

Azure Open AI – Level up hands-on labs.

Francis Simy Nazareth, Sher Shah Khan, Divi Mishra

Agenda

Presentation: Open AI Concepts (L-200)

Demonstration: Art of possibilities with Open AI

Introduction to Labs

Lab 1: Deploy a static website

Lab 2: Deploy OpenAI

Lab 3: Use your data with Open AI

Lab 4: Deploy & customize the web app.

Discussion / Labs.

Lab 5: Prompt engineering.

Lab 6: Using vector search

Lab 7: Semantic Search

Lab 8: Integrating Open AI with Azure APIM

Lab 9: Using Bing Search APIs

Workshop scenario

Contoso Bank is a retail bank. Contoso Bank wants to utilize generative AI capabilities through a chatbot in their website.

Your mission is to create an Open AI based chat bot that can converse with the user about Contoso Bank's products.

Lab 1: Deploy the static website

1. Fork the repository: <https://github.com/francisnazareth/contosobank/>
2. Deploy an Azure Static Website using the repository
 - a. In Azure Portal, create a new static web app
 - b. Select (or create) a resource group
 - c. Specify a name for the static webapp
 - d. For the hosting plan, select “Free”
 - e. Select a region for deployment
 - f. Select your github account and the cloned project to deploy (select the GitHub account, organization, repository and branch).
 - g. Click on “Review and Create” & “Create” to create the static webapp.

Microsoft Azure

Search resources, services, and docs (G+/I)

Home > Static Web Apps >

Create Static Web App

Project Details
Select a subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ

MCAPS-fnazaret-1

Resource Group * ⓘ

(New) rg-oai-workshop

[Create new](#)

Static Web App details

Name *

contosobank123

Hosting plan
The hosting plan dictates your bandwidth, custom domain, storage, and other available features. [Compare plans](#)

Plan type

☒ Free: For hobby or personal projects

☐ Standard: For general purpose production apps

Azure Functions and staging details
Region for Azure Functions API and staging environments *

West Europe

Deployment details

Source

☒ GitHub ☐ Azure DevOps ☐ Other

GitHub account

francisnazareth

[Change account](#) ⓘ

ⓘ If you can't find an organization or repository, you might need to enable additional permissions on GitHub. You must have write access to your chosen repository to deploy with GitHub Actions.

Organization *

francisnazareth

Repository *

contosobank

Branch *

main

Build Details

[Review + create](#)

[< Previous](#)

[Next: Tags >](#)

- h. Once the static webapp is created, click on “Settings -> Environments”. Wait for the status to be ready.

Production

There is one production environment per app. Check its status and last update time here.

Name	Branch	Last Update Time	Status	Traffic	Browse
Production	main	Thursday, September 7, 2023 at 10:41:...	⚙️ WaitingForDeployment	100%	Browse

Production

There is one production environment per app. Check its status and last update time here.

Name	Branch	Last Update Time	Status	Traffic	Browse
Production	main	Thursday, September 7, 2023 at 10:43:...	✅ Ready	100%	Browse

Once the status changes to ready, explore the static web app (Click on the URL from Overview tile)

contosobank123 Static Web App

Search « Browse Delete Manage deployment token Send us your feedback

Overview

Access control (IAM) Tags Diagnose and solve problems

Settings

Configuration Application Insights

Essentials

Resource group (move) : [rg-sai-workshop](#)

Subscription (move) : [MCAPS-fnazaret-1](#)

Subscription ID : [cb56ce60-e634-4951-8ac8-ec94c671d4e8](#)

Location : Global

Sku : Free

Tags (edit) : [Add tags](#)

URL : <https://black-sand-0de388203.3.azurestaticapps.net>

Source : [main \(GitHub\)](#)

Deployment history : [GitHub Action runs](#)

Edit workflow : [azure-static-web-apps-black-sand-0de388203.yml](#)

Explore through different tabs of the Contoso Bank website.



We can help you reach your savings goals faster

We have a range of options to match your style of saving and help make it easier for you to reach your goals.



Home Loans

Choose a payment plan that works for you

Basic

Fixed 12 month special rate
Min 20% equity required
T&C's apply

Pro

Fixed 24 month special rate
Min 20% equity required
T&C's apply

Premium


House variable rate
Min 20% equity required
T&C's apply

Lab 2: Provision Azure Open AI resource

Create the following resources:

1. Azure Open AI instance. (select the region as East US)

Create Azure OpenAI ...

