V2.1.0

Mar 2026

ASUS AI SuperBuild

User Guide

Table of Contents

Chapter 1: Getting Started 2
1.1 System Requirements 2
1.2 Installation 3
1.3 System Information & Updates 14
Chapter 2: The AI SuperBuild Interface 17
2.1 Main Chat Interface 17
2.2 Managing Chat History 18
Chapter 3: Assistant Configuration 19
3.1 AI Assistant List 19
3.2 Model Upload and Convert 19
3.3 Creating a New Assistant 27
3.4 Editing an Existing Assistant 33
3.5 Knowledge Base Management 33
3.6 Evaluation and Report 34
3.7 One-Key Import/Export 50
Chapter 4: MCP (Model-Context-Protocol) 54

Chapter 1: Getting Started

This chapter provides an overview of the system requirements and the installation process for AI SuperBuild.

1.1 System Requirements

Before installing AI SuperBuild, please ensure your system meets the minimum hardware and software requirements outlined below. Meeting the recommended requirements will provide a better user experience, especially when working with multiple or larger language models.

Hardware Requirements

Component	Minimum Requirements	Recommended Requirements
Processor	Intel® Core™ Ultra processor Series 1 (Meteor Lake)	Intel® Core™ Ultra 200V series (Lunar Lake)
Memory (RAM)	16GB	32GB
Storage	4GB for AI Assistant with 1 LLM	12GB for AI Assistant with 3 LLMs
Graphics	Integrated Intel® Graphics	Integrated Intel® Arc™ Graphics
Network	Broadband connection for LLMs and component downloads	

Note:

- *AI SuperBuild has been validated on limited Intel AIPC: NUC 14 Pro, NUC 14 Pro AI, NUC 14 Pro AI+, NUC 15 Pro and NUC 16 Pro.*
- *Minimum Intel Graphics driver version is 30.0.100.9955, and the minimum NPU driver version is 32.0.100.3714. Please visit the Intel Download Center for the latest drivers*

Software Requirements

Microsoft Windows 11 (Version 23H2 or newer) is required. During installation, the AI SuperBuild application may download and install additional required components.

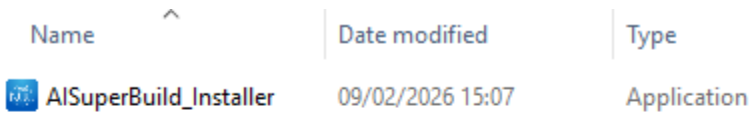
1.2 Installation

The following steps will guide you through the installation of AI SuperBuild. The process includes the setup of required prerequisite software, followed by the main application installation and first-time model download.

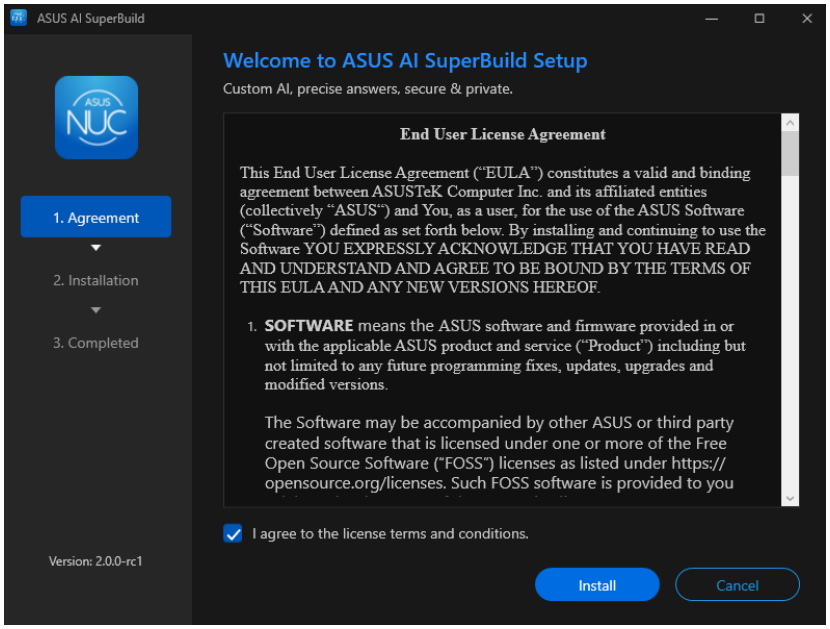
Installation Steps

1. Unzip the Package:
   Begin by unzipping the downloaded ASUS_AI_SuperBuild_x.x.x.

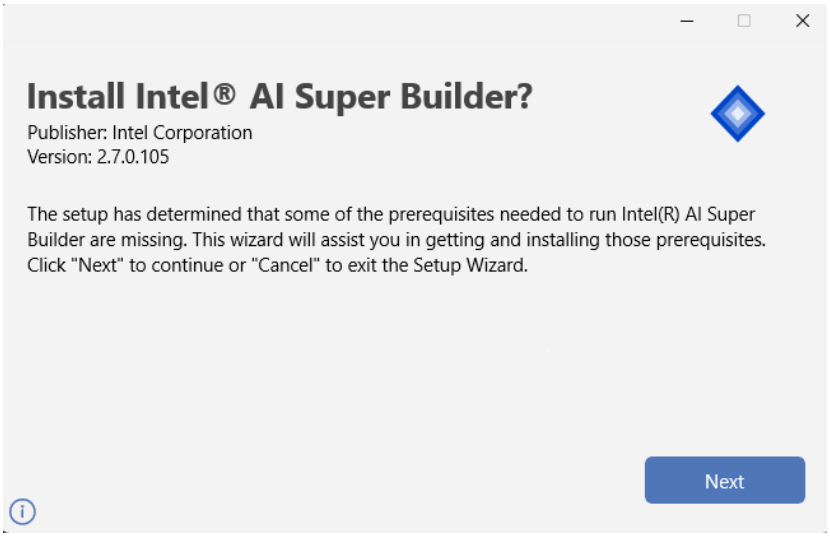
2. Run the Installer Script:
   Open the extracted folder and double-click “AISuperBuild_Installer”. This installer will initiate the entire installation process.



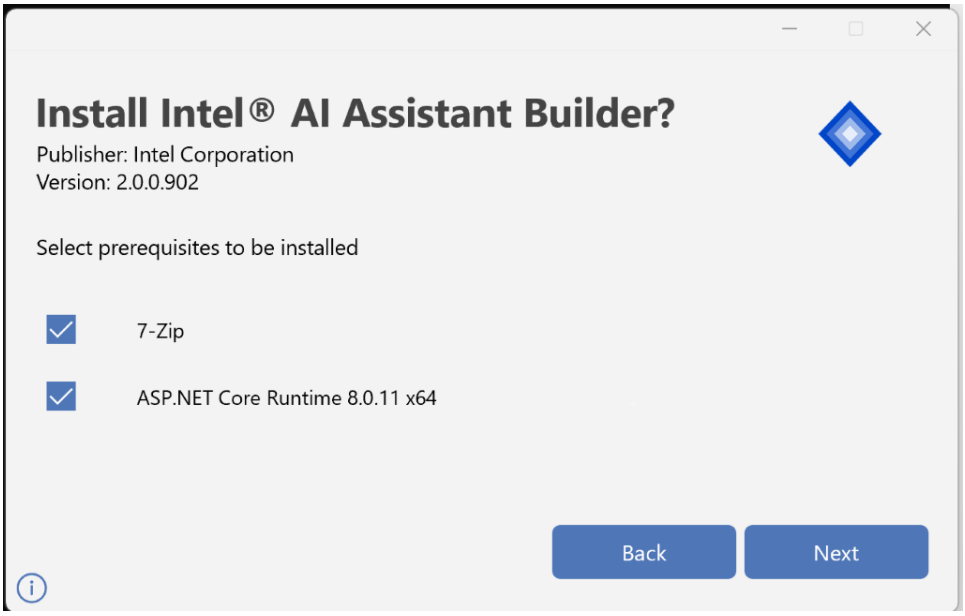
3. Follow the On-Screen Prompts:
   A series of installation wizards will appear. Follow the on-screen instructions, clicking Next, Install, and Accept as prompted to install the prerequisite software (Intel® AI Assistant Builder, 7-Zip, ASP.NET Core Runtime) and the main AI SuperBuild application.



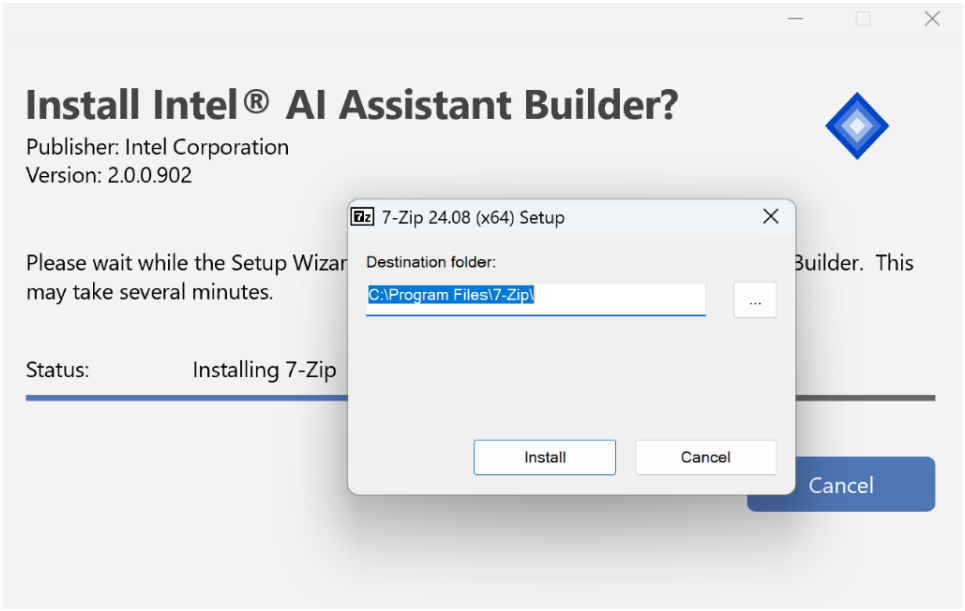
Select to agree to the license terms and conditions and select install



