

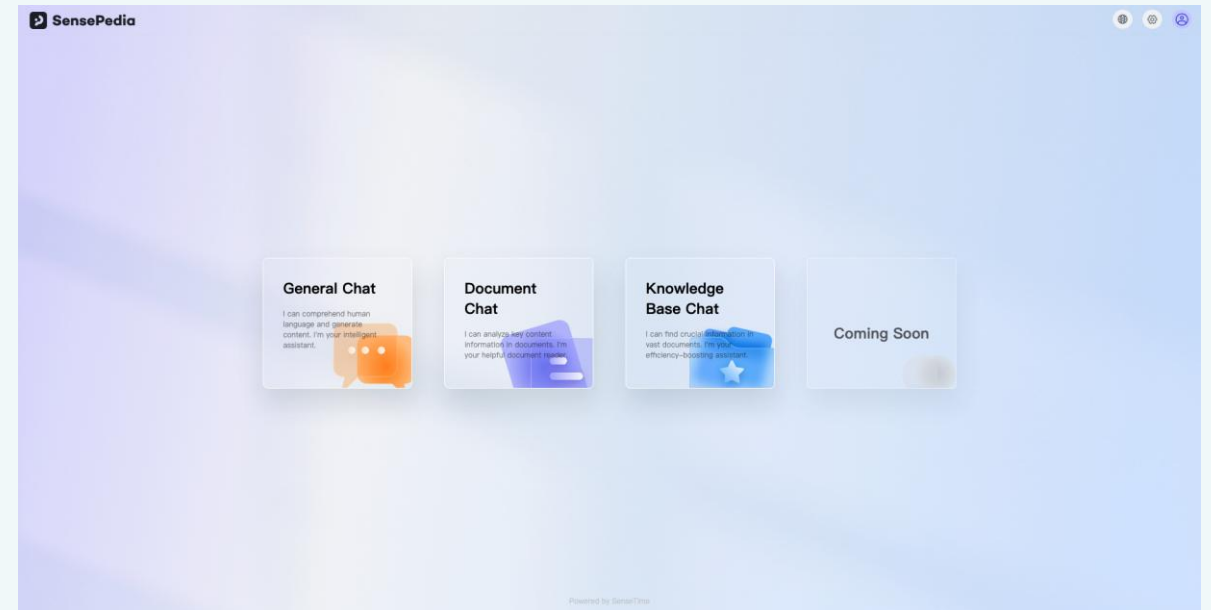
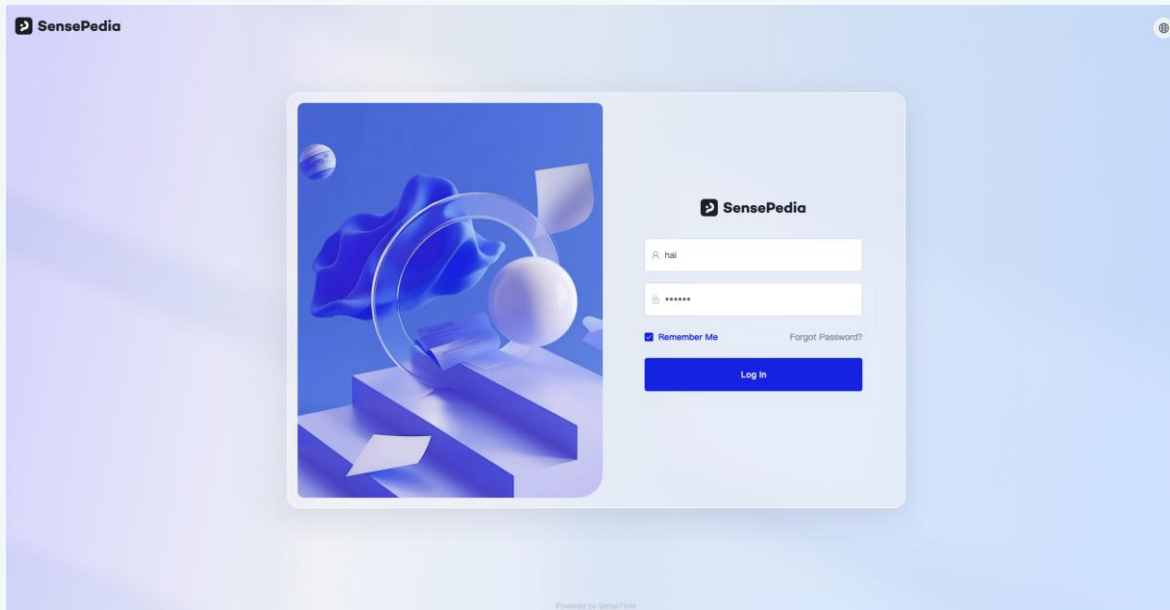
# SENSEPEDIA

## PRODUCT INTRODUCTION

# WHAT IS SENSEPEDIA

**SensePedia is a question-and-answer system based on documents and knowledge bases, capable of quick question answering, document analysis, and reference cross-checking**

