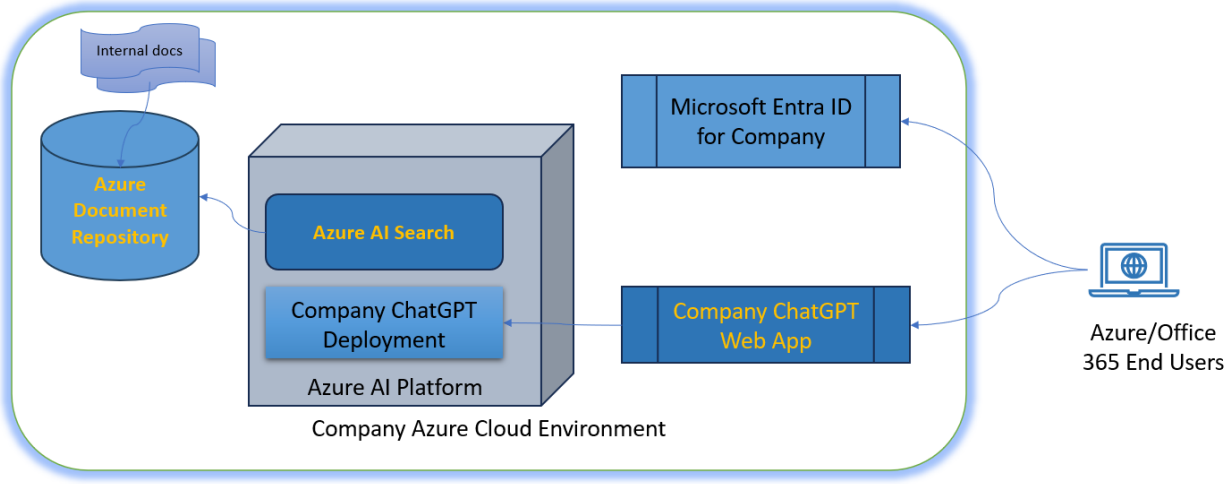


Figure 1. ChatGPT Deployment Diagram

CHATGPT SETUP & DEPLOYMENT IN AZURE

This article specifies the manual setup instructions by using Azure web page. You can use Microsoft Azure script to do these as well. The setup and deployment steps are:

- Prepare your Azure account,
- Create Azure storage account and document repository,
- Create Azure AI Search sevice,
- Create and deploy your OpenAI platform,
- Create your web application to connect to your OpenAI platform.
- Secure your ChatGPT web application and OpenAI platform.

The setup steps are using PharmaSUG ChatGPT configuration with PharmaSUG paper library as an example.

PREPARE YOUR AZURE ACCOUNT

You need the following to start your OpenAI/ChatGPT setup.

- Azure account
- Azure subscription with access enabled for the Azure OpenAI service.

- Azure account permissions
    - Your Azure account must have *Microsoft.Authorization/roleAssignments/write* permissions,
    - Your Azure account also needs *Microsoft.Resources/deployments/write permissions* on the subscription level.

CREATE AZURE STORAGE ACCOUNT AND DOCUMENT REPOSITORY

The document repository is used for internal document storage. Once you have any change in your document repository, you can trigger Azure AI search service to reindex (ingest) the document repository so that you can query new document content via your OpenAI platform.

The document repository configuration steps are:

- Go to Azure portal (https://portal.azure.com) and navigate to “Storage accounts” page.
- Click “create” to create a dedicated storage account (*pharmasug2024*) by following the creation wizard.
    - Make sure you enable “Enable storage account key access” and “Enable hierarchical namespace” options.
    - It is better to define the storage location as same as other services like Azure OpenAI.
- Go to the web page for storage account *pharmasug2024*, create a Blob storage container *pharmasug-doc*. Under the container *pharmasug-doc*, create subfolders like *2022*,*2023*,*2024* for later use.

