



DATA SATURDAYS

Data Saturdays

Community is important to Data Saturdays. It is a data technology focused training event held on Saturdays. Volunteer led, free local one day technical conferences that help people grow their career and knowledge/skills.

Email: adpead@gmail.com Twitter: [@adpead](https://twitter.com/adpead)



Integrating Azure OpenAI Into your .NET Applications

DATA #44 CAPE TOWN
— SATURDAYS —



DATA
SATURDAYS

Data Saturdays

Community is important to Data Saturdays. It is a data technology focused training event held on Saturdays. Volunteer led, free local one day technical conferences that help people grow their career and knowledge/skills.

Email: adpead@gmail.com

Twitter: [@adpead](https://twitter.com/adpead)

- Twitter: @adpead
- About.me: https://about.me/allan_pead
- LinkedIn: <https://www.linkedin.com/in/adpead/>
- Blog: <https://explorationspace.co.za>
- Raspberry Pi South Africa
- <https://www.facebook.com/groups/1493503984198019>
- Cape Town MS Developer User Group
- <https://www.meetup.com/Cape-Town-Ms-Dev-User-Group/>

Allan Pead



Microsoft Internet of Things Most Valuable Professional
Microsoft Developer Technologies Most Valuable Professional



Azure AI

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

Cognitive Services



Vision



Speech



Language



Decision



Azure OpenAI Service

Azure Machine Learning



Prepare & Preprocess | Build, Train & Consume | Deploy & Scale | Manage & Monitor

AI Infrastructure

Azure OpenAI Service

GPT-3

Codex

DALL·E (preview)

ChatGPT (preview)

GPT-4 (preview)



Deploy on your
own data



Provisioned
Throughput
SKU



Plugins for
Azure OpenAI
Service

<https://azure.microsoft.com/en-in/products/cognitive-services/openai-service/>

Azure OpenAI - Completions

- In the context of natural language processing and artificial intelligence, refers to the output generated by an AI model.
 - In response to a prompt or input
- **Text Generation**
 - **Translation**
 - **Summarization**
 - **Question Answering**
 - **Code Generation**
 - **Dialog and Chatbots**
 - **Content Creation**

Azure OpenAI Service Privacy

Prompts and completions (outputs), your embeddings, and your training data

- are NOT available to other customers.
- are NOT available to OpenAI.
- are NOT used to improve OpenAI models.
- are NOT used to improve any Microsoft or 3rd party products or services.
- are NOT used for automatically improving Azure OpenAI models for your use in your resource (The models are stateless, unless you explicitly fine-tune models with your training data).
- Your fine-tuned Azure OpenAI models are available exclusively for your use.

<https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy>

Demo

Azure OpenAI Studio



DATA
SATURDAYS

Data Saturdays

Community is important to Data Saturdays. It is a data technology focused training event held on Saturdays. Volunteer led, free local one day technical conferences that help people grow their career and knowledge/skills.