**Powered by Large Language Model (LLM) & Retrieval-Augmented Generation (RAG) technology.**





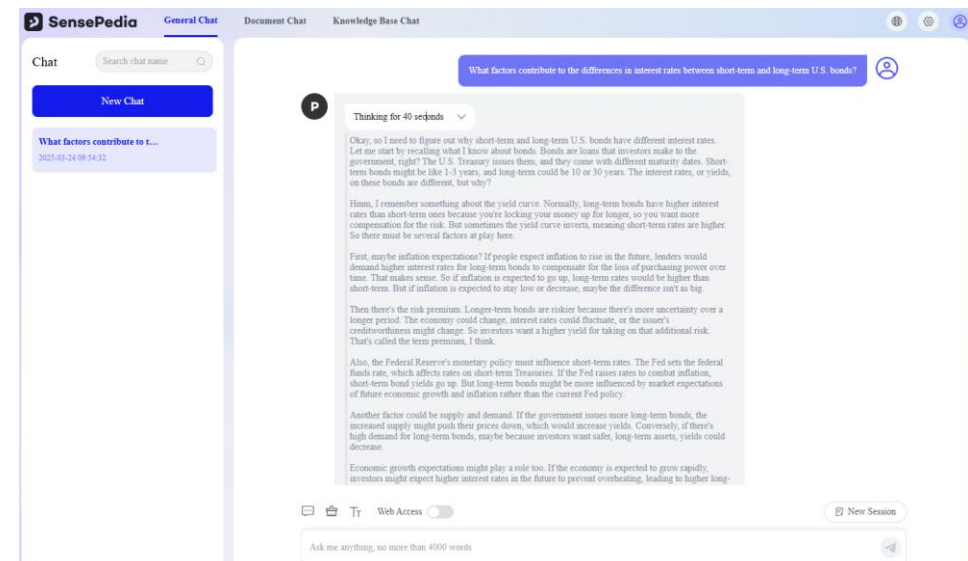
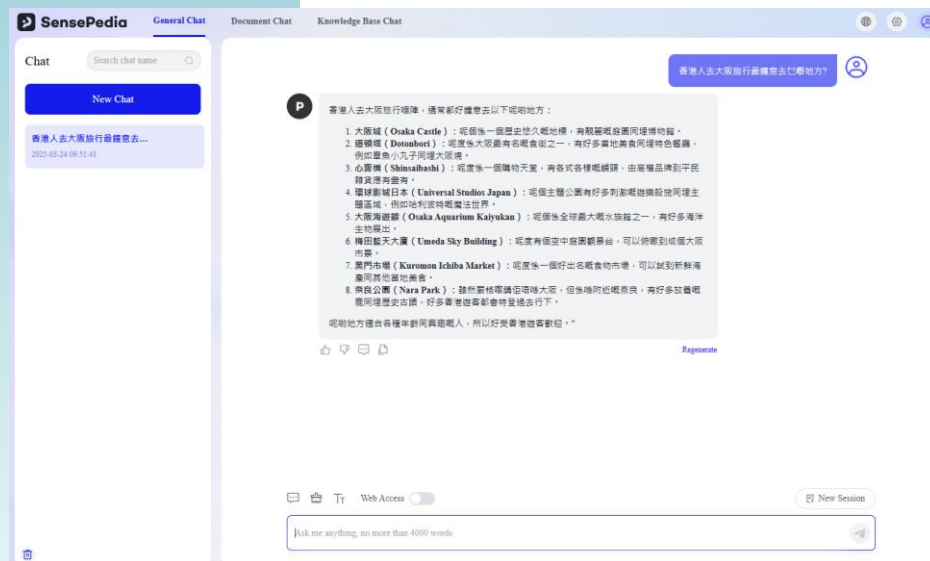
# PRODUCT FEATURE



## General Chat

Engage in conversations on a wide range of topics, including history, science, and more.

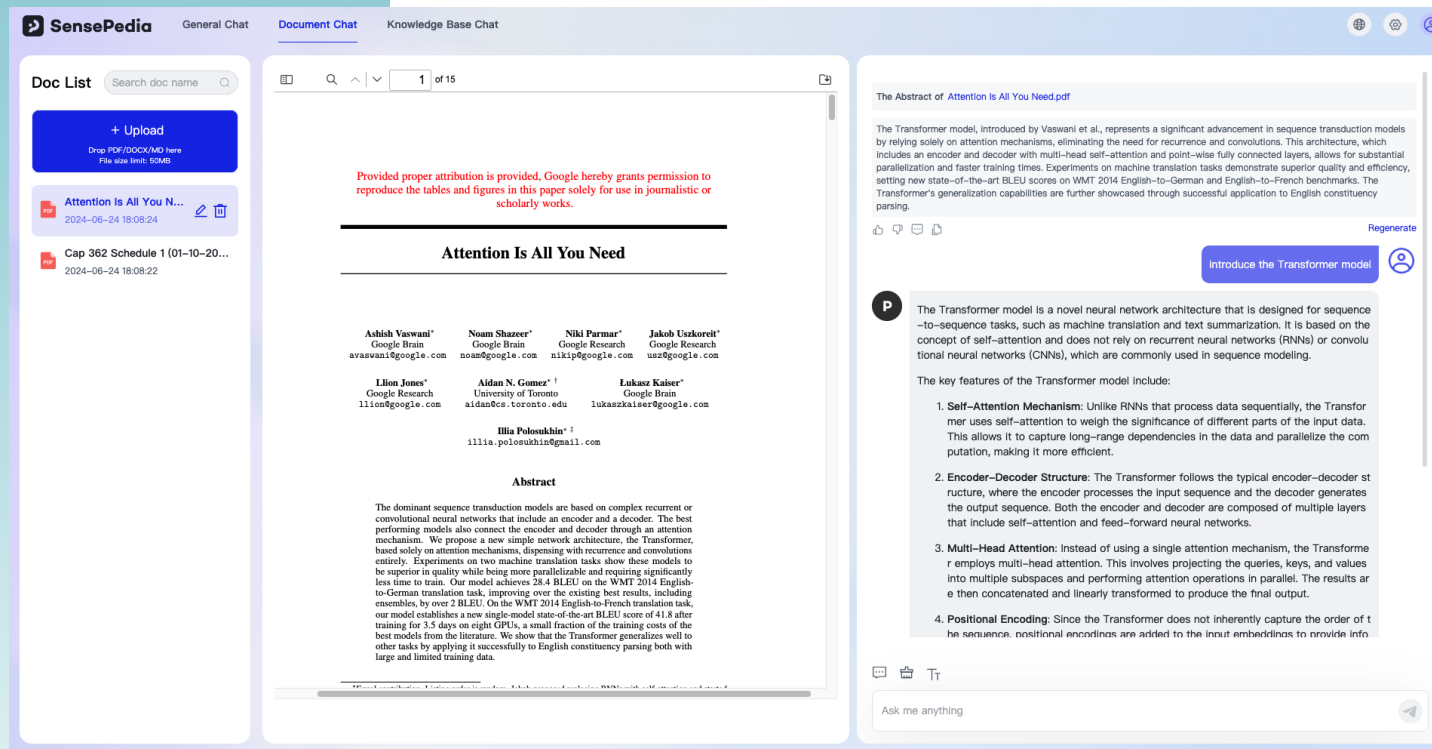
- Continuous Conversations
- Prompt Customization
- **Web search**
- **Supports Cantonese/Traditional Chinese/Simplified Chinese/English Input**
- **Multimodal Capabilities**
- **Chain-of-Thought**



# PRODUCT FEATURE

## Document Chat

Upload document, SensePedia will generate an abstract and allow users to ask questions about the document, providing detailed and relevant answers.