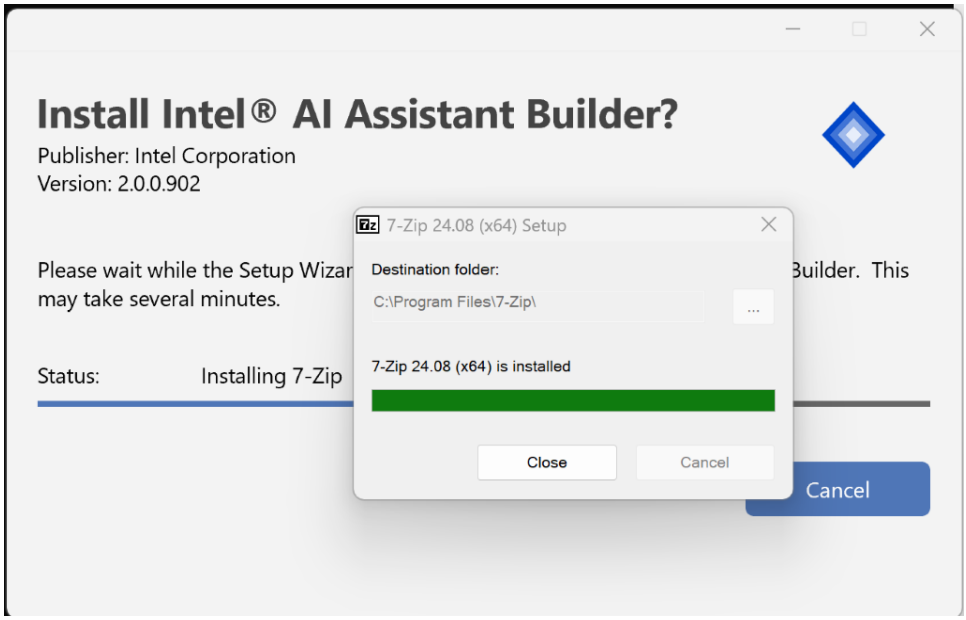
Select Next



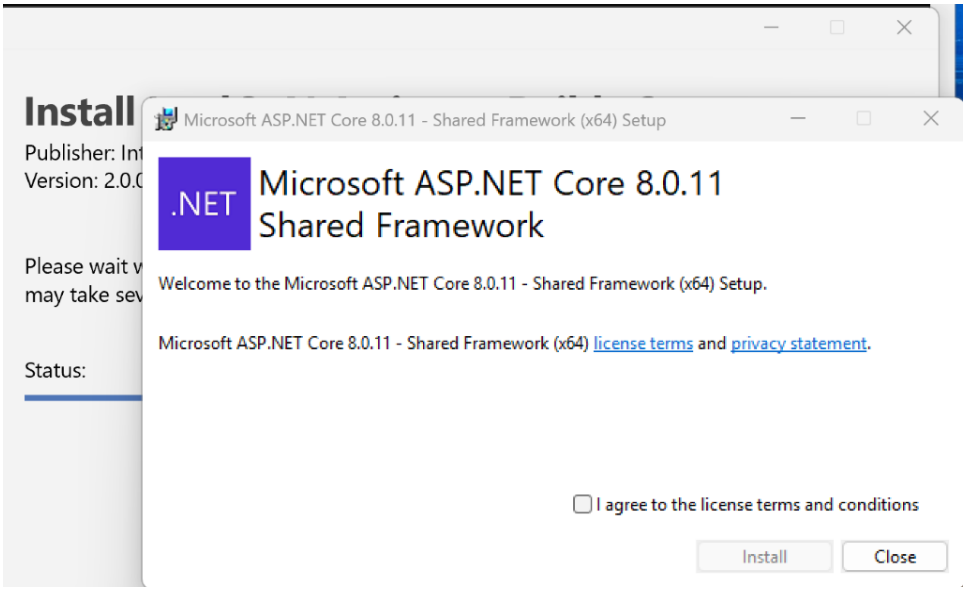
Select Next



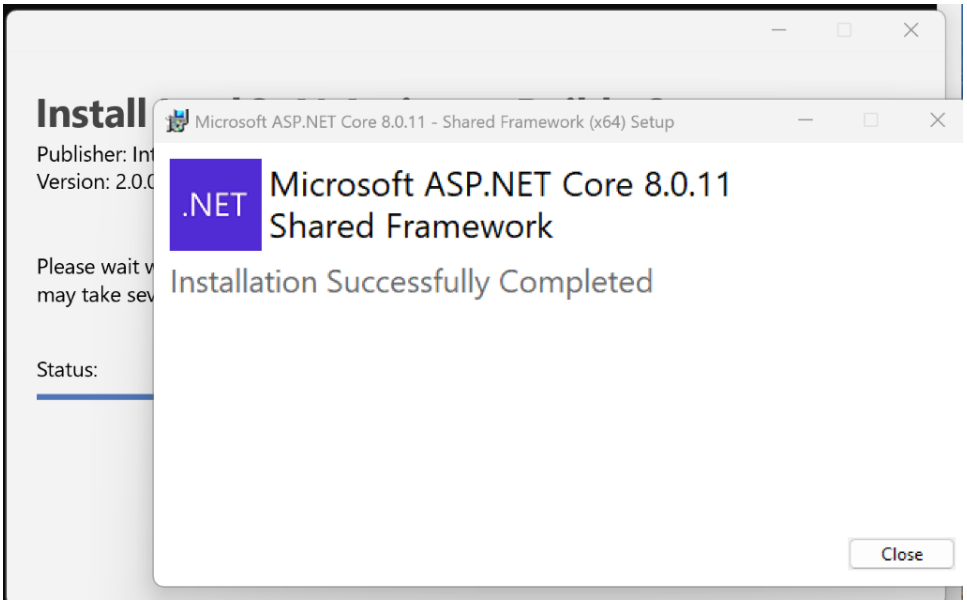
Select Install



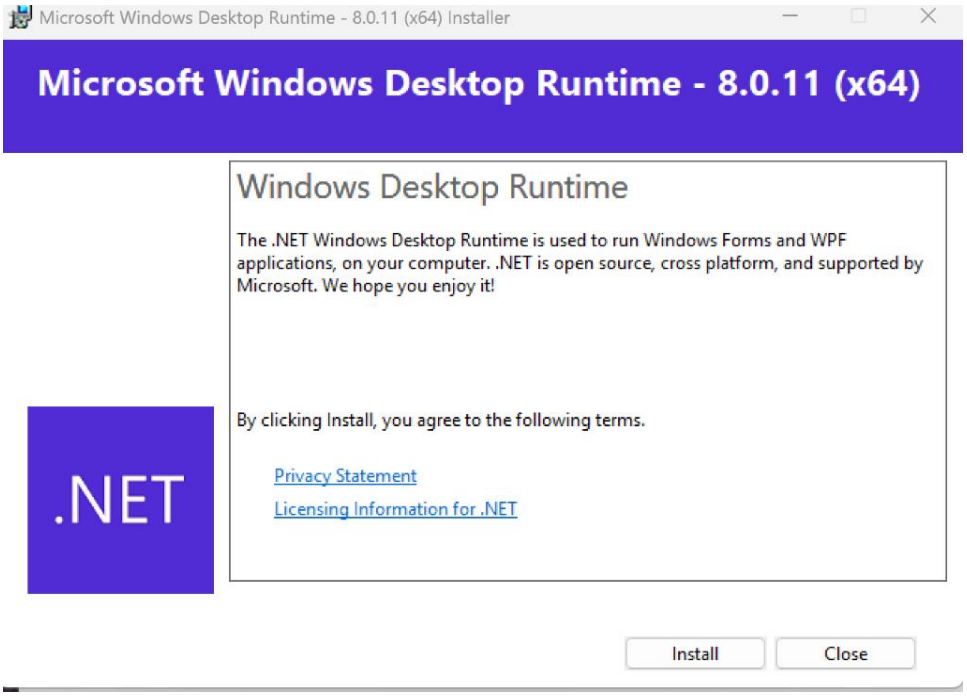
Select Close



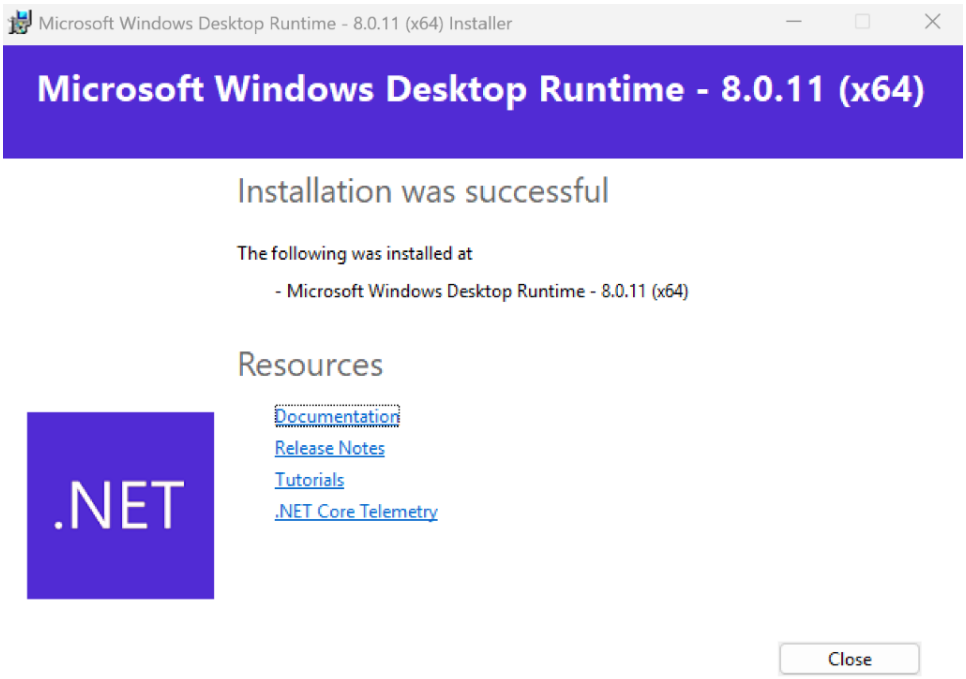
I agree, then install



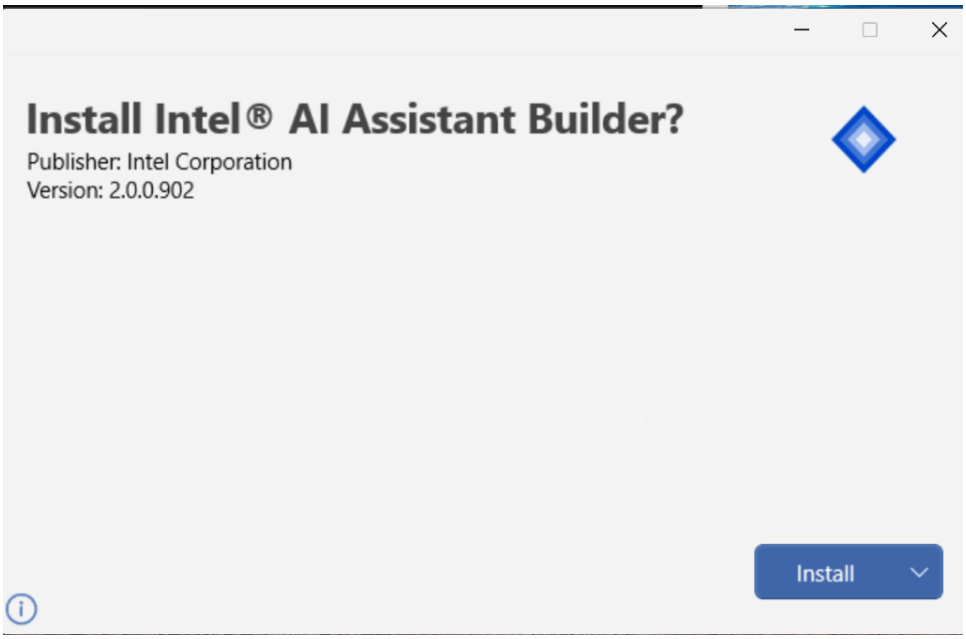
Select Close



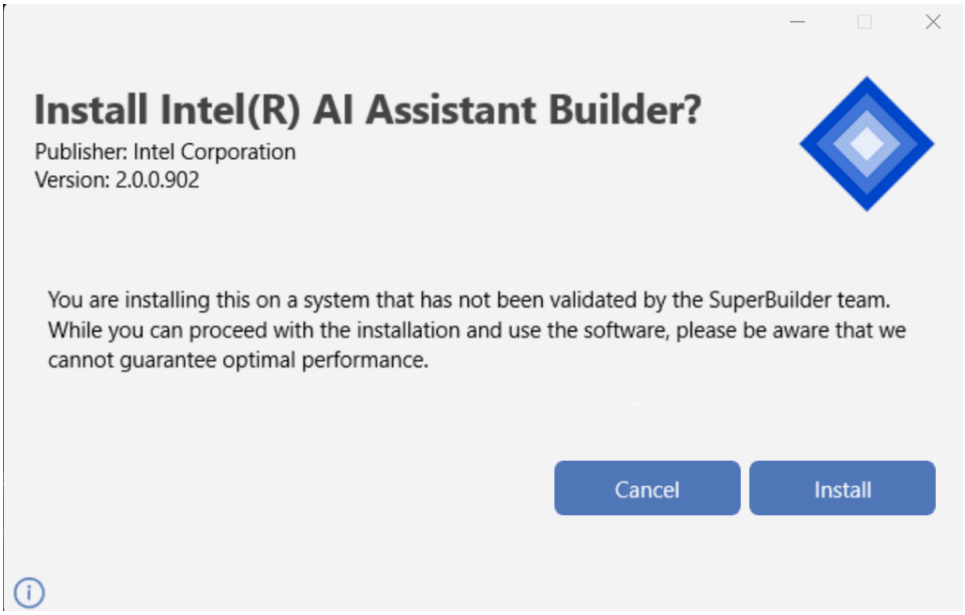
Select Install



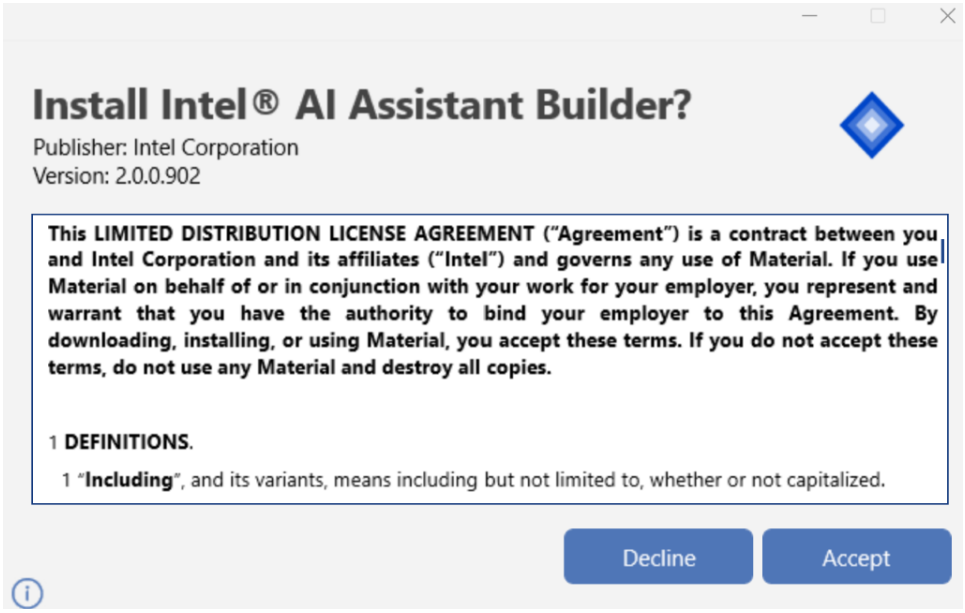
Select Close



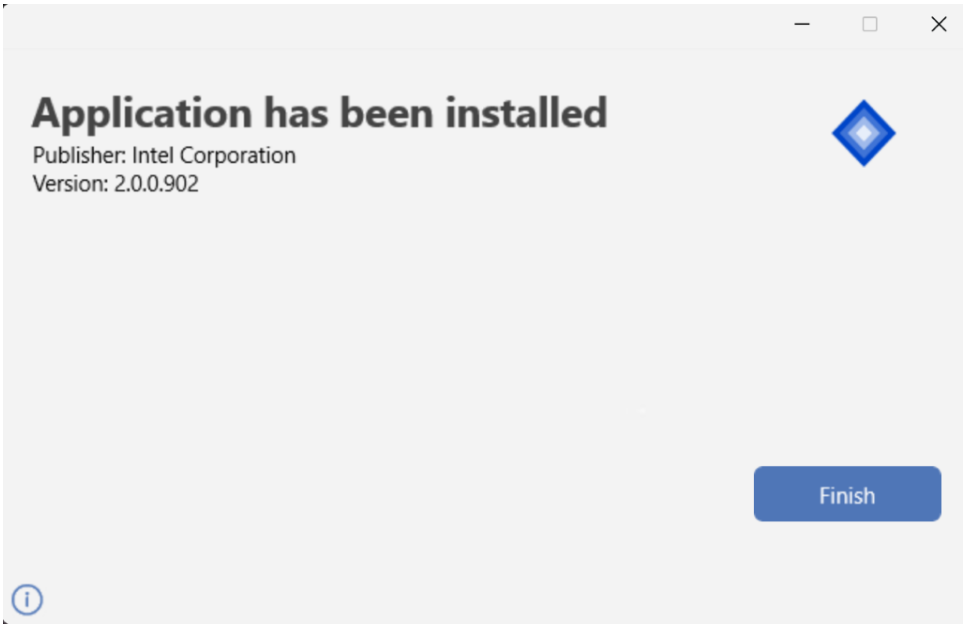
Install



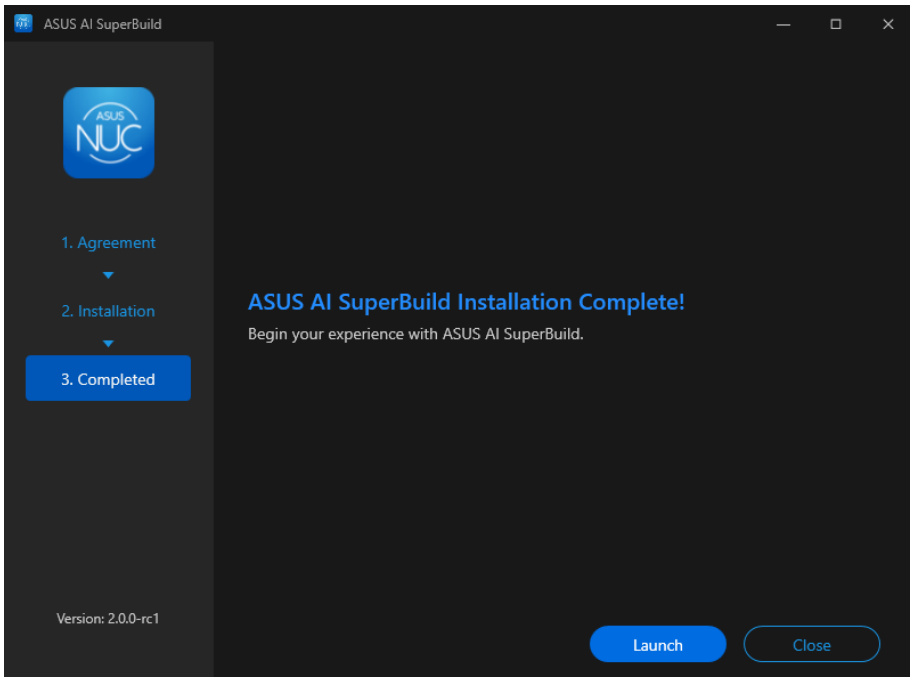
Accept



Finish



Select Finish

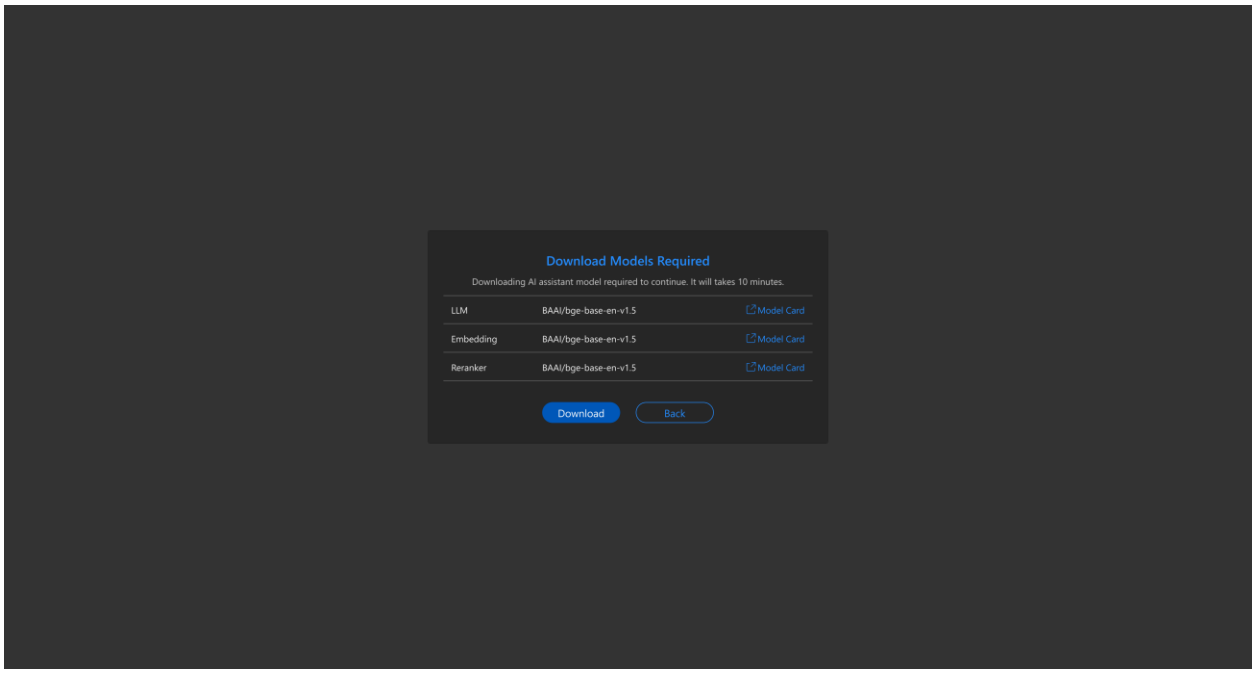


Installation Complete

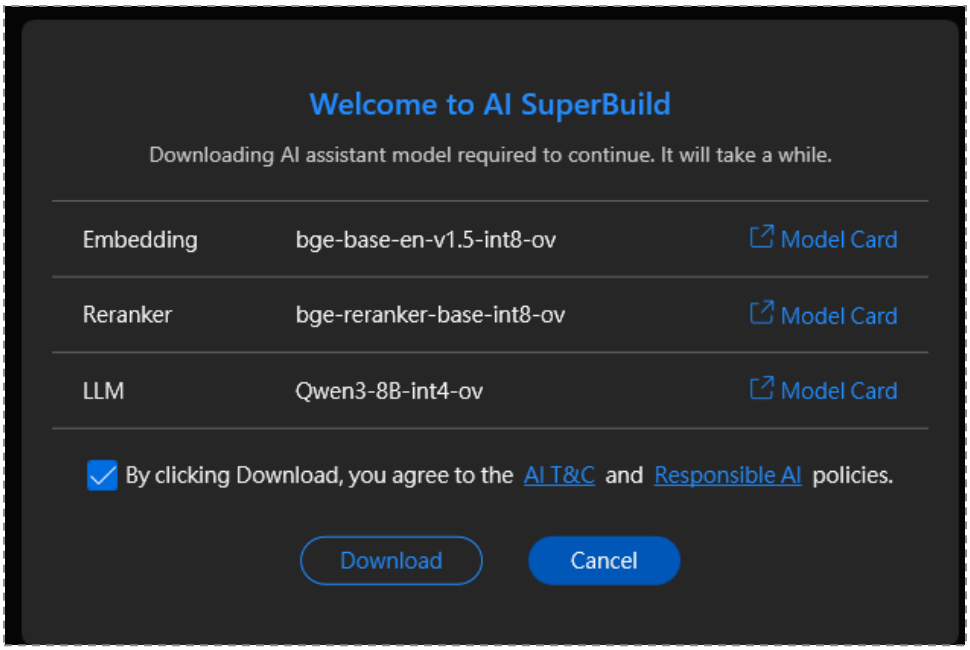
4. Launch the Application:
   Once the installation is complete if you want to launch ASUS AI SuperBuild. Click Lunch. A new desktop shortcut will also be created for future access.

First-Time Setup: Model Download

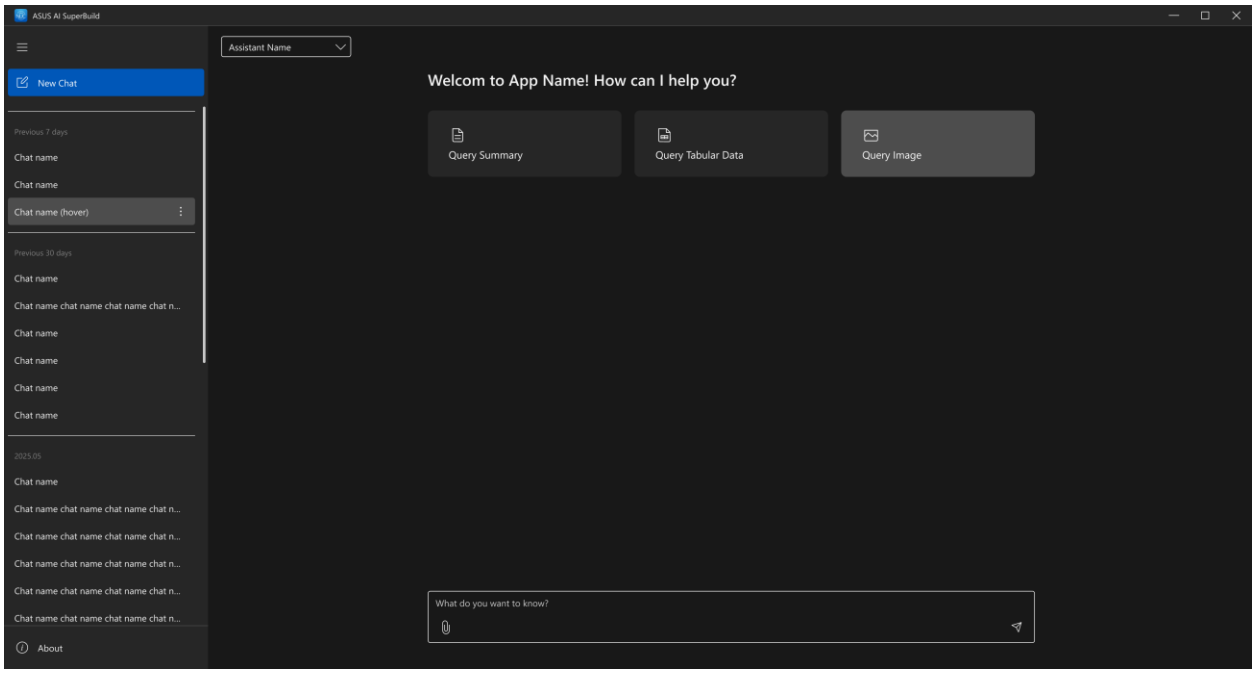
1. Initial Model Download:
   Upon the first launch, a "Welcome to AI SuperBuild" window will appear, prompting you to download the required AI assistant models (Embedding, Reranker, and LLM).



2. Agree and Download:
   Check the box to agree to the AI T&C and Responsible AI policies, then click Download.
   The process may take around 15 minutes, depending on your internet speed.



3. Configuration and Use:
   After the download completes, the application will configure the active assistant. You can then begin using your AI SuperBuild assistant.



Known Issues

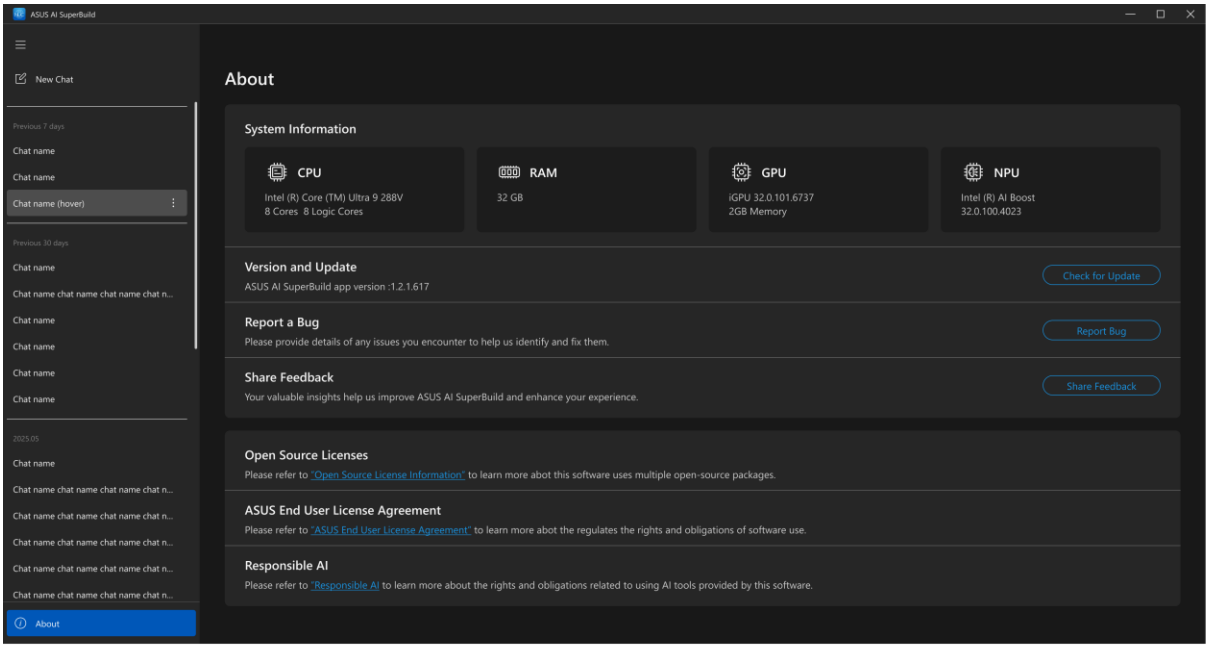
- When the software is used for the first time, it may take approximately 35 to 50 seconds to respond to the initial prompt.
- Text generation speed may be slow in the current version. This is slated for improvement in a future release.

1.3 System Information & Updates

This chapter explains how to access system information, check for software updates, and find legal and compliance information related to AI SuperBuild.

1.3.1 About Page

The "About" page provides a comprehensive overview of your system's hardware, the AI SuperBuild software version, and links to important documentation. It is the central place to verify your setup and ensure your software is up to date.



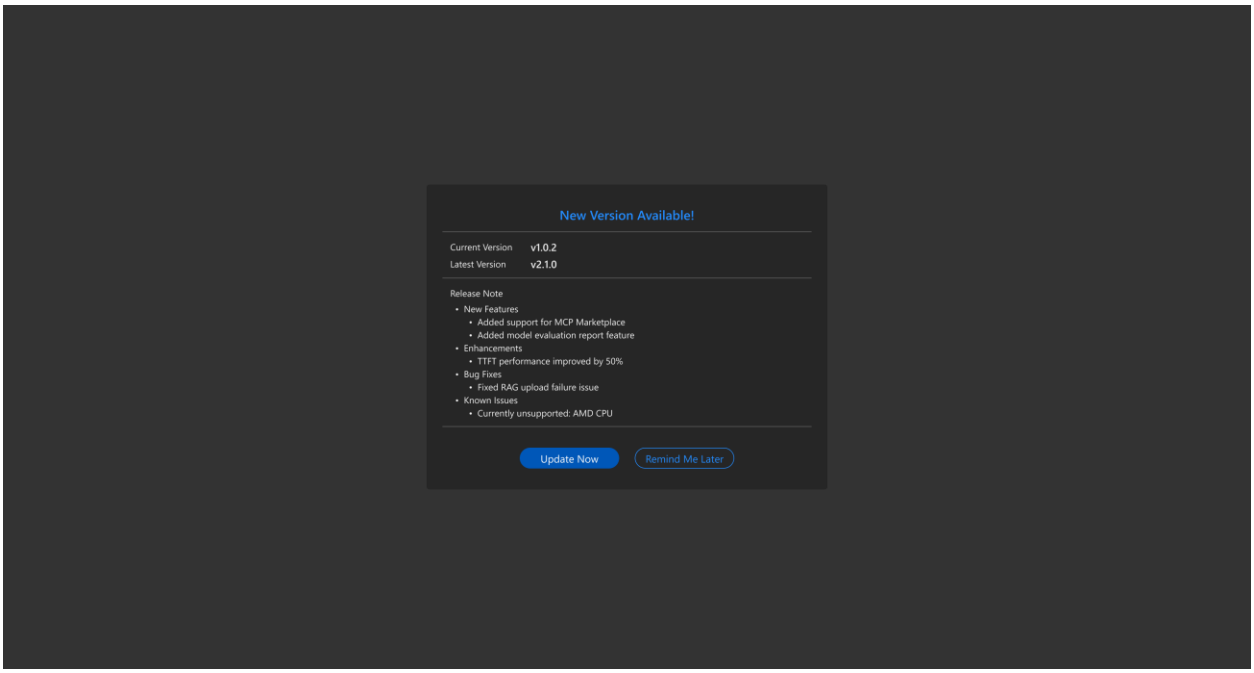
UI Element	Description
System Information	This section displays the key hardware components of your system.
CPU	Shows the processor model, number of cores, and logical processors. (e.g., Intel(R) Core(TM) Ultra 9 288V)
RAM	Shows the total amount of installed system memory (e.g., 32 GB).
GPU	Displays the graphics processing unit model, driver version, and memory (e.g.,

	GPU 32.0.101.6737, 2GB Memory).
NPU	Displays the Neural Processing Unit model and driver version, which is used for accelerating AI tasks (e.g., Intel(R) AI Boost 32.0.100.4023).
Version and Update	This section shows the currently installed application version of AI SuperBuild (e.g., 1.1.0)
Check for Update	Clicks to check for and download the latest version of the application.
Report a bug	Clicking this option will open a Google Form, allowing you to report any bugs or system issues directly to ASUS.
Share Feedback	We value your input! Click this button to send any feedback, ideas, or comments regarding your AI SuperBuild experience directly to the ASUS team.
Open Source Licenses	Opens a new window with information about the open-source software packages used in the application.
ASUS End User License Agreement	Opens the EULA, which details the rights and obligations of software use.
AI T&C	Opens the Terms and Conditions related to using the AI functions within the software.
Responsible AI	Opens documentation regarding the responsible and ethical use of the AI tools provided by the software.

1.3.2 Check for update

Each time you launch the AI SuperBuild, the system automatically checks for the latest software version. If an update is available, simply click Update Now, and the application will handle the download and installation automatically.

Alternatively, you can manually check for new versions at any time by navigating to the Settings page and clicking Check for update.

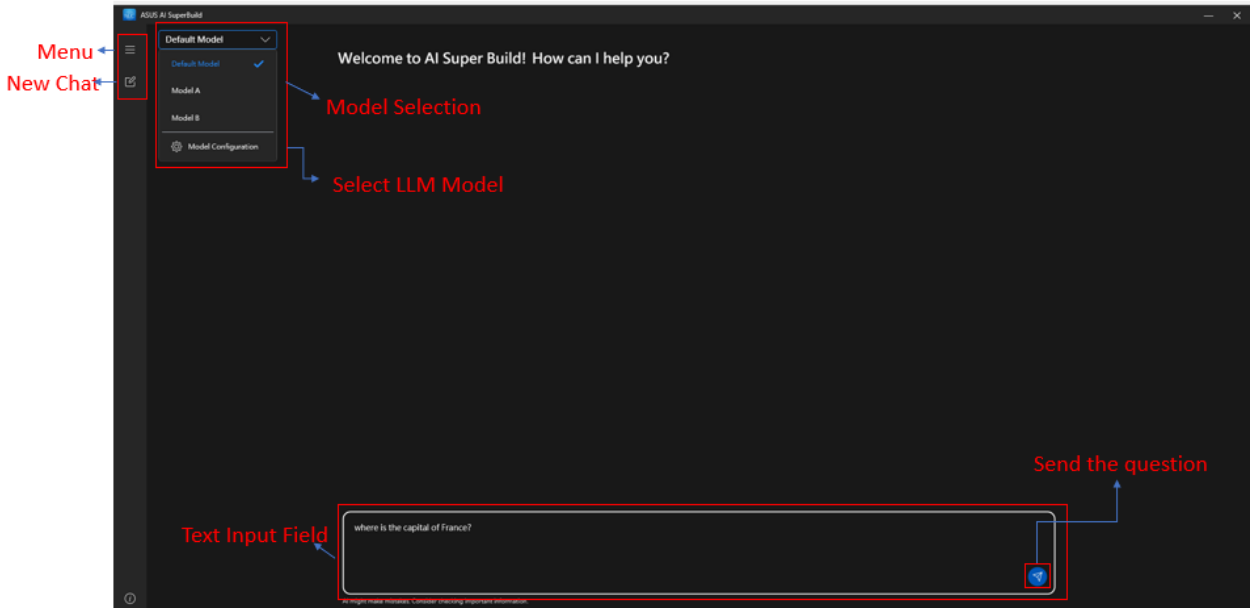


Chapter 2: The AI SuperBuild Interface

This chapter describes the main user interface and core functionalities for interacting with the AI Assistant, including how to start chats, select models, and manage your conversation history.

2.1 Main Chat Interface

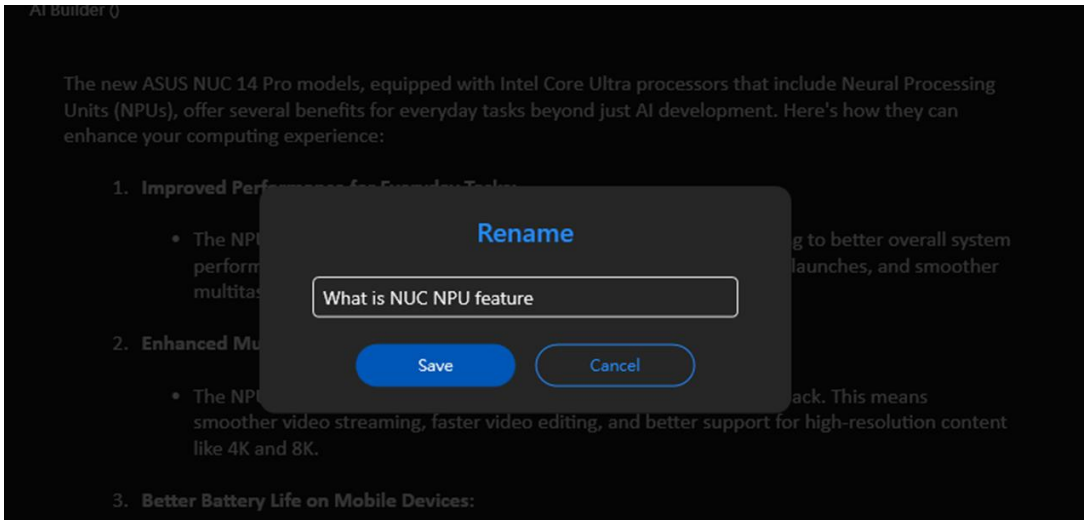
The main interface is your central hub for interacting with the AI SuperBuild assistant. The interface features an intuitive layout with a left-hand navigation panel for managing chats, a central conversation window, and input tools at the bottom.

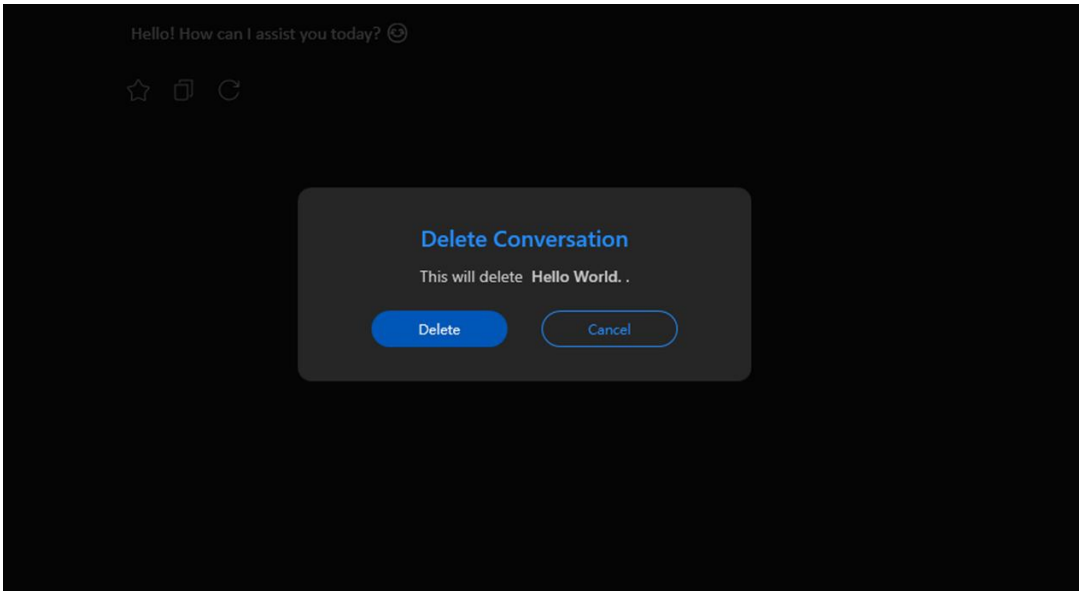


UI Element	Description
Menu (☰)	Toggles the visibility of the left-hand navigation panel.
New Chat	Submits your text prompt to the AI assistant.
Model Selection	Allows you to select the active AI model for your session from a list of available models like "Default Model", "Model A", etc.
Model Configuration	Navigates to the advanced settings page to create, edit, and manage your AI assistants.
Text Input Field	The main area where you type your questions or prompts for the AI assistant.
Send Icon	Clicks to submit your text prompt to the AI assistant for a response.

2.2 Managing Chat History

The left-hand navigation panel displays your conversation history, organized by date. You can easily revisit, rename, or delete past conversations.





Action	Description
Select a Chat	Click on any chat name in the history to load it into the main window and continue the conversation.
Rename a Chat	Right-click on a chat and select "Rename" to give it a more descriptive name for easier identification.

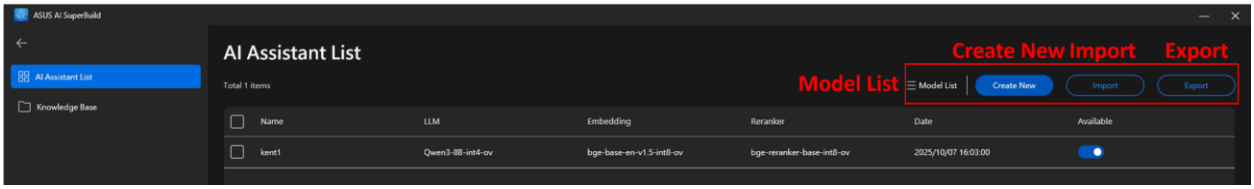
Delete a Chat	Right-click on a chat and select "Delete" to permanently remove the conversation.

Chapter 3: Assistant Configuration

This chapter covers the advanced settings for creating and managing your AI Assistants. You can access this section by navigating from the **Model dropdown list > Model Configuration**.

3.1 AI Assistant List

This page displays the configuration settings for all assistants. Each assistant is defined by a specific Large Language Model (LLM) and its associated parameters for embedding and reranking, which determines how it processes and responds to queries.



Button/UI Element	Description
Create New	Opens the "New Assistant" wizard to create a new, customized AI assistant profile from scratch.
Import	Allows you to import a previously exported assistant configuration file, making it easy to share or restore settings.
Export	Allows you to export the selected assistant's configuration as a file for backup or deployment on another machine.
Model List	Navigates to a screen where you can manage all the AI models (LLM, Embedding, Reranker) available in the system.
Available (Toggle)	Toggles the assistant's availability in the main chat Model Selection dropdown.

3.2 Model Upload and Convert

This chapter explains how to upload and convert models in AI SuperBuild. This feature allows you to import models from local storage or third-party platforms like **Hugging Face** and

ModelScope. This allows you to extend beyond the default 8 built-in LLM models.
AI SuperBuild currently supports the following:

- **Model Sites**
  - Hugging Face
  - ModelScope
  - Local Path (self-built or locally prepared models)
- **Model types**
  - LLM (Large Language Model)
  - Embedding model
  - Reranker model

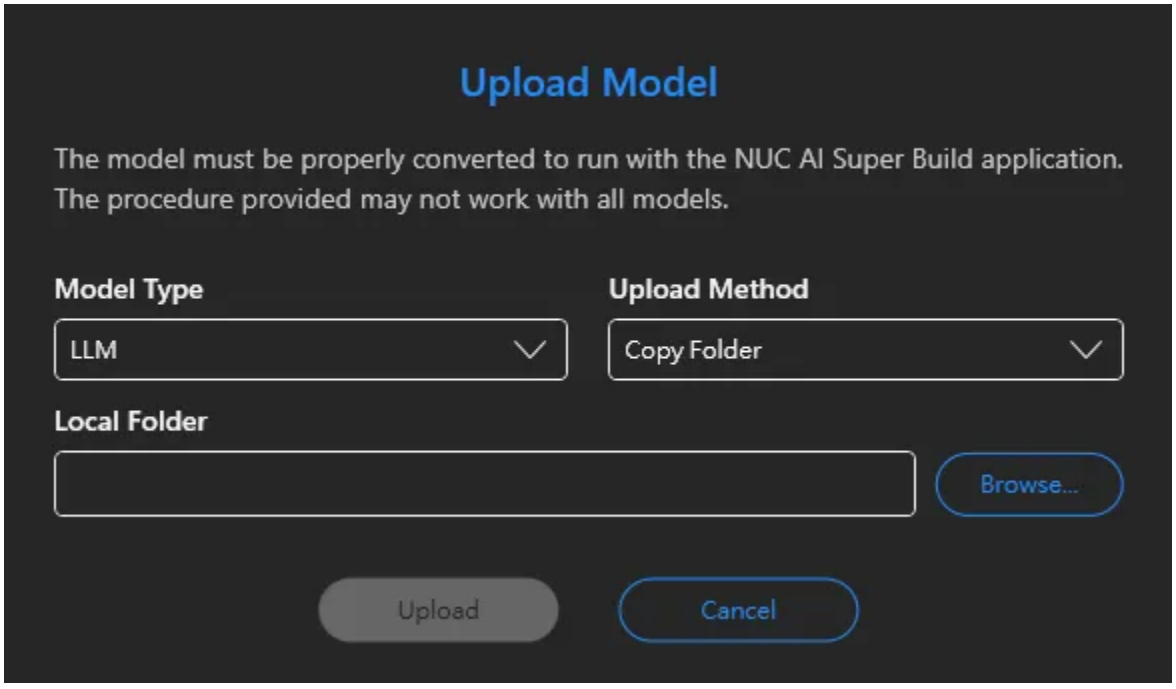
Note

Currently, AI SuperBuild only supports **text generation LLM models**. Make sure the model you choose is a text generation model.