 Changes on this step may reset later selections you have made. Review all options prior to deployment.

✓ Basics

✓ Network

✓ Tags

④ Review + submit

Enable new business solutions with OpenAI's language generation capabilities powered by GPT-3 models. These models have been pretrained with trillions of words and can easily adapt to your scenario with a few short examples provided at inference. Apply them to numerous scenarios, from summarization to content and code generation.

[Learn more](#)

Project Details

Subscription * ⓘ

MCAPS-fnazaret-1

Resource group * ⓘ

rg-openai

[Create new](#)

Instance Details

Region ⓘ

East US

Name * ⓘ

oaiworkshop

✓

Pricing tier * ⓘ

Standard S0

[View full pricing details](#)

Content review policy

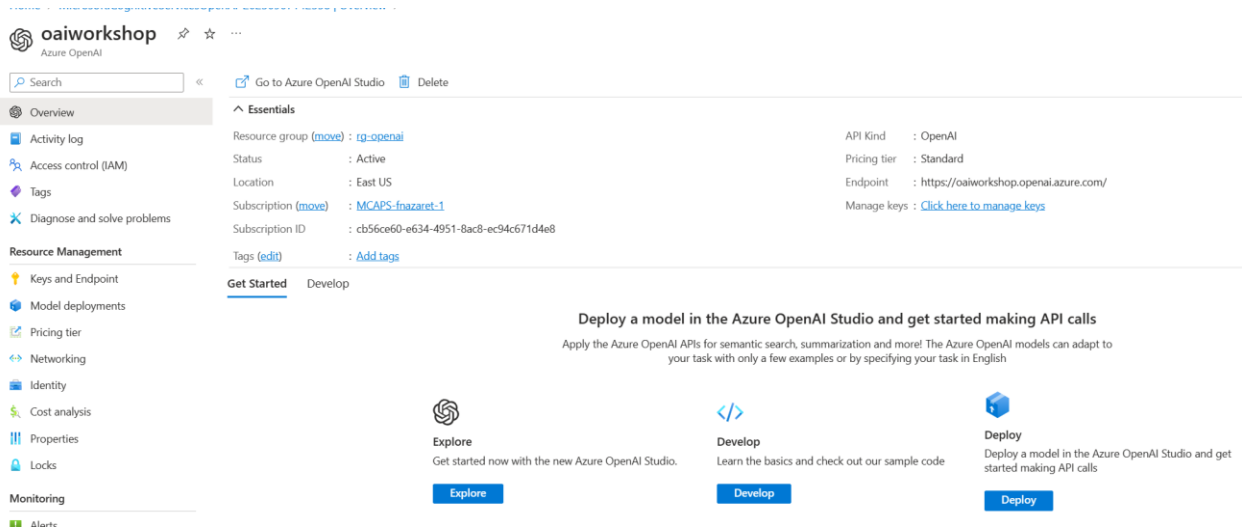
To detect and mitigate harmful use of the Azure OpenAI Service, Microsoft logs the content you send to the Completions and image generations APIs as well as the content it sends back. If content is flagged by the service's filters, it may be reviewed by a Microsoft full-time employee.

[Learn more about how Microsoft processes, uses, and stores your data](#)