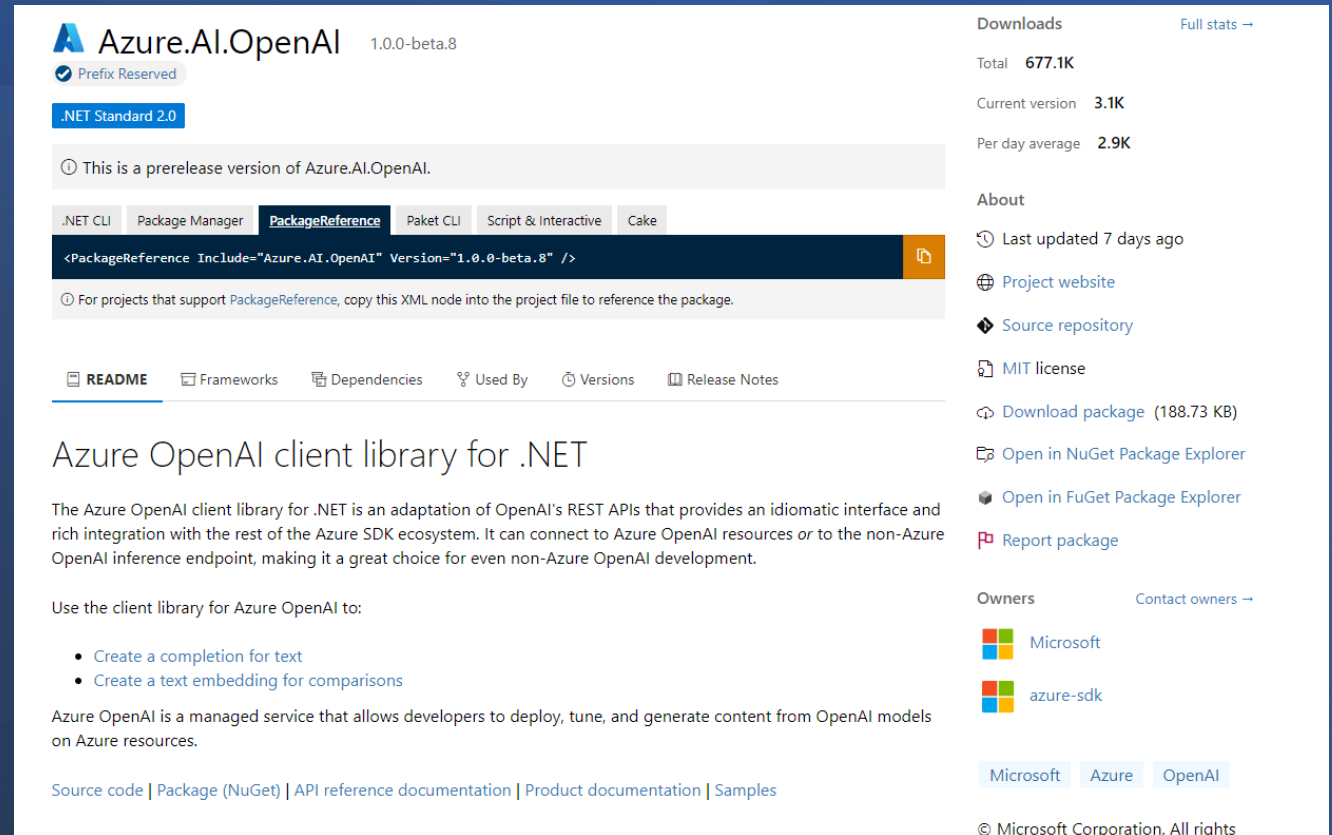
Azure OpenAI SDK

SDK Support

- C#
- Go
- Java
- JavaScript
- Python
- Powershell
- REST

Azure OpenAI SDK - .NET

Azure OpenAI client library for .NET



The screenshot shows the NuGet package page for Azure.AI.OpenAI. The package is version 1.0.0-beta.8 and is marked as a prerelease. It is compatible with .NET Standard 2.0. The page includes a 'PackageReference' tab with the XML code snippet: `<PackageReference Include="Azure.AI.OpenAI" Version="1.0.0-beta.8" />`. The 'About' section describes the library as an adaptation of OpenAI's REST APIs. The 'Downloads' section shows a total of 677.1K downloads. The 'Owners' section lists Microsoft and azure-sdk. The footer includes links to source code, package (NuGet), API reference documentation, product documentation, and samples.

Azure.AI.OpenAI 1.0.0-beta.8
Prefix Reserved
.NET Standard 2.0

This is a prerelease version of Azure.AI.OpenAI.

PackageReference
<PackageReference Include="Azure.AI.OpenAI" Version="1.0.0-beta.8" />

For projects that support PackageReference, copy this XML node into the project file to reference the package.

README Frameworks Dependencies Used By Versions Release Notes

Azure OpenAI client library for .NET

The Azure OpenAI client library for .NET is an adaptation of OpenAI's REST APIs that provides an idiomatic interface and rich integration with the rest of the Azure SDK ecosystem. It can connect to Azure OpenAI resources or to the non-Azure OpenAI inference endpoint, making it a great choice for even non-Azure OpenAI development.

Use the client library for Azure OpenAI to:

- Create a completion for text
- Create a text embedding for comparisons

Azure OpenAI is a managed service that allows developers to deploy, tune, and generate content from OpenAI models on Azure resources.