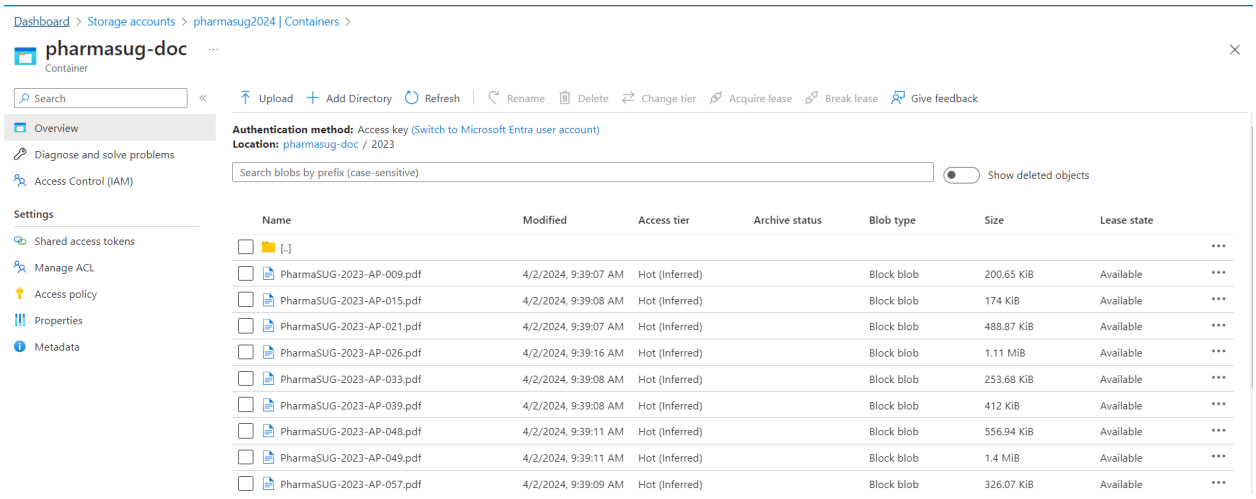


Figure 2. Document Repository in Azure Cloud

CREATE AZURE AI SEARCH SERVICE

Azure AI search service is used to do internal document semantics search service. It can be linked to ChatGPT service created later. Please follow steps below to set up Azure AI search service.

- Go to Azure portal (https://portal.azure.com) and navigate to “AI Search” page.
- Click “create” to create a dedicated Azure AI Search Service *pharmasug-ai-search* by following the creation wizard.
    - Choose pricing tier to basic or above to use semantics search capability.

- Go to “Data Sources” tab of *pharmasug-ai-search*, add a data source *pharmasug-doc-data-source* to link to storage container *pharmasug-doc*.
    - Make sure data source type is “Azure Data Lake Storage Gen2”.
- Go to “Overview” tab of *pharmasug-ai-search*, click “Import data” to add existing data source defined above. Once the search indexer is created, Azure AI will trigger document indexing automatically if you already upload some documents into the data source.
    - Use Index name - *pharmasug-doc-index*
    - Use Indexer name - *pharmasug-doc-indexer*