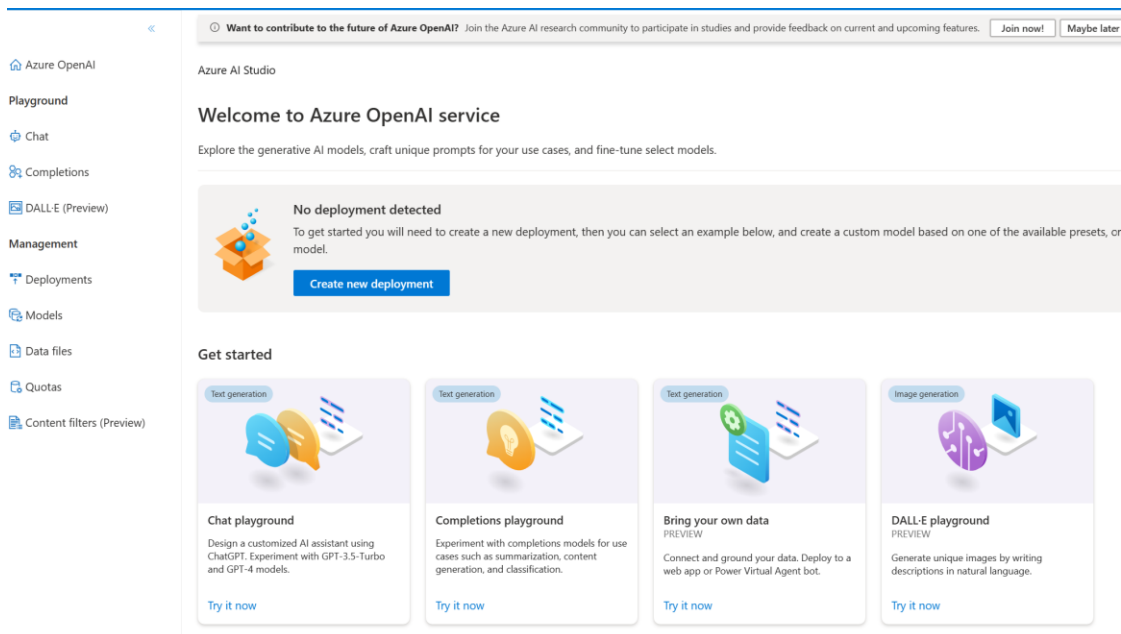
[Apply for modified content filters and abuse monitoring](#)

[Review the Azure OpenAI code of conduct](#)

2. In the Azure Open AI resource, click on “Explore” to open the Open AI studio.



In the Azure Open AI studio, create a new deployment



Select the model as “gpt-35-turbo-16k”

Specify a deployment name (for example, same name as model).

Click on “Create”.

Edit deployment

Model name
gpt-35-turbo-16k

Deployment name ⓘ
gpt-35-turbo-16k

⚙️ Advanced options ▾

Content Filter ⓘ
Default ▾

ⓘ 120K tokens per minute quota available for your deployment

Tokens per Minute Rate Limit (thousands) ⓘ

120K

Corresponding requests per minute (RPM) = 720

Save and close

Cancel

The deployment should be provisioned.

+ Create new deployment

Edit deployment

Delete deployment

Column options

Refresh

Open in Playground

Search

Deployment name ▾	Model name ▾	M... ▾	De... ▾	Capac... ▾	Sta... ▾	Model dep... ▾	Content Fil... ▾	Rate li... ▾	Updated at ▾
✔ gpt-35-turbo-16k	gpt-35-turbo-16k	0613	Standard	120K TPM	✔ Succ	1/15/2024	Default	120000	9/12/2023 12:35 I

Test out the chat using chat playground. (You may need to wait for few minutes for the model to be available).

Chat playground

↑ Impx

Assistant setup

System message

Add your data (preview)

Save changes

Specify how the chat should act

Use a template to get started, or just start writing your own system message below. Want some tips? [Learn more](#)

Use a system message template

Select a template

System message ⓘ

You are an AI assistant that helps people find information.

Examples ⓘ

+ Add an example

Chat session

Clear chat

View code

Show raw JSON

Hi

...

Hello! How can I assist you today?

Type user query here. (Shift + Enter for new line)

➤

↻

Try different system message templates

Specify how the chat should act

Use a template to get started, or just start writing your own system message below. Want some tips? [Learn more](#)

Use a system message template

Xbox customer support agent

System message ⓘ

You are an Xbox customer support agent whose primary goal is to help users with issues they are experiencing with their Xbox devices. You are friendly and concise. You only provide factual answers to queries, and do not provide answers that are not related to Xbox.

Examples ⓘ

User:

How much is a PS5?

Assistant:

I apologize, but I do not have information about the prices of other gaming devices such as the PS5. My primary focus is to assist with issues regarding Xbox devices. Is there a specific



Hello! How can I assist you today?

New chat session started

The assistant setup has been updated. Previous messages won't be used as context for new queries.

