AZURE OPENAI CHAT WEB PART FOR SHAREPOINT ONLINE

Paul Borisov, Cloud Architect

Data Access Diagram: the following slide provides details on data privacy

SharePoint Online web part with regular and Full-Screen appearance.
Optional internet and data connections, voice input, samples, AI images.

Alternative chat history storage options with global and private sharing capabilities for users

OR

OR

SharePoint List
(Optimal, OOTB)

Web Browser
(Quick demo)

AzureSQL database
(Large Productions)

App Service

Key Vault

Request enrichment with an API key after successful authorization.

Clean architecture without cumbersome function apps

REST API requests
Texts only, the API key is not exposed in the browser (the more secure setup)

API Management Services

Request validation and authorization

Azure AD
(Microsoft Entra ID)

Request validation and authorization

External Azure AD
Trusted tenant 1
...
Trusted tenant n

OR

Azure OpenAI

Default models:
GPT-3.5 Turbo 16k
GPT-4 32k
GPT-4 Turbo (v1106, 128k)
GPT-4 Vision
Dalle3 Imaging

OpenAI

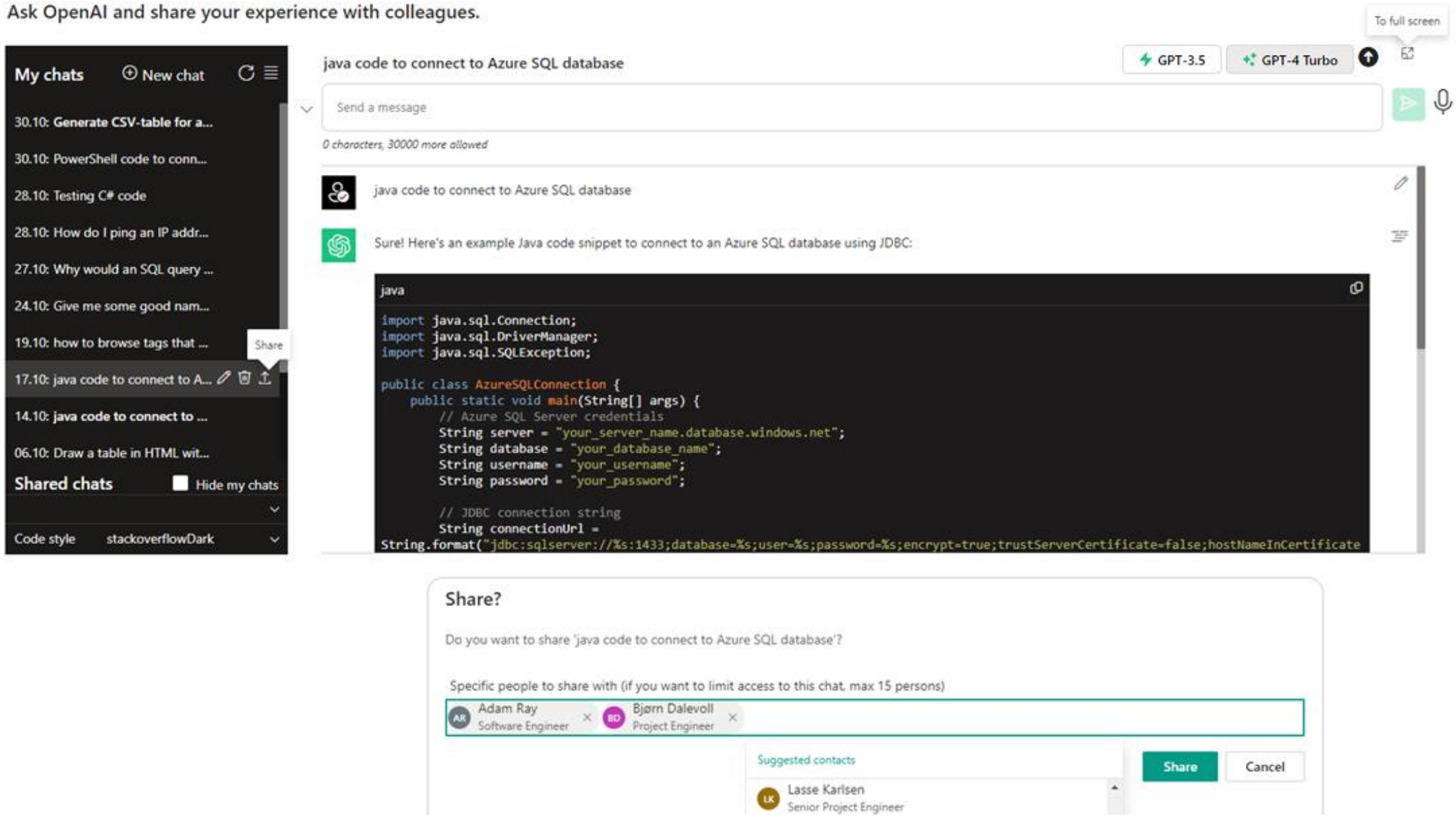
Some customers prefer using the Native OpenAI API subscription. The solution supports this option with the newest models available:
GPT 3.5 Turbo 16k
GPT-4 Turbo (v1106, 128k)
GPT-4 Vision
Dalle3 Imaging