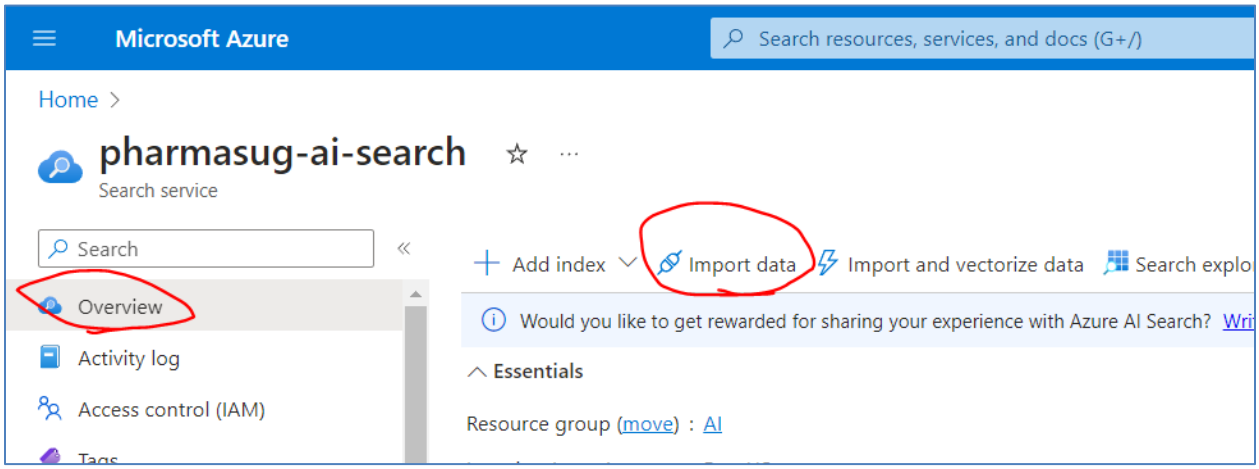


Figure 3. Azure AI Search Service

- Enter *pharmasug-doc-index*, select “semantic configurations” tab to create a semantic configuration.

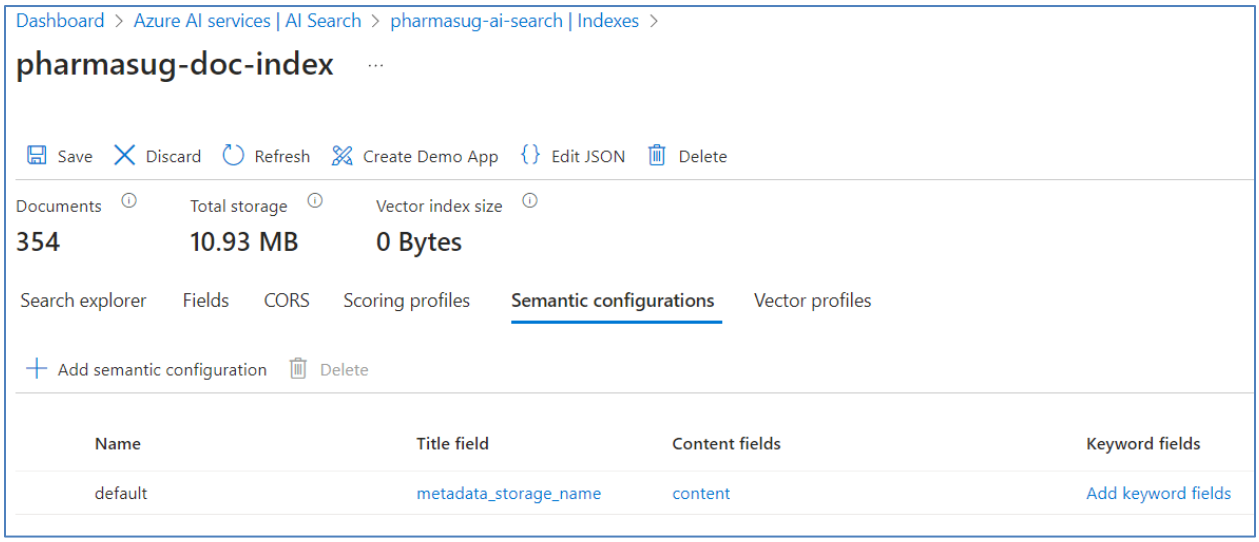


Figure 4. Azure AI Search Index Configuration

- If you have more documents to upload after initial indexing, you can go to *pharmasug-doc-indexer* and rerun document indexing.