Can you compare X-Box series S with X-box series X



Certainly! Here's a comparison between the Xbox Series S and Xbox Series X:

1. Performance: The Xbox Series X is more powerful than the Series S. It has a faster CPU, more RAM, and a more advanced GPU, allowing for better graphics and higher frame rates.
2. Resolution: The Series X supports gaming at up to 4K resolution, while the Series S supports up to 1440p resolution. This means the Series X can deliver a more detailed and visually immersive

Type user query here. (Shift + Enter for new line)

Lab 3: Use your data with Open AI

1. Provision a storage account to store your own data

- Create a storage account in the same region as Azure Open AI.
- Accept defaults for storage account settings,
- Review the settings and create the storage account.

Home > Storage accounts >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription * MCAPS-fhazaret-1

Resource group * rg-oai-workshop [Create new](#)

Instance details

Storage account name ⓘ * stoaiworkshop

Region ⓘ * (US) East US 2 [Deploy to an edge zone](#)

Performance ⓘ * ☒ Standard: Recommended for most scenarios (general-purpose v2 account)
☐ Premium: Recommended for scenarios that require low latency.

Redundancy ⓘ * Locally-redundant storage (LRS)

[Review](#) < Previous Next : Advanced >

- Once the storage account is created, create a container named “lab” within the storage account.(Data storage -> Containers)

New container

Name * lab

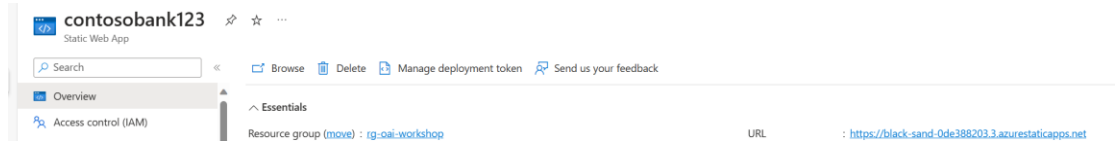
Anonymous access level ⓘ Private (no anonymous access)

i The access level is set to private because anonymous access is disabled on this storage account.

Advanced

2. Download html files from the static website.

- a. From the static website, copy the static website URL.



- b. Use the following command (in Azure cloud shell, or in any other Linux command line terminal) to download all HTML files
- ```
wget -m -p -k <<Static Website URL>>
```

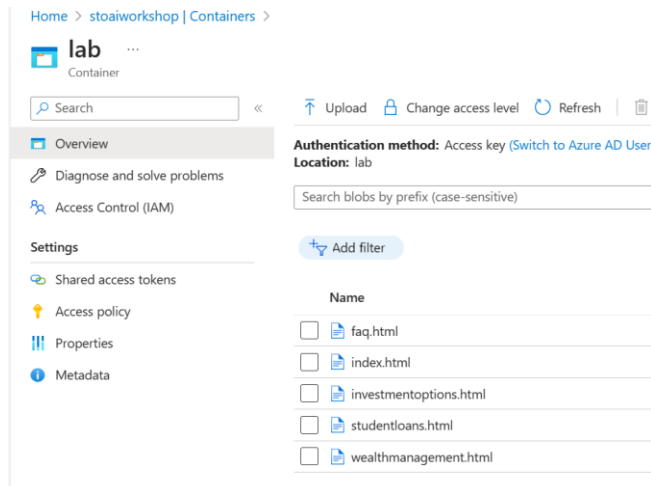
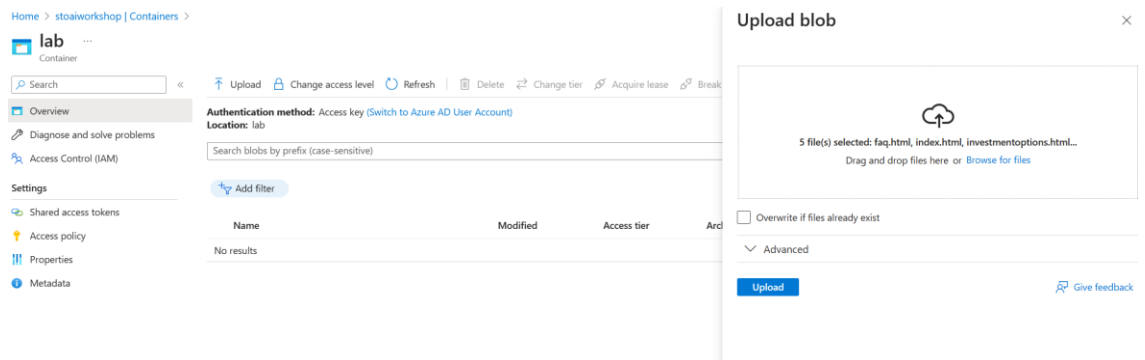
For example,

```
wget -m -p -k https://black-sand-0de388203.3.azurestaticapps.net
```

The command will create a folder by the name of the static website URL, and will download 5 HTML files there. You can manually download the HTML files as well.