Data Privacy

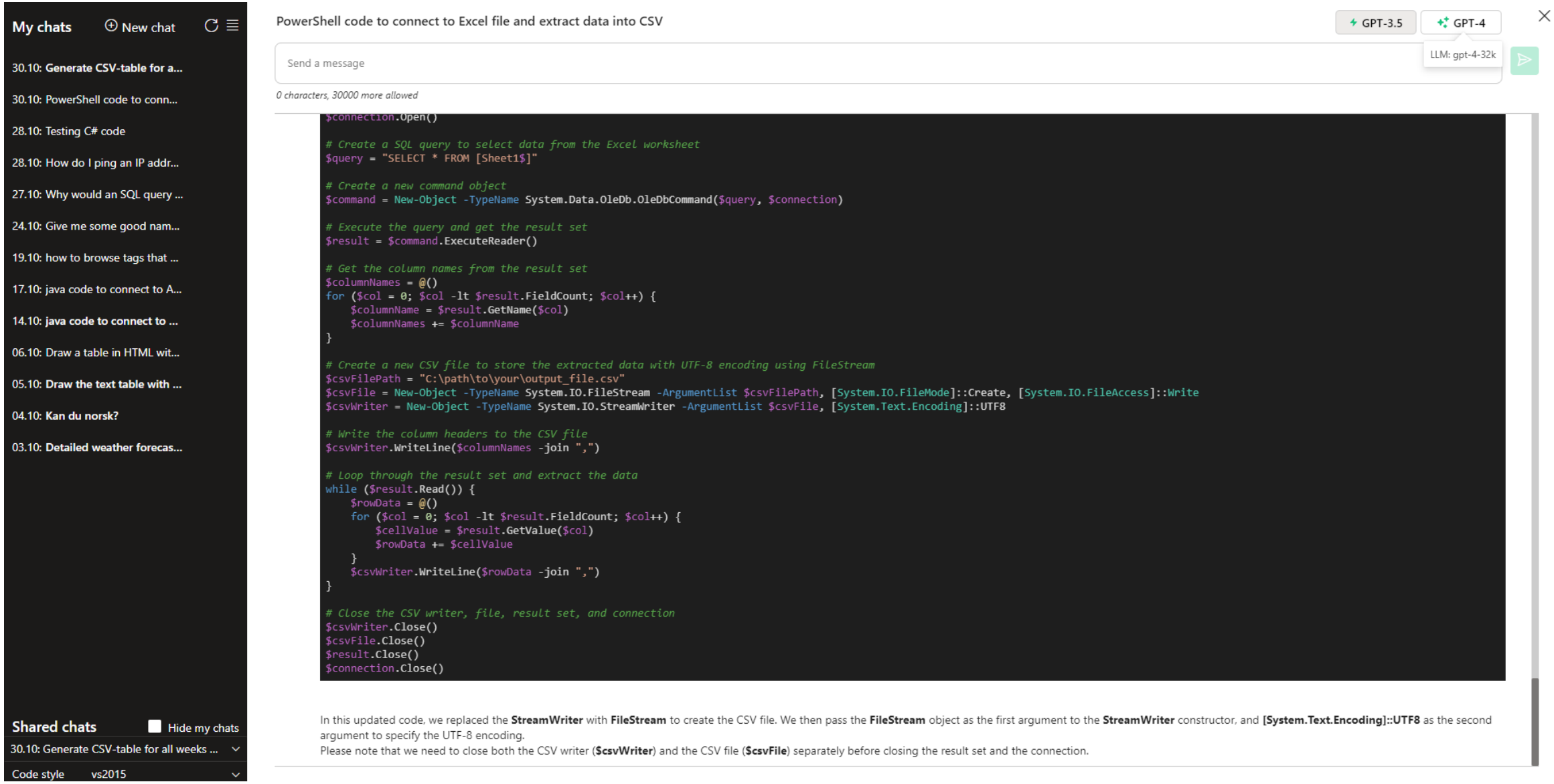
Azure OpenAI Chat web part interacts with private **Azure OpenAI** endpoints that are published via Azure API Management service (APIM).