- Support PDF, MD, And DOCX Formats
- Document Preview
- Automatic Abstract Generation
- Document-Based Conversation

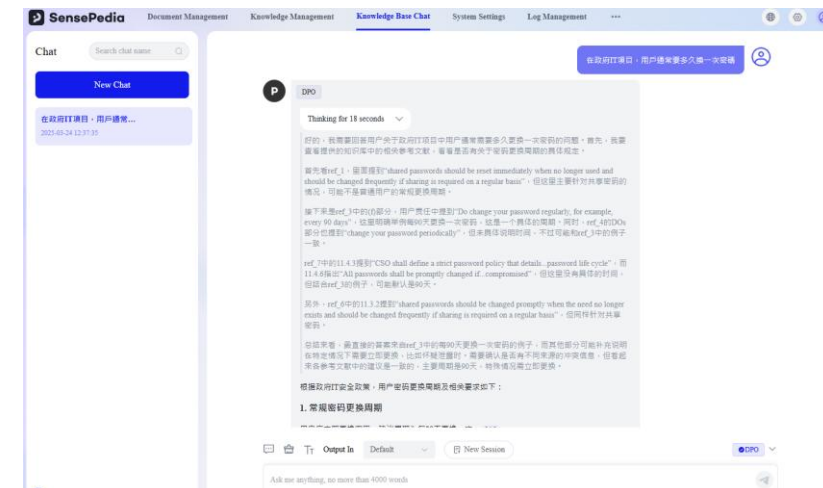
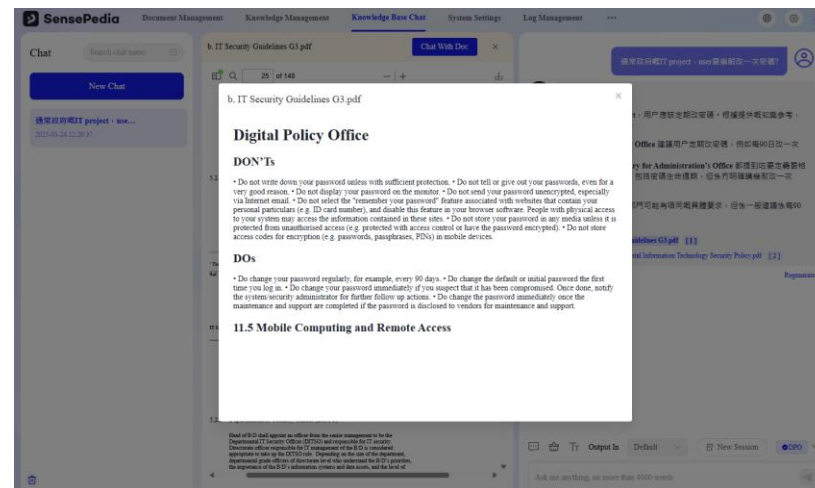
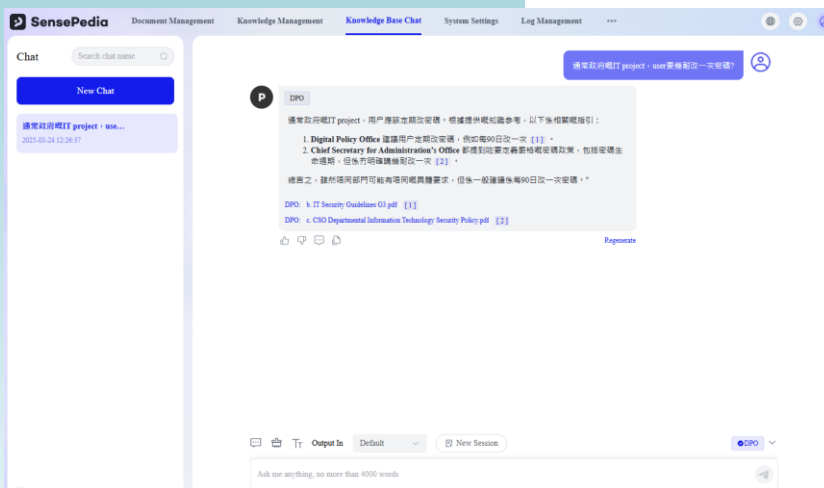
# PRODUCT FEATURE



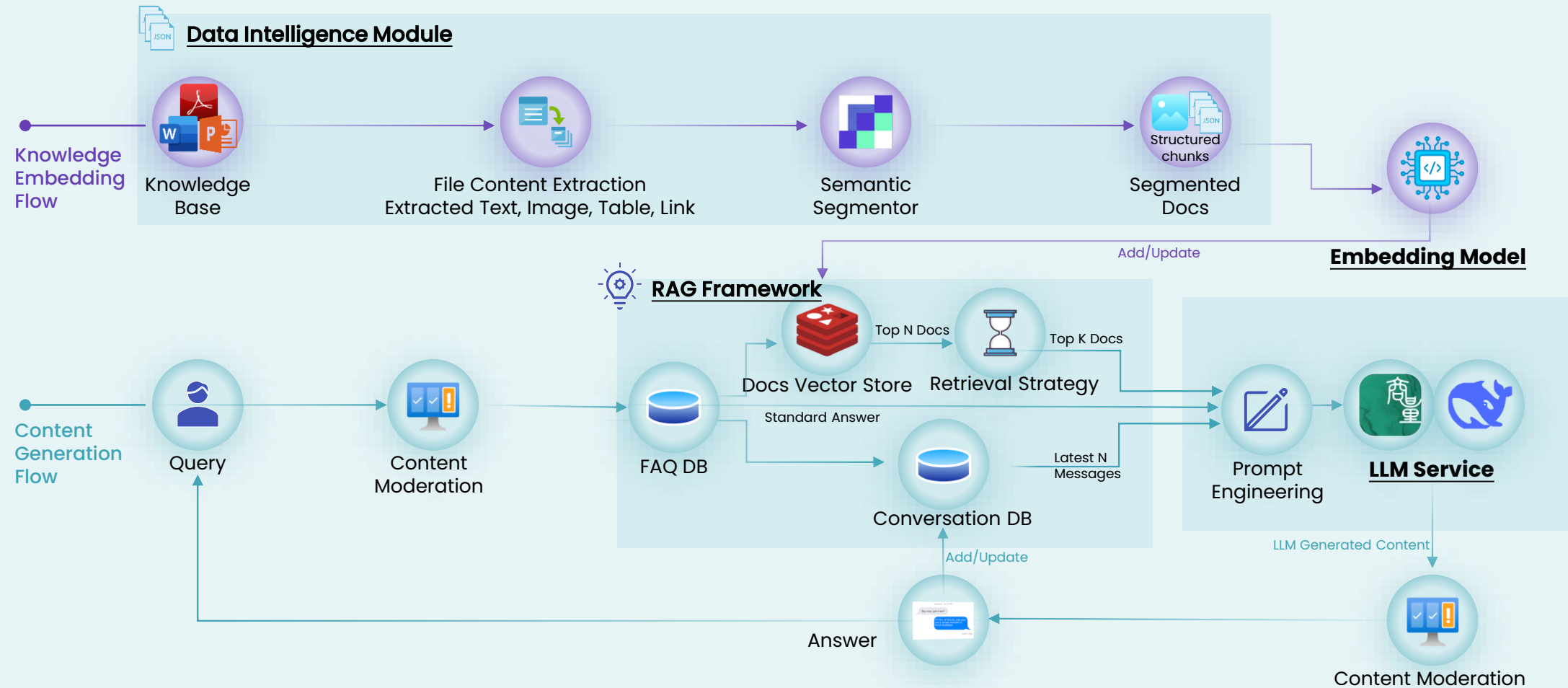
## Knowledge Base Chat

**Q&A based on defined knowledge base, provides answer references with document previews and relevant paragraphs**


- Create Custom Knowledge Base, Including PDF/MD/DOCX Files
- Start Chat By Selecting Knowledge Base
- Reference cross checking
- **Supports Cantonese/Traditional Chinese/Simplified Chinese/English Input**
- **Chain-of-Thought**





