

Guohe WiseCore

Enterprise Assistant Solution

Improving user efficiency is an important goal of the AI Agent

User pain points

The work efficiency is low, and it is inconvenient to use AI.

- Don't have an entrance to use the AI function.
- Lack of access to knowledge.
- Inefficient information retrieval.
- Lack of intelligent support.

Business scenarios

Intelligent Q&A, which extracts relevant information from documents to generate answers.

- Enterprise employees want fast real-time intelligent chat with gpt.
- Employees want to find information about company policies quickly.
- Employees want to quickly summarize the content of the document and extract the key points.
- Employees want to quickly generate BI views to quickly analyze data.
- Customer support staff who want to be able to quickly answer common customer questions and improve their productivity.

System function scenarios

Chat With GPT

Ask questions in natural language to get the knowledge you need.

Chat with private documents

Quickly get accurate, relevant information from your internal data.

GPT With BI

Quickly generate BI views so that users can quickly analyze data.

Document summary

Quickly summarize the content of the document, so that users can understand and grasp the core content and key points of the document faster.

Combine large models with private documents

With GPT based on your own data, you can run supported chat models based on your own data without training or fine-tuning the model. By running models with your own data, you can chat based on your own data and analyze your data more accurately and quickly.

Data customization

- Using your own data, you can provide GPT models with domain- or industry-specific data, making it more suitable for your business needs.
- Deliver more accurate results. This helps reduce errors and improve the performance of your model.
- You can use your own data to optimize your model to answer a specific domain or business question. This allows you to create a highly specialized Q&A system that can better meet the needs of your users or customers.

Data security and real-time

- With absolute control over your data, you can access, modify, or scale your dataset as often as you need it, without relying on external data providers.
- You have more control over your data to ensure that sensitive information doesn't leave your internal network. This helps protect your data privacy and compliance, and reduces the risk of data breaches.
- Your data can be updated as often as needed to ensure that your model is always up-to-date and adaptable to changing conditions.

Diversity and professionalism of the answers

- Analyze data more accurately and quickly.
- Based on the most recent information available in the specified data source.
- Create a highly specialized Q&A system.
- Responses can be not only text, but also BI charts.

Flexibility in management

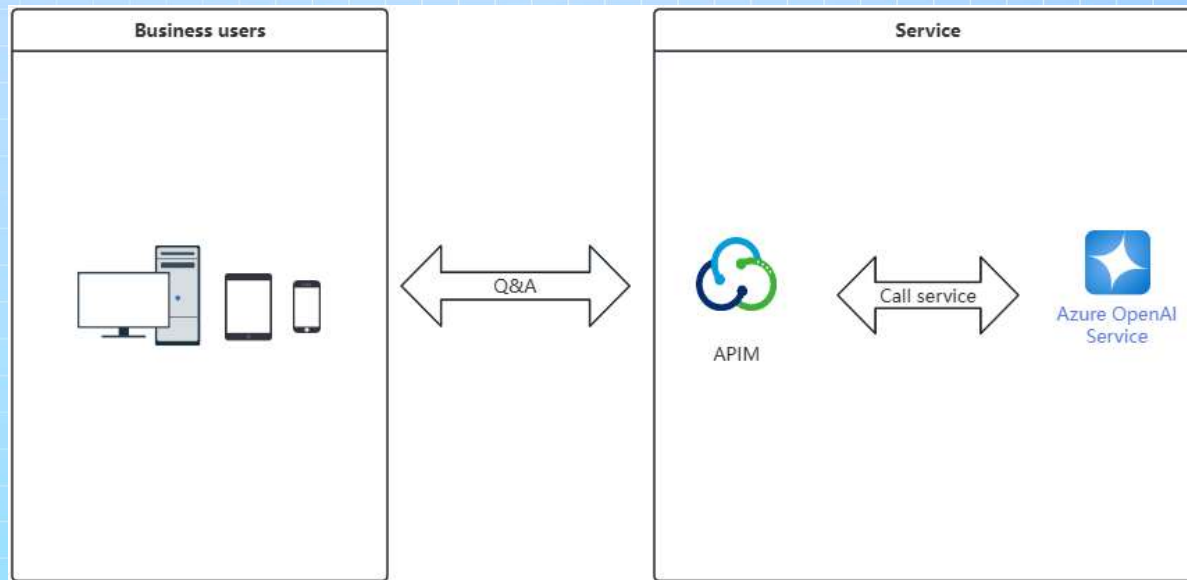
- Centralized OpenAI service management and monitoring.
- Dynamic request distribution optimizes OpenAI quota usage.
- Integrate AAD authentication and access control.
- Real-time user/department usage analytics for easy cost allocation.