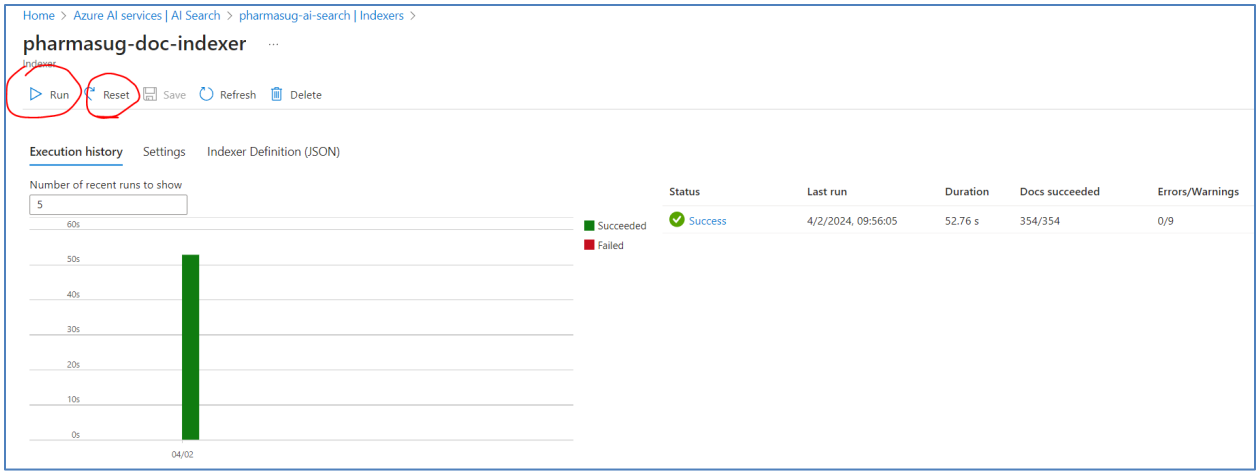


Figure 5. Azure AI Search Indexer

CREATE AND DEPLOY AZURE OPENAI PLATFORM

The core component Azure OpenAI platform is created and deployed as below.

- Go to Azure portal (https://portal.azure.com) and navigate to “Azure OpenAI” page.
- Click “Create” to create a dedicated Azure OpenAI Service *pharmasug-chatgpt* by following the creation wizard.
- Go to “Overview” tab of Azure OpenAI Service *pharmasug-chatgpt,* click “Go to Azure OpenAI Studio” to open Azure OpenAI Studio.
- Select “Chat” tab in Azure OpenAI Studio, click “Create new deployment” to create “gpt-35-turbo” or “gpt-4” with model deploy name “pharmasug-chat”. Then you use ChatGPT chat and completion playground to ask ChatGPT general or SAS/R/Python questions.