# APPLICATION FLOW




## KEY COMPONENTS

 **Retrieval Strategy**  
Role assignment, tone tuning

 **Data Processing Toolkit**  
Efficiently to turn unstructured documents into structured chunks

 **Embedding Model**  
Private knowledge database  
Hallucination elimination

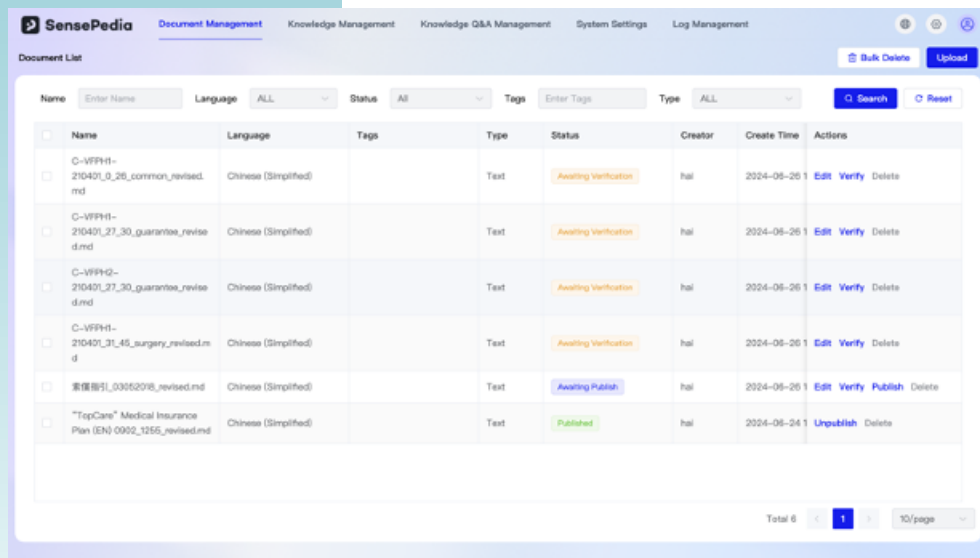
 **Large Language Model**  
Personal customer service  
Sentiment management

# PRODUCT FEATURE

## Create customized knowledge base before utilizing the Knowledge Base Chat

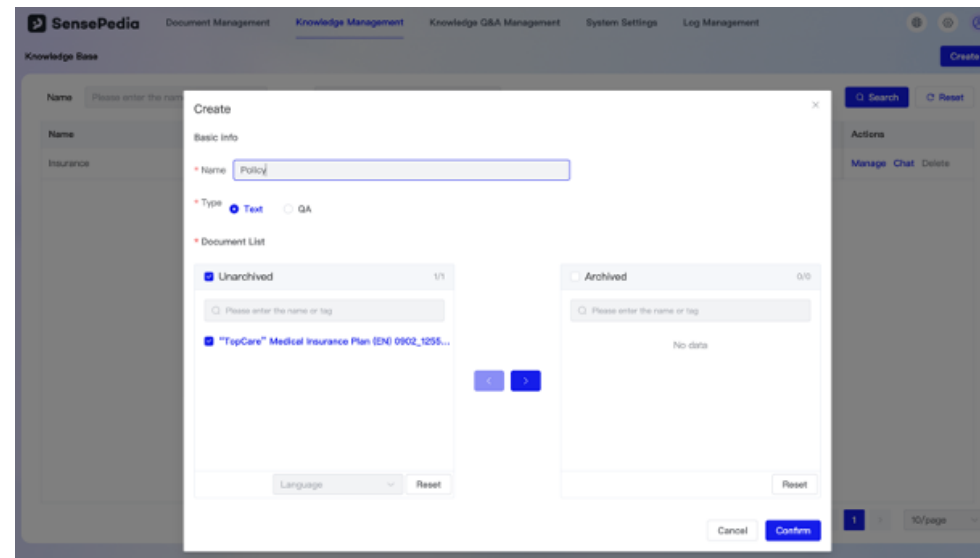
- Support PDF, MD, And DOCX Formats
- Supports automatic document parsing and segmentation with manual verification.
- Supports uploading specified format XLSX files for Questions & Answers list.
- Supports selecting uploaded documents for knowledge base creation.

## Knowledge Base Management



The screenshot shows the 'Document Management' interface of SensePedia. It features a navigation bar with 'Document Management' selected. Below the navigation bar, there are filters for Name, Language, Status, Tags, and Type. A table lists documents with columns for Name, Language, Tags, Type, Status, Creator, Create Time, and Actions. The table contains six rows of documents, mostly in 'Awaiting Verification' status. At the bottom, it shows 'Total 6' documents and a pagination control for '10/page'.

Name	Language	Tags	Type	Status	Creator	Create Time	Actions
C-VFPH-210401_0_26_common_revised.md	Chinese (Simplified)		Text	Awaiting Verification	hai	2024-06-26 1	Edit Verify Delete
C-VFPH-210401_27_30_guarantee_revised.md	Chinese (Simplified)		Text	Awaiting Verification	hai	2024-06-26 1	Edit Verify Delete
C-VFPH-210401_27_30_guarantee_revised.d.md	Chinese (Simplified)		Text	Awaiting Verification	hai	2024-06-26 1	Edit Verify Delete
C-VFPH-210401_31_45_surgery_revised.md	Chinese (Simplified)		Text	Awaiting Verification	hai	2024-06-26 1	Edit Verify Delete
索健指引_03052016_revised.md	Chinese (Simplified)		Text	Awaiting Publish	hai	2024-06-26 1	Edit Verify Publish Delete
"TopCare" Medical Insurance Plan (EN) 0902_1255_revised.md	Chinese (Simplified)		Text	Published	hai	2024-06-24 1	Unpublish Delete



The screenshot shows the 'Knowledge Management' interface of SensePedia. It features a navigation bar with 'Knowledge Management' selected. A 'Create' modal is open, showing a form for creating a new knowledge base. The form includes fields for Name, Type (Text or QA), and Document List. The Document List section shows a list of documents, with one document selected: "TopCare" Medical Insurance Plan (EN) 0902\_1255... The modal also includes a 'Language' dropdown and a 'Reset' button. At the bottom, there are 'Cancel' and 'Confirm' buttons.

Create

Basic Info

Name: Insurance

Type: ☒ Text ☐ QA

Document List

☒ Unarchived 1/1

☐ Archived 0/0

☒ "TopCare" Medical Insurance Plan (EN) 0902\_1255...