Solution introduction

Solution 1: Chat With GPT

Chat With GPT

Standardized AI work assistant to improve the work & productivity of enterprise employees.



Contrast

Access is difficult

On-premise, low latency

Request Delay or Queued

Enterprise-grade services

The token limit is low

Enterprise-level token management

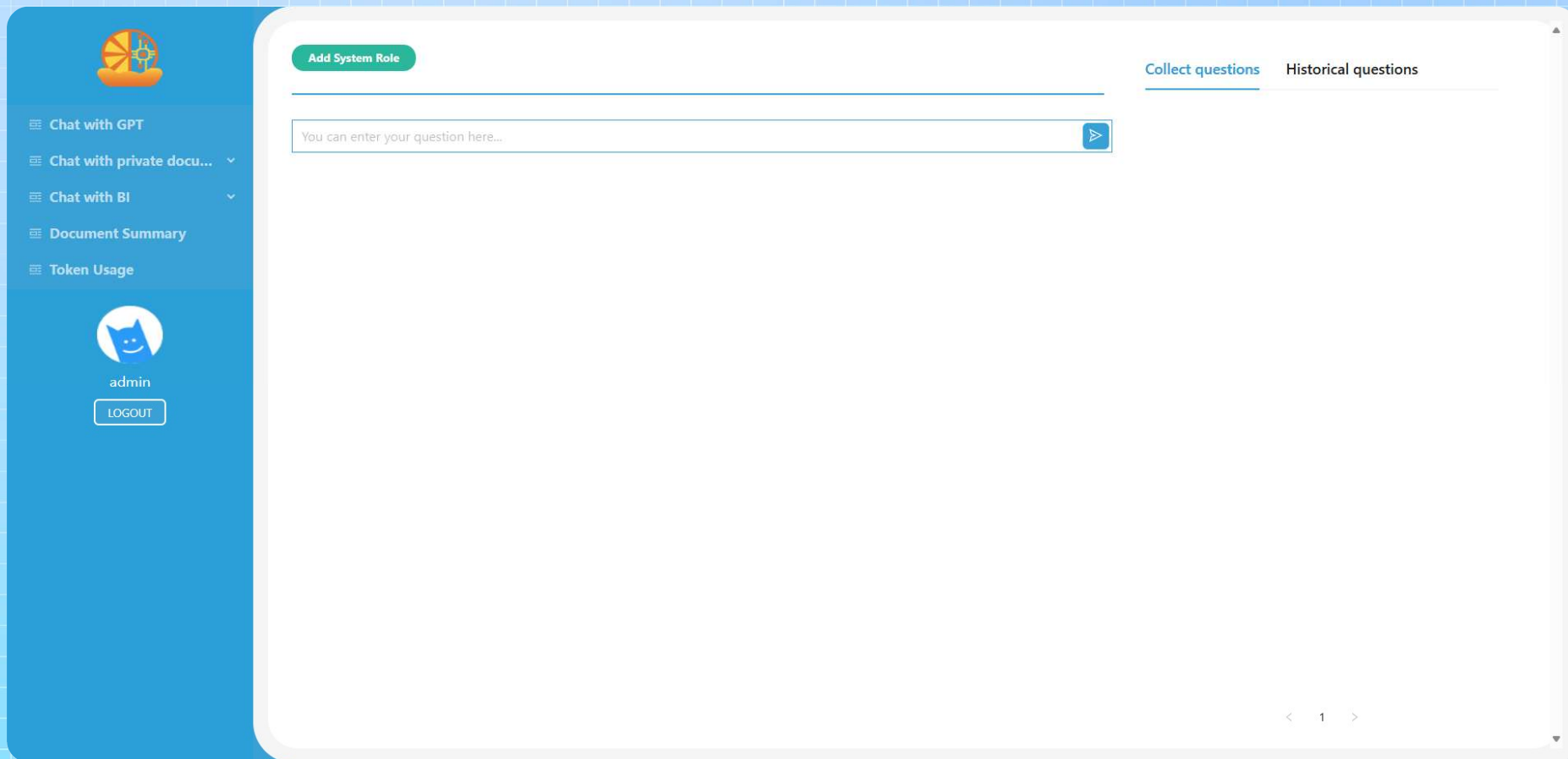
Difficult to manage

Professional operation and maintenance team

Solution 1: Chat With GPT

Chat With GPT


Online intelligent Q&A, ask questions through natural language to obtain the knowledge you need.