- By default, this setup provides enhanced data privacy. In this configuration, requests to AI do not travel outside your Azure tenant.
- APIM consistently validates the identities of SharePoint users for each individual request. If the request originates from authorized domains, APIM retrieves the ****api-key**** from the secure vault and injects it into the request before forwarding it to the AI endpoint. This process ensures that the api-key does not get exposed in the browser.
- Chats are private and visible only to their creators. Creators have the option to share their chats when this feature is enabled in the web part settings (disabled by default). Creators can share their chats with everyone or only with specific people in the company.
- The web part incorporates tampering prevention logic to guard against unauthorized access to another user's data by their GUID.
- In addition to the default configuration, you have the option to publish the Native Open AI endpoint in APIM. You can find instructions in the project documentation.
  - **CONS**: Granting access to the Native Open AI endpoint requires a separate **api-key** for it and could potentially compromise data privacy, as requests might travel outside your Azure tenant under this setup.
  - **PROS**: Using the Native OpenAI endpoint could grant you access to the newly released language models that are not yet available in Azure OpenAI.

User Interface



Full-Screen Mode



Features

Key features

- Support for the latest text language models provided by Microsoft and Native OpenAI: GPT-4 32k, GPT 3.5 16k, GPT-4 Vision, GPT-4 Turbo (1106, 128k)
- Unlimited Chats per user and the option to support Unlimited Chat History Length
- Chat Sharing: Public sharing with Everyone in the company and Private sharing for a limited number of company users
- Support for the Full-Screen Mode. Support for Regular and Full-Width Site Page Sections
- Mobile-Friendly UI with simple customizations (CSS)
- Event Streaming to provide consecutive outputs of AI-responses
- Code Highlighter with options to select Highlight Styles.
- Upper and Lower Side positioning for the Prompt Area Textbox
- Full-Scale Localizations: English, Finnish, and Norwegian
- Optional integrations with company data. For security reasons, these integrations are disabled by default and must be explicitly enabled in the web part settings.
    - AI-Image generations with Dalle3. High-definition images available out-of-the-box
    - Search on the Internet with Bing and Google
    - SharePoint Search
    - Company Users
    - Local Date and Time
    - Analysis of an uploaded PDF and summarization of its content
    - Analysis of uploaded images and description of their content
    - Voice input and examples for prompt texts

Advanced features

- Built-in integration with API Management Service.
    - For customers that do not have it, the web part also supports direct queries to API with an api-key (configurable)
- Default support for Azure OpenAI and Native OpenAI API
- Alternative storage types to manage Chat History
    - SharePoint List: Supports up to 5000 Chats, Built-in “One-Click” list provisioning, Chat Privacy and Security Trimming for SharePoint Search
    - Azure SQL Database: Provides more advanced storage capabilities tested on 1 Million+ Chat records
    - Local Storage: it can be used to provide quick demo without long-time configurations

Optional Integrations: SharePoint Search, Company Users, Local Date & Time

How it works

The Function Calling feature is available in both Azure OpenAI and Native OpenAI (more advanced version). The flow for using this feature typically looks like:

1. In the process of interacting with OpenAI, you can send a request with options to call available integration functions that you have provided.
2. If OpenAI determines that it requires specific data, it will request to call a function with parameters that it deems suitable.
3. To meet OpenAI's demand, you can call the function with the proposed parameters on behalf of OpenAI. This allows you to obtain raw results locally, which you can then submit to OpenAI. You have control over what data you submit.
4. If OpenAI determines that it needs more information, it may demand calling another function, and this process can continue as necessary.
5. OpenAI handles the received data and transforms it as you have requested. The final response from OpenAI will include the results based on the processed data.

```
// 3 show cases available in the web part:
// Search in SharePoint for “your text“. Format the results as an HTML table.
// Get company users that have names starting with P. Format the results as an HTML table.
// Date, Time, Date and time.
export default class FunctionHelper {
  private static available: { [key: string]: any } = {
    companyUsers: this.companyUsers,
    currentDateOrTime: this.currentDateOrTime,
    currentDateAndTime: this.currentDateAndTime,
    peopleSearch: this.companyUsers,
    searchSharepoint: this.searchSharepoint,
  };
```

Integrations

Find in SharePoint "IDD Questionnaire" and show search results as HTML table

Find in SharePoint "IDD Questionnaire" and show search results as HTML table

Here are the search results for "IDD Questionnaire" in SharePoint:

Title	Author	Size	Path
Annex 3 - IDD Questionnaire for Corporates	Olivia Roca;SP Admin	85,822 bytes	Link
Annex 4 - IDD Questionnaire for State Owned Entities (SOEs)	Olivia Roca;SP Admin	85,302 bytes	Link
Annex 5 - IDD Questionnaire for Municipalities, Ministries and Other Government Agencies	Olivia Roca;SP Admin	84,004 bytes	Link

Show me company employees with names that start with p

Here is a list of company employees with names that start with the letter "P":

- Patti Fernandez
- Pradeep Gupta

Show me company employees with names that start with H

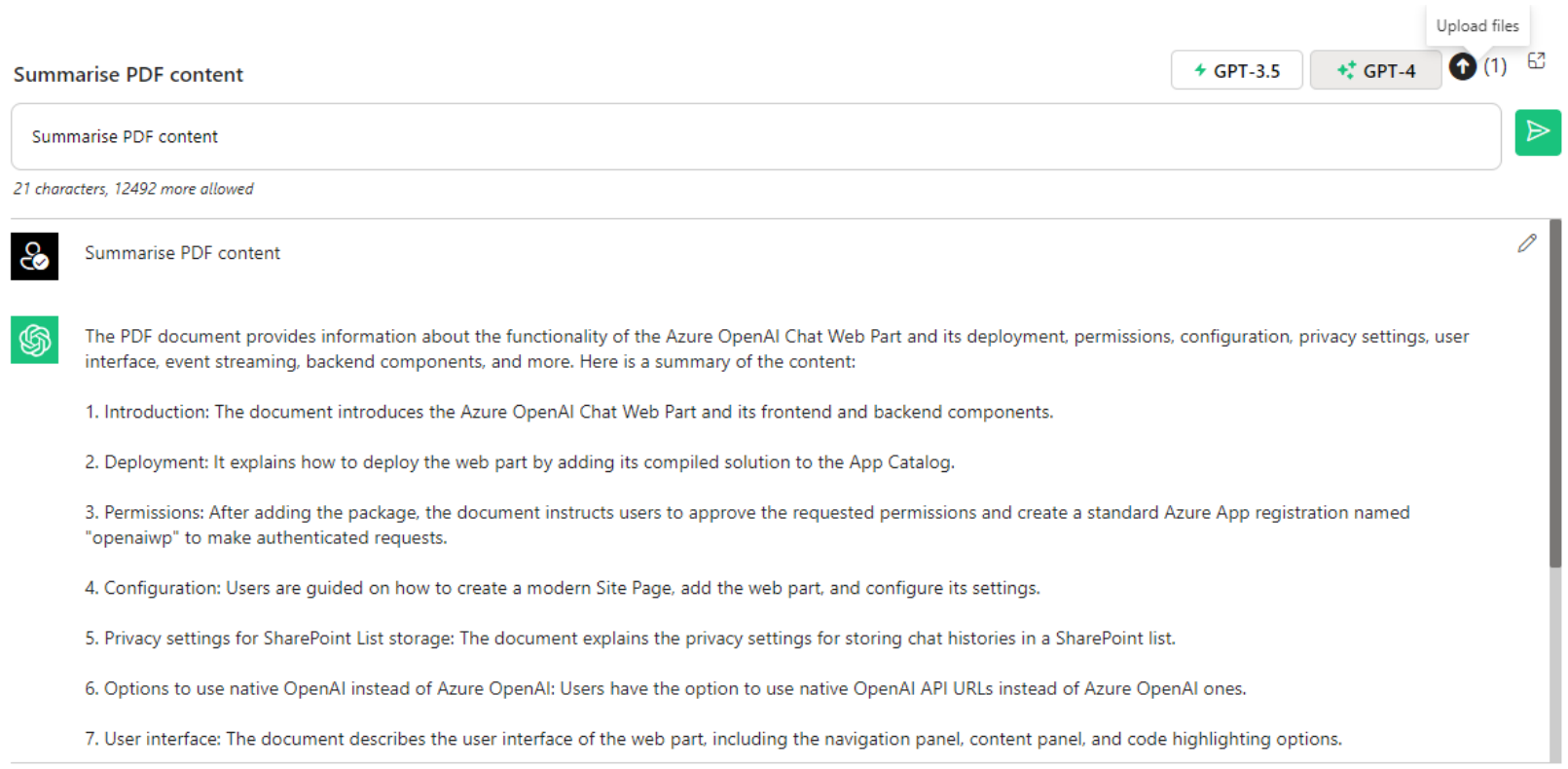
Here are the company employees with names that start with the letter "H":