3.2.1 Model Upload

If you already have models, you can make them available to AI SuperBuild in two ways:

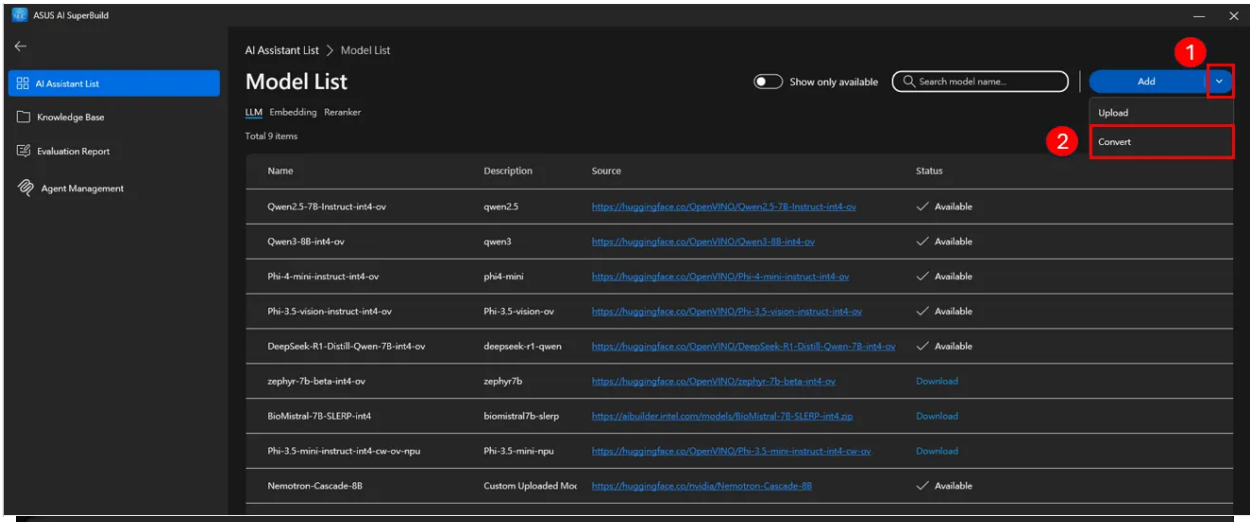
- **Upload models by folder**
  - In AI SuperBuild, navigate to **AI Assistant List** → **Model List** → **ADD** → **Upload**.
  - **Copy** or **move** the entire model folder into the designated AI SuperBuild model directory.
  - This is recommended when you have already downloaded models (for example from Hugging Face, ModelScope, or other sources) and just want AI SuperBuild to recognize them.
  - Follow the on-screen instructions to select and upload your prepared model folder.

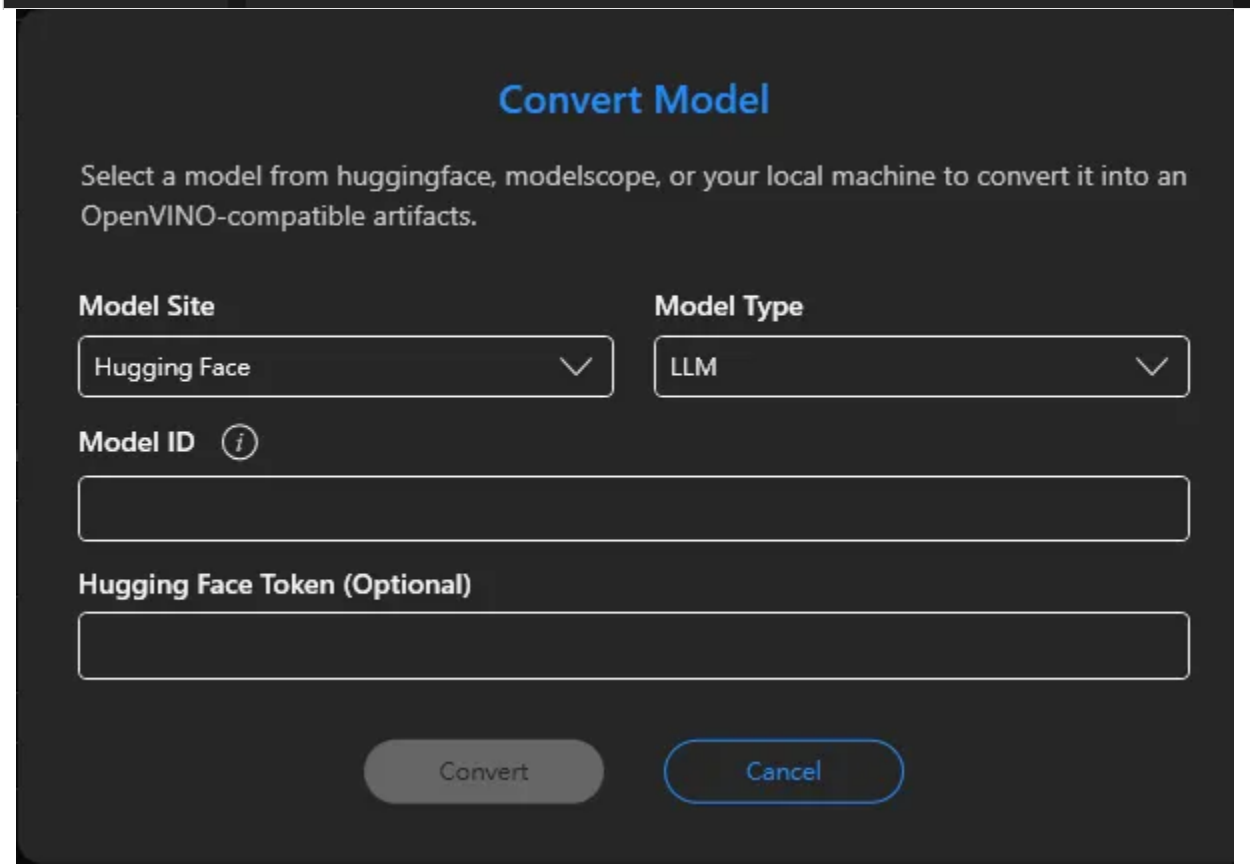


3.2.2 Model Convert

To download, convert models from Hugging Face, ModelScope, or a local path, use the **Model Convert** feature.

1. Navigate to **AI Assistant List** → **Model List** → **ADD** → **Convert**.
2. You will see:
    - A selector for **Model Site** (Hugging Face / ModelScope / Local Path)
    - A selector for **Model Type** (LLM / Embedding / Reranker)
    - An input field for **Model ID** or local path
    - A **Convert** button and related status information

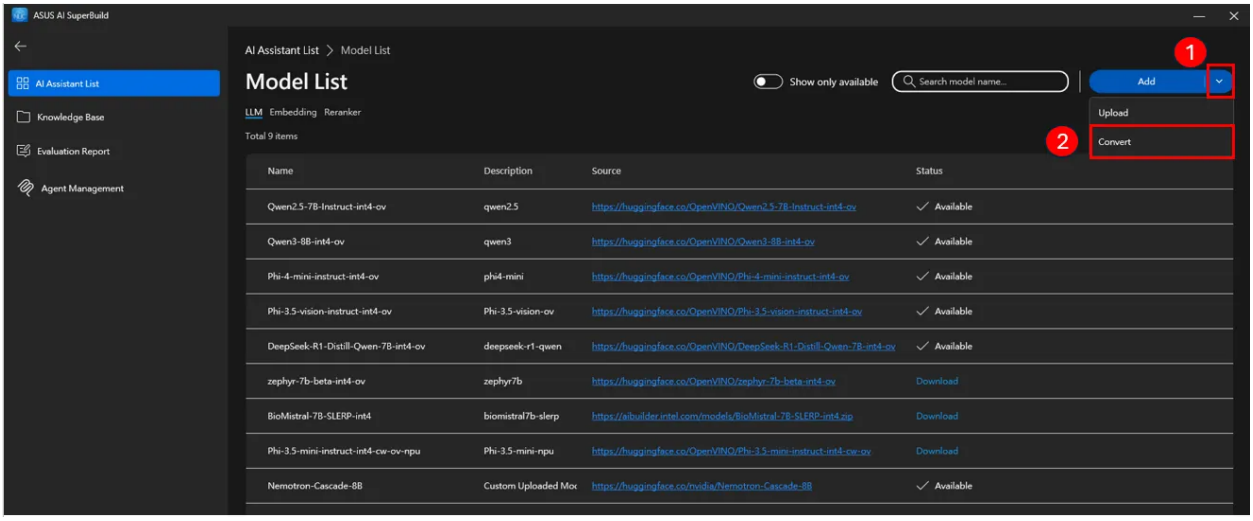




3.2.3 Model Convert

To download, convert models from Hugging Face, ModelScope, or a local path, use the **Model Convert** feature.

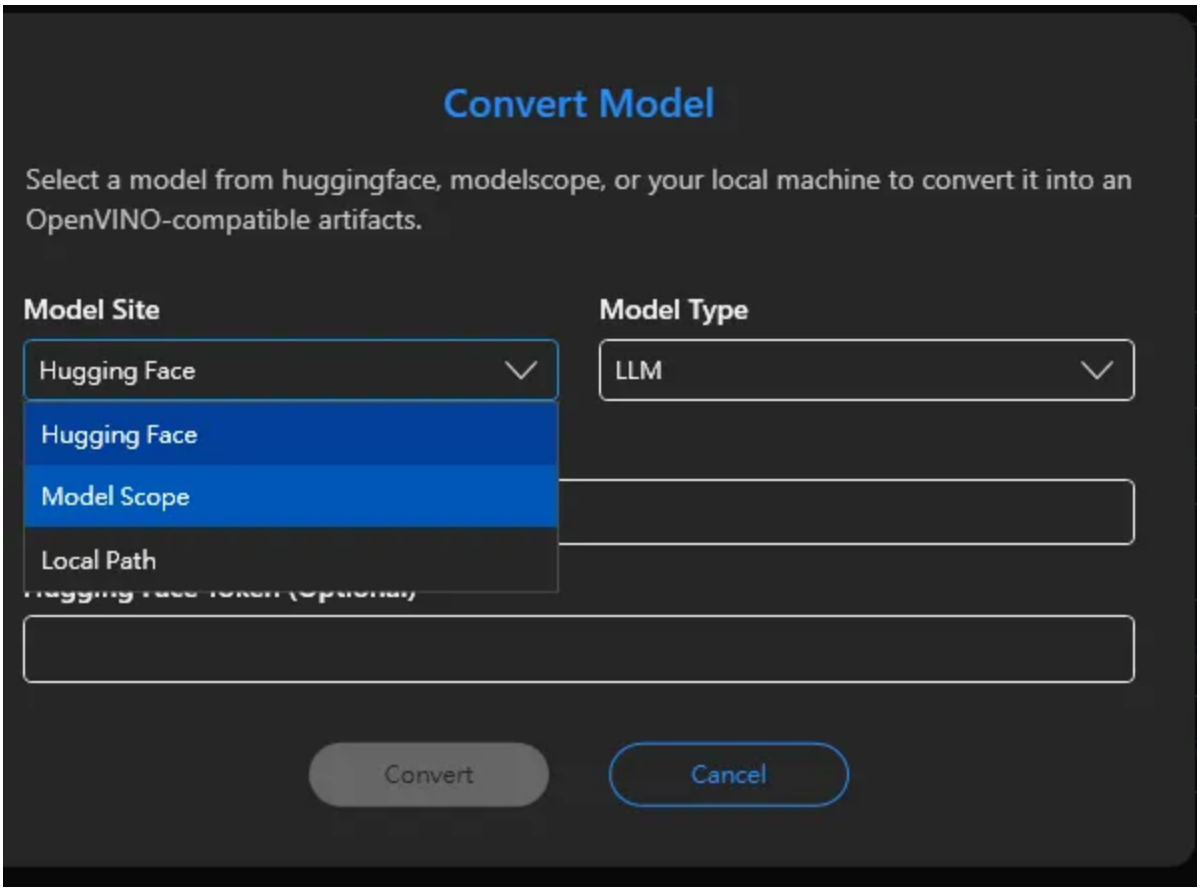
1. Navigate to **AI Assistant List** → **Model List** → **ADD** → **Convert**.
2. You will see:
    - A selector for **Model Site** (Hugging Face / ModelScope / Local Path)
    - A selector for **Model Type** (LLM / Embedding / Reranker)
    - An input field for **Model ID** or local path
    - A **Convert** button and related status information



In the **Model Site** section, choose where you want to get the model from:

- **Hugging Face** – Download and convert a model directly from the Hugging Face Hub.
- **ModelScope** - Use a model hosted on the ModelScope platform.
- **Local Path** – Use a model that has already been prepared on your local environment.

After you select a source, the corresponding input field (**Model ID** or **local path**) will be enabled.



Using Hugging Face

If you choose **Hugging Face** as the model source, follow these steps to obtain the correct model ID:

1. Open the Hugging Face website: huggingface.co
2. Use the search bar to find the model you want to use (for example: meta-llama/Meta-Llama-3-8B-Instruct).
3. In the search results, click the target model to open its detail page.
4. On the model page, click the **Copy** button next to the model name to copy the full **model ID**.

Important

Confirm that the model is a **Text Generation** model so that it can work correctly in AI SuperBuild.

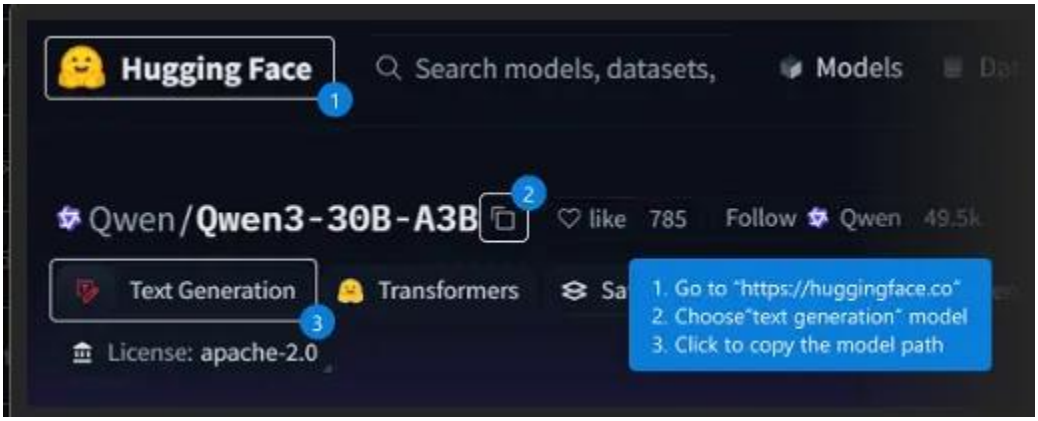
Using ModelScope

If you choose **ModelScope** as the model source:

- Go to the ModelScope website and locate the target model.
- Copy the **model ID** from the ModelScope page (similar to Hugging Face).
- Confirm that the model type and task are supported by AI SuperBuild (text generation LLM, embedding, or reranker).
- Paste the ModelScope model ID into the **Model ID** field in AI SuperBuild.

Note

The overall process is the same as Hugging Face: copy the model ID from ModelScope and use it in the **Model ID** field.



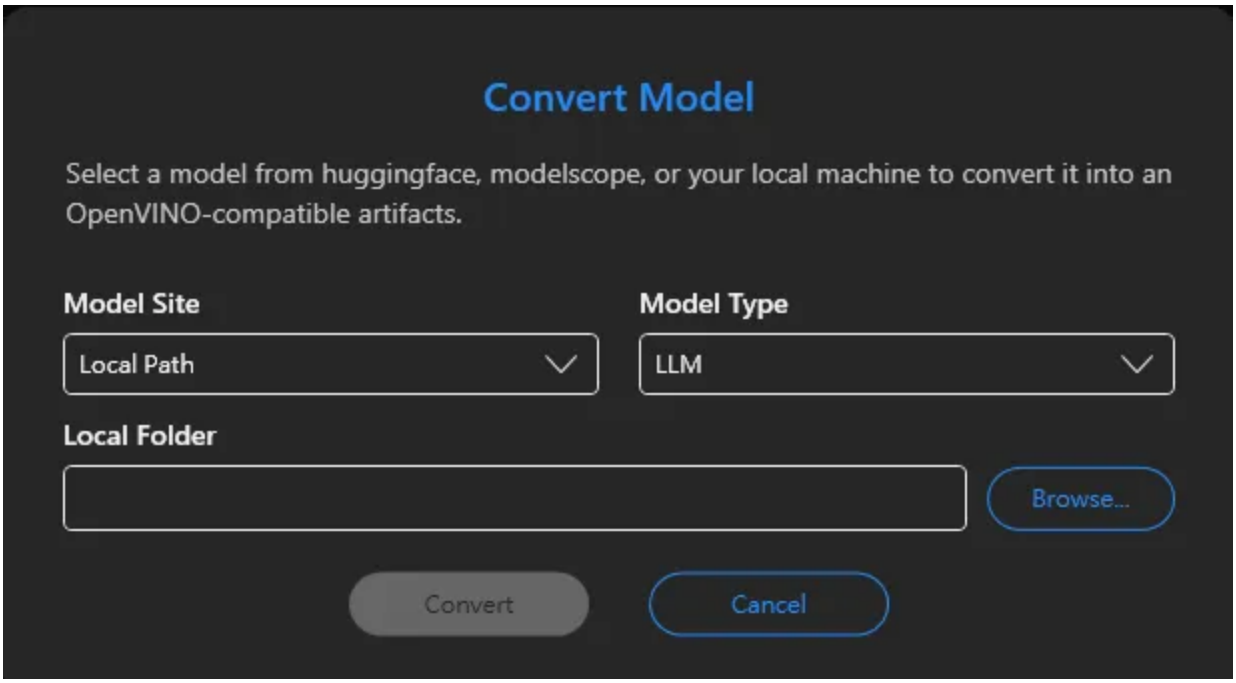
Using Local Path models

If you choose **Local Path** as the model source:

- Prepare the model files in a local directory.
- Set **Model Site** to **Local Path** and specify the local directory path in the **Model ID / Path** field.

Tip

This is recommended when your models are hosted on an internal file server or have already been converted by your ML/infra team.

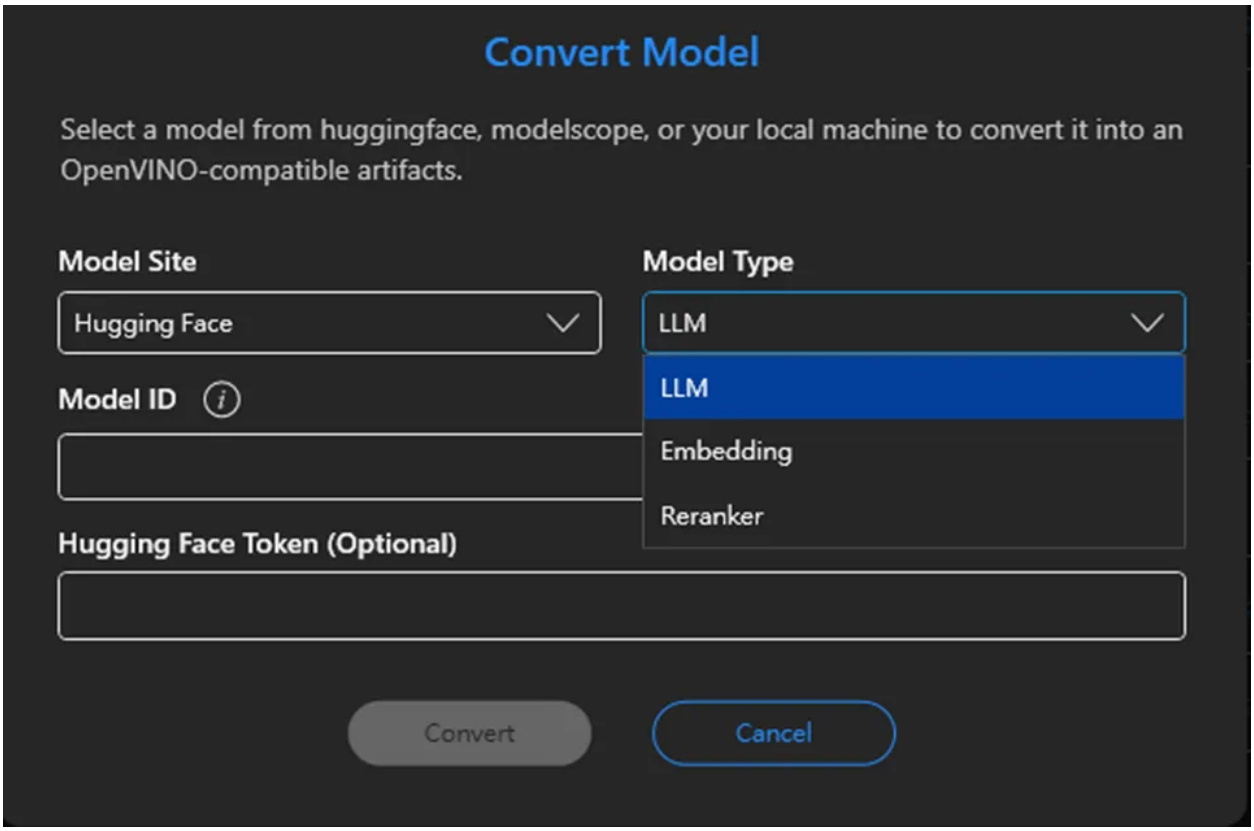


Select the Model Type

In the **Model Type** section, choose the type of model you want to convert:

- **LLM** – For text generation language models.
- **Embedding** – For models that generate vector embeddings from text.
- **Reranker** – For models used to re-rank search or retrieval results.

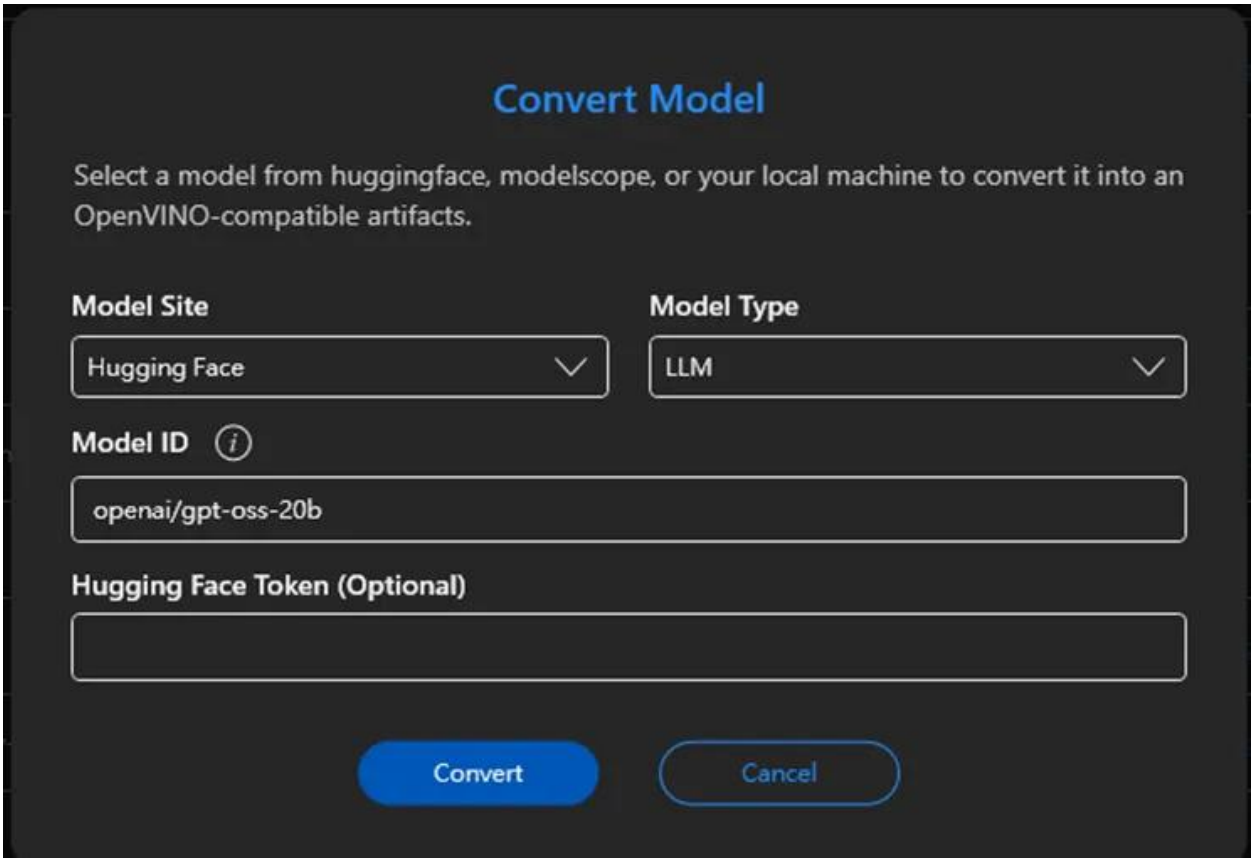
Make sure the selected model type matches the actual purpose and architecture of the model. For example, text generation models should be configured as **LLM**.



Start conversion in AI SuperBuild

After you set the **Model Site** and **Model Type**, and provide the correct **Model ID / Path**:

1. Go back to the **Model Upload & Convert** page in AI SuperBuild.
2. Confirm the following settings:
    - **Model Site** is set to **Hugging Face**, **ModelScope**, or **Local Path** as needed.
    - **Model Type** is correctly selected (LLM / Embedding / Reranker).
    - The **Model ID / Path** field contains the copied model ID or local directory path.
3. Click **Convert**.

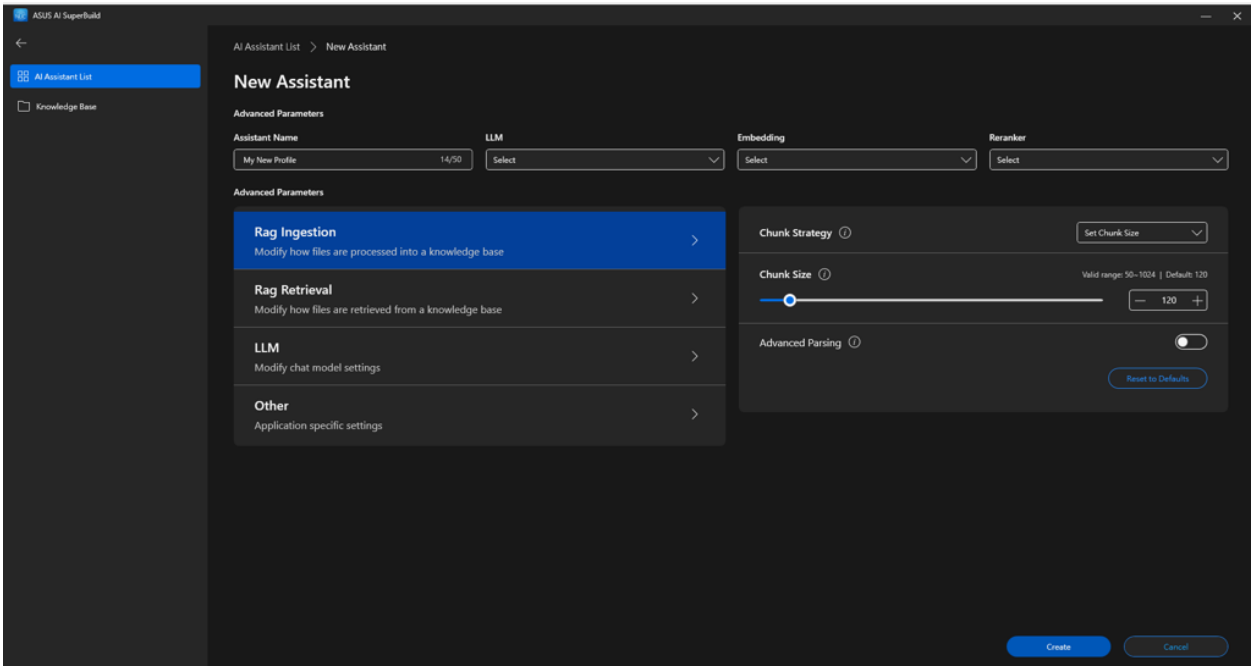


AI SuperBuild will automatically:

- Download the model from the selected platform (Hugging Face / ModelScope), or load it from the local path.
- Convert it into a format supported by AI SuperBuild.
- Optimize the model for inference (depending on internal configuration).
- Register the converted model into AI SuperBuild so it can be used in projects.

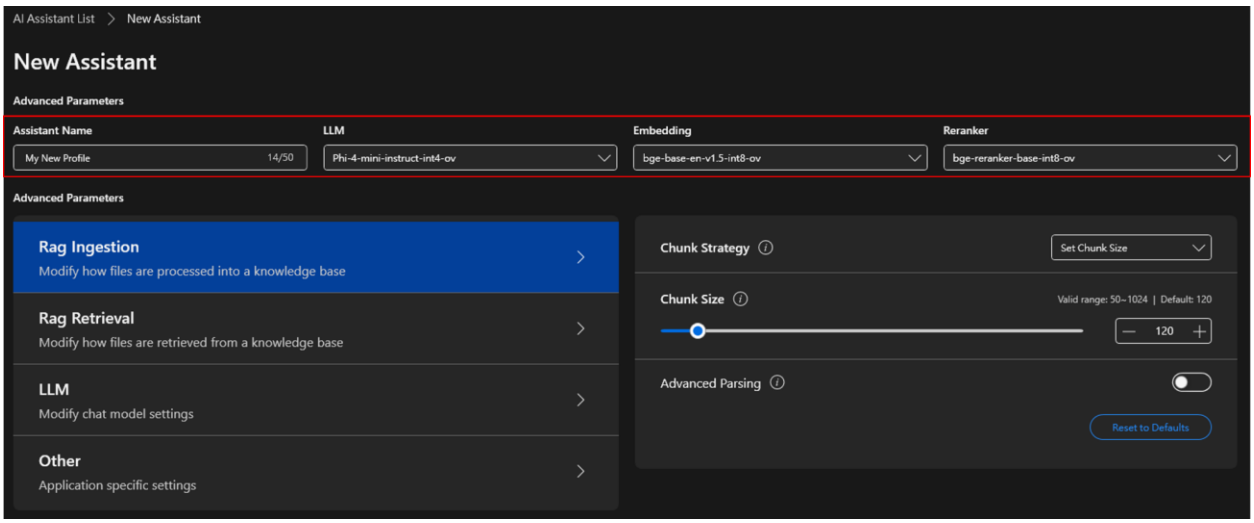
3.3 Creating a New Assistant

When you click "Create New," you are taken to the New Assistant configuration page. This is where you define the core components and behaviors of your new AI assistant, from the base models to the fine-grained parameters that control its performance.