Solution 1.1: Document Summary

Document Summary

Customize the criteria to summarize the text content.



- Chat with GPT
- Chat with private docu...
 - Embedding Chat
 - Embedding Queries
 - Document Management
- Chat with BI
- Document Summary**
- Token Usage



admin

LOGOUT

Summarization

Select a type of summarization

☒ Basic Summary

☐ Bullet Points

☐ Custom

Enter some text to summarize

A neutron star is the collapsed core of a massive supergiant star, which had a total mass of between 10 and 25 solar masses, possibly more if the star was especially metal-rich.[1] Neutron stars are the smallest and densest stellar objects, excluding black holes and hypothetical white holes, quark stars, and strange stars.[2] Neutron stars have a radius on the order of 10 kilometres (6.2 mi) and a mass of about 1.4 solar masses.[3] They result from the supernova explosion of a massive star, combined with gravitational collapse, that compresses the core past white dwarf star density to that of atomic nuclei.

Summarize

Summary result

A neutron star is the dense, compact core left after a massive supergiant star (10–25+ solar masses) undergoes a supernova explosion and gravitational collapse. These small stellar objects, about 10 kilometers in radius and 1.4 solar masses, are among the densest in the universe, surpassed only by black holes and certain hypothetical stars.

Input

- News
- Competitor announcements
- Market Policy Report
- . . .



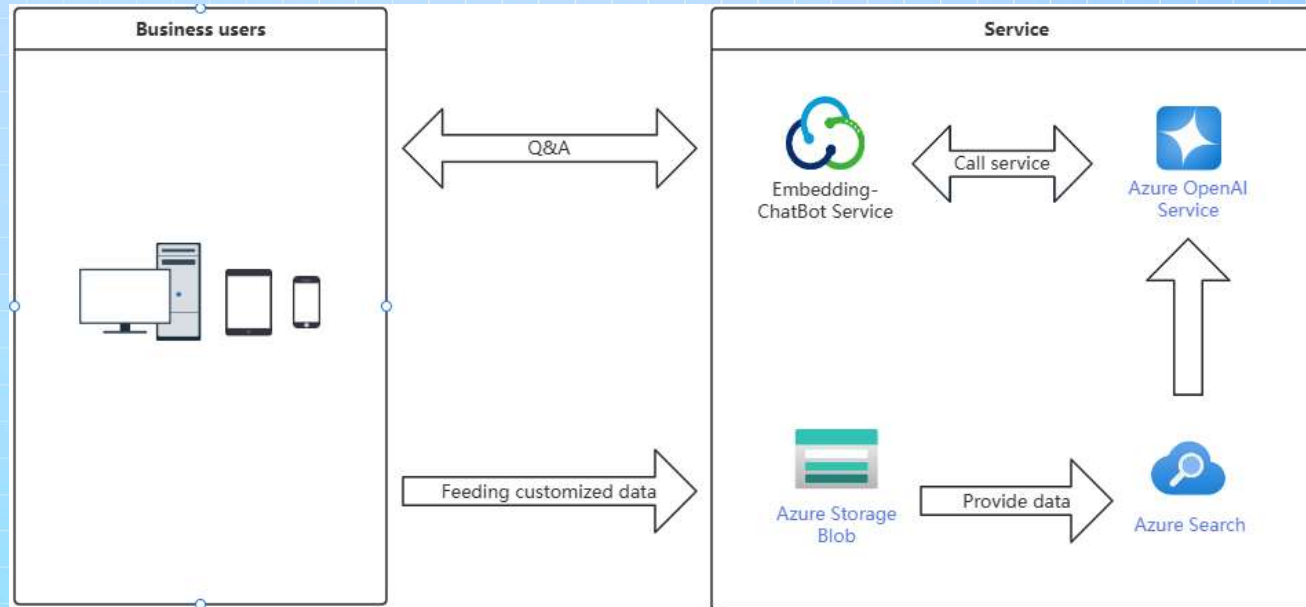
Output

- Summary
- Refinement of the main points
- Customized output
- . . .