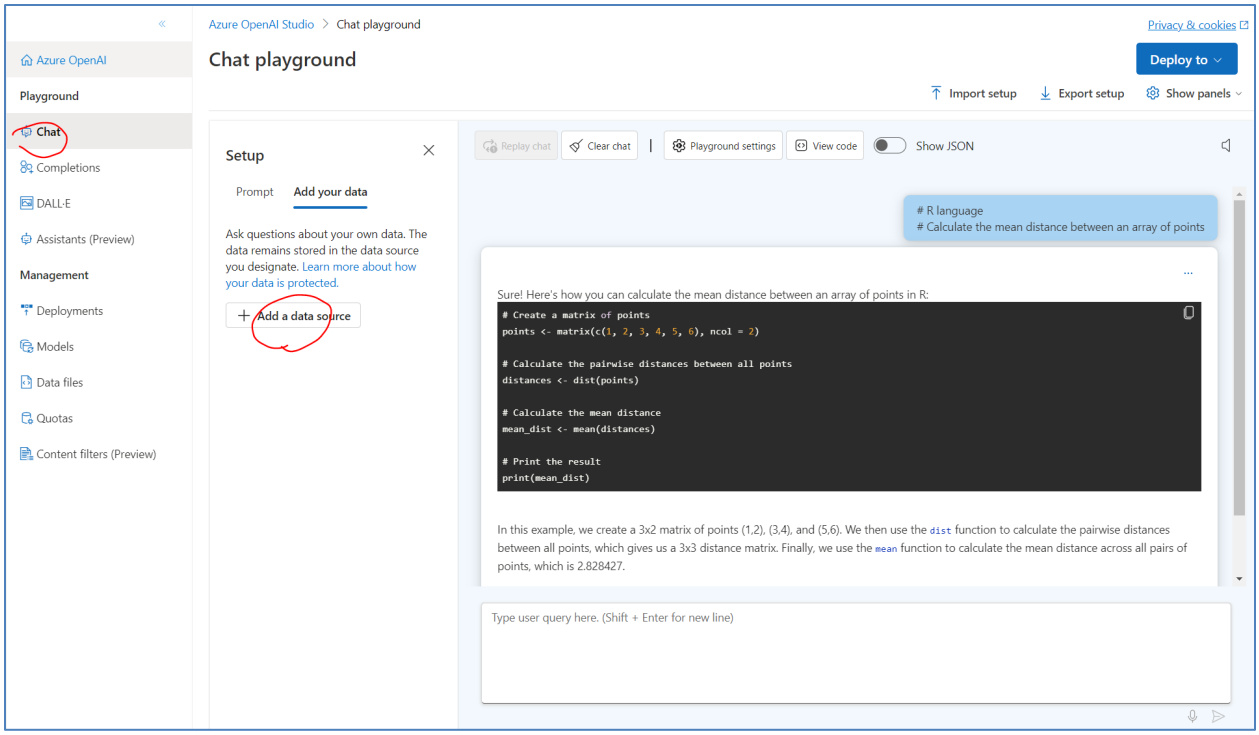


Figure 6. Azure OpenAI Playground

- In “Chat” tab, find “Setup” window, click “Add a data source” to link ChatGPT with Azure AI Search Service “*pharmasug-ai-search*” defined before. Once it is connected, you can query your internal documents with natural language.

CREATE CHATGPT WEB APPLICATION TO CONNECT TO OPENAI PLATFORM

Once a dedicated ChatGPT platform is created and deployed, you can deploy a web application for users to access ChatGPT environment in addition to built-in playground. You can have two ways to do it.

- Deploy an Azure OpenAI built-in web application - In Azure OpenAI Studio web page, click “Deploy to” button with “A new web app…” to create an Microsoft Entra ID protected web application for other users to access your ChatGPT environment.
- Deploy a customized web application linked to dedicated ChatGPT environment. The web application could be created with Python/ or C# based on OpenAI API packages.

SECURE YOUR CHATGPT WEB APPLICATION AND OPENAI PLATFORM

After you deploy OpenAI platform and associated ChatGPT web application, you can utilize Azure cloud app security features to enhance data privacy and security of ChatGPT environment. The main security function is “Access Control (IAM)” which applies to Azure resources defined above. And you can integrate with Azure Single Sign-On (SSO) authentication for your web application.

CONFIGURATION FOR DIFFERENT USE CASES

If you have your customized web application to link to your own ChatGPT environment, you can configure deployed ChatGPT model for different purposes by using different API calls and parameters. The following lists configuration for 3 typical ChatGPT use cases.

- general chat and completion

- programming copilot
- internal document semantics search

The sample web application is as below.