Language:

Reset

Cancel Confirm

# PRODUCT FEATURE



## System Config

### User Permissions System

- **Superadmin:** Can create admin accounts, configure system parameters, and set welcome messages.
- **Admin:** Can manage documents and knowledge bases, and create regular user accounts.
- **User:** Can use the knowledge bases created by Admins for Q&A.

The system allows flexible configuration of welcome messages and quick question cards.

SensePedia

Document ManagementKnowledge ManagementKnowledge Q&A ManagementSystem SettingsLog Management

User Management

User List

NamePlease enter namePhonePlease enter numberRoleALLStatusALLSearchReset

ID	Login Account	User Name	Phone Number	Role	Status	Create Time	Actions
1	admin	SuperAdmin	15822222	超级管理员	Enable	2023-04-04 16:49:07	Edit Delete
2	paky	测试	1566666666	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
3	jiewei	测试	1780909876	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
4	baojin	保安	1588888888	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
5	dengxiang	邓祥	1566666666	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
6	huanjie	欢	1780909	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
7	hai	小海	1588888888	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
8	ding	小丁	1566666666	普通管理员	Enable	2023-04-04 16:49:07	Edit Delete
9	CSAtest			普通管理员	Enable	2024-07-10 19:39:18	Edit Delete
10	CSAtest1			操作员	Enable	2024-07-11 16:41:35	Edit Delete

Total 151210/page

SensePedia

Document ManagementKnowledge ManagementKnowledge Q&A ManagementSystem SettingsLog Management

User Management

Prompt List

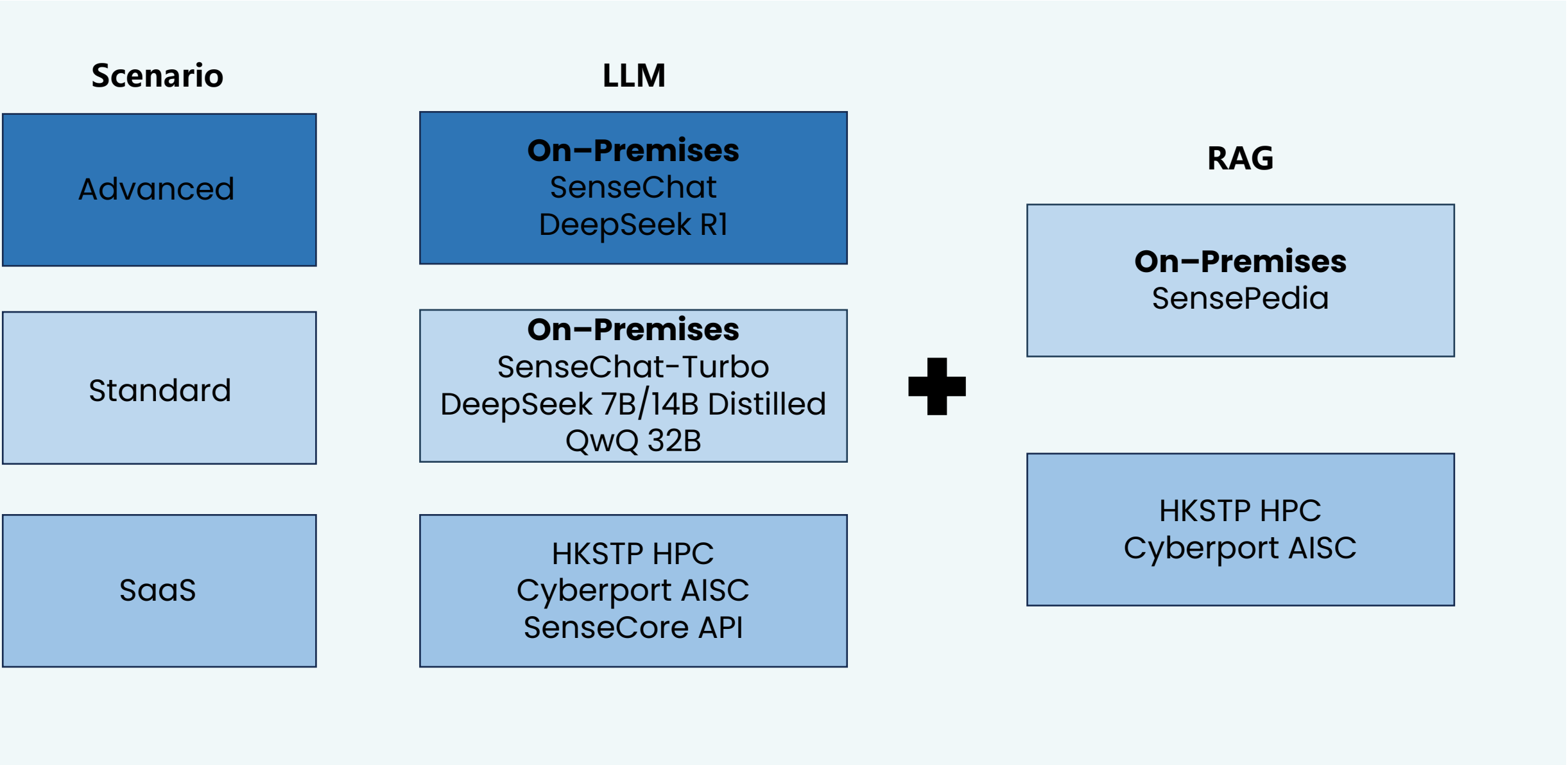
TitlePlease enter titleLanguageEnglishSearchReset

Title	Language	Description	Icon	Prompt	Create Time	Actions
Generate a Story	English	Generate a whimsical story about Whiskers the cat in a magical forest.		Write a short story about an adventurous cat named Whiskers who discovers a hidden magical forest. The story should be whimsical and have a surprise ending. Use line breaks or separate into paragraphs to divide different parts of the story.	2024-04-19 07:47:12	Edit
Write a notice	English	Write a notice about an event with details.		Write a short notice informing the community about an upcoming event. Include the event name, date, time, location, and a brief description. Use line breaks to clearly separate each piece of information.	2024-04-19 07:47:12	Edit
Generate a Code	English	Generate code to implement a simple algorithm.		Write code to implement the bubble sort algorithm and provide an example snippet.	2024-04-19 07:47:12	Edit
Daily report generation	English	Write a brief report summarizing daily tasks and accomplishments.		Write a brief work report summarizing your tasks and accomplishments for the day. Include what tasks were completed, any challenges faced, and plans for the next day.	2024-04-19 07:47:12	Edit