Solution 2: Chat with private documents

Embedding Chat

Relying on the ability of large models, it provides Q&A services based on internal data of enterprises. It can be deployed independently to ensure data security.



Merit

- ① Q&A is highly customizable and can answer user questions based on the information provided.
- ② PaaS services are highly available.
- ③ Azure services are highly integrated to ensure data security.
- ④ You don't need to know the technical documentation for each Azure product.

Chat with private documents - Embedding GPT

Embedding GPT

Quickly get accurate, relevant information from your internal data.

The screenshot displays the Embedding GPT web application interface. On the left is a blue sidebar with a top icon of a gear and a document. Below it are menu items: 'Chat with GPT', 'Chat with private docu...' (with a dropdown arrow), 'Embedding Chat', 'Embedding Queries', 'Document Management', 'Chat with BI' (with a dropdown arrow), 'Document Summary', and 'Token Usage'. At the bottom of the sidebar is a user profile section with a blue circular avatar icon, the name 'admin', and a 'LOGOUT' button. The main content area has a white background. At the top left of this area is a 'Library:' dropdown menu showing 'admin/em/Dous'. At the top right are two tabs: 'Collect questions' (active) and 'Historical questions'. Below the tabs is a question 'Q How to write a good composition' with a 'Collect' button. Below the question is an answer 'A' containing a paragraph and a numbered list of 8 steps for writing a good composition. The steps are: 1. Understanding the Prompt, 2. Planning and Organizing Ideas, 3. Crafting a Strong Introduction, 4. Developing Body Paragraphs, 5. Using Effective Techniques, 6. Concluding Thoughtfully, 7. Revising and Editing, and 8. Seeking Feedback. Below the list is a concluding sentence. At the bottom left of the main area is a timestamp '11:50 AM'. At the bottom center is a red 'Clear Chat' button. At the bottom is a text input field with the placeholder 'You can enter your question here...' and a blue send button. At the bottom right are navigation arrows and the page number '1'.

Library: admin/em/Dous

Collect questions Historical questions

Q How to write a good composition [Collect](#)

A

To write a good composition, it is essential to follow a structured approach that includes the following key elements:

1. **Understanding the Prompt:** Analyze the topic or question you need to address. Make sure you fully understand what is being asked before you start writing.
2. **Planning and Organizing Ideas:** Spend time brainstorming and organizing your thoughts. Create an outline to structure your composition. This will help you determine the main points you want to convey and the order in which to present them.
3. **Crafting a Strong Introduction:** Start with an engaging opening that captures the reader's attention. You can use techniques like an interesting fact, a question, or a vivid description. Clearly state your thesis or the main idea of your composition.
4. **Developing Body Paragraphs:** Each paragraph should focus on a single idea that supports your thesis. Start with a topic sentence, provide evidence or examples, and elaborate on your points. Use transitions between paragraphs to maintain the flow of your writing.
5. **Using Effective Techniques:** Incorporate literary techniques and stylistic devices to enhance your writing. This includes using sensory details, metaphors, and varying sentence structures to make your composition more engaging.
6. **Concluding Thoughtfully:** Summarize the main points of your composition in the conclusion and restate the significance of your thesis. Leave the reader with a final thought or call to action.
7. **Revising and Editing:** After completing your first draft, take time to revise your work. Check for clarity, coherence, and logical flow. Edit for grammar, punctuation, and spelling errors.
8. **Seeking Feedback:** If possible, get feedback from peers or instructors. They can provide valuable insights and suggest improvements.

By following these steps and practicing regularly, you can improve your writing skills and produce compelling compositions.