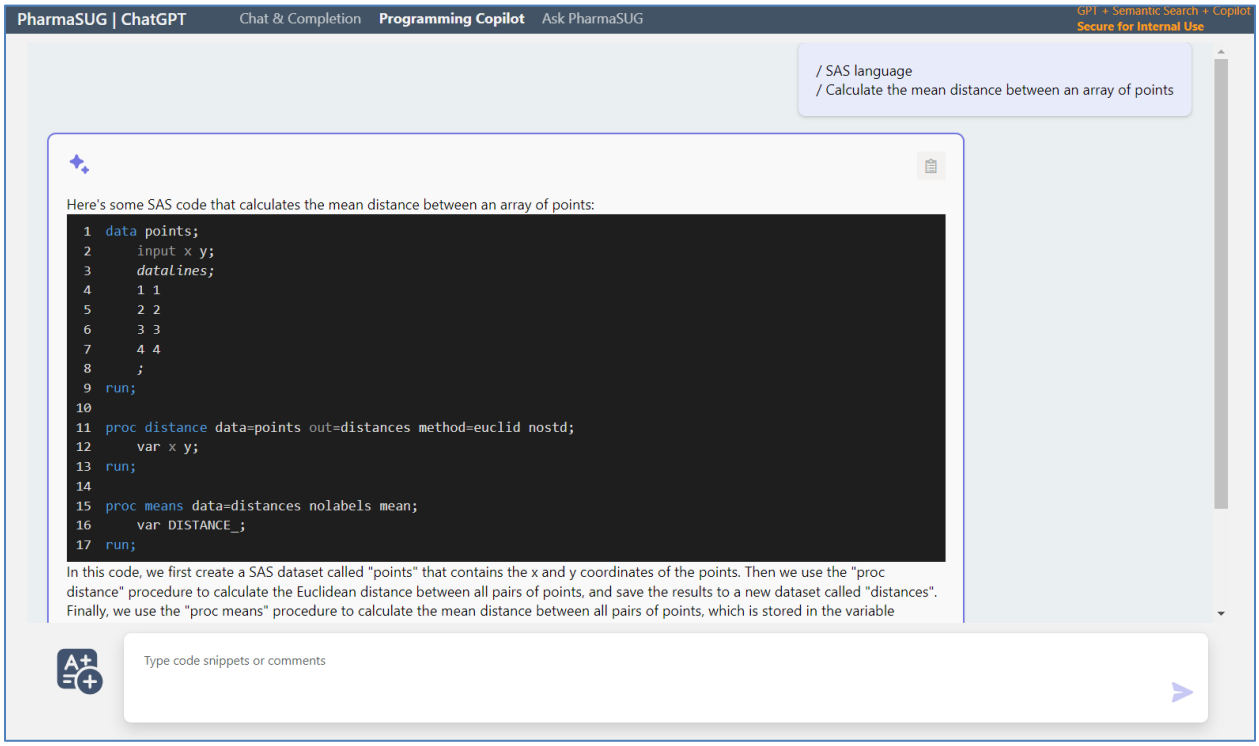


Figure 7. Sample ChatGPT Web Application

CONFIGURE FOR GENERAL CHAT AND COMPLETION

General chat and completion capability allows the user to ask general questions or ask ChatGPT to complete remaining request. The sample code block to invoke OpenAI platform from web application is listed as below.

```
            system_message = """You are an AI assistant that helps people find
information."""
            messages = self.get_messages_from_history(
                self.prompt_templates.get(chat_mode, system_message),
                self.chatgpt_model,
                history,
                None,
                self.completion_prompt_few_shots,
                self.chatgpt_token_limit
                )
            chat_completion_res = self.openai_client.chat.completions.create(
                messages=messages,
                model=self.chatgpt_deployment,
                stream=False
```

```
        )
```

Where `system_message` can be “You are an AI assistant that helps people find information.”

CONFIGURE FOR PROGRAMMING COPILOT AGAINST OPENAI

Programming copilot capability allows the user to ask programming related questions. ChatGPT supports most of programming languages like R, Python, Java, SAS, HTML, JavaScript etc.

The code block to invoke ChatGPT platform from web application is same as above, but with different `system_message` like “You are an AI assistant that helps people generate code with codex model.”