(If you are using Azure cloudshell, you may want to bring the files to your machine – you can zip all HTML files using the command **zip htmlfiles.zip \*.html** , download the zip file to your machine, and unzip the zip file).

3. Upload the files to the container you created in Azure Storage Account.



4. Create an Azure Cognitive Search service instance.  
Select the same resource group and region.

Please note that the free tier is not supported with Azure Open AI

Also, though basic tier can be used with Open AI, semantic search is not supported in Basic tier.

## Create a search service ...

Basics   Scale   Networking   Tags   Review + create

### Project details

|                  |                   |
|------------------|-------------------|
| Subscription *   | MCAPS-fnazaret-1  |
| Resource Group * | rg-openaiworkshop |

[Create new](#)

### Instance Details

|                  |                                                                                                                                     |
|------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| Service name * ⓘ | oaisearch125                                                                                                                        |
| Location *       | East US                                                                                                                             |
| Pricing tier * ⓘ | <b>Standard</b><br>25 GB/Partition*, max 12 replicas, max 12 partitions, max 36 search units<br><a href="#">Change Pricing Tier</a> |

5. Create an index and add your data.
  - a. From the Open AI studio, in the “Chat Playground”, select “Add your Data (Preview)”

Playground

Chat

Completions

DALL-E (Preview)

Management

Deployments

Models

### Assistant setup

System message   Add your data (preview)

Ask questions about your own data. Your data is stored securely in your Azure subscription. [Learn more about how your data is protected.](#)

+ Add a data source

- b. In the “+ Add a data source”, select the option “Azure blob storage”.
  - c. Select the storage account you created in previous step.
  - d. Select the container you created in previous step.

- e. Select the cognitive service you provisioned in previous step.
- f. Provide a name for the index (for example, index1)
- g. For the indexer schedule, select “Hourly”.
- h. Leave the check box for “Add vector search” unchecked.
- i. Select the check box for “I acknowledge that connecting to an Azure Cognitive Search account will incur charges to my account”.

Add data

×

● Data source

○ Data management

○ Review and finish

Select or add data source

Your data source is used to help ground the model with specific data. Select an existing Azure Cognitive Search index, Azure Storage container, or upload local files as the source we will build the grounding data from. Your data is stored securely in your Azure subscription.

Select data source \*

Azure Blob Storage

Subscription \*

MCAPS-fnazaret-1

Select storage container ⓘ \*

lab

Select Azure Blob storage resource ⓘ \*

stoaiworkshop

Select Azure Cognitive Search resource ⓘ \*

oaidemosearch123

Enter the index name ⓘ \*

index1

Indexer schedule ⓘ \*

Hourly

☐ Add vector search to this search resource.

☒ I acknowledge that connecting to an Azure Cognitive Search account will incur usage to my account.

[View Pricing](#)

Next

Cancel

- j. Click “Next”.
- k. For the search type, select “Keyword”.

Add data

✓ Data source

● Data management

○ Review and finish

Data management

Set up specific configurations for your data and how the model will respond to requests.


Search type ⓘ \*


Keyword

- l. Select the checkbox to accept cost, and click Next.
- m. Click on “Save & Close”.
- n. Wait till the data is added. This may take a few minutes.

## Assistant setup




 System message

 Add your data (preview)

Gain insights into your own data source. Your data is stored securely in your Azure subscription. [Learn more about how your data is protected.](#)




☒ Limit responses to your data content ⓘ

Data source: Azure Blob Storage  
Search Resource: oaidemosearch123  
Index: [index1](#)


 Remove data source


- o. Try asking questions about the FAQs, student loans, etc. in the chat window. (It may take a few minutes for the indexed content to be available in the chat).