Total 0110/page



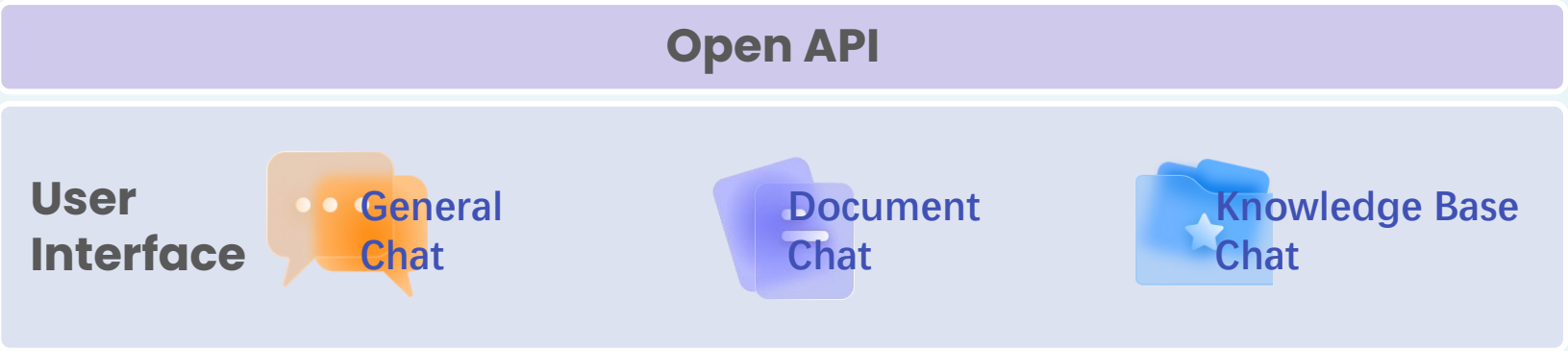
# COMBINATION



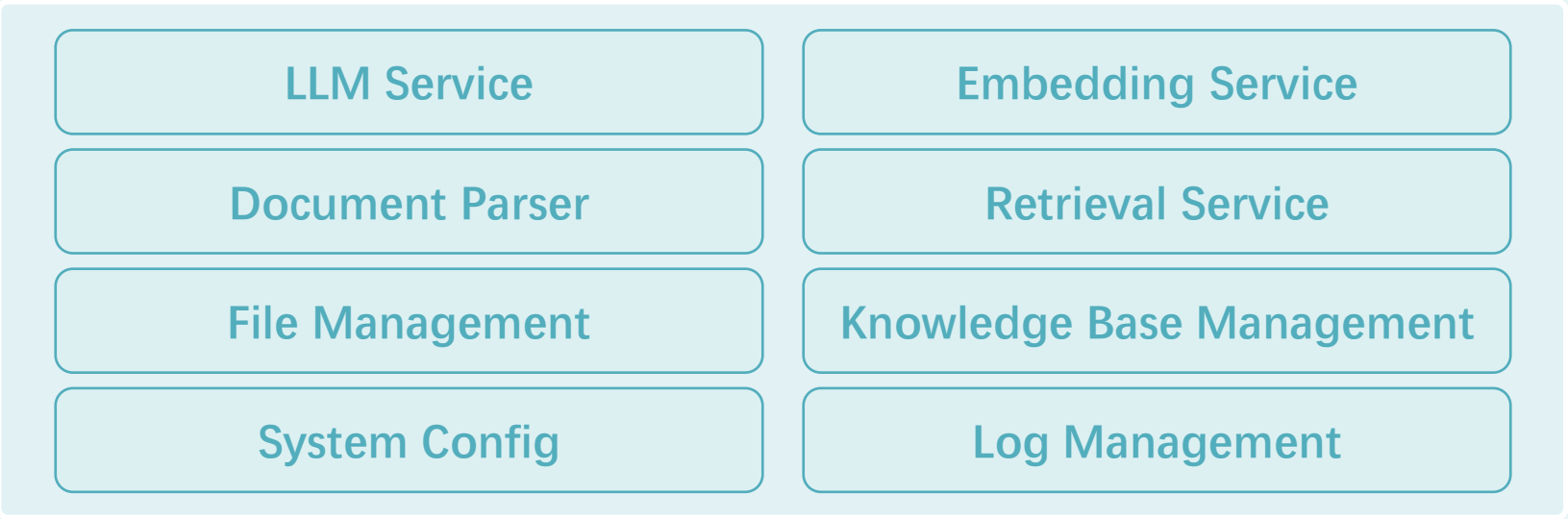
# PRODUCT ARCHITECTURE



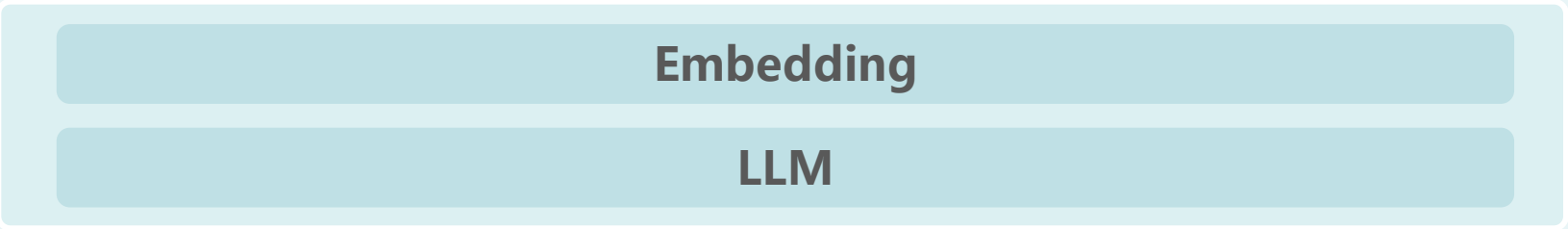
Application



RAG



Model



# HARDWARE CONFIGURATION

## Nvidia A100 for LLM

### Operation System:

Ubuntu 20.04.6 LTS

### GPU:

NVIDIA A100-SXM4-80GB

- SenseChat-Turbo/ DeepSeek 7B/14B Distilled/QwQ 32B: 2-card
- SenseChat LLM: 8-card

### CPU:

Intel(R) Xeon(R) Platinum 8358P  
CPU @ 2.60GHz \* 2  
sockets 32\*2

### Disk

- Disk1 for operating system: 512GB SSD \*2 (RAID 1)
- Disk2 for model: 1T NVME SSD \* 3 (RAID 5)

### Memory:

- 1T DDR4

## Huawei 910B for LLM

### Operation System:

EulerOS 2.0 (SP10)

### GPU:

910B 64G VRAM \* 8

- SenseChat-Turbo/ DeepSeek 7B/14B Distilled/QwQ 32B: 2-card
- SenseChat LLM: 8-card