3.3.1 Basic Settings

These are the fundamental components of your AI assistant.



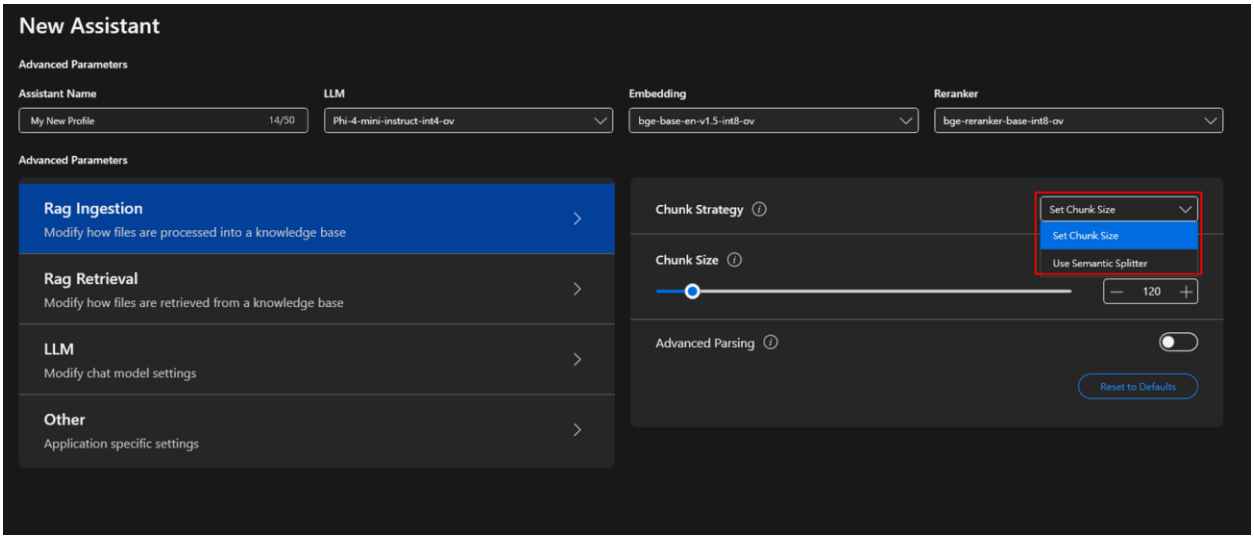
Parameter	Description
Assistant Name	A custom name for your new assistant profile. This name will appear in the Model Selection dropdown on the main chat screen.
LLM (Large	The core AI engine that generates human-like text to answer questions,

Language Model)	summarize documents, and engage in conversation. From the dropdown, select the base language model that will power this assistant.
Embedding	The model ~~used~~is used to convert text into numerical vectors, allowing the AI to understand semantic meaning and relationships. This is crucial for retrieving relevant information from the knowledge base.
Reranker	A secondary model that refines search results by re-evaluating the top documents retrieved by the Embedding model. It improves the accuracy of the context provided to the LLM for generating answers.

3.3.2 Advanced Parameters

This section allows you to fine-tune how the assistant ingests and retrieves information, and how the language model generates its final response.

Rag Ingestion
This section controls how your documents are processed and added to the knowledge base.

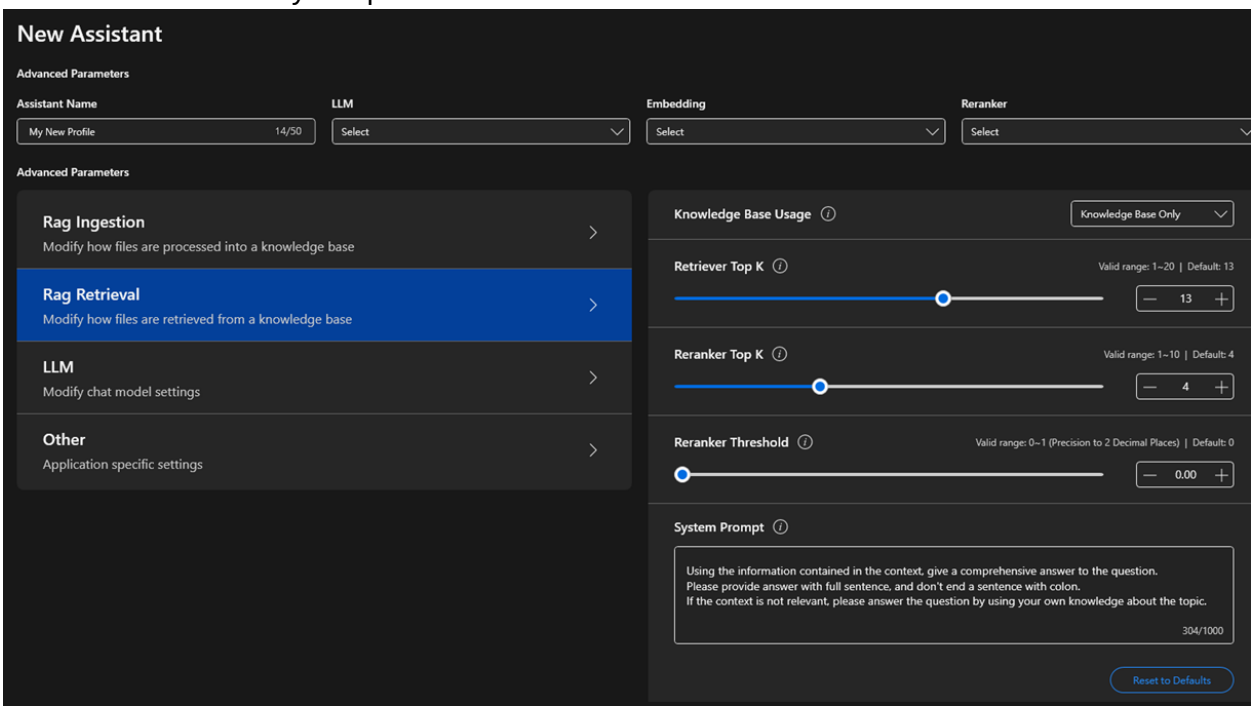


Parameter	Description
Chunk Strategy	Defines how large documents are split into smaller pieces ("chunks"). This is critical for how the material is indexed and searched. Options include "Set Chunk Size" or "Use Semantic Splitter".
Chunk Size	Sets the size of each text chunk. A smaller size may yield more granular results, while a larger size retains more context. The valid range is 50-

	1024 (Default: 120).
Advanced Parsing	When enabled, this improves data extraction from complex documents like PDFs by specifically identifying and processing tables, though it may increase upload time.

Rag Retrieval

This section defines how the AI searches the knowledge base to find the most relevant information to answer your queries.

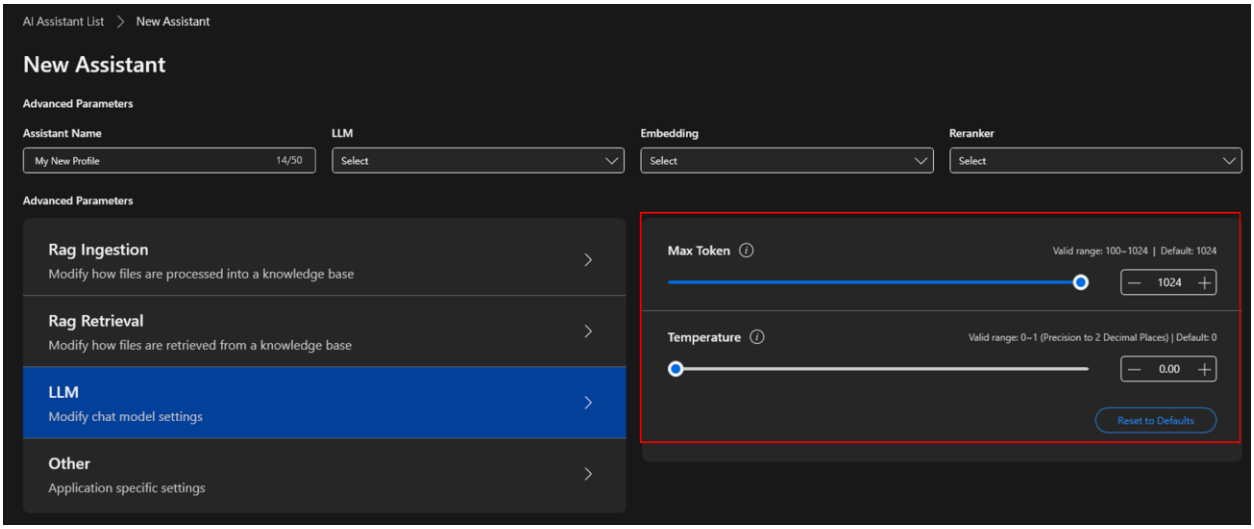


Parameter	Description
Retriever Top K	Controls how many of the most relevant documents ("chunks") are initially fetched from the knowledge base. A higher number provides more context but may increase processing time.
Reranker Top K	Determines how many of the initially retrieved documents are passed to the reranker for a more detailed relevance scoring. The reranker selects the best documents from this smaller subset.
Reranker Threshold	Sets a minimum relevance score for a reranked document to be considered valid context. Documents scoring below this are discarded.
System Prompt	A set of instructions that guides the LLM on its behavior, such as how

	to use the retrieved context, what tone to adopt, or what to do if the context is not relevant.

LLM

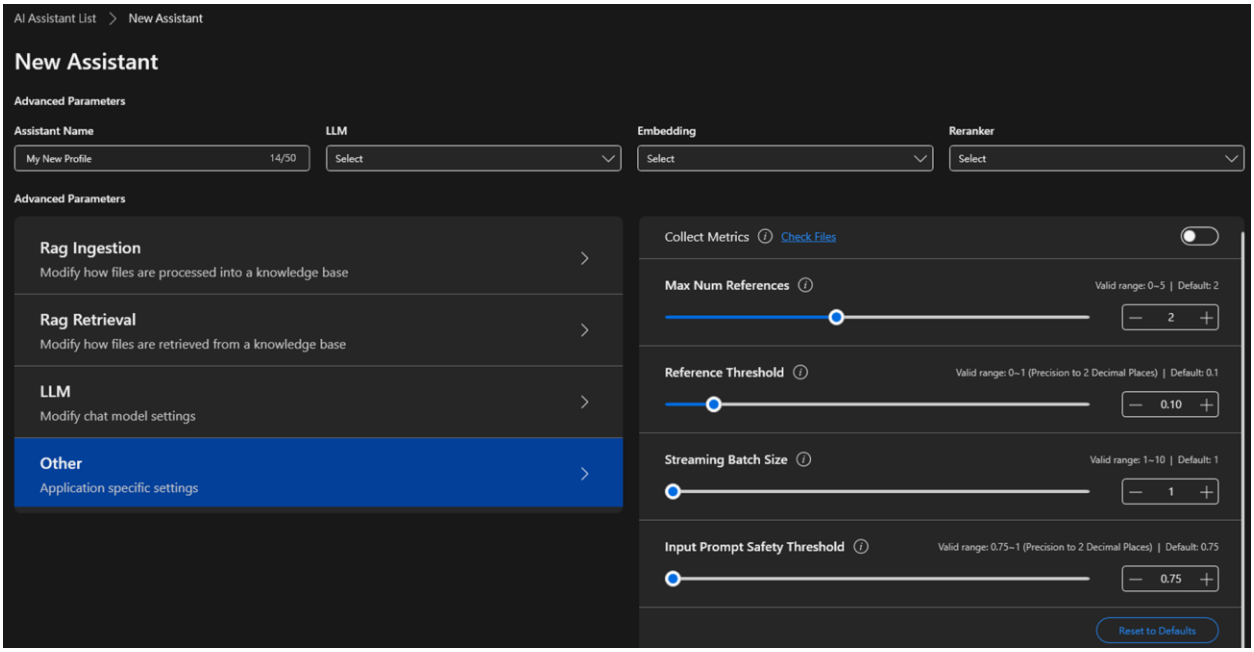
Here you can modify the chat model's settings to control the style and length of the AI's responses.



Parameter	Description
Max Token	Sets the maximum number of tokens (words or parts of words) that the model can generate in a single response, effectively controlling the answer's maximum length.
Temperature	Controls the randomness of the model's output. A lower value (e.g., 0.1) makes responses more deterministic and factual, while a higher value (e.g., 0.9) encourages more creativity.

Other

This section contains various application-specific settings for advanced control over the assistant's behavior and performance monitoring.



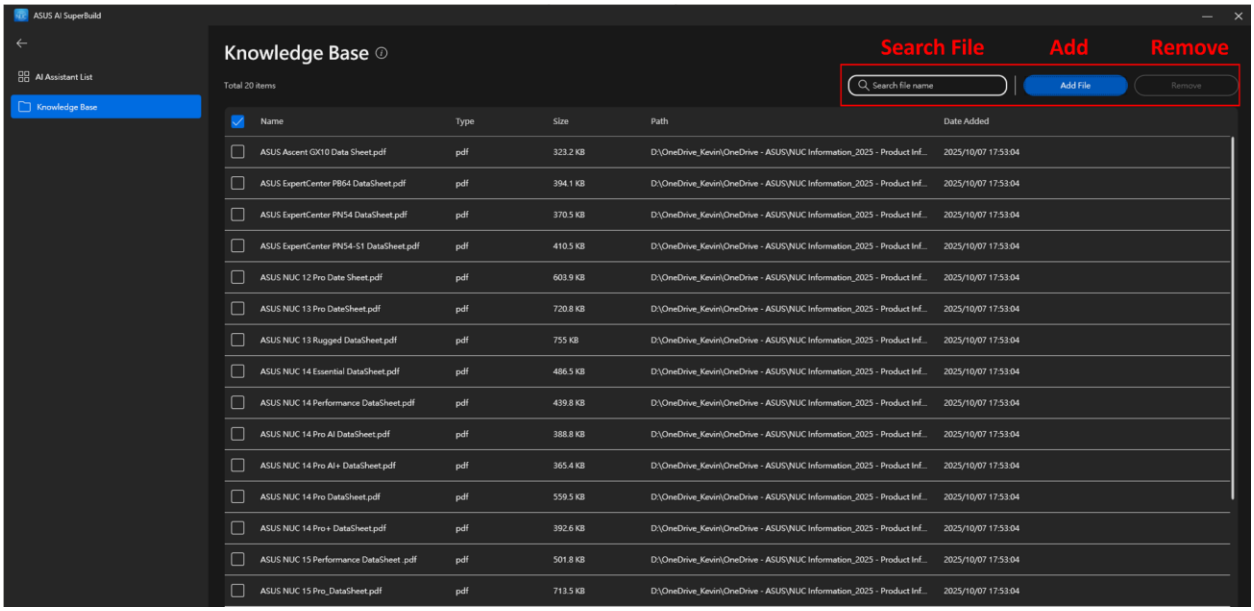
Parameter	Description
Collect Metrics	Enables or disables the collection of performance and usage data for analysis.
Max Num References	Sets the maximum number of source documents from the knowledge base that can be cited in a single response.
Reference Threshold	Defines the minimum relevance score a document must have to be considered as a citable reference in an answer.
Streaming Batch Size	Controls the number of tokens processed in each batch when generating a response in streaming mode.
Input Prompt Safety Threshold	Sets the confidence level for the content safety filter applied to user inputs. Prompts deemed unsafe will be blocked.

3.4 Editing an Existing Assistant

In addition to creating new assistants, you can easily modify existing configurations. On the **AI Assistant List** page, hover over the assistant you wish to modify, click the "..." icon that appears on the right, and select **Edit**. This will open the same detailed parameter page as when creating a new assistant, allowing you to fine-tune all basic and advanced settings for that specific assistant.

3.5 Knowledge Base Management

The Knowledge Base is where you manage the source files for Retrieval-Augmented Generation (RAG). By uploading documents (PDF, DOCX, TXT, PPTX.), you provide the AI assistant with a specific set of information to draw from, ensuring more accurate and context-aware responses grounded in your data.



Button/UI Element	Description
Add File	Opens a file browser to upload one or more supported documents to the knowledge base.
Remove	Deletes the selected file(s) from the knowledge base.
Search for file name...	A search bar to quickly find specific files in the knowledge base by name.

3.6 Evaluation and Report

Determining the optimal combination of LLMs and Knowledge Base files can be challenging. The Evaluation feature streamlines this process by benchmarking your system's performance against a standard dataset.

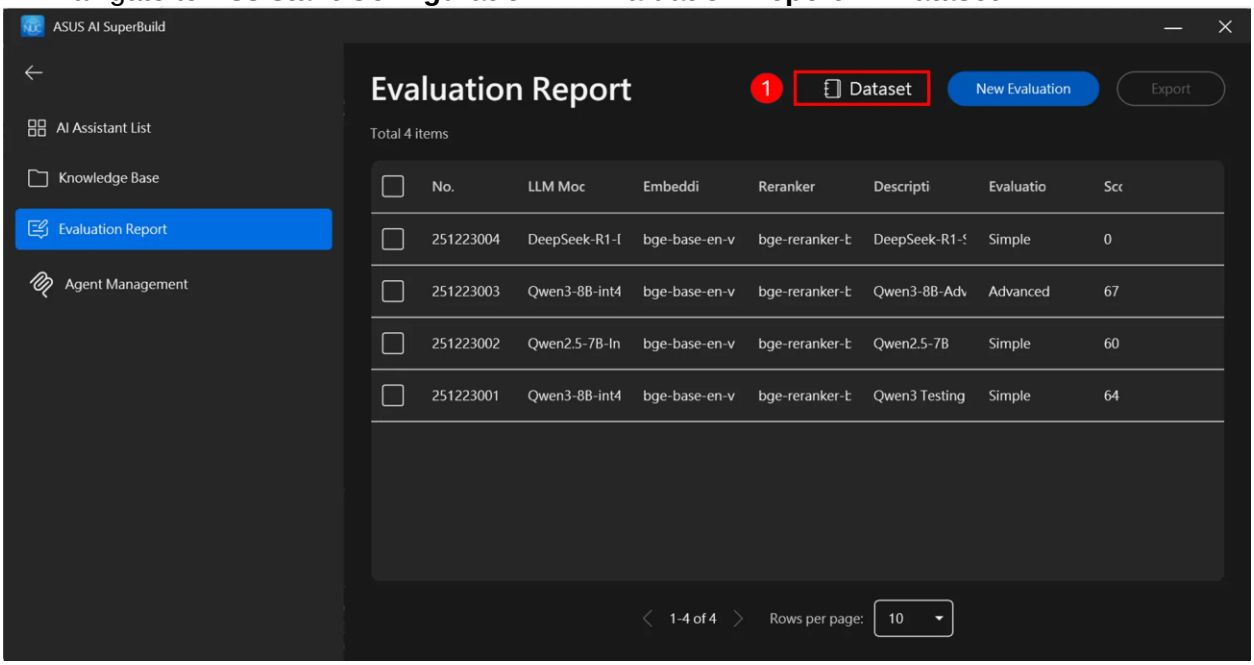
Two evaluation modes are available:

- **Simple Mode:** Quickly calculates an accuracy score based on uploaded question-answer pairs to benchmark the current LLM and Knowledge Base.
- **Advanced Mode:** Leverages cloud-based AI to generate comprehensive reports with actionable improvement suggestions.

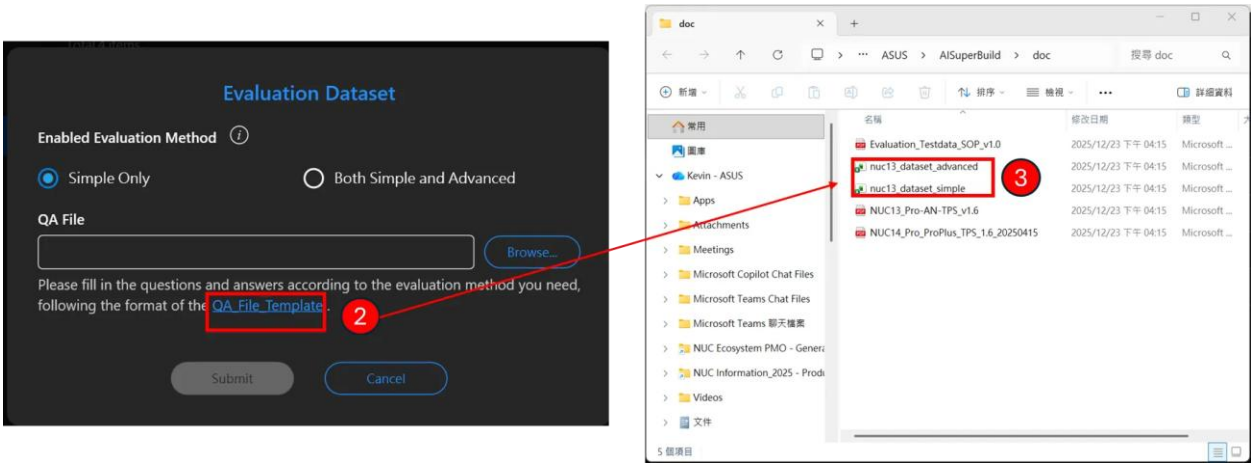
3.6.1 Preparing the Evaluation Dataset

To ensure data privacy and relevance, you must provide a custom evaluation dataset containing domain-specific questions and ground-truth answers. This dataset serves as the benchmark for performance evaluation.

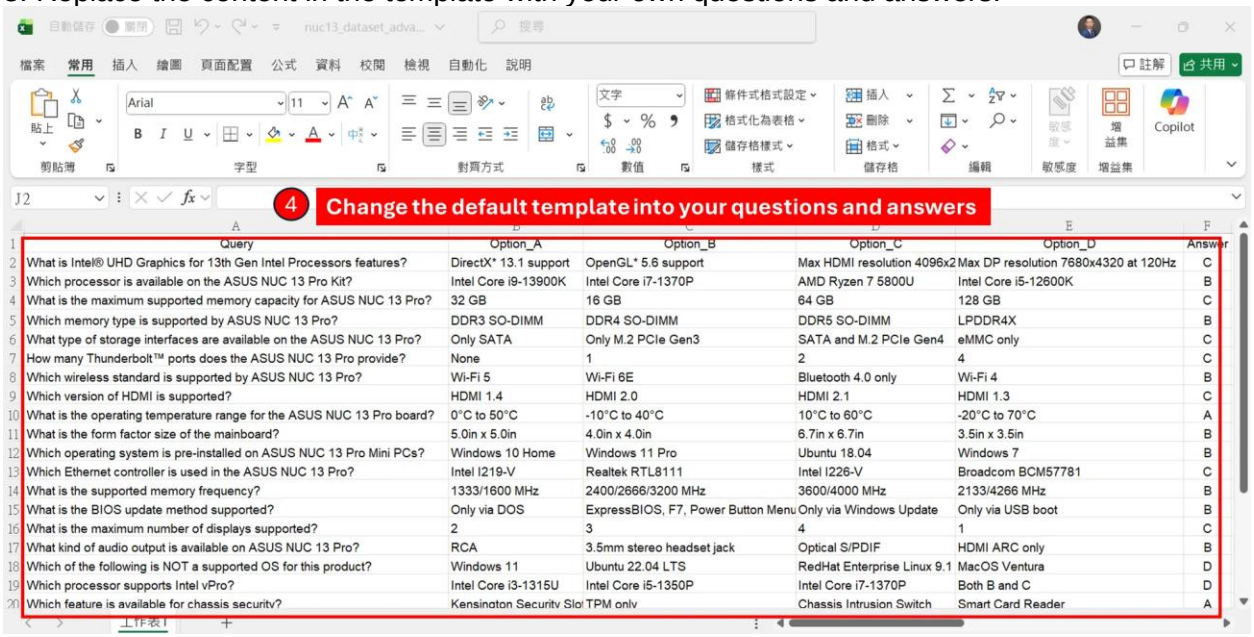
1. Navigate to **Assistant Configuration >> Evaluation Report >> Dataset**



2. Click **QA_File_Template** to download the example file.

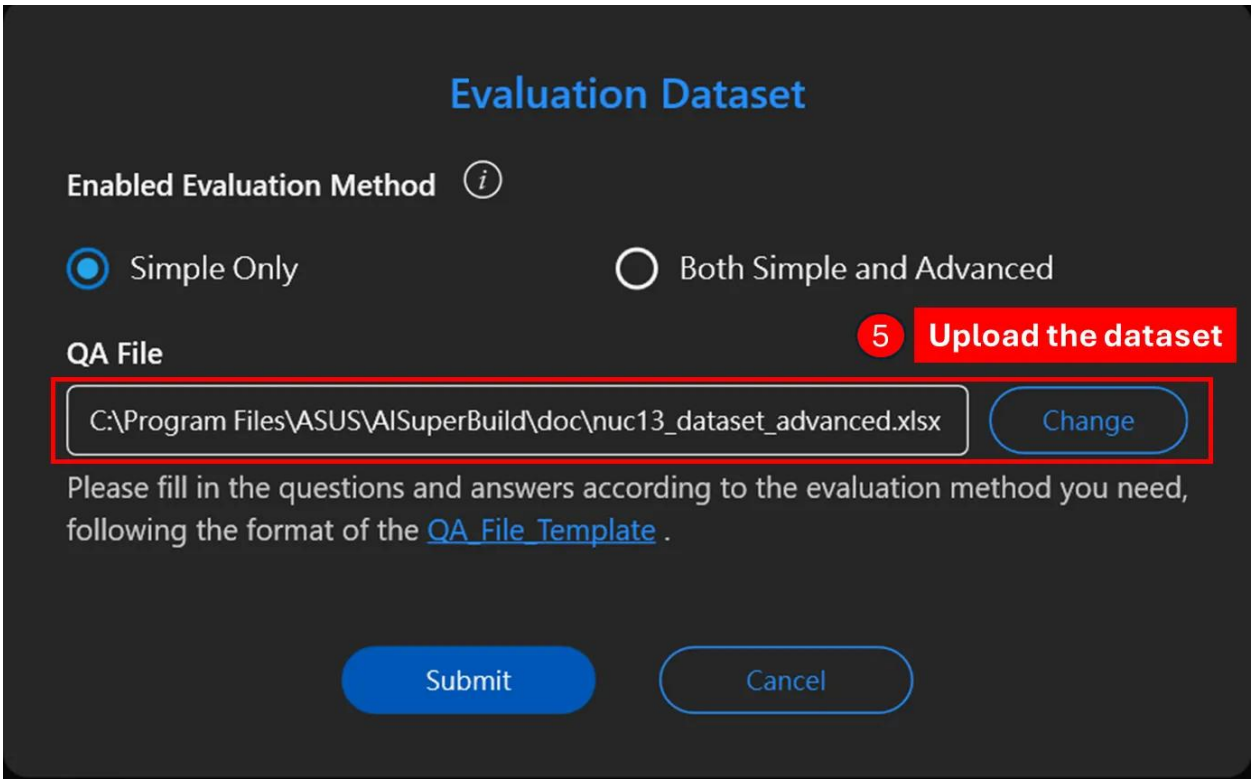


3. Replace the content in the template with your own questions and answers.



Query	Option_A	Option_B	Option_C	Option_D	Answer
What is Intel® UHD Graphics for 13th Gen Intel Processors features?	DirectX* 13.1 support	OpenGL* 5.6 support	Max HDMI resolution 4096x2	Max DP resolution 7680x4320 at 120Hz	C
Which processor is available on the ASUS NUC 13 Pro Kit?	Intel Core i9-13900K	Intel Core i7-1370P	AMD Ryzen 7 5800U	Intel Core i5-12600K	B
What is the maximum supported memory capacity for ASUS NUC 13 Pro?	32 GB	16 GB	64 GB	128 GB	C
Which memory type is supported by ASUS NUC 13 Pro?	DDR3 SO-DIMM	DDR4 SO-DIMM	DDR5 SO-DIMM	LPDDR4X	B
What type of storage interfaces are available on the ASUS NUC 13 Pro?	Only SATA	Only M.2 PCIe Gen3	SATA and M.2 PCIe Gen4	eMMC only	C
How many Thunderbolt™ ports does the ASUS NUC 13 Pro provide?	None	1	2	4	C
Which wireless standard is supported by ASUS NUC 13 Pro?	Wi-Fi 5	Wi-Fi 6E	Bluetooth 4.0 only	Wi-Fi 4	B
Which version of HDMI is supported?	HDMI 1.4	HDMI 2.0	HDMI 2.1	HDMI 1.3	C
What is the operating temperature range for the ASUS NUC 13 Pro board?	0°C to 50°C	-10°C to 40°C	10°C to 60°C	-20°C to 70°C	A
What is the form factor size of the mainboard?	5.0in x 5.0in	4.0in x 4.0in	6.7in x 6.7in	3.5in x 3.5in	B
Which operating system is pre-installed on ASUS NUC 13 Pro Mini PCs?	Windows 10 Home	Windows 11 Pro	Ubuntu 18.04	Windows 7	B
Which Ethernet controller is used in the ASUS NUC 13 Pro?	Intel I219-V	Realtek RTL8111	Intel I226-V	Broadcom BCM57781	C
What is the supported memory frequency?	1333/1600 MHz	2400/2666/3200 MHz	3600/4000 MHz	2133/4266 MHz	B
What is the BIOS update method supported?	Only via DOS	ExpressBIOS, F7, Power Button Menu	Only via Windows Update	Only via USB boot	B
What is the maximum number of displays supported?	2	3	4	1	C
What kind of audio output is available on ASUS NUC 13 Pro?	RCA	3.5mm stereo headset jack	Optical S/PDIF	HDMI ARC only	B
Which of the following is NOT a supported OS for this product?	Windows 11	Ubuntu 22.04 LTS	RedHat Enterprise Linux 9.1	MacOS Ventura	D
Which processor supports Intel vPro?	Intel Core i3-1315U	Intel Core i5-1350P	Intel Core i7-1370P	Both B and C	D
Which feature is available for chassis security?	Kensington Security Slot	TPM only	Chassis Intrusion Switch	Smart Card Reader	A