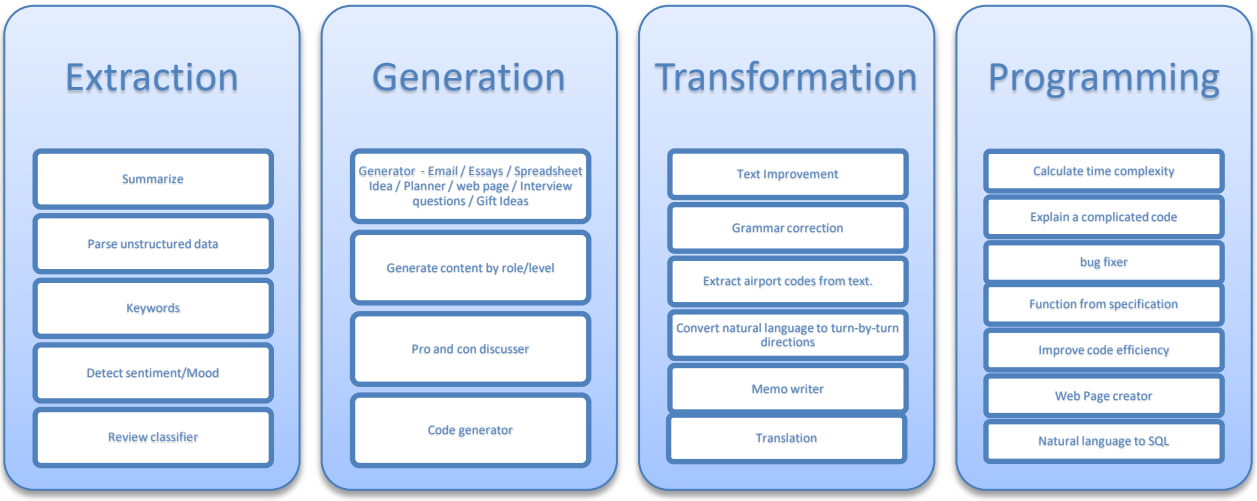
CONFIGURE FOR INTERNAL DOCUMENT SEMANTICS SEARCH

To support Internal document semantics search, the following python code block with additional “dataSources” information needs to be added. The system message could be assigned as “You are an AI assistant for PharmaSUG that helps people search PharmaSUG papers with PharmaSUG search service.”

```
            extra_body=dict(
              {
                  "dataSources":[{
                      "type": "AzureCognitiveSearch",
                      "parameters": {
                          "endpoint": self.search_client._endpoint,
                          "key": self.search_key,
                          "indexName": self.search_client._index_name,
                          'semanticConfiguration':'default',
                          "queryType": 'semantic',
                          "topNDocuments": 5,
                          "fieldsMapping": {
                              "contentFields": [],
                              "titleField": 'metadata_storage_name',
                              "urlField": 'url',
                              "filepathField": 'metadata_storage_path'
                          }
                      }
                  }]
              })
          chat_completion_res = self.openai_client_ext.chat.completions.create(
              messages=messages,
              model=self.chatgpt_deployment,
              stream=False,
              extra_body=extra_body
          )
        chat_content = chat_completion_res.choices[0].message.content
```

CHATGPT ESSENTAIL USAGE

As a user of ChatGPT, there are some essential functions (as below) to enhance your experience and make the most of this powerful tool. Most importantly, you can ask ChatGPT more SAS/R programming questions without comprise of your project and data information.



FUTURE ENHANCEMENT AND EXTENSION

Once your ChatGPT environment is created, you can explore OpenAI features and APIs to extend ChatGPT environment. Some enhancements and extensions are:

- Refine ChatGPT prompt flow to produce more accurate answers.
- Ingest more domain knowledges, SAS/R coding block for company-wide sharing.
- Add more specific functions utilizing large language model to provide business functions, for instance MedTRA coding query.
- Combine with Other document intelligence services.

CONCLUSION

OpenAI/ChatGPT is a great platform for generative AI. Its content generation and natural language features can help us all to improve productivity of our daily work. Through dedicated ChatGPT environment and additional access control, data security and privacy for ChatGPT are enforced firmly. With Microsoft Azure cloud, you have easier way to deploy your own ChatGPT to provide users full capabilities of ChatGPT. Furthermore, you can easily customize and extend your ChatGPT environment to serve your business better.

REFERENCES

Website ChatGPT overview https://openai.com/chatgpt

Website Azure OpenAI Service https://openai.com/chatgpt

Website OpenAI API Documents https://platform.openai.com/docs

Website ChatGPT + Enterprise data with Azure OpenAI and AI Search https://github.com/Azure-Samples/azure-search-openai-demo

Website Codex models and Azure OpenAI Service https://learn.microsoft.com/en-us/azure/ai-services/openai/how-to/work-with-code

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Bill Zhang
ClinChoice Inc.
Bill.zhang@clinchoice.com

Jun Yang
ClinChoice Inc.
jun.yang@clinchoice.com