### CPU:

Kunpeng-920\* 4

### Disk

- Disk1 for operating system: 512GB SSD \*2 (RAID 1)
- Disk2 for model: 1T NVME SSD \* 3 (RAID 5)

### Memory:

- 1T DDR4

## NVIDIA for RAG & document parsing

### Operation System:

Ubuntu 20.04.6 LTS

**GPU:** NVIDIA A100-SXM4-80GB \* 1/  
NVIDIA A40-48GB \* 1/ NVIDIA-L20  
\* 1/ NVIDIA A10-24GB \* 2/ NVIDIA  
T4-16GB \* 4

**CPU:** Intel(R) Xeon(R) Gold 6278C  
CPU @ 2.60GHz \* 1 Sockets 26\*2

### Disk

- Disk1 for operating system: 512GB SSD \*2 (RAID 1)
- Disk2 for model: 2T NVME SSD \* 3 (RAID 5)

### Memory:

- 1T DDR4

# USE CASE: AI ASSISTANT @ HK INSURANCE COMPANY



## Value Proposition

- **Methodology**

Leveraging customer files to develop an all-in-one knowledge center and activate the data in daily use

- **Competitive Advantage**



Strong capability of Cantonese with robustness



Professional data process, model training and deployment



Profound understanding of GenAI with professional support

- **Project Value**

Vertical LLM application helps decrease professional labor cost.

## Product Feature and SPEC

- **Deliverable**

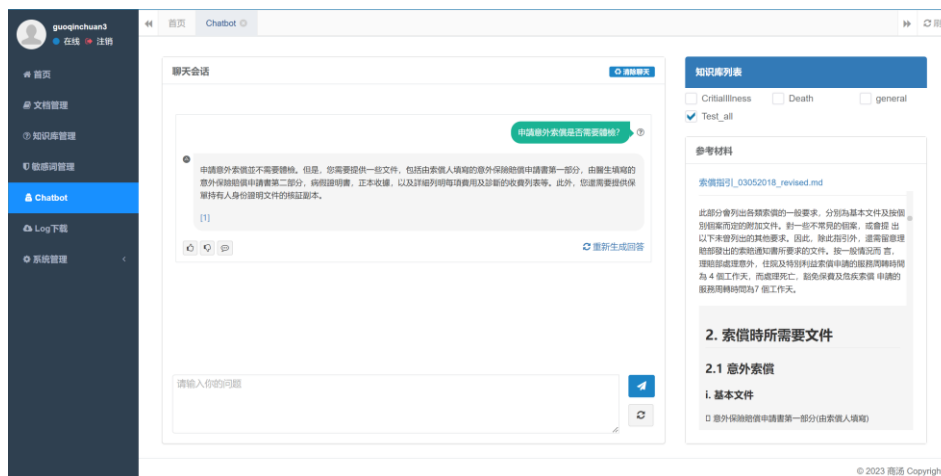
POC: **Conduct a knowledge center based on the files provided, and build up an AI Bot using RAG**

- **Deployment model**

Cloud based

- **Capability**

User will ask questions in natural language and get answers and precise reference from the AI Bot without illusion.



- The customer is one of the most prestigious insurance companies in Hong Kong.
- Starting from a single file, the project is gradually involving 115 files covering main types of Insurances like Accident, Health, Disability, Critical Illness and Death.
- The project will potentially benefit thousands of the employees and agents as the scope is expanding.



# USE CASE: AI BOT @ TELECOMMUNICATIONS COMPANY



## Value Proposition

- **Methodology**

Leveraging AI BOT to generate quick and accurate answers to gradually decrease human labor at the call center

- **Competitive Advantage**



Strong capability of Cantonese with robustness



Professional data process, model training and deployment



Profound understanding of GenAI with professional support

- **Project Value**

Purely based on call center scenarios, increase modal understanding towards Cantonese, raise the direct usability rate.

## Product Feature and SPEC

- **Deliverable**

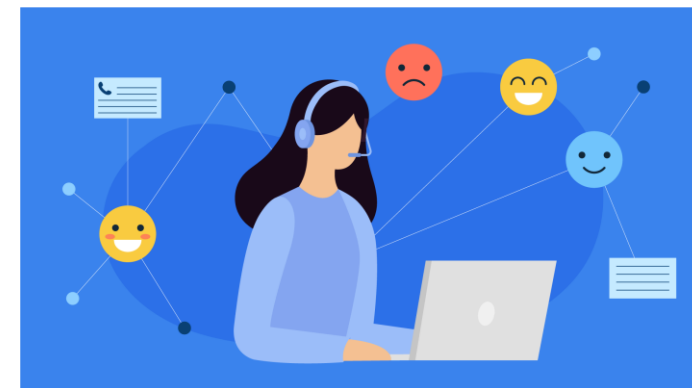
POC: **Q&A AI BOT for the call center at Whatsapp**

- **Deployment model**

Cloud based

- **Capability**

By finetuning LLM, the direct usability rate of Generative AI was increased significantly, surpassing GPT4 and it's still growing.



- The customer is one of the largest telecommunications companies in Asia.
- Their call center team comprises over 6,000 service personnel, serving not only their own company but also providing services to other companies, well-known among customers in the Asia-Pacific region.

Methodology | LLM Evaluation Round 4

Date of Evaluation	Dec 27 <sup>th</sup> , 2023	Test set	~520 real enquiries in written Cantonese from Netvigitor LiveChat			
Generative AI	RAG based on LLM as-is	Methodology	Compare LLM-generated answers with actual answers from Netvigitor LiveChat given the SAME selected context			
LLM Model	GPT 3.5 (4k context)	GPT 4.0 (8k context)	SenseNova	VertexAI		
Processing time (sec)	Average	2.0	15	8.7	3.6	
	Median	2.0	12.7	7.1	3.4	
Quality (0-100)	Average	28.0	29.7	44.6	6.8	
	Median	9.52	13.9	10.8	5.4	
Direct usability % (Edit distance-based)						
		Direct usable: 9%	Direct usable: 12%	Direct usable: 39.2%	Due to languages and context understanding	
Relevancy (Cosine similarity)						
		88.35	89.34	89.69	82.96	

# USE CASE: DOCUMENT PREPARATION @ HK GOVT

1

## Objectives

They face heavy workloads of text processing in reading and writing documents. Seek to enhance their document preparation process through the adoption of AI Assistant