4. Click **"Change"** to upload your updated dataset file, then click **"Submit"**.

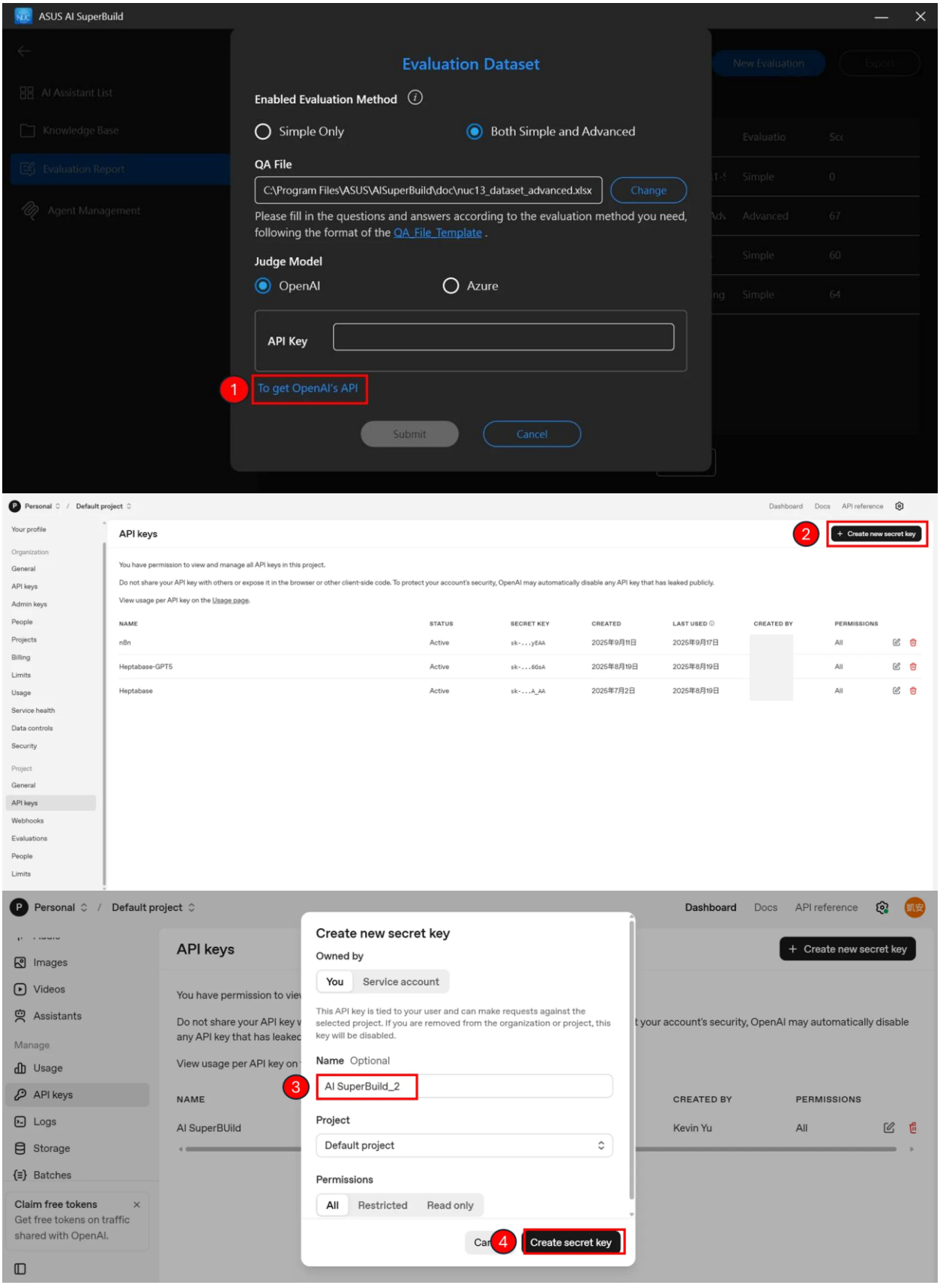


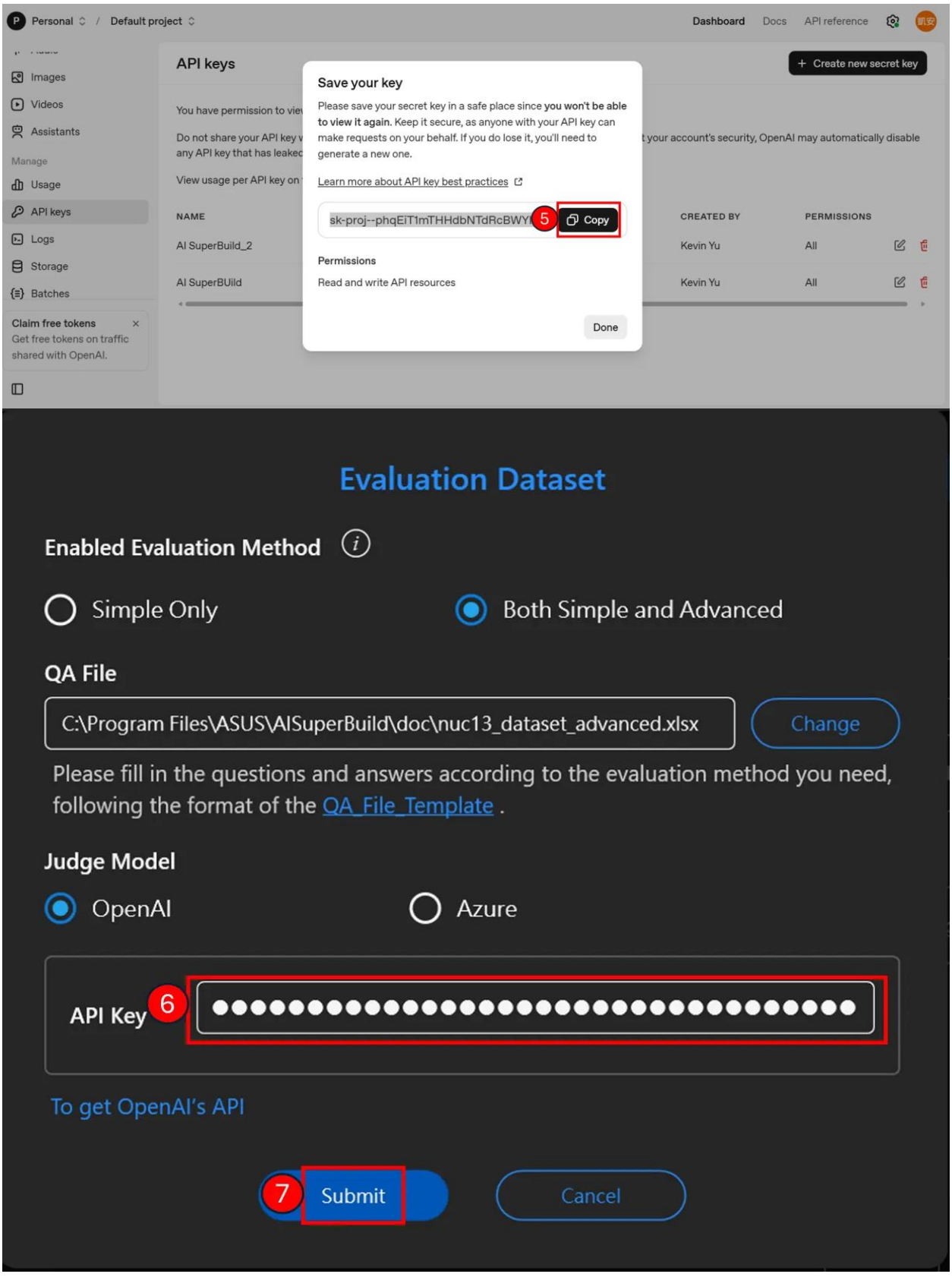
Evaluation Configuration

- For basic benchmarking, select **Simple Only**.
- For detailed analysis, select **Both Simple and Advanced**. This requires a cloud-based LLM API key (OpenAI or Azure OpenAI).

OpenAI API Configuration

1. Click the **To get OpenAI's API** link to obtain your API key.
2. Enter the key into the **API Key** field in AI SuperBuild.


ASUS AI SuperBuild
Evaluation Dataset
Enabled Evaluation Method
Simple Only
Both Simple and Advanced
QA File
C:\Program Files\ASUS\AISuperBuild\doc\nuc13_dataset_advanced.xlsx
Change
Please fill in the questions and answers according to the evaluation method you need, following the format of the QA_File_Template .
Judge Model
OpenAI
Azure
API Key
1
To get OpenAI's API
Submit
Cancel
New Evaluation
Export
AI Assistant List
Knowledge Base
Evaluation Report
Agent Management
Evaluatio
Sc
Simple
0
Advanced
67
Simple
60
Simple
64
Personal / Default project
Dashboard
Docs
API reference
Your profile
Organization
General
API keys
Admin keys
People
Projects
Billing
Limits
Usage
Service health
Data controls
Security
Project
General
API keys
Webhooks
Evaluations
People
Limits
API keys
2
+ Create new secret key
You have permission to view and manage all API keys in this project.
Do not share your API key with others or expose it in the browser or other client-side code. To protect your account's security, OpenAI may automatically disable any API key that has leaked publicly.
View usage per API key on the Usage page.
NAME
STATUS
SECRET KEY
CREATED
LAST USED
CREATED BY
PERMISSIONS
n8n
Active
sk-...yEAA
2025年9月11日
2025年9月17日
All
Heptabase-GPT5
Active
sk-...6GsA
2025年8月19日
2025年8月19日
All
Heptabase
Active
sk-...A_AA
2025年7月2日
2025年8月19日
All
Personal / Default project
Dashboard
Docs
API reference
Images
Videos
Assistants
Manage
Usage
API keys
Logs
Storage
Batches
Claim free tokens
Get free tokens on traffic shared with OpenAI.
API keys
+ Create new secret key
You have permission to vie
Do not share your API key v
any API key that has leaked
View usage per API key on
your account's security, OpenAI may automatically disable
NAME
CREATED BY
PERMISSIONS
AI SuperBUild
Kevin Yu
All
Create new secret key
Owned by
You
Service account
This API key is tied to your user and can make requests against the selected project. If you are removed from the organization or project, this key will be disabled.
Name Optional
3
AI SuperBuild_2
Project
Default project
Permissions
All
Restricted
Read only
Can
4
Create secret key

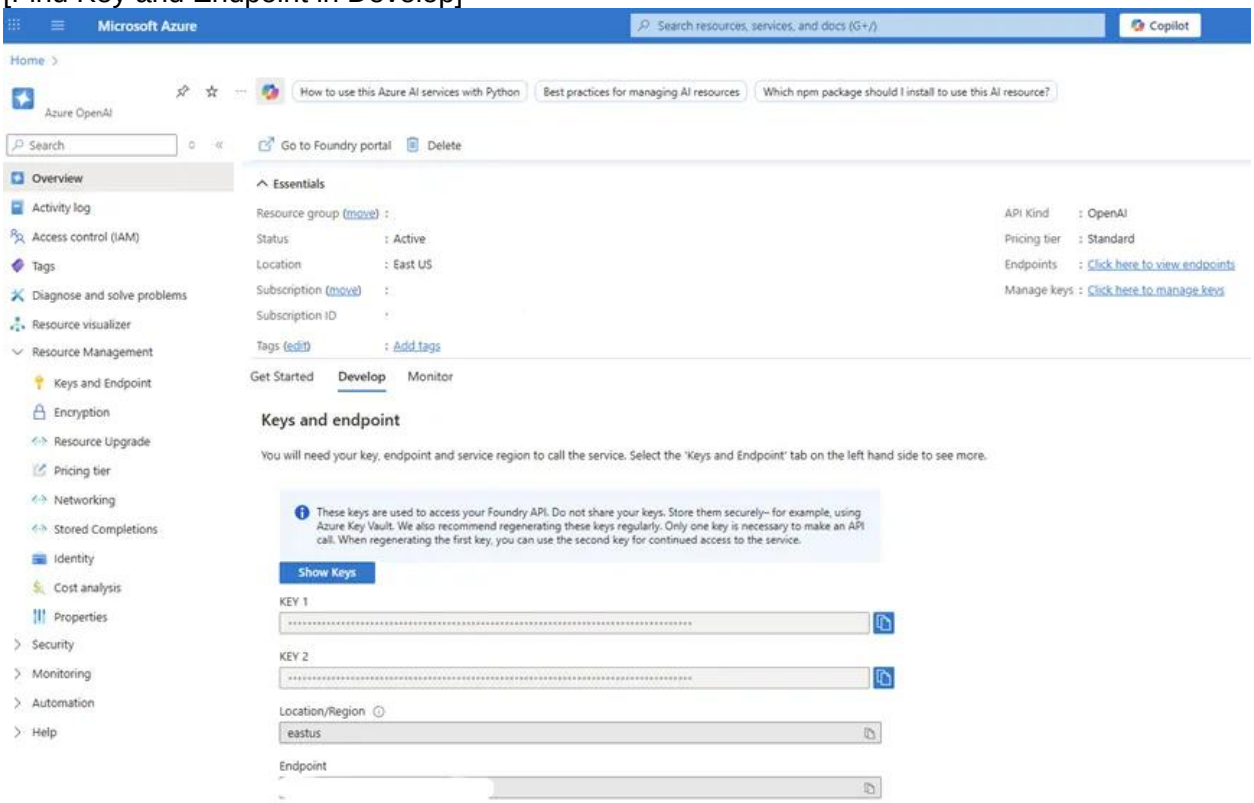

Personal / Default project
Dashboard
Docs
API reference
API keys
+ Create new secret key
Images
Videos
Assistants
Manage
Usage
API keys
Logs
Storage
Batches
Claim free tokens
Get free tokens on traffic shared with OpenAI.
Save your key
Please save your secret key in a safe place since you won't be able to view it again. Keep it secure, as anyone with your API key can make requests on your behalf. If you do lose it, you'll need to generate a new one.
Learn more about API key best practices
sk-proj--phqEiT1mTHHdbNTdRcBWY
5
Copy
Permissions
Read and write API resources
Done
NAME
CREATED BY
PERMISSIONS
AI SuperBuild_2
Kevin Yu
All
AI SuperBUild
Kevin Yu
All
Evaluation Dataset
Enabled Evaluation Method
Simple Only
Both Simple and Advanced
QA File
C:\Program Files\ASUS\AISuperBuild\doc\nuc13_dataset_advanced.xlsx
Change
Please fill in the questions and answers according to the evaluation method you need, following the format of the QA_File_Template .
Judge Model
OpenAI
Azure
API Key
6
To get OpenAI's API
7
Submit
Cancel

Azure OpenAI

To use Azure OpenAI with AI SuperBuild, you must provide both your **API key** and **endpoint**. You can obtain this information from the Azure OpenAI portal (ai.azure.com). After that, create and deploy the required model (currently, only **gpt-4.1-mini** is supported for evaluation). Follow the steps below in the Azure AI portal:

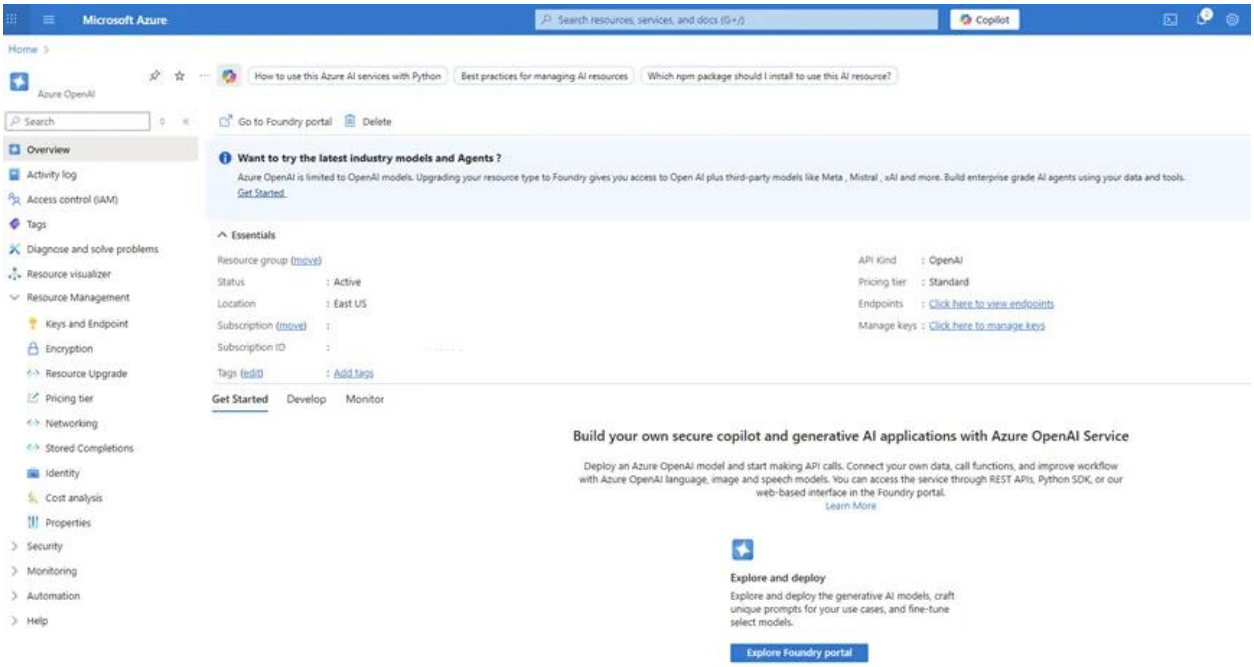
1. Go to the **Develop** section to find your **Key** and **Endpoint**.

[Find Key and Endpoint in Develop]



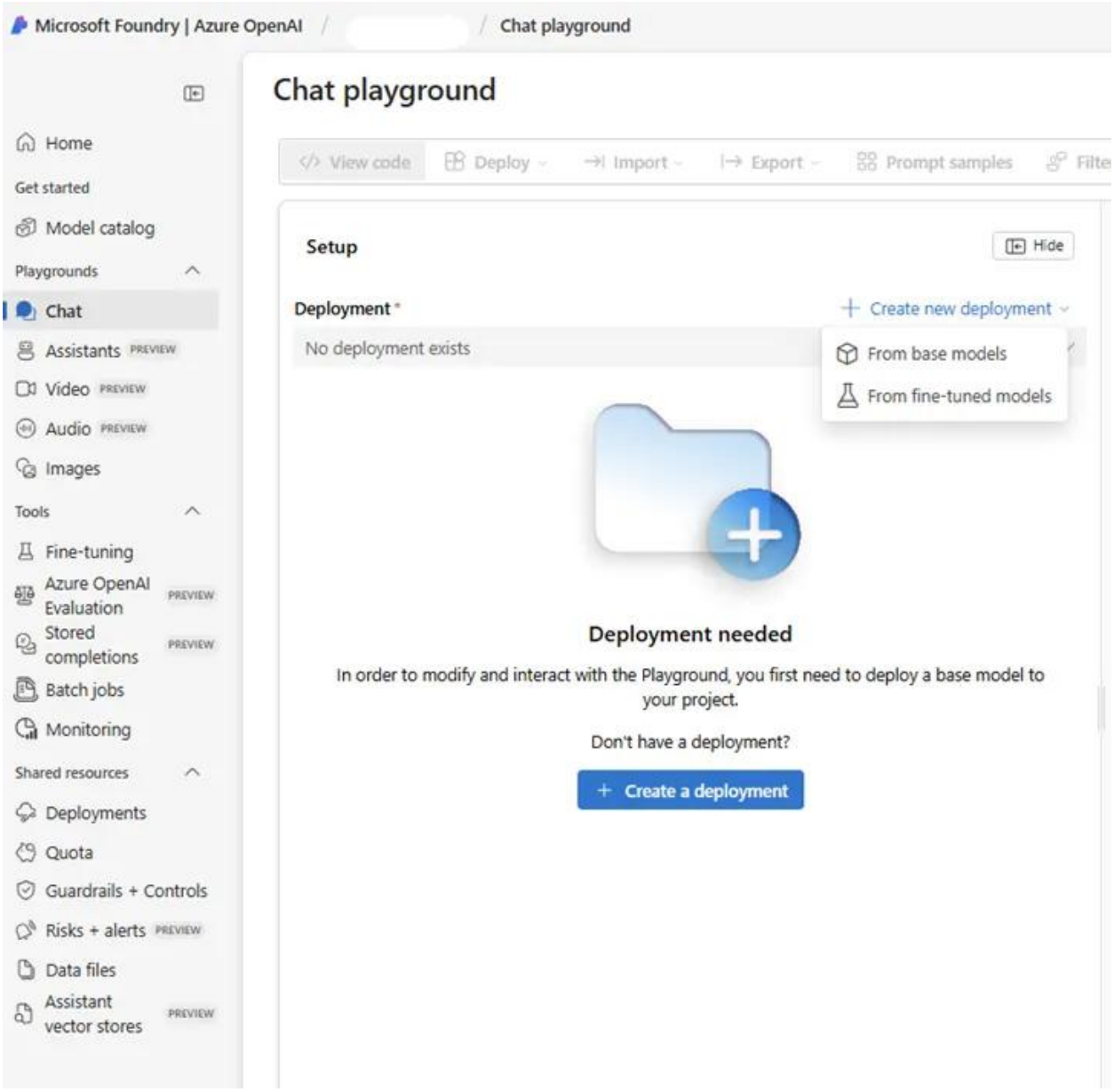
2. Go to **Get started** and enter the **Explore** experience to deploy a model.

[Enter Explore and deploy in Get started]

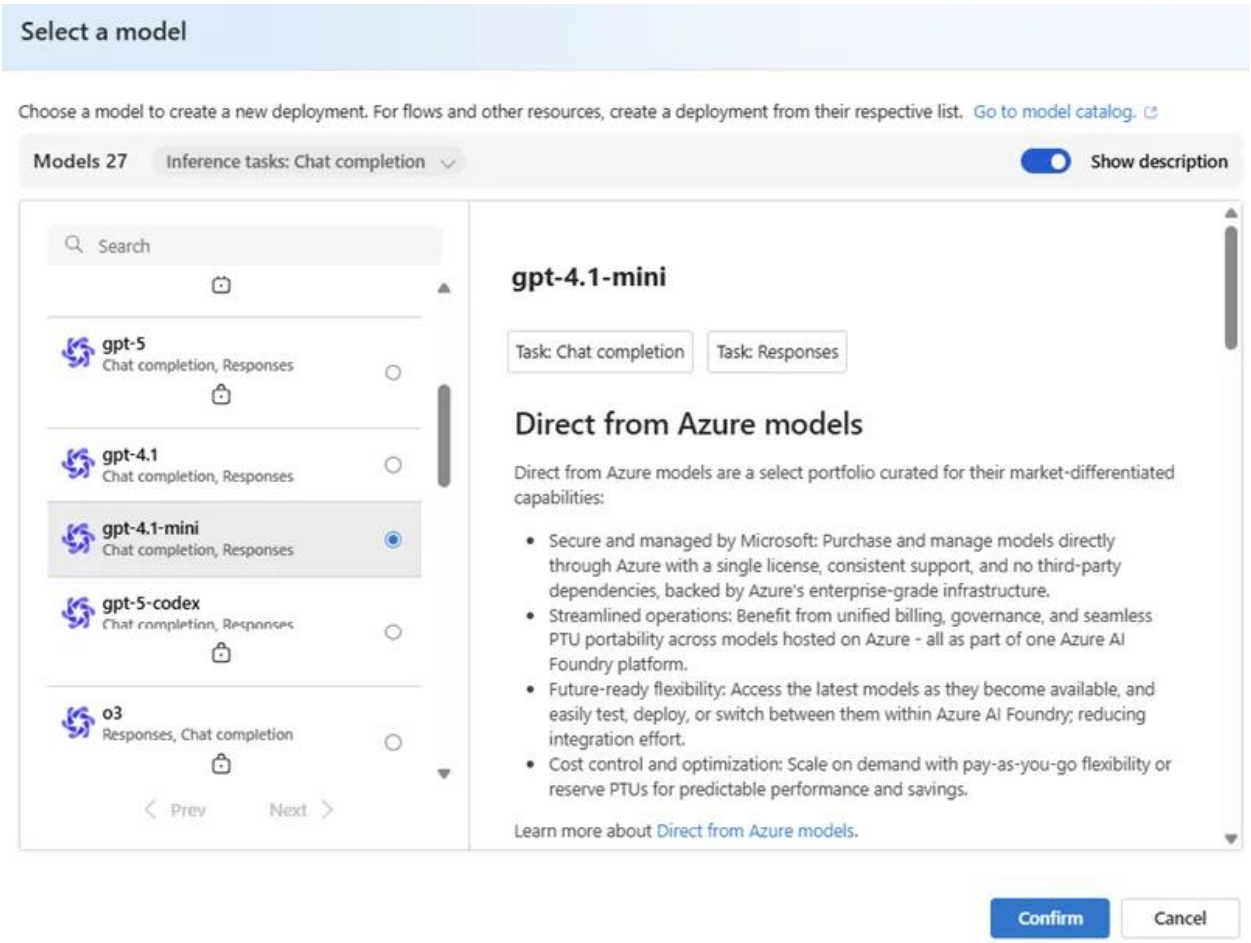


3. In the **Chat** section, select **Create a deployment**.

[Tap Create a deployment in Chat]