Chat session

 Clear chat  View code  Show raw JSON

1 [wealthmanagement.html - Part 1](#)

what are your wealth management offers 





Contoso Bank offers a range of wealth management solutions to help individuals achieve their financial goals <sup>1</sup>. Here are some of the wealth management offers provided by Contoso Bank:

1. Personalized Retirement Planning: Contoso Bank offers retirement income strategies tailored to your specific financial situation. This includes analyzing your income sources, investments, and goals to create a customized retirement income strategy <sup>1</sup>.

2. Estate and Legacy Planning: Contoso Bank can assist you in creating a comprehensive estate plan that preserves and distributes your wealth according to your wishes. They also offer trust management services to ensure the smooth transfer of assets to beneficiaries <sup>1</sup>.

3. Healthcare Financial Planning: Contoso Bank

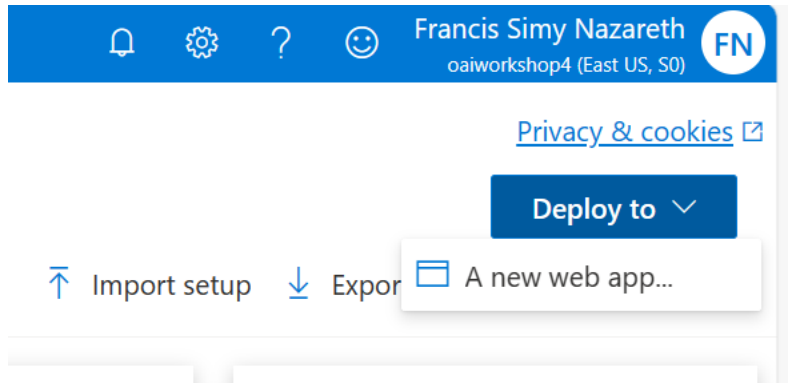
Type user query here. (Shift + Enter for new line)  



## Lab 4: Deploy & customize the AI chat webapp

### 4.1 Deploy the web app.

- a. In the chat playground, click the “Deploy to” button on the top-right corner, and select the option “A new web app”.



- b. In the “Deploy to a web app” screen
  - a. Provide a name for the web app
  - b. Select the subscription on to which the web app will be deployed.
  - c. Select an existing resource group to which the web app will be deployed.
  - d. Select the location (for example, East US)
  - e. Select a pricing plan (for example, Premium P1V3)
  - f. Uncheck “Enable chat history in web app”
  - g. Check “I acknowledge that web apps will incur usage to my account.
  - h. Deploy the web app.

Deploy to a web app

ⓘ

Your web app will be configured with Azure Active Directory authentication enabled. It may take a few minutes to apply after deployment completes, during which time you will not be able to chat in the app. Please wait 10 minutes, then reload the app and log in to begin chatting.

×

Pick your configurations to deploy a web app. [Learn more about web apps](#)

☒ Create a new web app

☐ Update an existing web app

Name ⓘ \*

oachat2423

Subscription ⓘ \*

MCAPS-fnazaret-1

Resource group ⓘ \*

rg-openaiworkshop

Location ⓘ \*

East US

Pricing plan ⓘ \*

Premium v3 (P1v3)

☐ Enable chat history in the web app

☒ I acknowledge that web apps will incur usage to my account [View Pricing](#)

Deploy

Cancel

- c. Wait till the status displays “Web App deployed”.Click on the button to launch the web app.