- Harri Niemeläinen
- Henrietta Mueller

Date and time are...

The current date is November 10, 2023, and the current time is 12:02:22. If you have any more questions or need further assistance, feel free to ask!

Upload and Summarize PDF content



Optional integrations: Search on the Internet with the summary for Bing and Google results

Search on Bing and Google: Open AI vs Gemini. Show 5 most recent results from each.

GPT-3.5 | GPT-4 | GPT-4 Turbo

Send a message

0 characters, 15000 more allowed

Search on Bing and Google: Open AI vs Gemini. Show 5 most recent results from each.

Here are the 5 most recent search results from Bing:

1. Generative AI is off to a rough start - InfoWorld: As the big vendors overstate the capabilities of their generative AI projects, maybe it's time to use open source to keep them honest. Published on 2023-12-11.

2. Google finally joins the gen AI race in earnest, and now it's really wide open - SiliconANGLE: With its new AI model Gemini, Google this week cracked open the generative AI race even wider — as we said would happen following that pre-Thanksgiving OpenAI debacle. Published on 2023-12-09.

3. Google may use Gemini AI to tell users their life story through phone data, photos - Indiatimes: Google plans to use artificial intelligence technology, including the Gemini AI model and large language models (LLMs), such as those used in Project. Published on 2023-12-11.

4. Open-source generative AI startup Mistral AI raises $415M in funding - SiliconANGLE: French generative artificial intelligence startup Mistral AI announced today it has closed on its second bumper funding round of the year, raising 385 million euros, or around $415 million. Published on 2023-12-11.

5. Google Launches New Multi-modal Gemini AI Model - InfoQ: Alphabet released the first phase of its next-generation AI model, Gemini. Gemini is the first model to outperform human experts on MMLU (Massive Multitask Language Understanding). Published on 2023-12-11.

And here are the 5 most recent search results from Google:

1. Gemini live today in Bard - Community - OpenAI Developer Forum: Discussion thread on the OpenAI Developer Forum about the launch of Gemini.

2. OpenAI's GPT-Vision vs. Google's Gemini: What To Expect - EnCord: A comparison between OpenAI's GPT-Vision and Google's Gemini AI models.

3. Gemini live today in Bard - Community - OpenAI Developer Forum: Continued discussion on the OpenAI Developer Forum about the launch of Gemini.

4. Google Nears Release of Gemini AI to Challenge OpenAI — The Information: Article from The Information discussing Google's upcoming release of Gemini AI to compete with OpenAI.

5. Google's Gemini is so powerful that OpenAI probably won't release - Reddit: A Reddit thread discussing the power of Google's Gemini AI model and its potential impact on OpenAI.