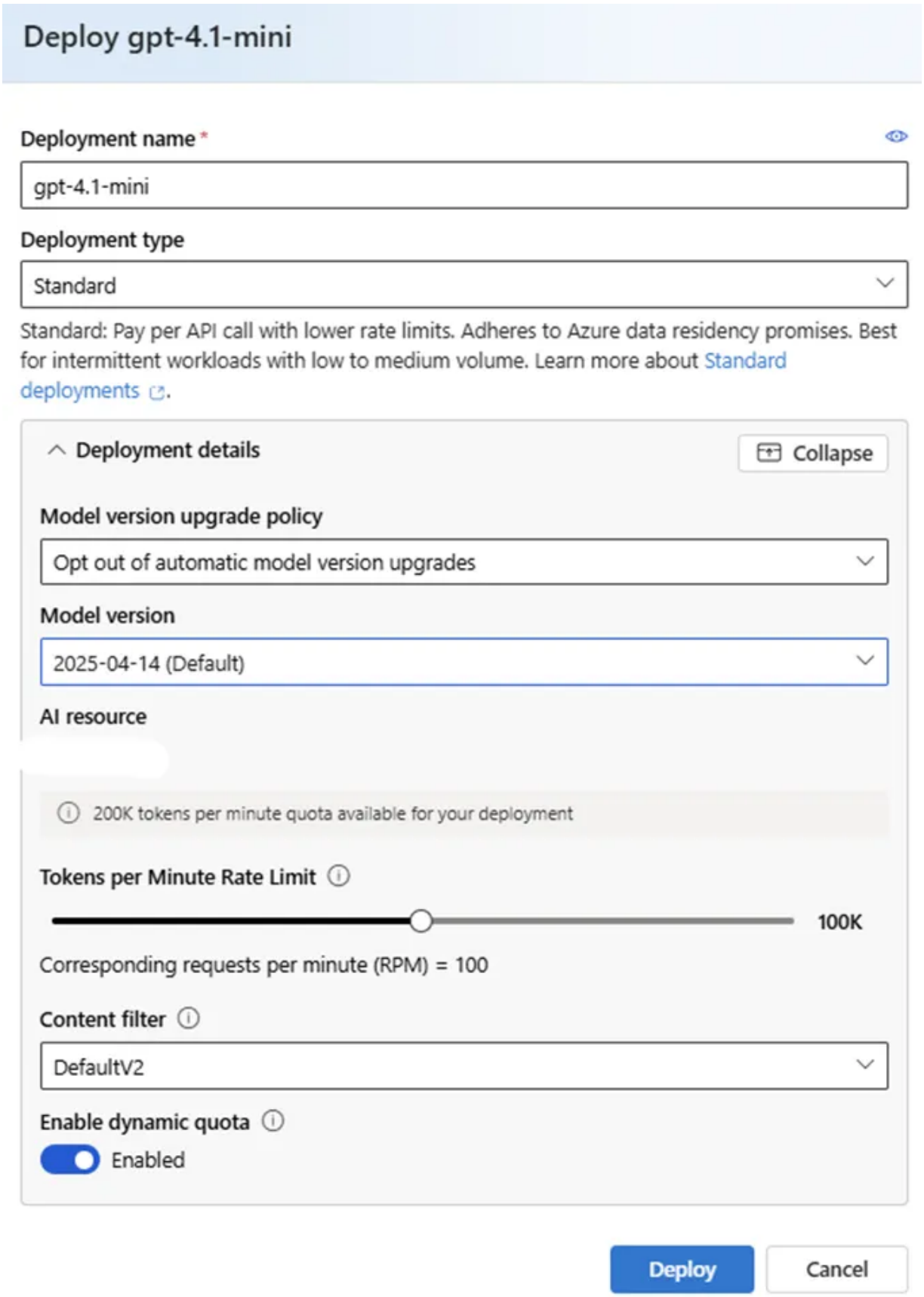


4. Choose **gpt-4.1-mini** as the model and confirm.
[Select gpt-4.1-mini and tap Confirm]

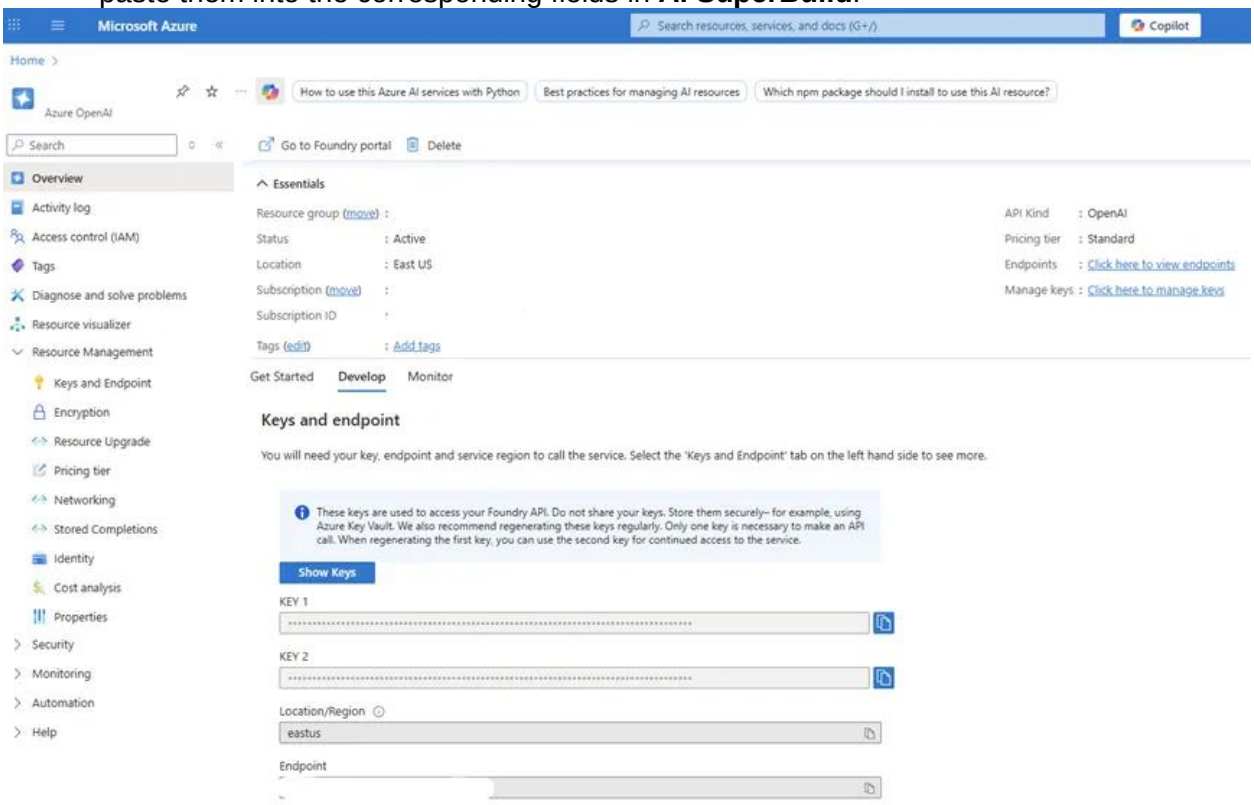


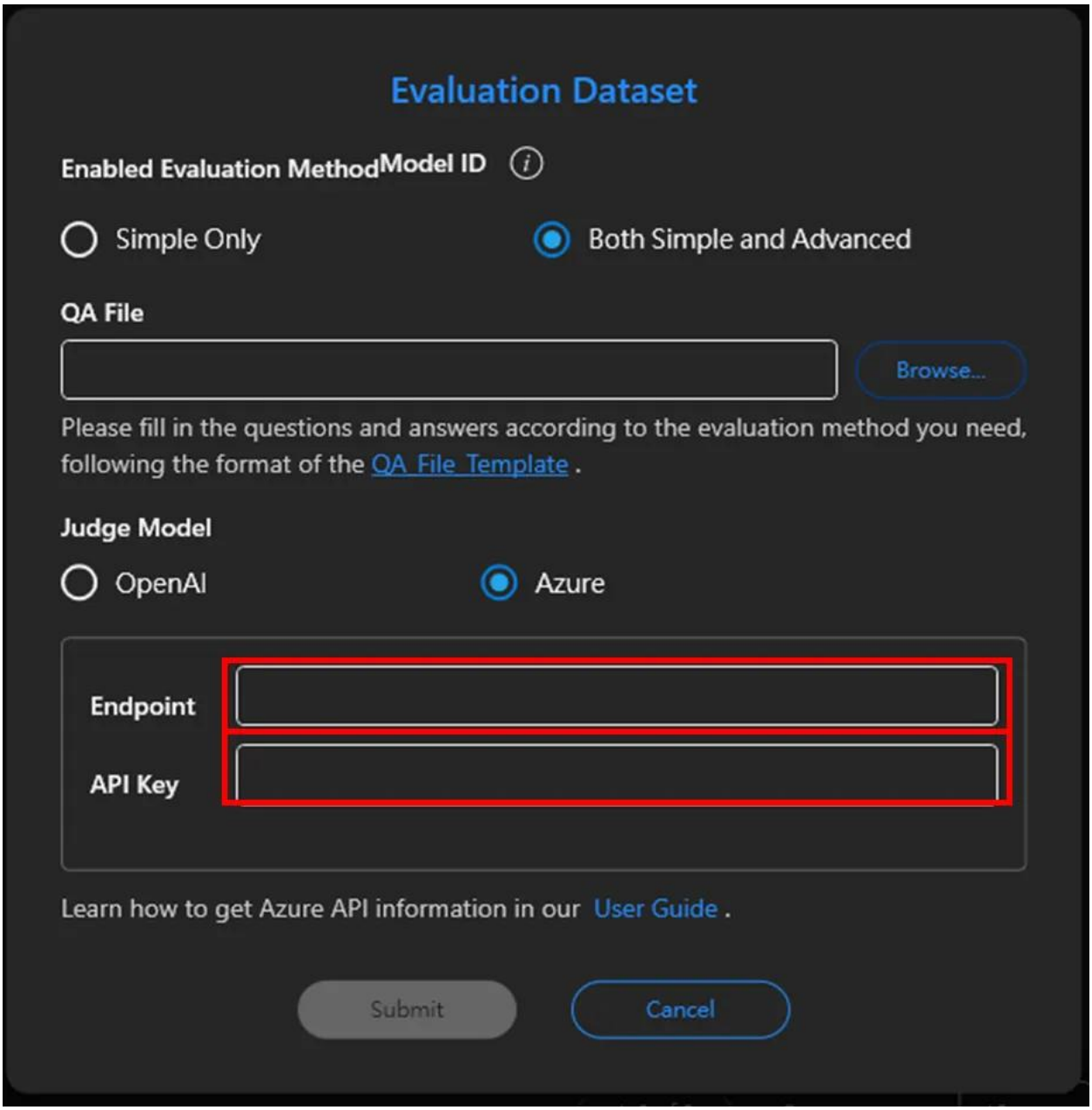
5. Configure the deployment:
   - **Deployment type** → *Standard*
   - **Model version upgrade policy** → *Opt out of automatic model version upgrades*

Then select **Deploy**.
[Select Deployment type → Standard, Model version upgrade policy → Opt out of automatic model version upgrades, Tap Deploy]


Deploy gpt-4.1-mini
Deployment name *
gpt-4.1-mini
Deployment type
Standard
Standard: Pay per API call with lower rate limits. Adheres to Azure data residency promises. Best for intermittent workloads with low to medium volume. Learn more about Standard deployments.
Deployment details
Collapse
Model version upgrade policy
Opt out of automatic model version upgrades
Model version
2025-04-14 (Default)
AI resource
200K tokens per minute quota available for your deployment
Tokens per Minute Rate Limit
100K
Corresponding requests per minute (RPM) = 100
Content filter
DefaultV2
Enable dynamic quota
Enabled
Deploy
Cancel

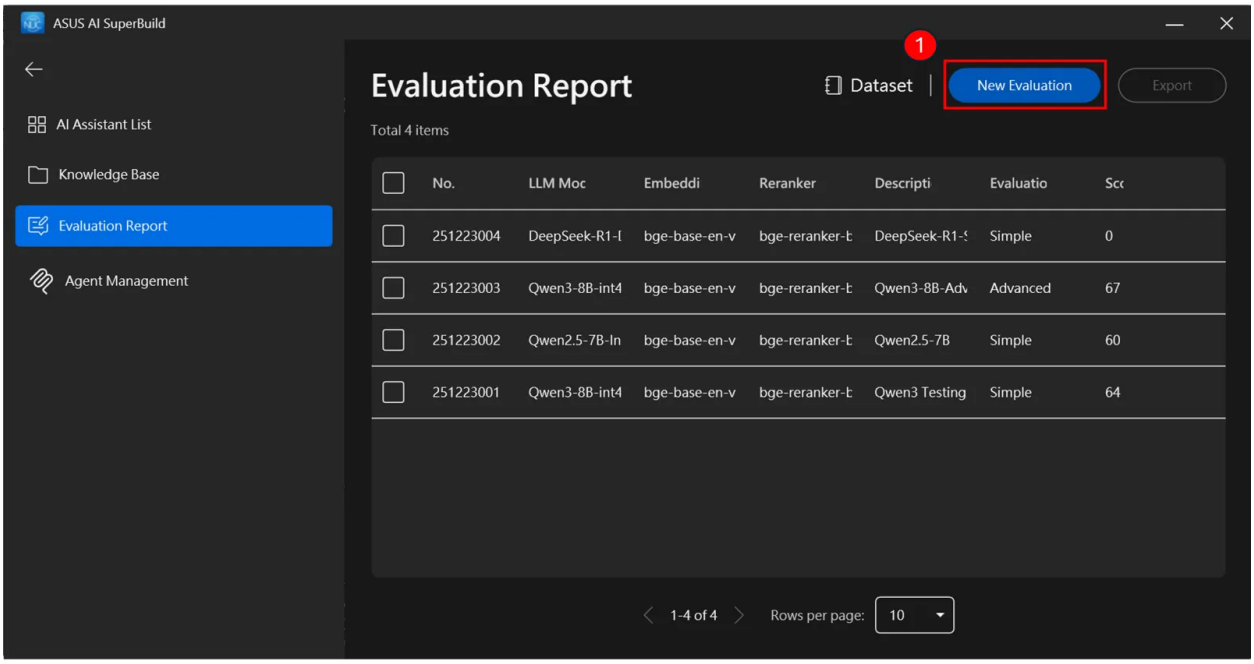
6. After the deployment is created, copy the **Key** and **Endpoint** from the Azure portal and paste them into the corresponding fields in **AI SuperBuild**.



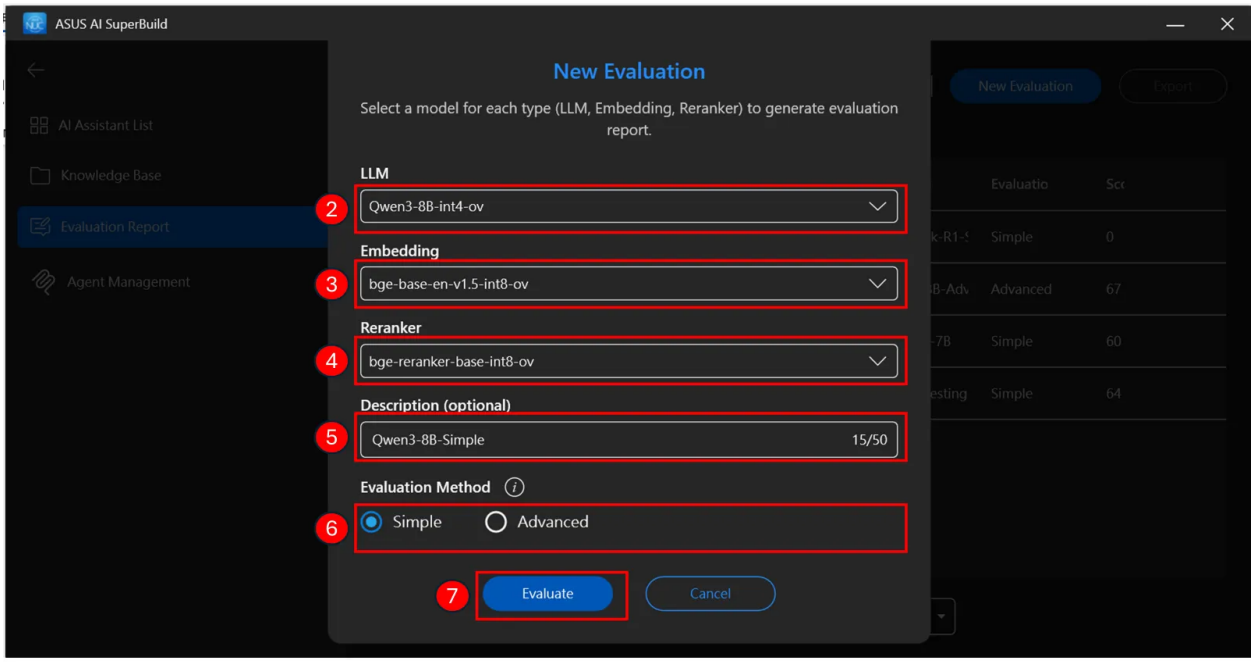


3.6.2 Starting the Evaluation

After the dataset is configured, you can evaluate whether the **LLM**, **Embedding**, and **Reranker** models—together with the knowledge base—are sufficient for the target customer scenario. Navigate to **"New Evaluation"**.

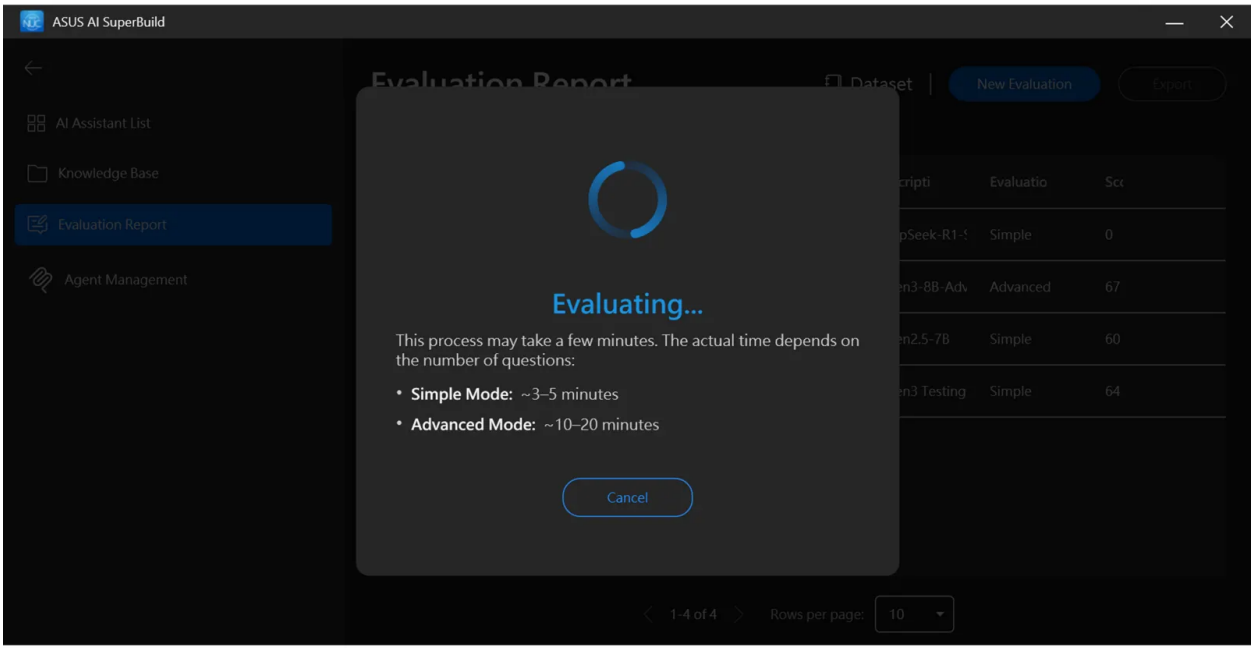


Select the **LLM**, **Embedding**, and **Reranker** models you want to use.

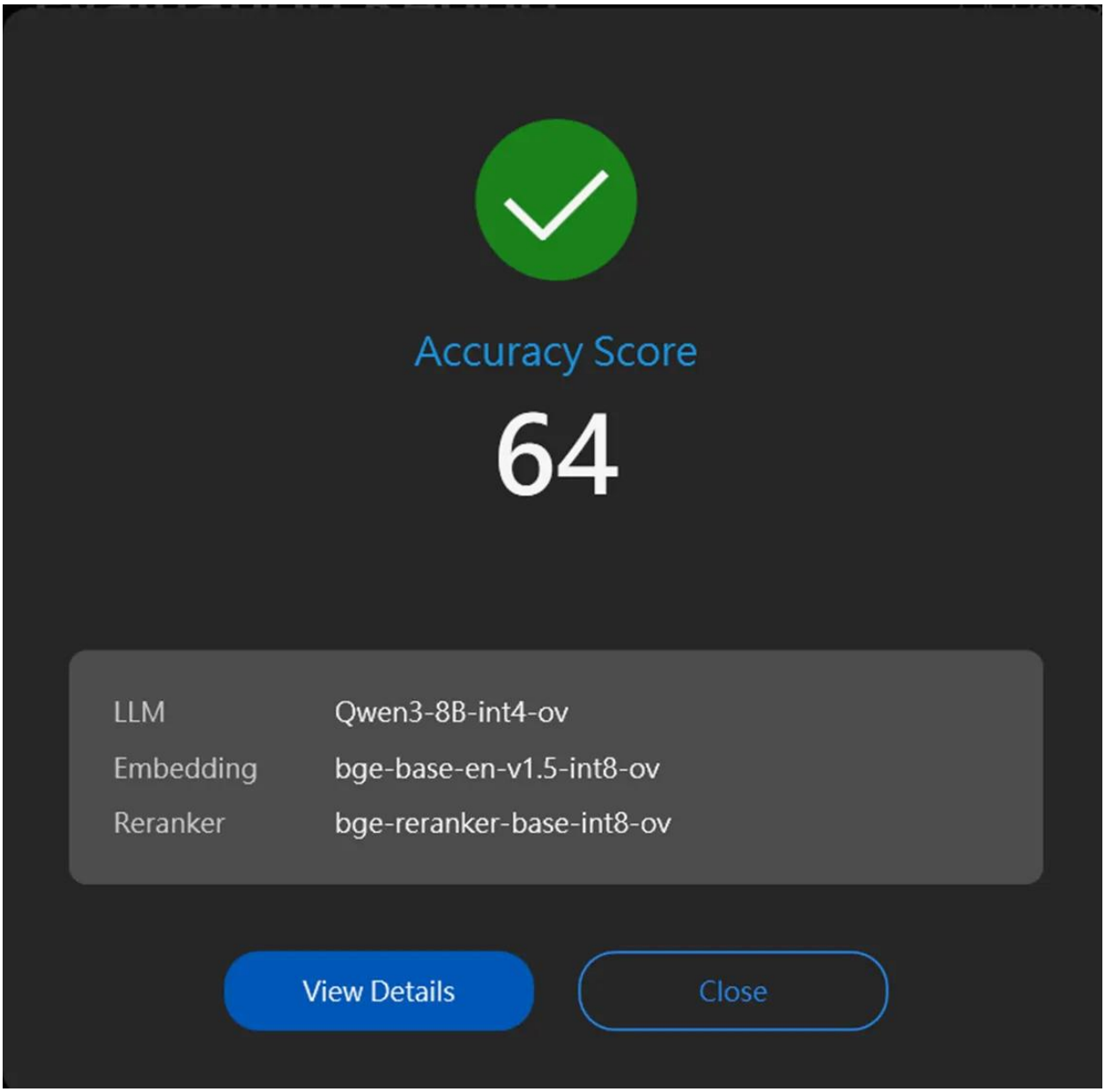


Based on the selected evaluation mode, the system begins processing the results.

- **Simple mode** typically takes around **3–5 minutes**.
- **Advanced mode** typically takes around **10–20 minutes**, depending on the number of questions and the selected models.

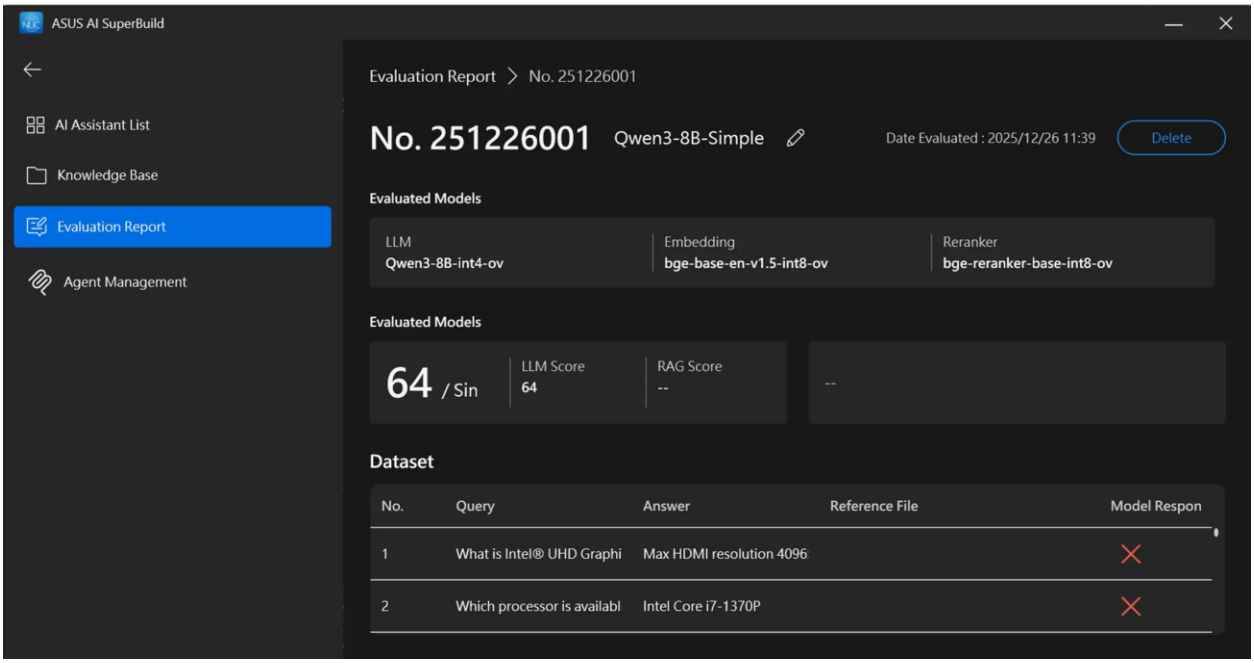


After the evaluation is complete, AI SuperBuild returns an accuracy score that summarizes the overall performance.



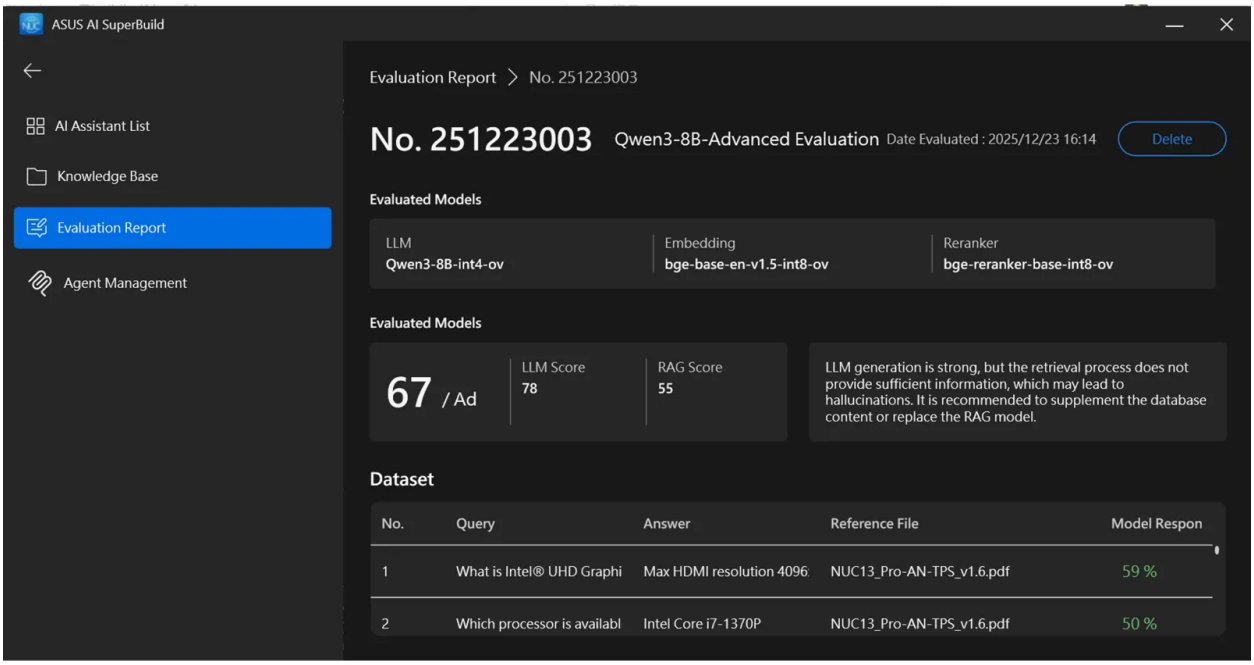
Simple mode results show:

- The overall score
- Whether each question is answered correctly or not



Advanced mode results show:

- The overall accuracy
- Detailed improvement suggestions, such as:
  - Switching to a more powerful model
  - Adding more or better-curated knowledge base files

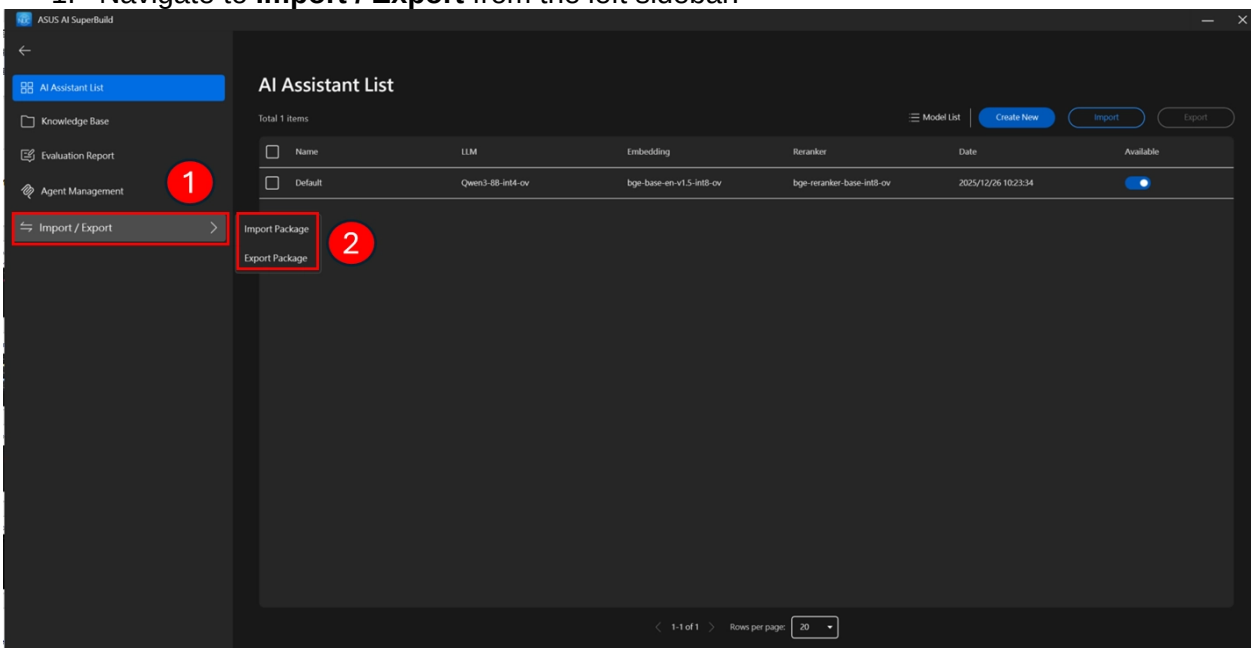


3.7 One-Key Import/Export

This feature allows users to back up and restore all configuration files with a single click, making it easy to migrate settings across devices or recover from changes.