✓

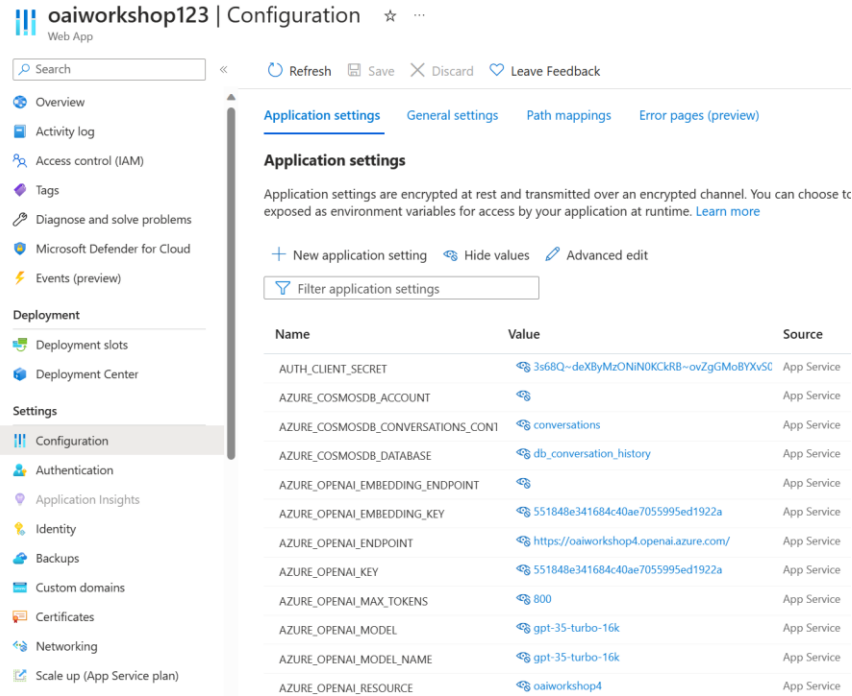
Web app deployed

×

oaiworkshop123

Launch web app

- d. Grant admin consent for the app and authorize the app.  
e. Try out conversing with the web app.  
f. From Azure Portal, verify that the web app’s configuration values are valid.



- g. Explore the source code for the web app here -> [microsoft/sample-app-aoai-chatGPT](https://github.com/microsoft/sample-app-aoai-chatGPT):  
[\[PREVIEW\] Sample code for a simple web chat experience targeting chatGPT through AOAI.](https://github.com/microsoft/sample-app-aoai-chatGPT)  
[\(github.com\)](https://github.com/microsoft/sample-app-aoai-chatGPT)

## 4.2 Customize the web app

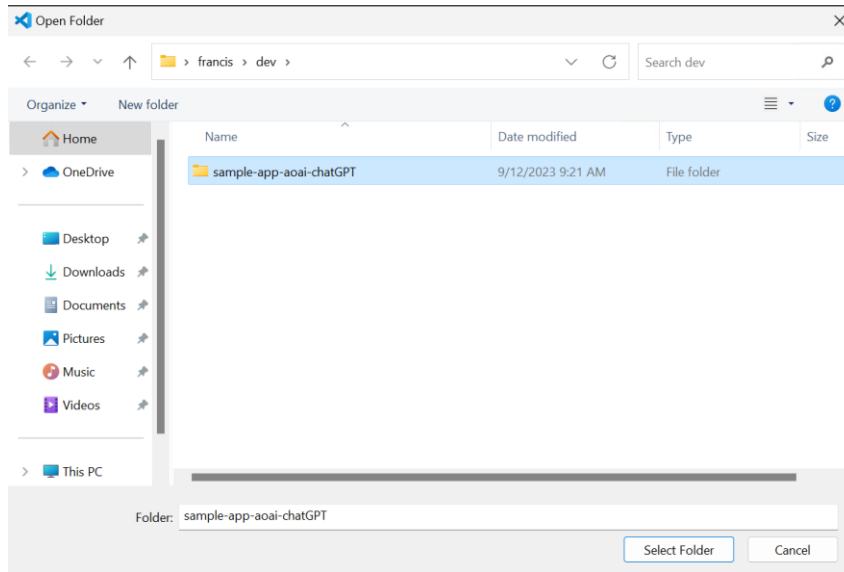
The following tools are recommended in your development machine to customize the sample web app.

1. Visual studio code (from [Visual Studio Code - Code Editing. Redefined](https://code.visualstudio.com/) )
2. Git (from [Git - Downloads \(git-scm.com\)](https://git-scm.com/) )
3. Python ([Download Python | Python.org](https://python.org/) )
4. set python and python\Scripts to OS path
5. Nodejs ([Node.js \(nodejs.org\)](https://nodejs.org/))
6. azure cli ([Install the Azure CLI for Windows | Microsoft Learn](https://learn.microsoft.com/en-us/cli/azure/install-azure-cli-windows?tabs=windows-powershell))