2

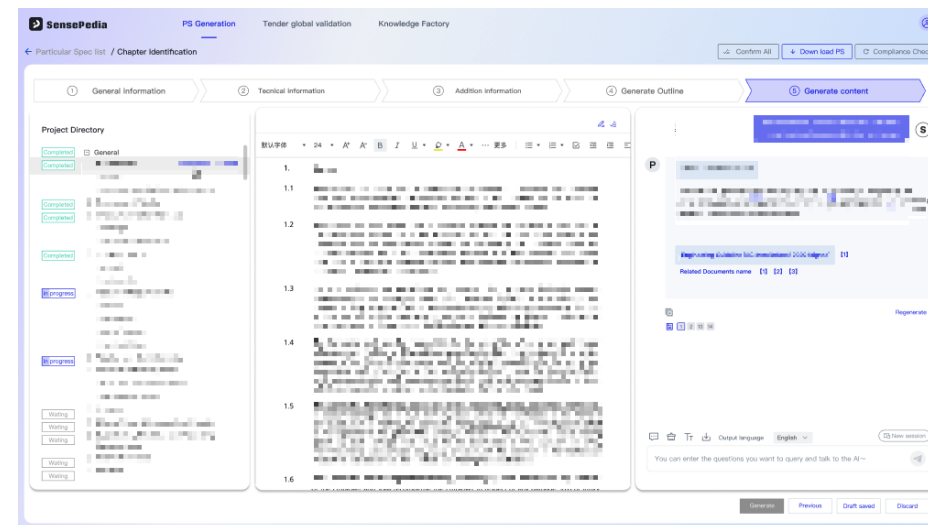
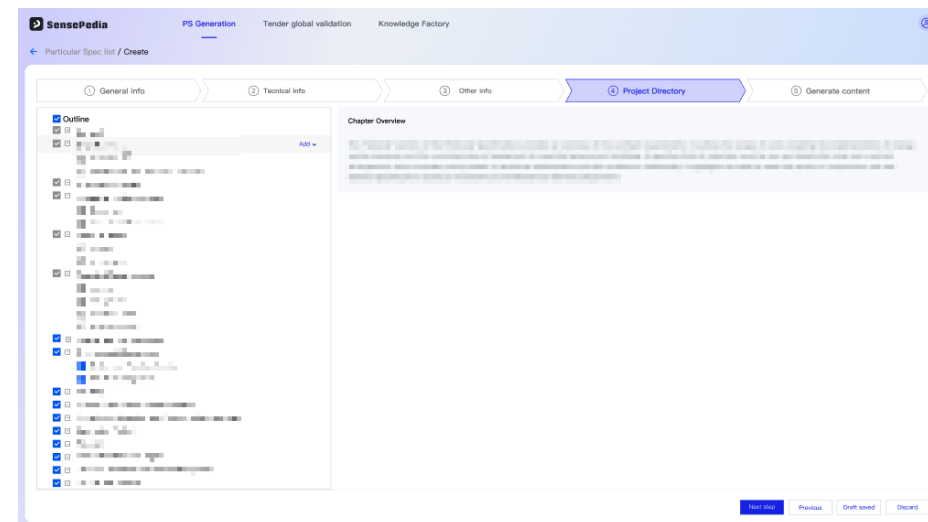
## Challenges

Need to increase productivity by automating the generation of documents while ensuring compliance with the latest standards and maintaining confidentiality

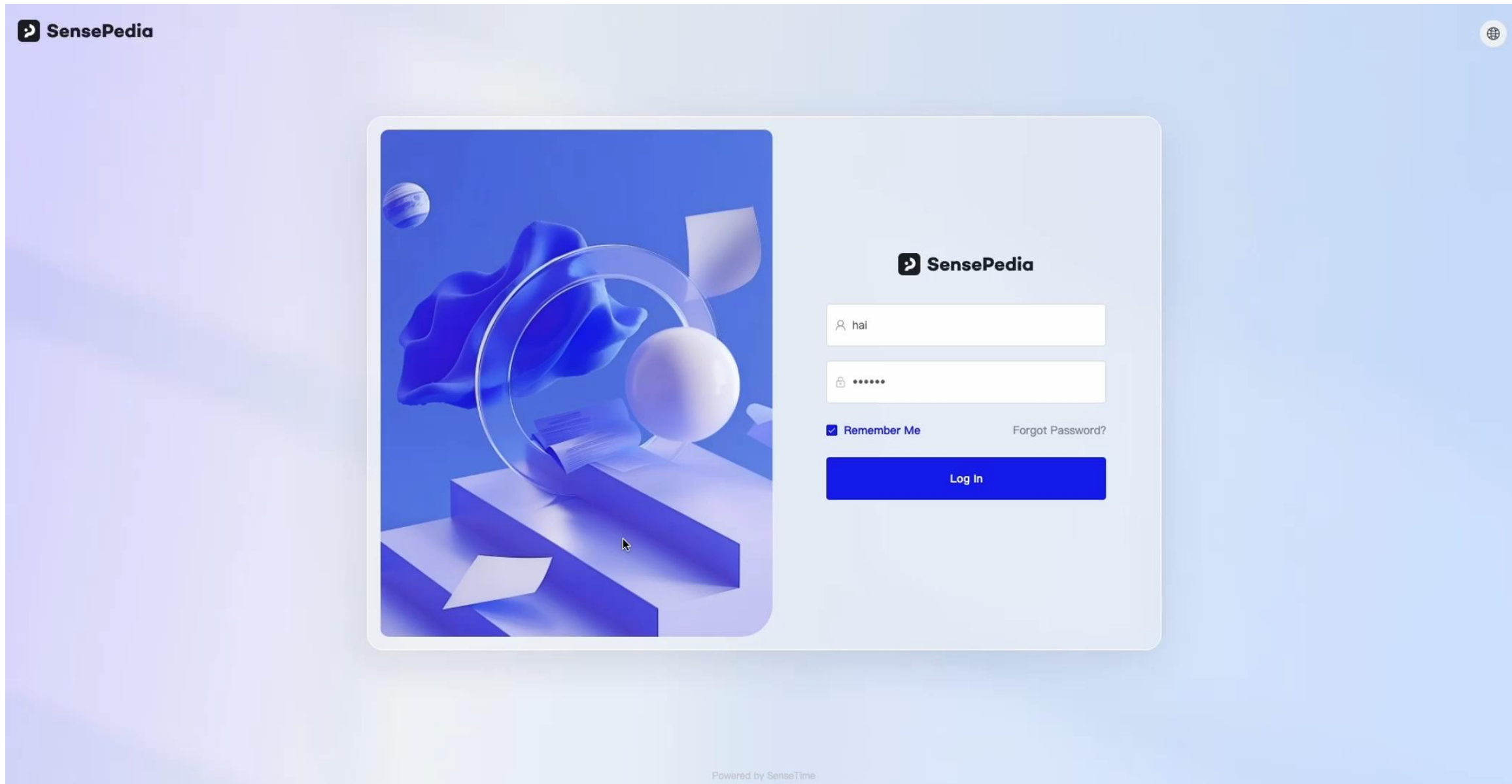
3

## Solutions

Built a comprehensive knowledge database trained on relevant documents with LLM model, and features for compliance checking and document drafting



# SENSEPEDIA DEMO





THANK YOU