[Source code](#) | [Package \(NuGet\)](#) | [API reference documentation](#) | [Product documentation](#) | [Samples](#)

Downloads [Full stats →](#)
Total **677.1K**
Current version **3.1K**
Per day average **2.9K**

About
Last updated 7 days ago
[Project website](#)
[Source repository](#)
[MIT license](#)
[Download package](#) (188.73 KB)
[Open in NuGet Package Explorer](#)
[Open in FuGet Package Explorer](#)
[Report package](#)

Owners [Contact owners →](#)
Microsoft
azure-sdk

Microsoft Azure OpenAI

© Microsoft Corporation. All rights

<https://www.nuget.org/packages/Azure.AI.OpenAI/>

OpenAIClient

- Adaptation of OpenAI's REST APIs
- Integration with the rest of the Azure SDK ecosystem
- Connect to Azure OpenAI resources *or* to the non-Azure OpenAI inference endpoint

```
OpenAIClient client = new OpenAIClient(  
    new  
    Uri("https://datasatcptopenai.azure.  
com/"),  
    new AzureKeyCredential(apiKey));
```

<https://learn.microsoft.com/en-us/dotnet/api/overview/azure/ai.openai-readme>

ChatCompletions

- Choices
 - The collection of completions choices associated with this completions response.
- Created
 - The first timestamp associated with generation activity for this completions response, represented as seconds since the beginning of the Unix epoch of 00:00 on 1 Jan 1970.
- Id
 - A unique identifier associated with this chat completions response.
- PromptFilterResults
 - Content filtering results for zero or more prompts in the request.
- Usage
 - Usage information for tokens processed and generated

OpenAIClient

- Return Completions
- Stream Completions

```
Response<ChatCompletions> responseWithoutStream  
= await client.GetChatCompletionsAsync
```

```
Response<StreamingChatCompletions>  
responseWithoutStream = await  
client.GetChatCompletionsStreamingAsync
```


Streaming Completions

- Continuous feed of Completions

```
Response<StreamingChatCompletions>
response = await
client.GetChatCompletionsStreamingAsync(
    deploymentOrModelName: "gpt-3.5-
turbo",
    chatCompletionsOptions);
using StreamingChatCompletions
streamingChatCompletions =
response.Value;

await foreach (StreamingChatChoice
choice in
streamingChatCompletions.GetChoicesStreaming())
{
    await foreach (ChatMessage message
in choice.GetMessageStreaming())
    {
        Console.Write(message.Content);
    }
    Console.WriteLine();
}
```

client.GetChatCompletionsAsync

- ChatCompletionsOptions

- Temperature
- MaxTokens
- NucleusSamplingFactor
- FrequencyPenalty
- PresencePenalty

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

Temperature

- Controls randomness.
- Lowering the temperature means that the model will produce more repetitive and deterministic responses.
- Increasing the temperature will result in more unexpected or creative responses.

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

MaxTokens

- Set a limit on the number of tokens per model response. The API supports a maximum of 32768 tokens shared between the prompt.
- One token is roughly 4 characters for typical English text.

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

NucleusSamplingFactor

- Similar to temperature, this controls randomness but uses a different method.
- Lowering will narrow the model's token selection to likelier tokens.
- Increasing will let the model choose from tokens with both high and low likelihood.

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

Never change both Temperature and NucleusSamplingFactor

FrequencyPenalty

- Reduce the chance of repeating a token proportionally based on how often it has appeared in the text so far.
- This decreases the likelihood of repeating the exact same text in a response.

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

PresencePenalty

- Reduce the chance of repeating any token that has appeared in the text at all so far.
- This increases the likelihood of introducing new topics in a response.

```
Response<ChatCompletions>
responseWithoutStream = await
client.GetChatCompletionsAsync(
    "datasatcptgpt35eu",
    new ChatCompletionsOptions()
    {
        Messages =
        {
            new ChatMessage(ChatRole.System,
                @"return the sentiment of the text in
                only one word. The options are bad, ok
                and great"),
            new ChatMessage(ChatRole.User,
                comment),
        },
        Temperature = (float)0.7,
        MaxTokens = 800,
        NucleusSamplingFactor = (float)0.95,
        FrequencyPenalty = 0,
        PresencePenalty = 0,
    });
```

Messages

- Prompt Input Messages
- Roles
 - System
 - User
 - Agent

```
new ChatCompletionOptions()  
{  
    Messages =  
    {  
        new  
        ChatMessage(ChatRole.System,  
            @"return the sentiment of the  
            text in only one word. The  
            options are bad, ok and great"),  
        new  
        ChatMessage(ChatRole.User,  
            comment),  
    },  
}
```