Once the tools are installed, open Visual Studio Code

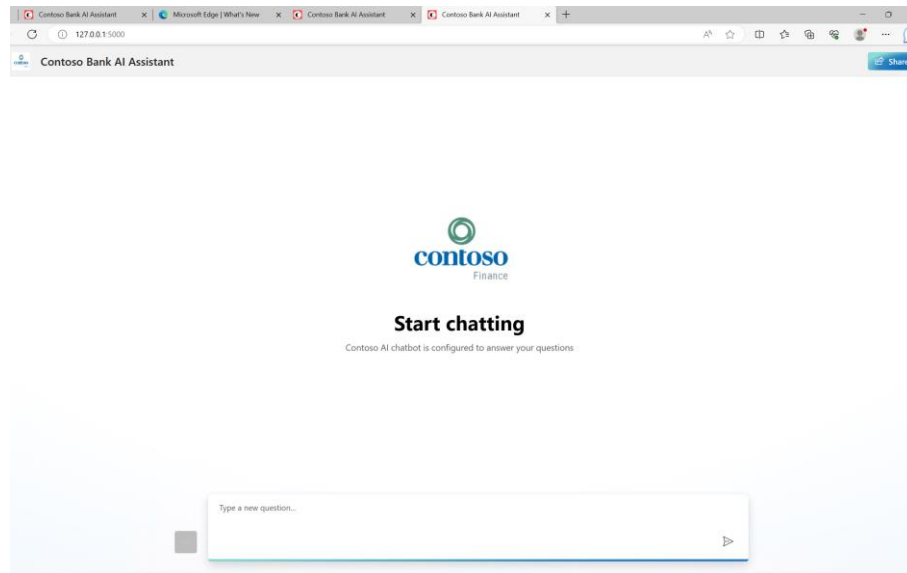
1. from the menu, open Terminal
2. `git clone https://github.com/microsoft/sample-app-aoai-chatGPT`
3. `cd sample-app-aoai-chatGPT`

4. Make sure that you are able to start the application, by typing `.\start.cmd`
5. Press Ctrl+C to close the running application.
6. Open the folder in visual studio code (File -> Open Folder, navigate to sample-app-aoai-chatGPT, and click on “Select Folder”).



7. Select “Trust authors of this code” in Visual Studio Code.
8. In the **frontend\public** folder, replace favicon.ico file with contoso bank’s favicon -> [favicon.ico \(32x32\) \(raw.githubusercontent.com\)](https://raw.githubusercontent.com/francisnazareth/contosobank/main/public/favicon.ico)
9. In visual studio code, change line #7 in **frontend\index.html**. Replace Azure AI with “Contoso Bank” (`<title>Contoso Bank</title>`)
10. Download the contoso logo from [contosobank/files/contoso-logo.png at main · francisnazareth/contosobank \(github.com\)](https://raw.githubusercontent.com/francisnazareth/contosobank/main/assets/contoso-logo.png)
11. Copy the contoso-logo.png to **frontend\src\assets\** folder.
12. Edit the file **frontend\src\pages\layout\Layout.tsx**
  - a. In line #3, change Azure.svg to **contoso-logo.png**  
(import Azure from "../../assets/contoso-logo.png";)
  - b. In line #82, change “Azure AI” to “Contoso Bank AI Assistant”
  - c. Save the file.
13. Edit the file **frontend\src\pages\chat\Chat.tsx** file
  - a. In line #11, change Azure.svg to contoso-logo.png
  - b. In line #552, change “this chatbot is configured to answer your questions” to “Contoso AI chatbot is configured to answer your questions”
  - c. Save the file.
14. Edit the file **frontend\src\pages\chat\Chat.module.css**
  - a. In line 64, change the height to 161px
  - b. In line 65, change the width to 216px
  - c. Save the file.
15. Build the front end
  - a. In visual studio code ->terminal, run “.\start.cmd”

You should be seeing the modified UI once the build finishes.



## 4.3 Publish the Web App to Azure.

1. From Azure CLI, login to your azure subscription (az login)
2. Change the app settings for your existing web app to allow local deployments.

```
az webapp config appsettings set -g <resource-group-name> -n <existing-app-name> --settings WEBSITE_WEBDEPLOY_USE_SCM=false
```

For example,

```
az webapp config appsettings set -g rg-openaiworkshop -n oaiworkshop123 --settings WEBSITE_WEBDEPLOY_USE_SCM=false
```

3. From the sample-app-aoai-chatGPT folder, run the following command to update the existing web application with your modified code.

```
az webapp up --runtime <runtime-stack> --sku <sku> -g <resource-group-name> -n <existing-app-name>
```

For example,

```
az webapp up --runtime "PYTHON:3.11" --sku P1V3 -g rg-openaiworkshop -n oaiworkshop123
```