Image Generations: Dalle-3

Create an image: an illustration of a business meeting

GPT-3.5 GPT-4 GPT-4 Turbo

Send a message

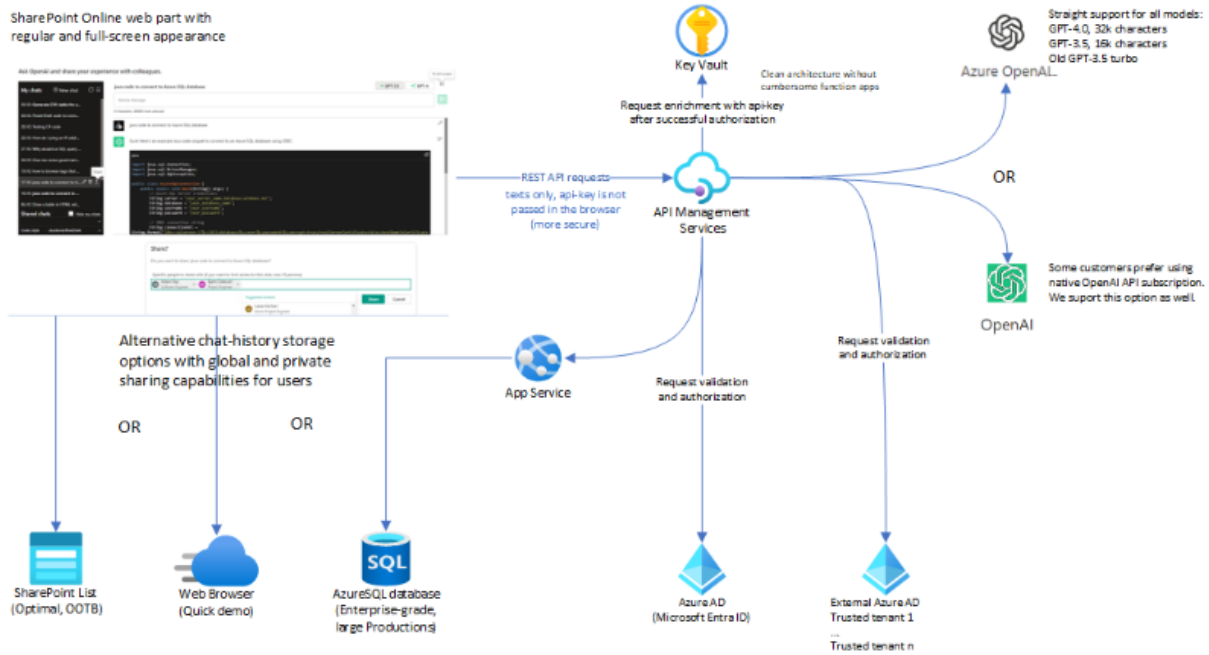
0 characters, 15000 more allowed

The images depict illustrated scenes of professional business environments.

Using GPT-4 Vision for Image Analysis: a Technical Diagram

What is in this image?

What is in this image?



The image appears to be a flowchart or architectural diagram relating to a software or web application. It includes various components and services that interact with each other to support the application's functionality. Here are the components and descriptions visible in the diagram:

1. **SharePoint Online web part**: The image at the top left shows a SharePoint Online web part that can be displayed with both regular and full-screen appearance. This is likely a user interface component for displaying content.

2. **API Management Services**: Positioned in the center-right of the image, this service seems to handle REST API requests, request validation, and authorization, acting as a gateway between clients and backend services.

3. **App Service**: A central component on the left side of the diagram, identified with the Azure logo, which suggests that it's an Azure App Service used to host web applications.

4. **Key Vault**: Shown linking to the API Management Services and marked with text indicating that it enriches requests with an API key after successful authorization, which implies security and configurations management.

5. **OpenAI**: Mentioned on the right side of the image with a note that some customers prefer using their OpenAI subscription, indicating integration with AI services provided by OpenAI.

6. **Azure AD (Microsoft Entral ID)** and **External Azure AD**: Components for identity and access management.

7. **AzureSQL database (Enterprise-grade, large Productions)**: A cloud-based database service indicated for production-grade, larger-scale usage.

8. **Storage options**: At the bottom left, there are three alternatives presented for chat-history storage: SharePoint List, Web Browser (Quick demo), and AzureSQL database.