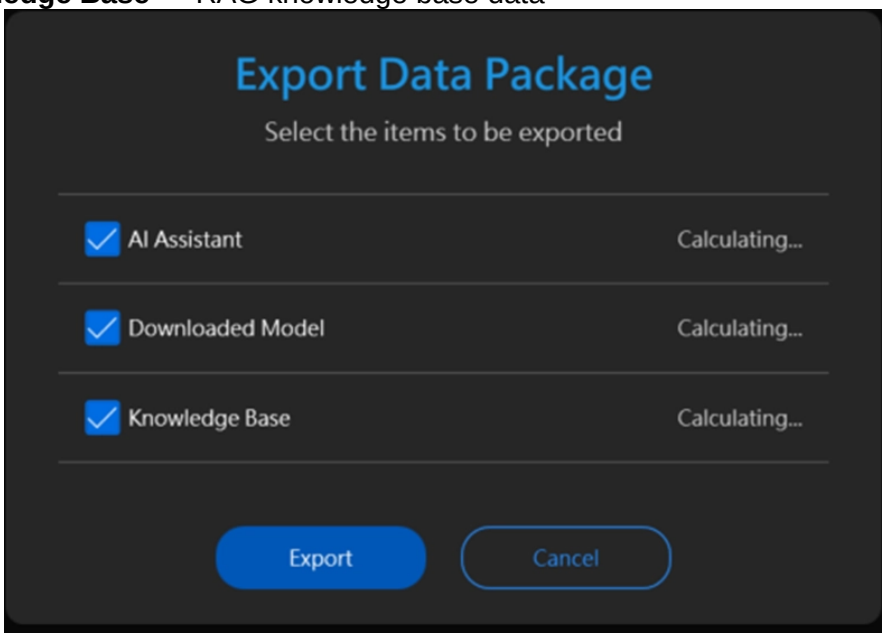
3.7.1 Export Configuration

1. Navigate to **Import / Export** from the left sidebar.

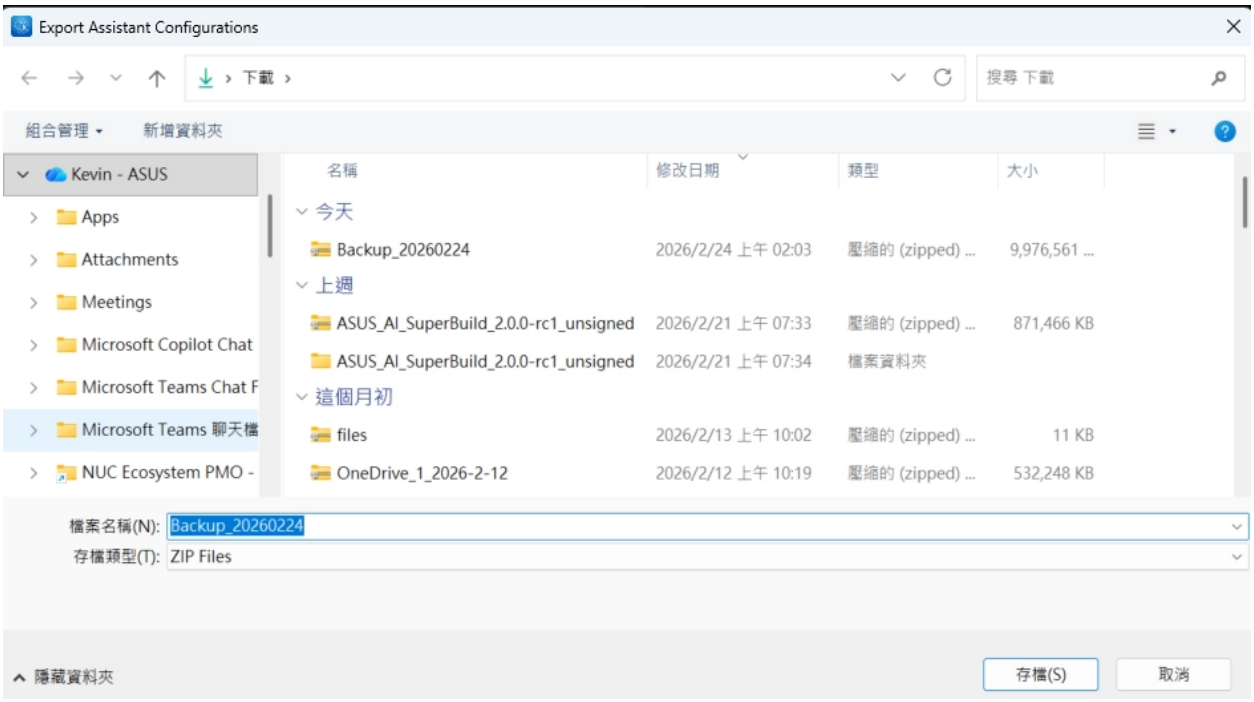


2. Click **“Export Package”** to back up your configuration files. By default, the system will export the following items:

- **AI Assistant** — all agent configurations
- **Downloaded Model** — locally downloaded LLM models
- **Knowledge Base** — RAG knowledge base data



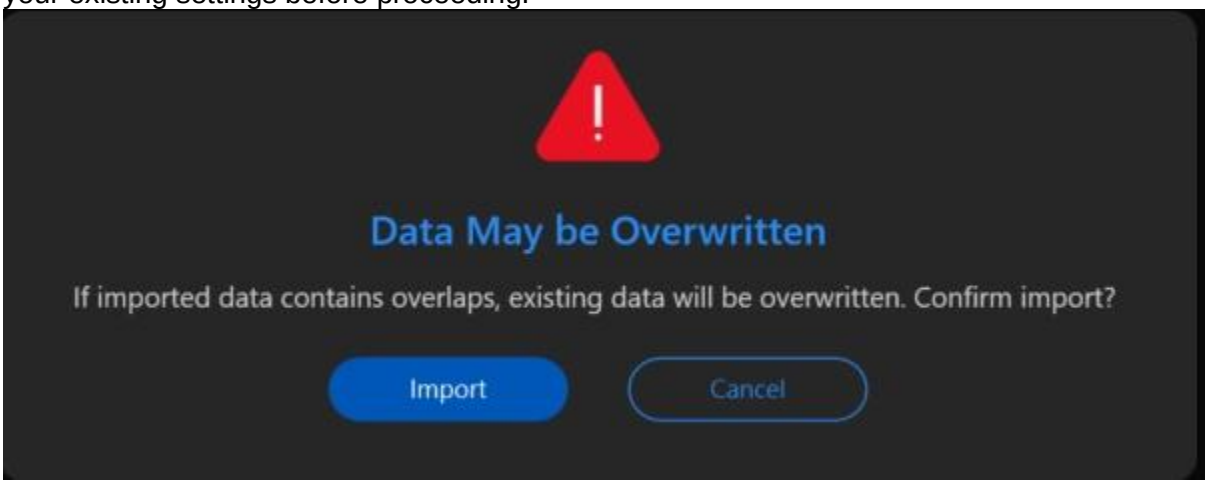
3. Choose a destination folder, then click **“Export”**. The configuration package will be saved to the selected location.



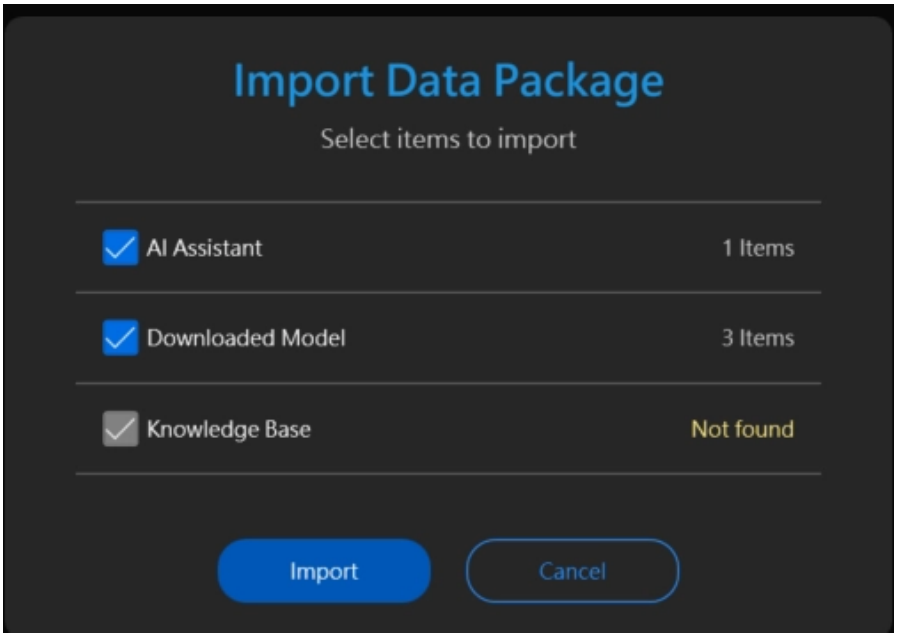
3.7.2 Import Configuration

1. Click **“Import Package”** and select an existing configuration file to restore.

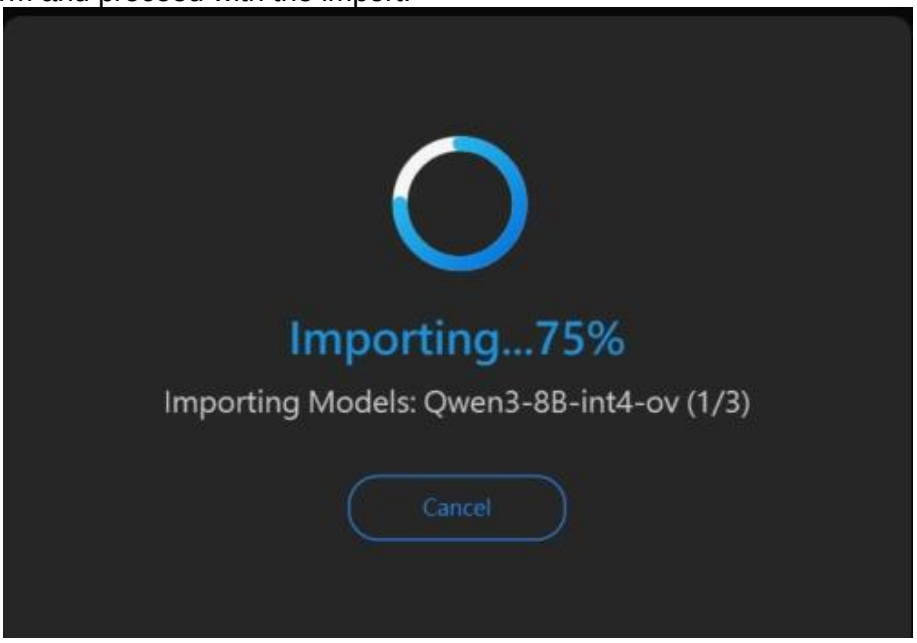
Note: Importing a configuration file will **overwrite** the current data. Please make sure to back up your existing settings before proceeding.



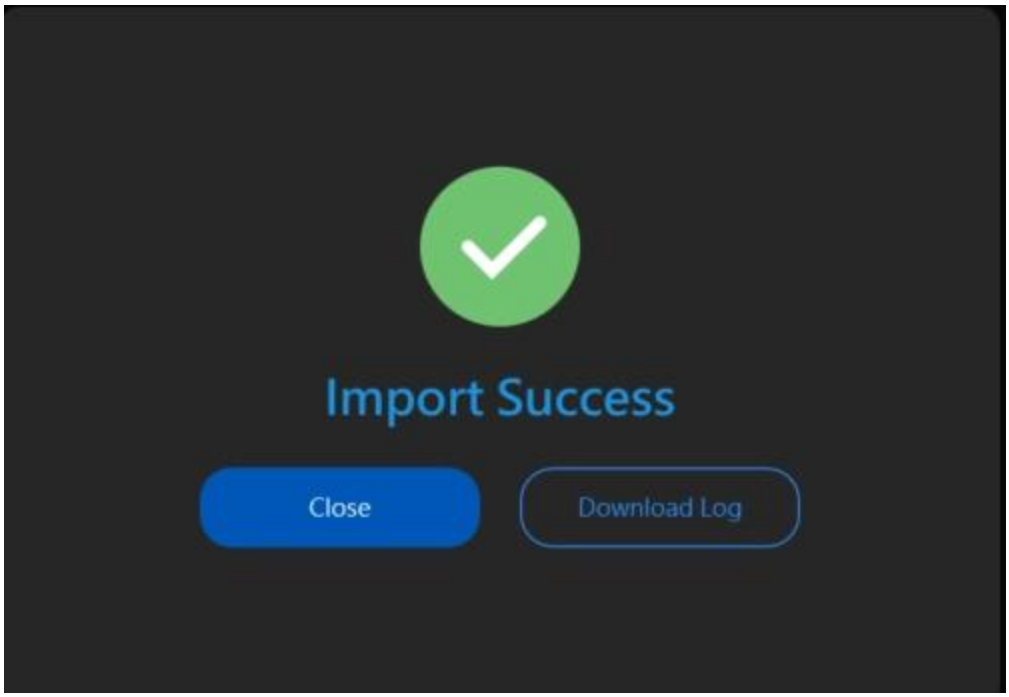
2. After analyzing the file, the system will display a summary of how many configuration items will be overwritten.



3. Confirm and proceed with the import.



4. Once the import is complete, a pop-up window will display the results.

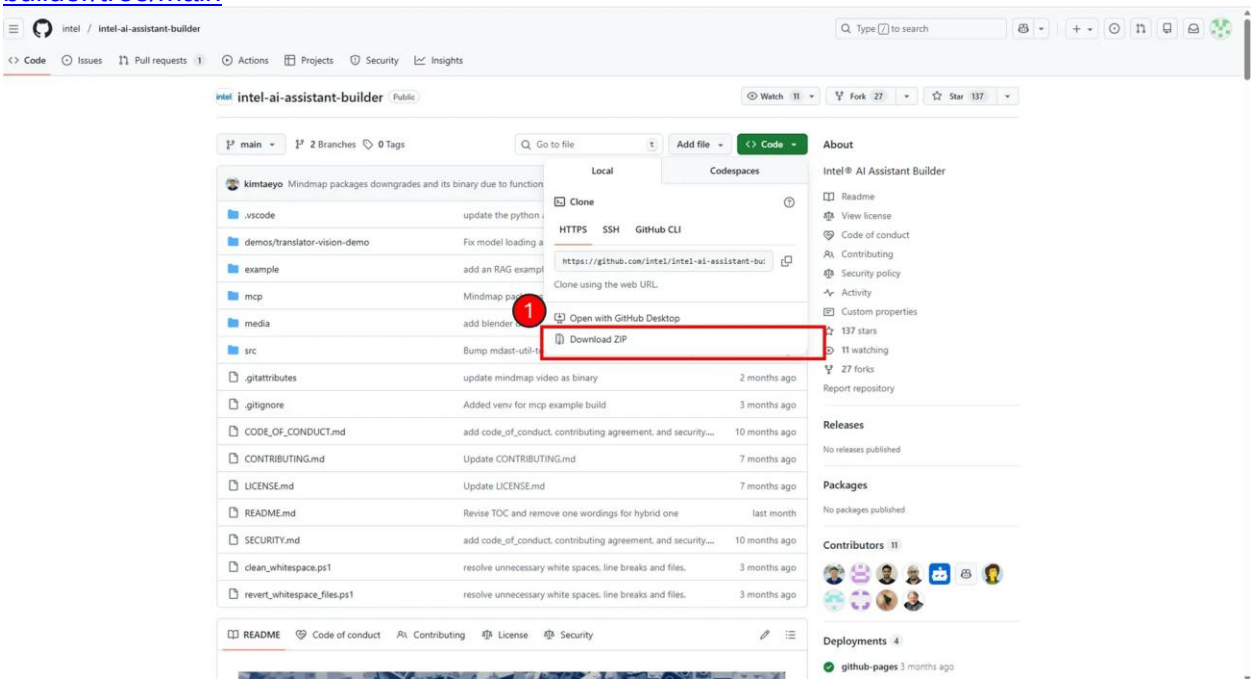

Import Success
Close
Download Log

Chapter 4: MCP (Model-Context-Protocol)

You can extend the LLM's capabilities using MCP (Model Context Protocol) by connecting it to your own MCP server to access custom services.

4.1 Example: Build a Hotel Assistant MCP Server

1. Install Python v3.10.x or above.
2. Verify that Python is installed correctly with the command:
   python --version
3. Download the example files from GitHub https://github.com/intel/intel-ai-assistant-builder/tree/main

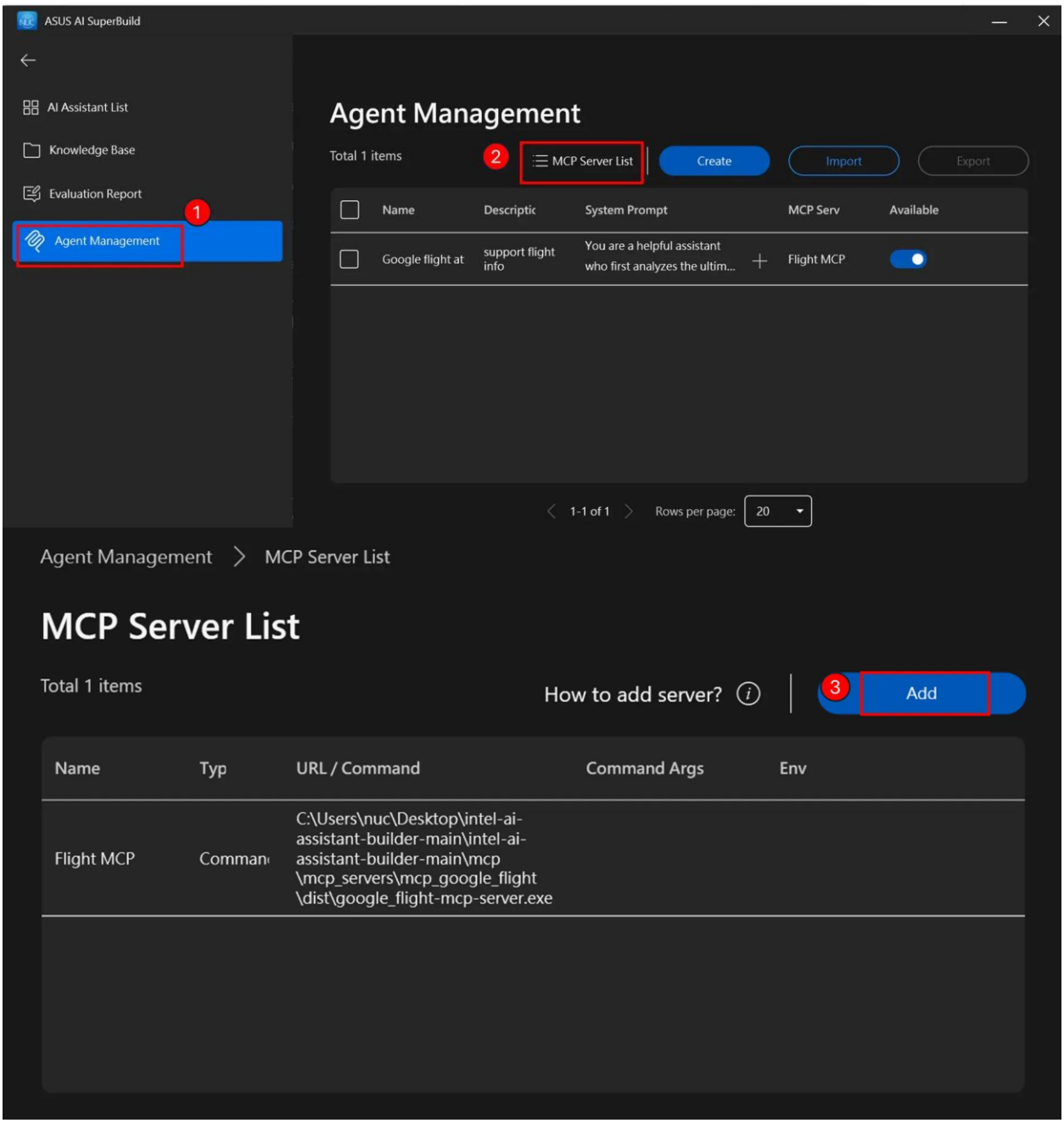


4. Open a command line window and navigate to the project folder. (The exact path will be different depending on your environment and where you saved the files.)
   cd C:\\Users\\nuc\\Desktop\\intel-ai-assistant-builder-main\\intel-ai-assistant-builder-main\\mcp\\mcp_servers\\mcp_google_hotel

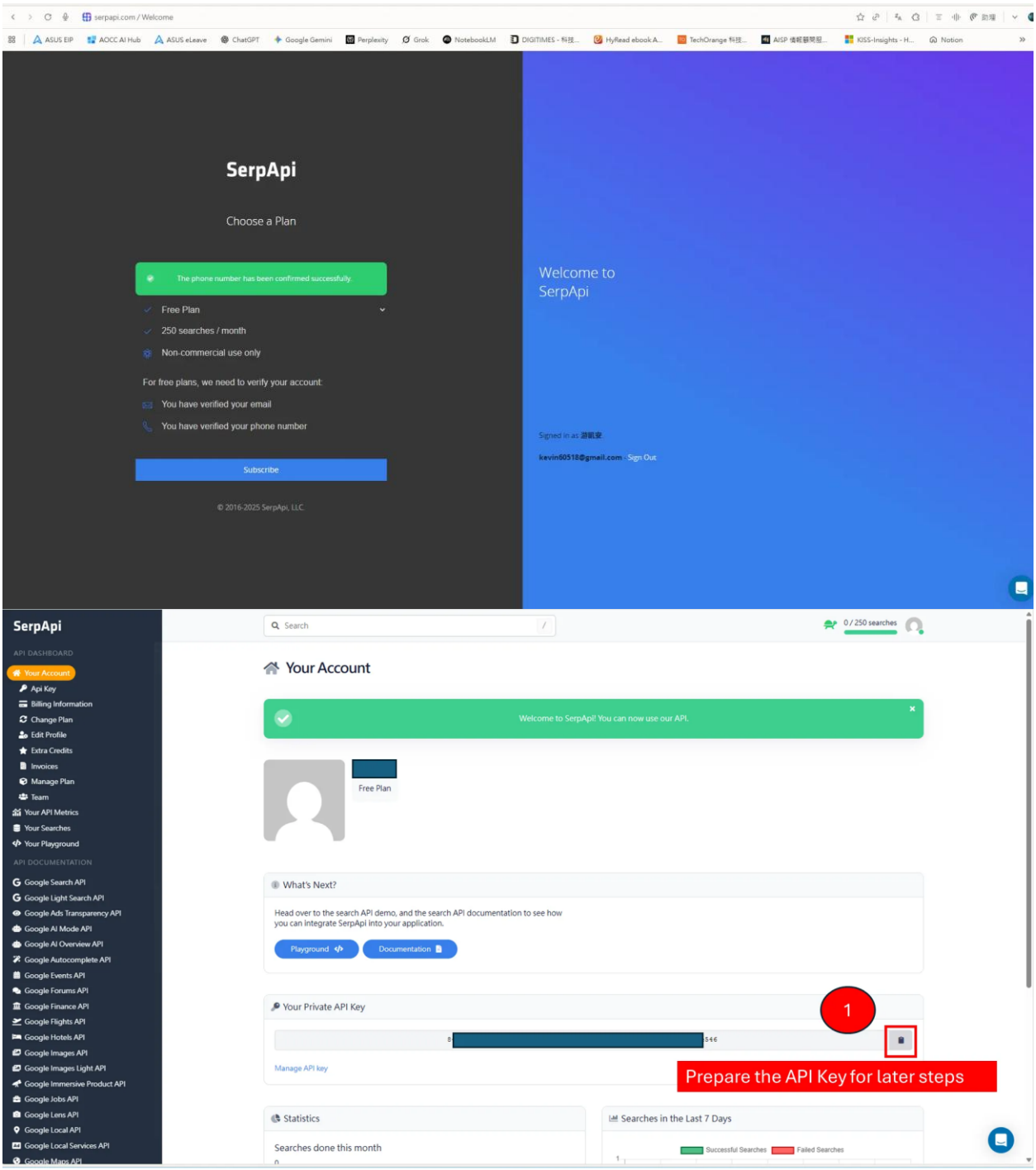
```
PS C:\Users\nuc> cd C:\Users\nuc\Desktop\intel-ai-assistant-builder-main\intel-ai-assistant-builder-main\mcp\mcp_servers
\mcp_google_hotel
```

5. Install the dependencies:
   python -m pip install -r requirements.txt
6. Build the executable file:
   build.bat
7. After the build is complete, you will see the .exe file in the same folder under the dist directory:
   C:\Users\nuc\Desktop\intel-ai-assistant-builder-main\intel-ai-assistant-builder-main\mcp\mcp_servers\mcp_google_hotel\dist

8. In **AI SuperBuild**, go to **Assistant Configuration → Agent Management → MCP Server List → Add**.

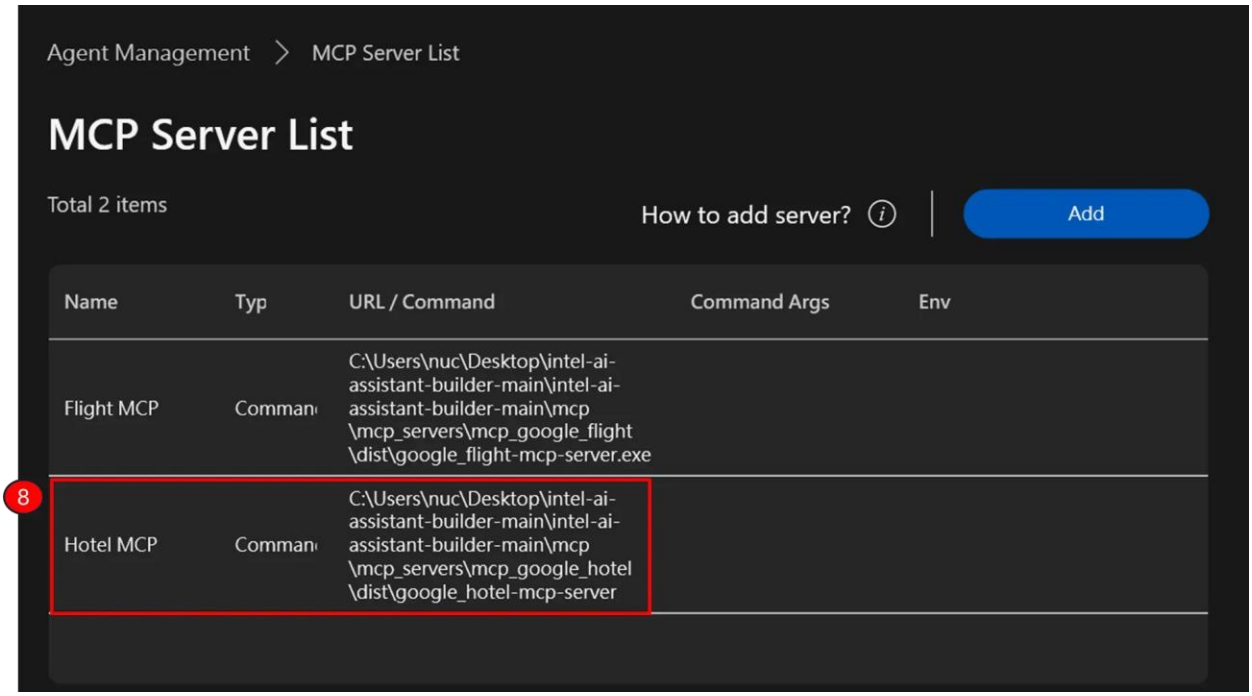


9. Use the **Command** type to configure the MCP server name and the path to the .exe file. You can get the API Key from the SerpAPI website (https://serpapi.com/users/sign_up).