11:50 AM

Clear Chat

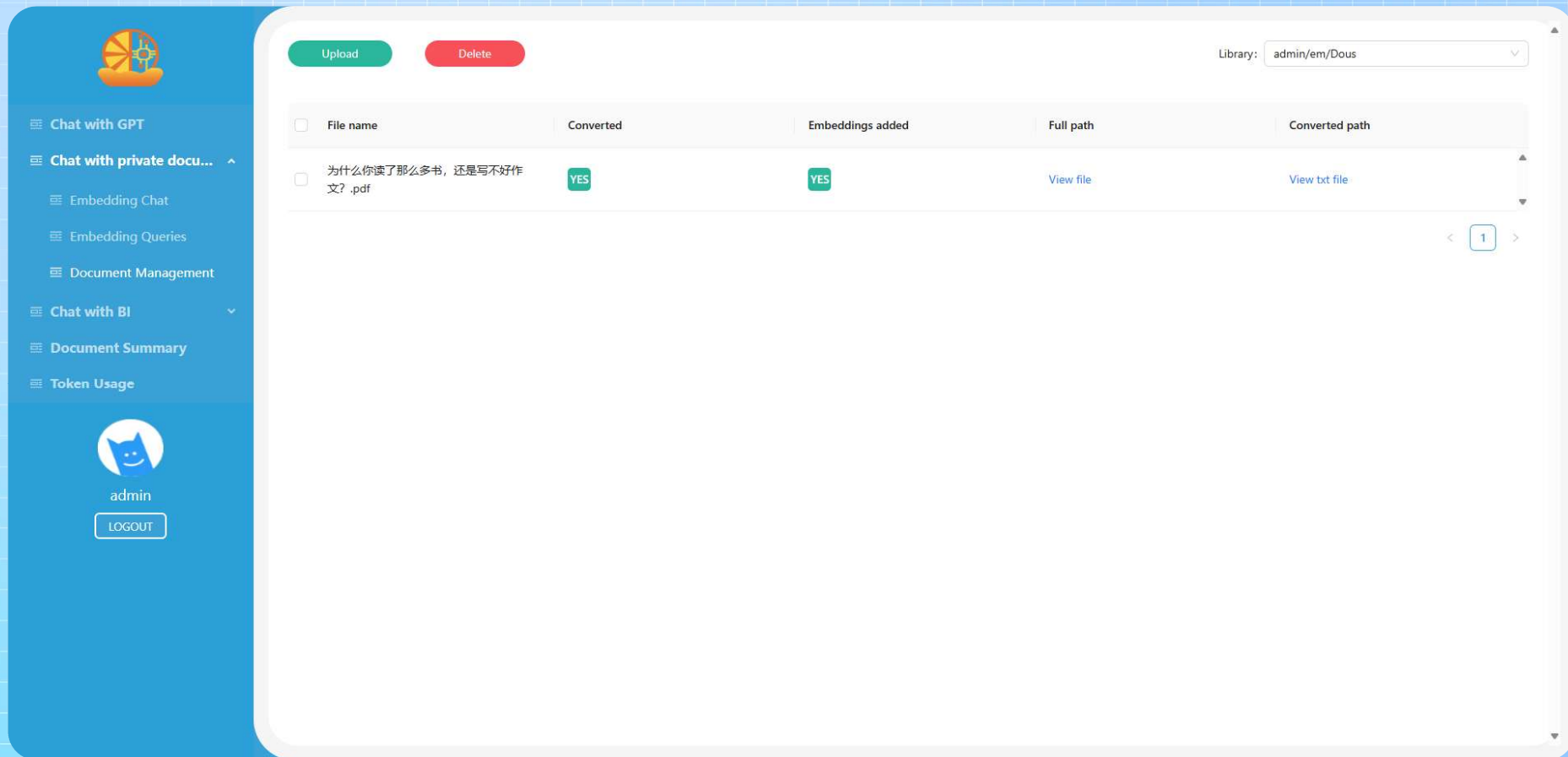
You can enter your question here...

< 1 >

Chat with private documents - Document Management

Document Management

Administrators can flexibly manage internal data and preview files.