The image also contains annotations related to support for different GPT models, such as GPT-4.0, and the mention of secure transmission of API keys. Overall, the diagram seems to describe a web application architecture that leverages Azure cloud services and offers flexibility in terms of AI integration and user history storage.

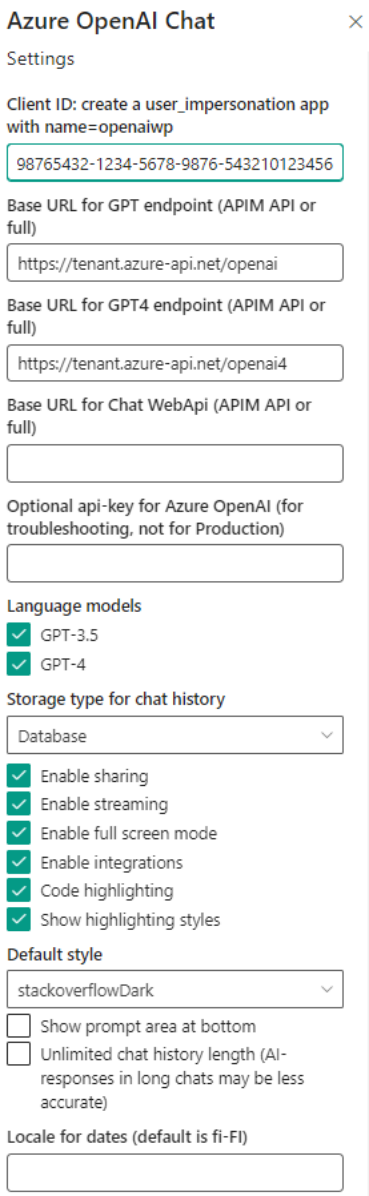
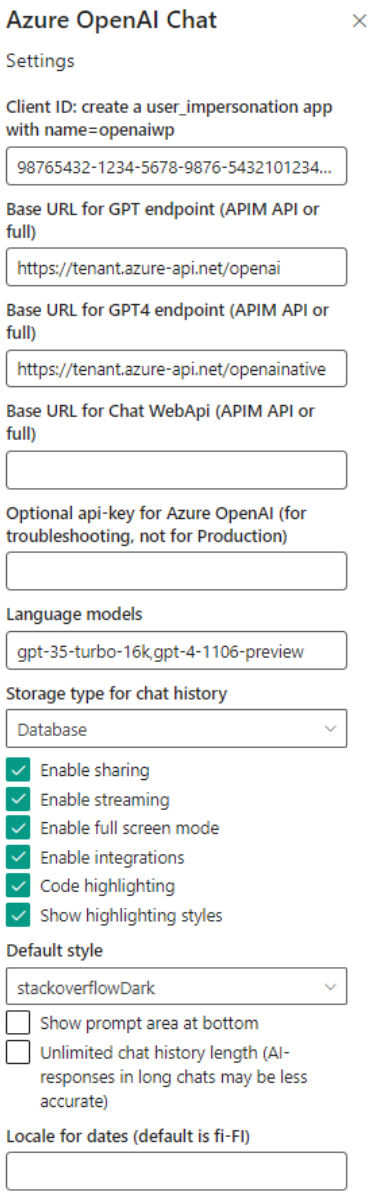
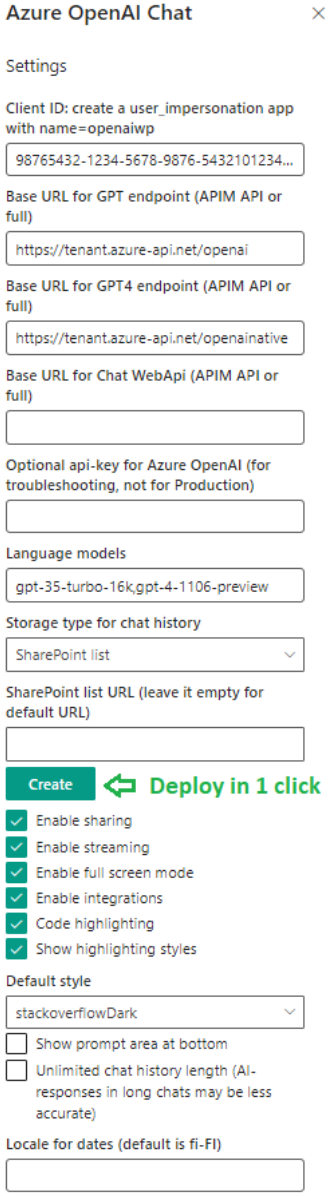
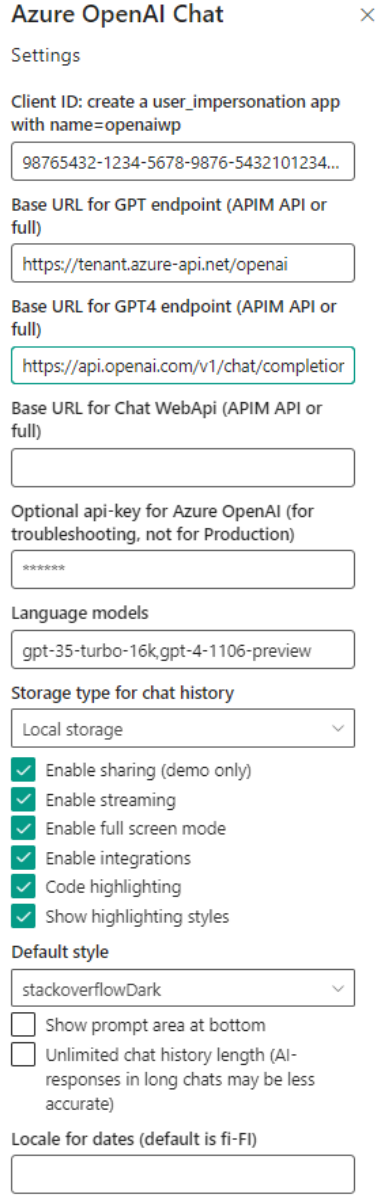
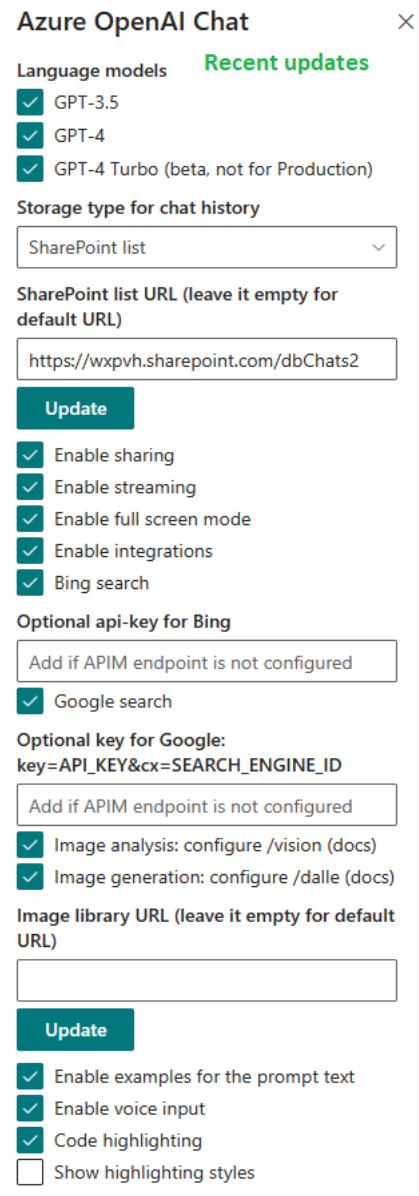
Using GPT-4 Vision for Image Analysis: a Photo

What is in this image? Can you recognize this place?

GPT-3.5 GPT-4

Send a message

0 characters, 15000 more allowed

This image features the 25 de Abril Bridge in Lisbon, Portugal. You can recognize it by its distinctive red color, similar in style to the Golden Gate Bridge in San Francisco, USA. The bridge spans the Tagus River, and in this image, you can also see a dry, grassy foreground that gives a golden hue to the lower part of the photograph. Additionally, there are two sailboats visible on the water, providing a sense of leisure and maritime activity. The sky is clear, which suggests that the photo was taken on a day with good weather.

Azure OpenAI Chat web part for SharePoint Online, configurable options