System message

- Included at the beginning of the prompt and is used to prime the model with context, instructions, or other information relevant to your use case.
- Describe Personality
- Specify Format

```
new ChatMessage(ChatRole.System,  
@"return the sentiment of the  
text in only one word. The  
options are bad, ok and great"),
```

<https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/advanced-prompt-engineering>

User message

- The chat messages a user would type
- Results in a completion

```
new ChatMessage(ChatRole.User,  
comment),
```

<https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/advanced-prompt-engineering>

Response

- Contains List of Completions

```
ChatCompletions completions =  
responseWithoutStream.Value;
```

```
    if ((completions != null) &&  
        (completions.Choices.Count > 0))  
        return  
completions.Choices[0].Message.Content  
;  
    else return "na";
```

Demo

Azure OpenAI .NET Demo



DATA
SATURDAYS

Data Saturdays

Community is important to Data Saturdays. It is a data technology focused training event held on Saturdays. Volunteer led, free local one day technical conferences that help people grow their career and knowledge/skills.

Sentiment

```
Messages =  
{  
    new ChatMessage(ChatRole.System, @"return the sentiment of the text in only  
one word. The options are bad, ok and great"),  
    new ChatMessage(ChatRole.User, comment),  
},
```

Olivia Clark Bad

My recent visit to your restaurant was a complete disappointment. The ambiance was far from inviting, with a lack of proper lighting and decor. The menu options were limited and uninspiring. When the dishes arrived, they were poorly prepared and lacked flavor. The service was equally lackluster, with inattentive staff. This experience was a stark contrast to my expectations, and I won't be recommending your restaurant to others.

The restaurant visit was disappointing due to poor ambiance, limited menu options, poorly prepared dishes, and inattentive service.

Summary

```
Messages =  
{  
    new ChatMessage(ChatRole.System, @"return only a highlight summary of the text  
provided in one sentence"),  
    new ChatMessage(ChatRole.User, comment),  
},
```

Michael Brown OK

My dining experience at your restaurant can be described as neutral. The atmosphere was pleasant, and the service was adequate. The food was decent, but it didn't leave a strong impression. It was an average meal, and I neither regret nor particularly cherish the experience.

The dining experience at the restaurant was average with a pleasant atmosphere, adequate service, and decent food.

Azure OpenAI Service

Request Access to Azure OpenAI Service

Prerequisites

- An Azure subscription - [Create one for free](#).
- Access granted to Azure OpenAI in the desired Azure subscription.

Currently, access to this service is granted only by application. You can apply for access to Azure OpenAI by completing the form at <https://aka.ms/oai/access>. Open an issue on this repo to contact us if you have an issue.

- An Azure OpenAI Service resource with either the `gpt-35-turbo` or the `gpt-4`¹ models deployed. For more information about model deployment, see the [resource deployment guide](#).

¹ GPT-4 models are currently only available by request. To access these models, existing Azure OpenAI customers can [apply for access by filling out this form](#).

<https://aka.ms/oai/access>.

Pricing

Models	Context	Prompt (Per 1,000 tokens)	Completion (Per 1,000 tokens)
GPT-3.5-Turbo	4K	\$0.0015	\$0.002
GPT-3.5-Turbo	16K	\$0.003	\$0.004
GPT-4	8K	\$0.03	\$0.06
GPT-4	32K	\$0.06	\$0.12

Image models

Models	Per 100 images
	Standard
Dall-E	\$2

Excludes other services eg Azure Cognitive Search

- Twitter: @adpead
 - About.me: https://about.me/allan_pead
 - LinkedIn: <https://www.linkedin.com/in/adpead/>
 - Blog: <https://explorationspace.co.za>
-
- Cape Town MS Developer User Group
 - <https://www.meetup.com/Cape-Town-Ms-Dev-User-Group/>

Thank you!!

Resources

GitHub Copilot

<https://github.com/features/copilot/>

Github Copilot X

<https://github.com/features/preview/copilot-x>

Getting Started with Azure OpenAI

<https://learn.microsoft.com/en-us/azure/cognitive-services/openai/overview>

Thank you to our Sponsors

DATA #44 CAPE TOWN
— SATURDAYS —
30 SEP 2023




NORTHERN DATA

 **InfiniteDBA**