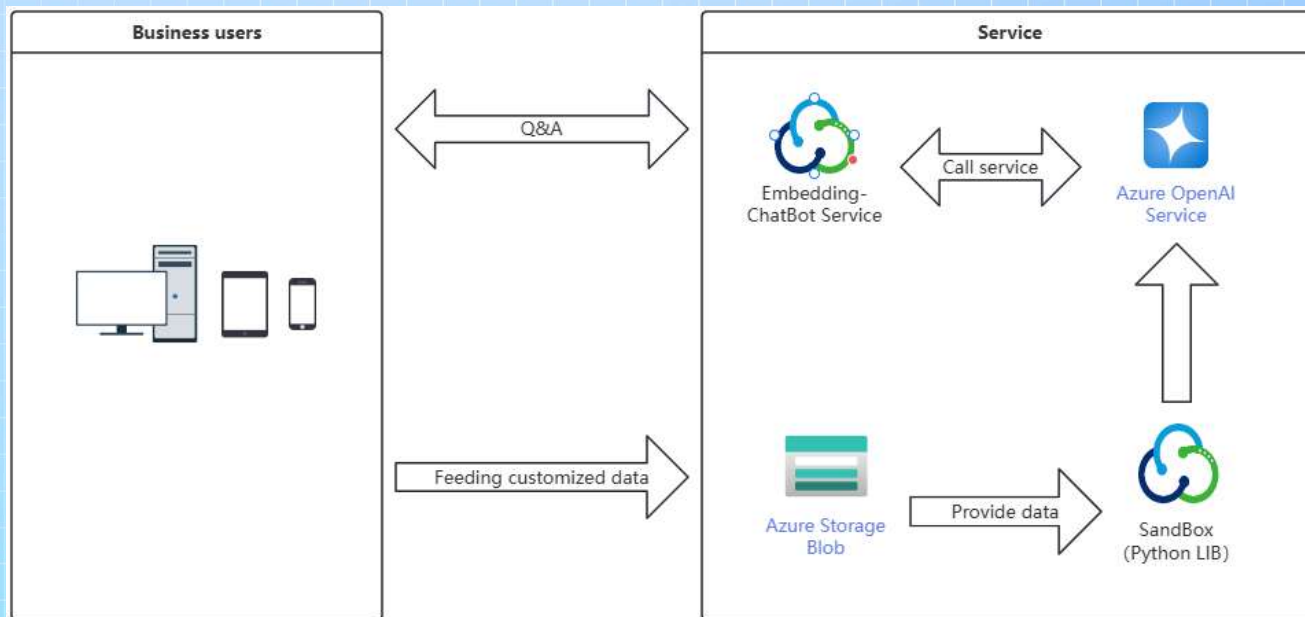
The screenshot displays the Document Management interface. On the left is a blue sidebar with a top section containing a gear icon and a bottom section with a GitHub Octocat icon and the text 'admin' and 'LOGOUT'. The sidebar menu includes: 'Chat with GPT', 'Chat with private documents' (expanded), 'Embedding Chat', 'Embedding Queries', 'Document Management', 'Chat with BI', 'Document Summary', and 'Token Usage'. The main content area has a top bar with 'Upload' and 'Delete' buttons, a 'Library:' dropdown set to 'admin/em/Dous', and a table of documents. The table has columns: 'File name', 'Converted', 'Embeddings added', 'Full path', and 'Converted path'. It contains one row with a checkbox, the file name '为什么你读了那么多书，还是写不好作文? .pdf', 'YES' in green, 'YES' in green, a 'View file' link, and a 'View txt file' link. A pagination bar at the bottom shows '< 1 >'.

<input type="checkbox"/>	File name	Converted	Embeddings added	Full path	Converted path
<input type="checkbox"/>	为什么你读了那么多书，还是写不好作文? .pdf	YES	YES	View file	View txt file

Solution 3: GPT With BI

GPT With BI

Relying on the ability of large models, it provides Q&A services based on internal data of enterprises. It can be deployed independently to ensure data security. Structured data can be processed and charted.




Merit

- ① Q&A is highly customizable and can answer user questions based on the information provided.
- ② It can process structured data and return BI charts, eliminating the tedious data processing process for users.
- ③ PaaS services are highly available.
- ④ Azure services are highly integrated to ensure data security.
- ⑤ You don't need to know the technical documentation for each Azure product.


GPT With BI - BI-GPT Queries

BI-GPT Queries

Quickly generate BI views so that users can quickly analyze data.



- Chat with GPT
- Chat with private docu... ^
- Embedding Chat
- Embedding Queries
- Document Management
- Chat with BI ^
- BI-GPT Queries
- Document Management
- Document Summary
- Token Usage



admin

LOGOUT

文档库: admin/bi/dous_1

Collect questions Historical questions

Q

List the time points of the work plan

★ Collect

A

Time Points

0	2024-12-26
1	2024-12-26
2	2025-01-03
3	2025-01-08
4	2025-01-10
5	2025-01-14
6	2025-01-16
7	2025-01-17
8	2025-01-27
9	2025-02-10
10	2025-02-17

11:32 AM

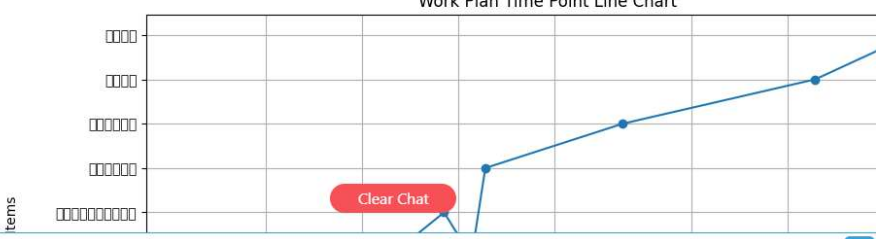
Q

Display the work plan time point line chart

★ Collect

A

Work Plan Time Point Line Chart




You can enter your question here...

1


GPT With BI - Document Management

Document Management

Administrators have the flexibility to manage the files they want to work with structured data and can preview the files.



- Chat with GPT
- Chat with private docu... ^
 - Embedding Chat
 - Embedding Queries
 - Document Management
- Chat with BI ^
 - BI-GPT Queries
 - Document Management
- Document Summary
- Token Usage



admin

LOGOUT

Upload

Delete

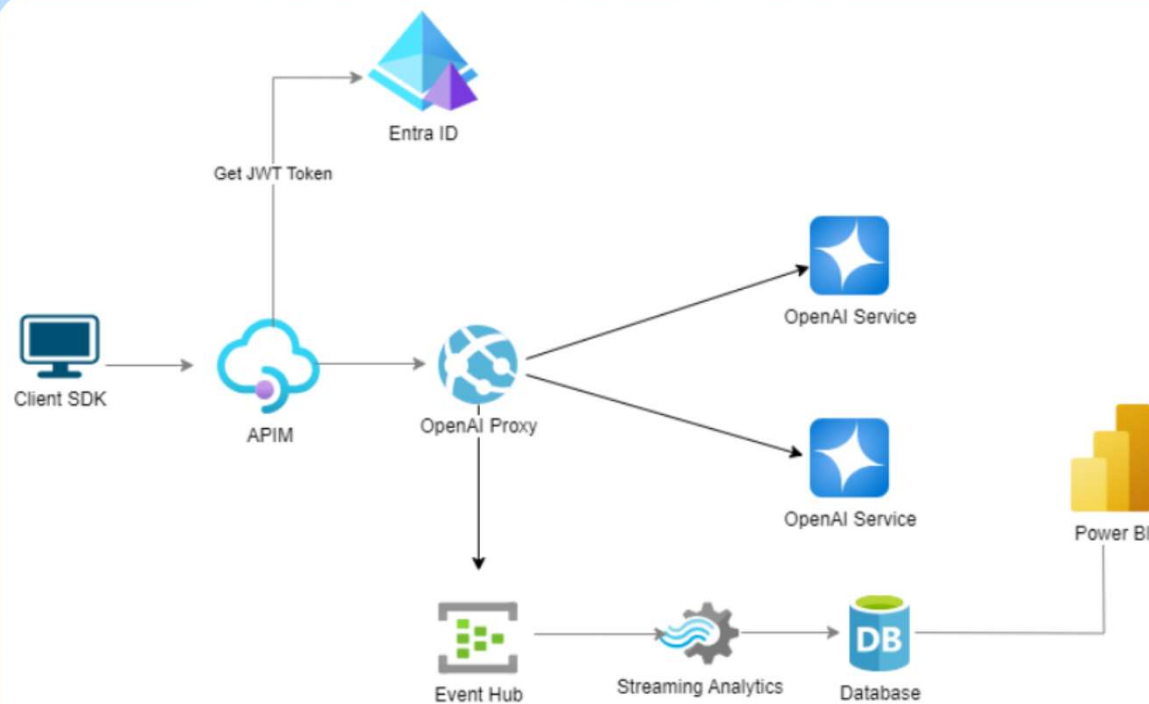
Library: admin/bi/dous_1

<input type="checkbox"/> File name	Full path
<input type="checkbox"/> Work plan.xlsx	View file

< 1 >

Solution 4: Monitor and Management

A centralized monitoring solution for OpenAI services based on Yarp's high-performance gateway, capable of real-time token usage statistics and load balancing.



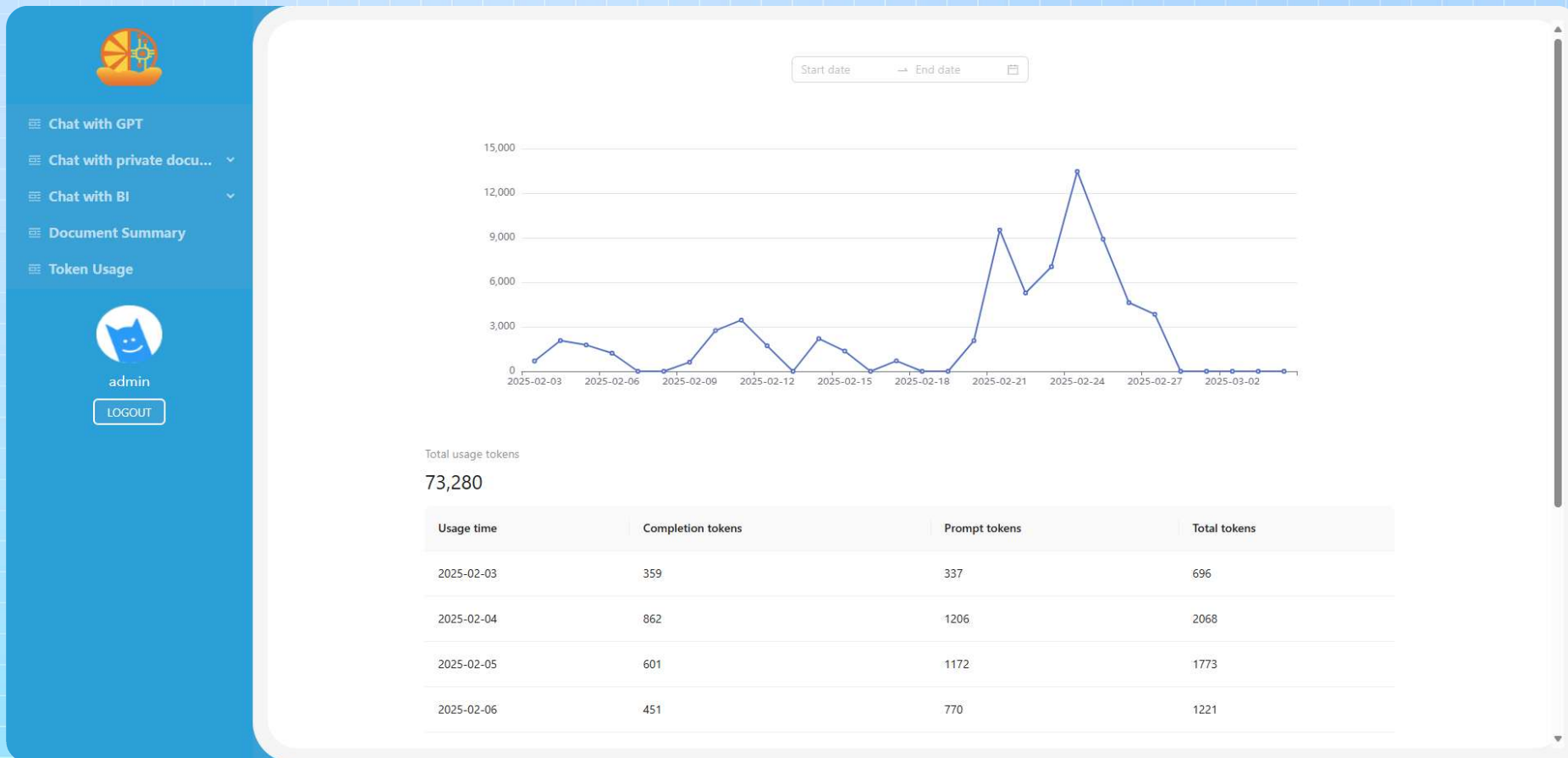
Merit

- Implement centralized OpenAI service management and monitoring.
- Dynamically distribute requests, optimizing OpenAI Quota usage.
- Integrate AAD identity authentication and access control, protecting OpenAI services.
- Real-time analysis of user/department usage for cost allocation.
- Support dynamic flow control of TPM/RPM, enhancing the user experience of PTU/PAYG/Mixed mode.
- Support OpenAI Stream mode fully, improving response speed.
- Support Circuit Breaker mode with Azure APIM.

Monitor and Management

Token Usage

Graphs and tables are displayed intuitively, allowing users to quickly understand the usage.



Quotation of products and services

| Product offers

Resource type	Expenses (RMB)	Billing method	Remark
Products services	200,000.00	One-time fee	Product Service Fees
Basic cloud resources	50,000.00	Year	Basic cloud resources and O&M services
Azure OpenAI service	Billed based on actual usage	Token usage	Reference Links
Azure AI search	Billed based on actual usage	Month	Reference Links