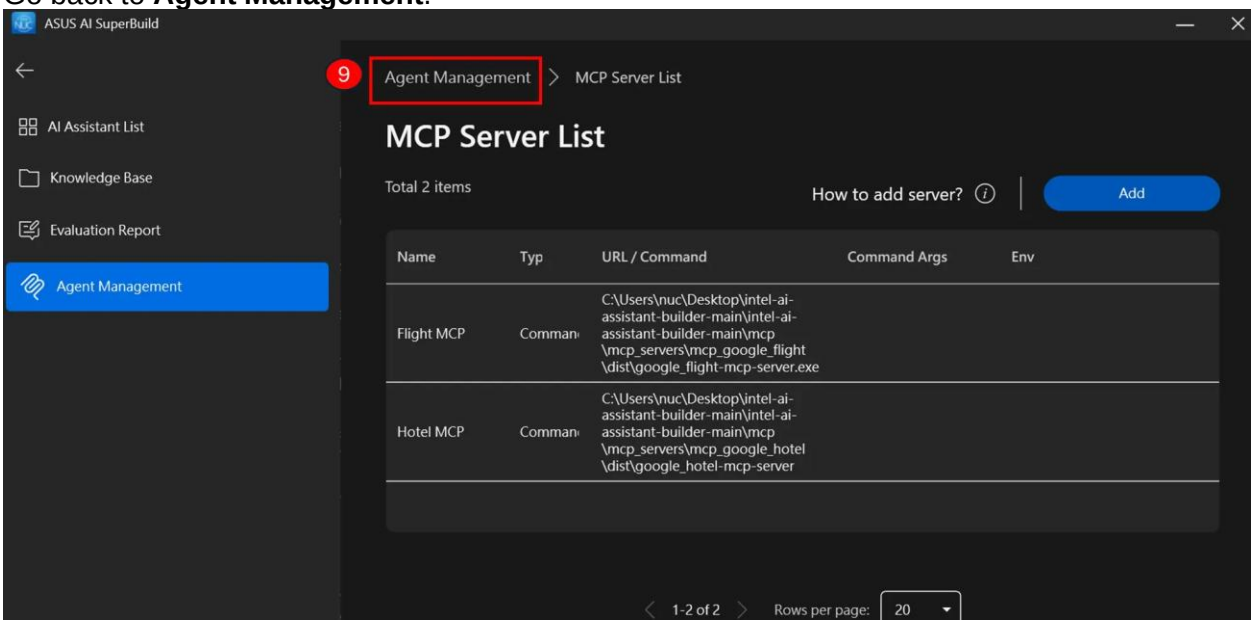


Important: For MCP Server ENV (optional), combine "SERP_API_KEY=" with your API key from SerpAPI. Here's an example:
SERP_API_KEY=8470813f5442b85b22309fc87f65**********546

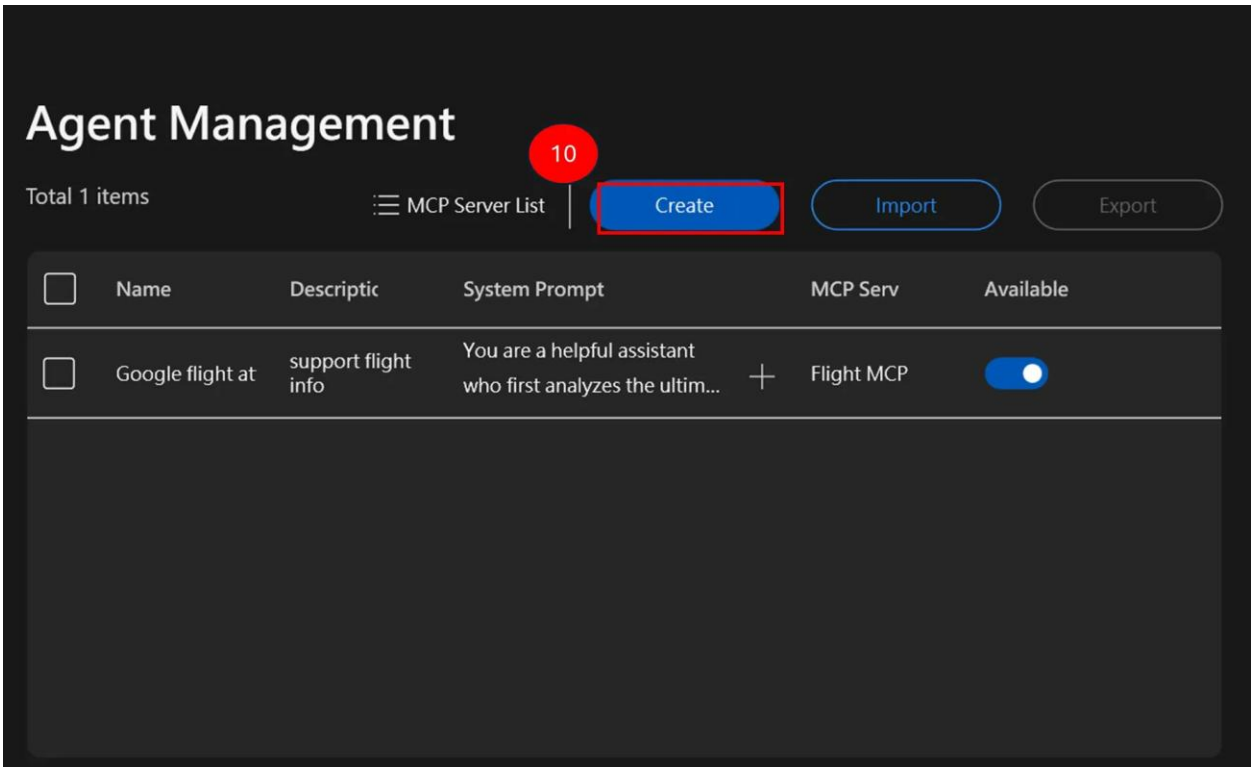
10. The new server will appear in the list.



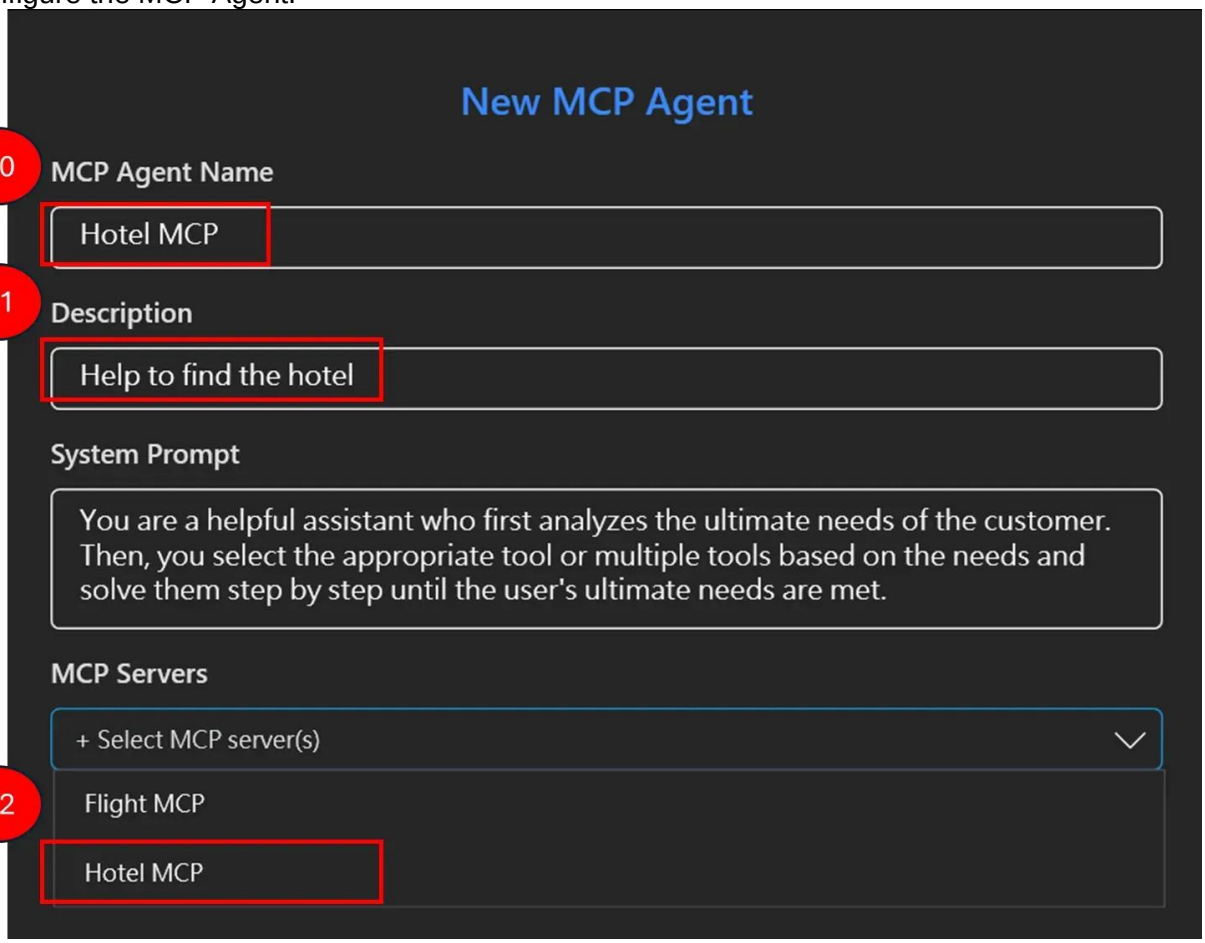
11. Go back to **Agent Management**.



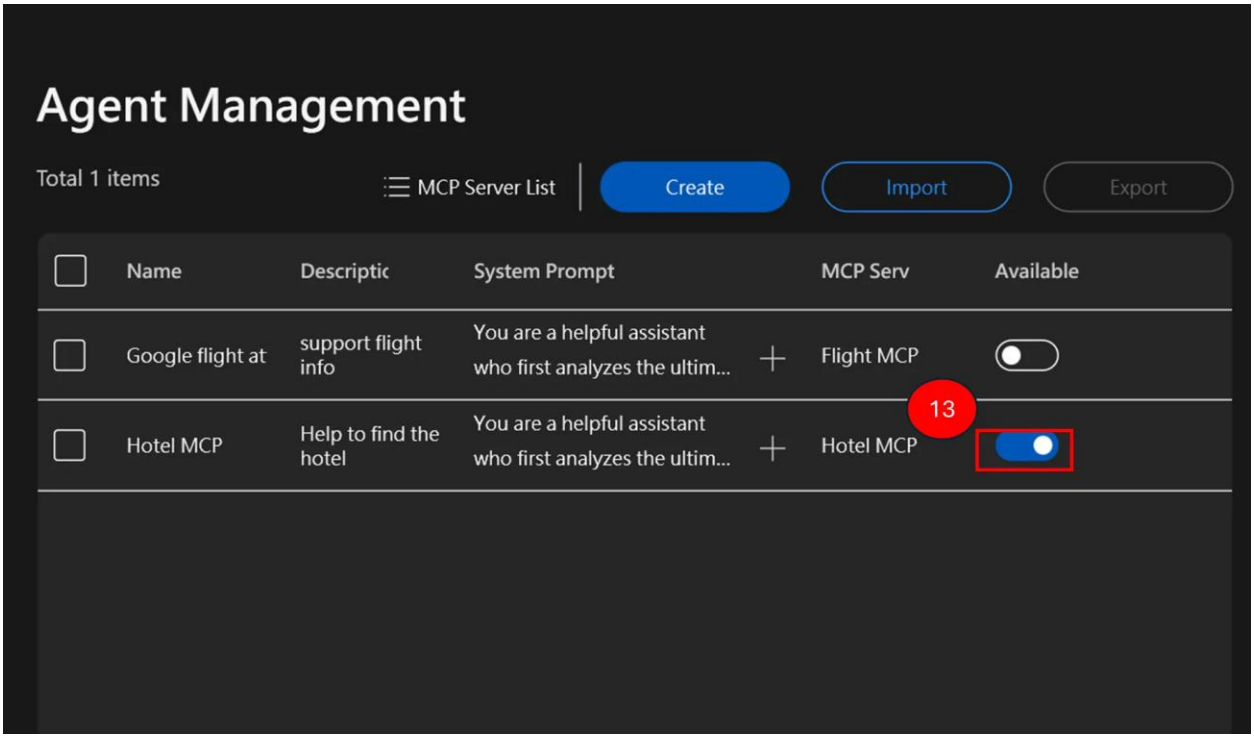
12. Create an MCP Agent task.



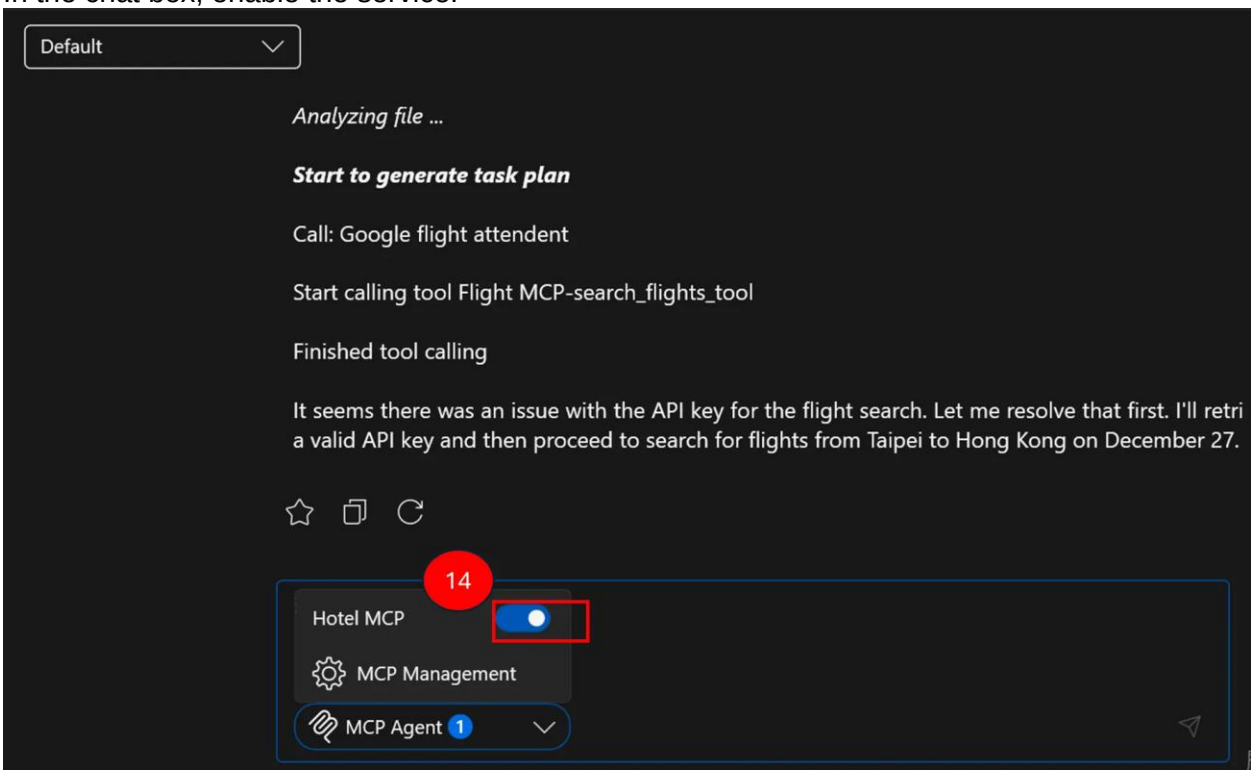
13. Configure the MCP Agent.



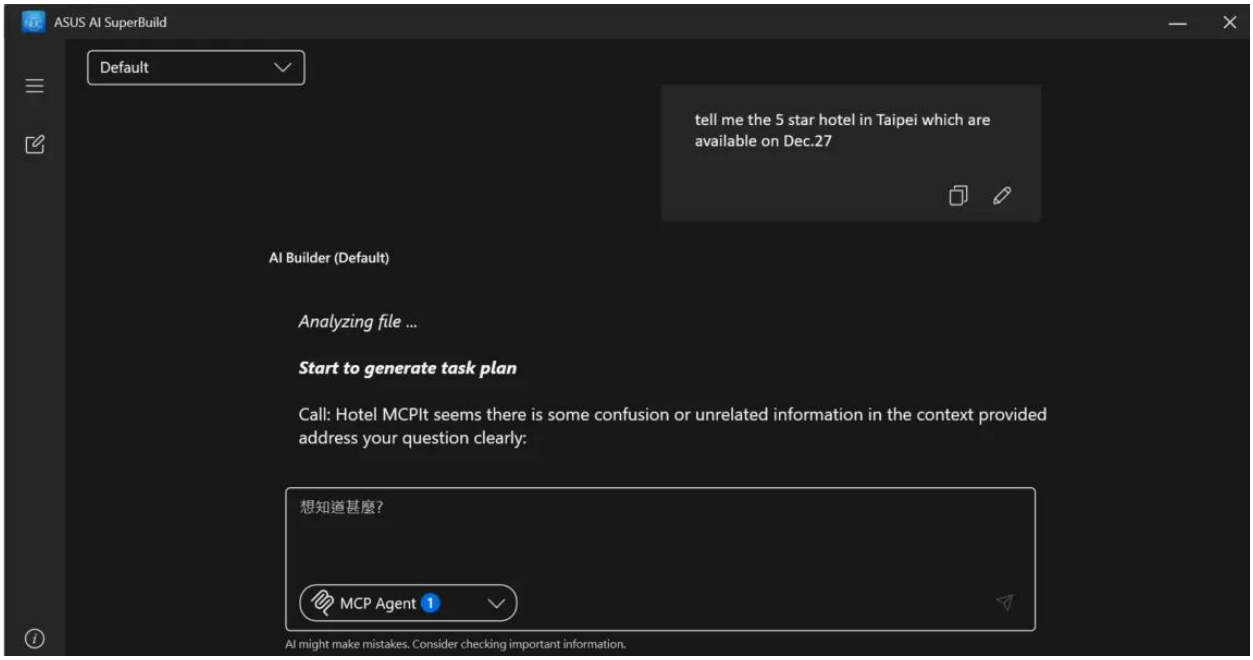
14. Enable the Hotel MCP server.



15. In the chat box, enable the service.



Start chatting and use the MCP tool calling in the conversation.

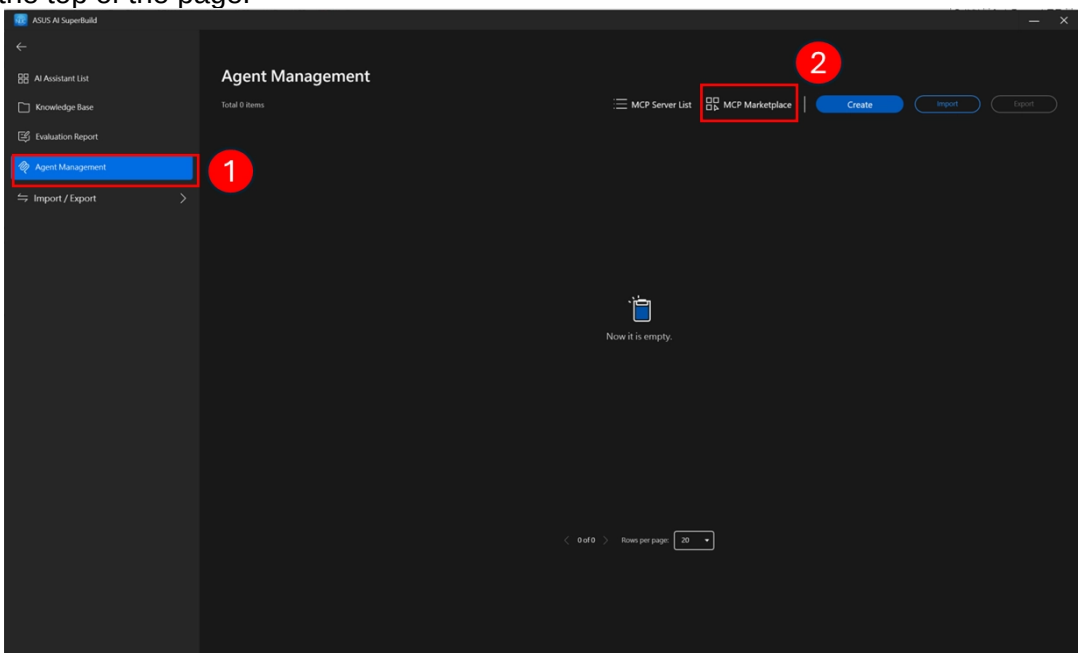


4.2 MCP Marketplace

To help users easily configure MCP servers, AI SuperBuild now supports connecting to MCP with pre-built versions. Navigate to the **MCP Marketplace** to browse and select the MCP tools you need.

1. Navigate to the MCP Marketplace.

   Go to **Agent Management** from the left sidebar ①, then click the **MCP Marketplace** tab ② at the top of the page.

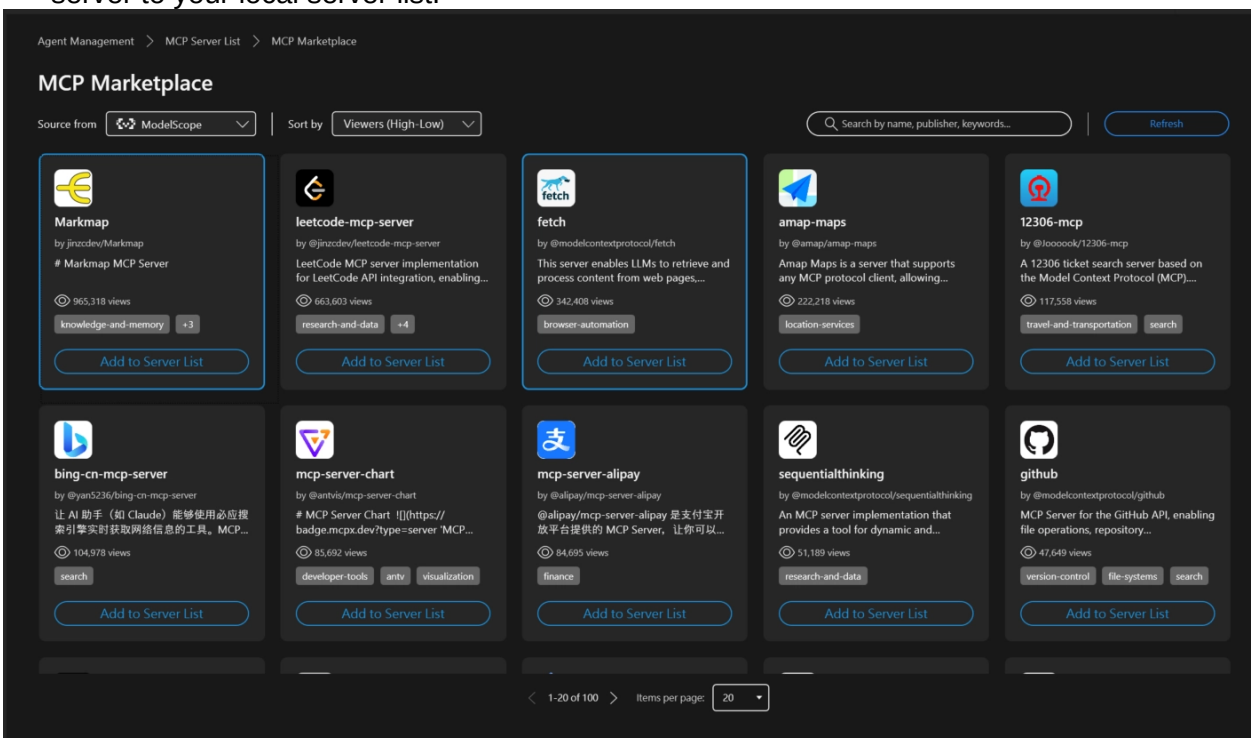


2. Browse and add MCP servers from the Marketplace.

   In the MCP Marketplace, you can browse a variety of pre-built MCP servers. Use the following options to find the right tool:

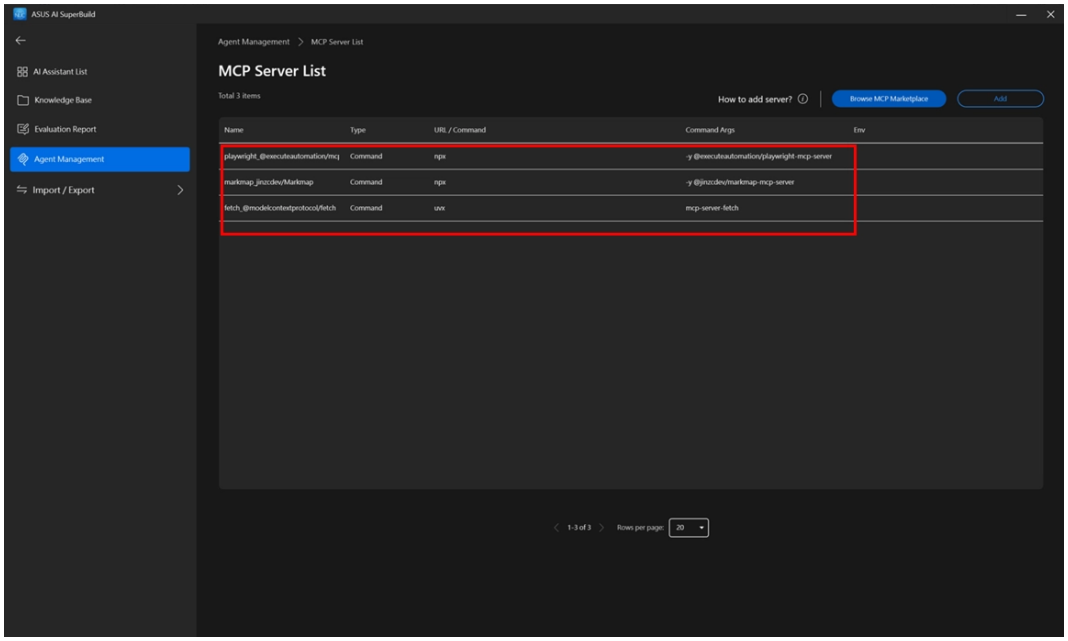
- **Source from**: Select the source repository — **ModelScope** (default) or **Docker MCP Hub**.
- **Sort by**: Sort the results by popularity (e.g., Viewers High–Low).
- **Search**: Use the search bar to find servers by name, publisher, or keywords.

Each card displays the server name, publisher, description, view count, and category tags. Hover over a card to see more details. Click **"Add to Server List"** to add the desired MCP server to your local server list.



3. Verify the added MCP servers.

Go back to **Agent Management > MCP Server List** to view the MCP servers you have added. The list shows each server's **Name**, **Type**, **URL / Command**, **Command Args**, and **Env** settings. You can also click **"Browse MCP Marketplace"** to add more servers, or click **"Add"** to manually configure a new server.

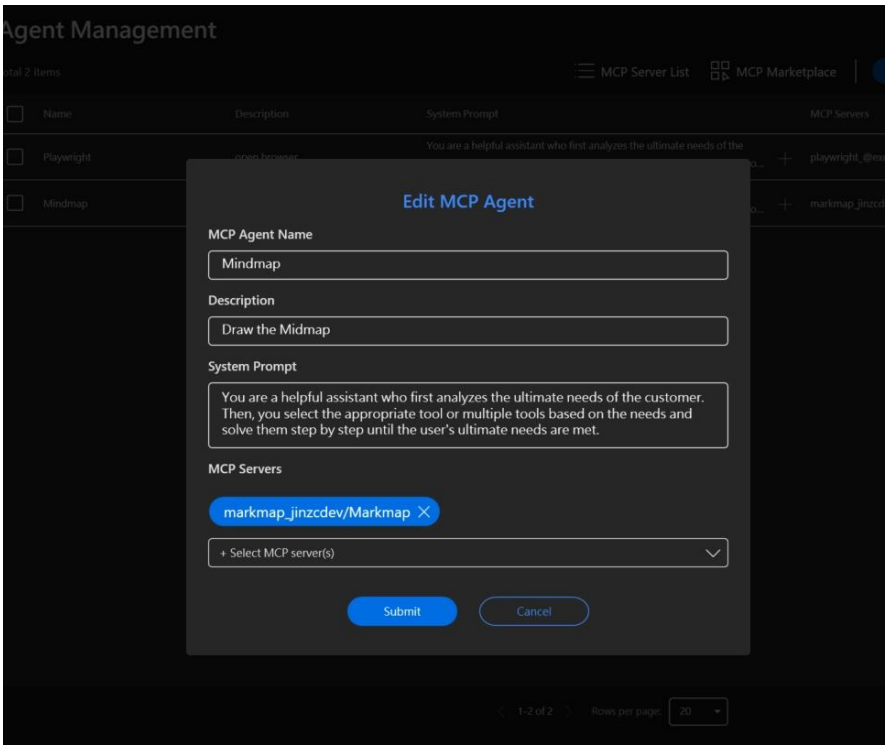


4. Create an MCP Agent with the selected servers.

Return to the **Agent Management** main page and click **"Create"** to set up a new MCP Agent. In the **Edit MCP Agent** dialog, fill in the following fields:

- **MCP Agent Name**: A descriptive name for the agent (e.g., *Mindmap*).
- **Description**: A brief summary of what this agent does (e.g., *Draw the Mindmap*).
- **System Prompt**: Define the agent's behavior and instructions.
- **MCP Servers**: Select one or more MCP servers from the dropdown to assign to this agent.

Click **"Submit"** to save. The agent will now be able to leverage the selected MCP tools to accomplish complex, multi-step tasks.


Agent Management
MCP Server List
MCP Marketplace
Name
Description
System Prompt
MCP Servers
Playwright
Mindmap
Edit MCP Agent
MCP Agent Name
Mindmap
Description
Draw the Midmap
System Prompt
You are a helpful assistant who first analyzes the ultimate needs of the customer. Then, you select the appropriate tool or multiple tools based on the needs and solve them step by step until the user's ultimate needs are met.
MCP Servers
markmap_jinzcdev/Markmap
+ Select MCP server(s)
Submit
Cancel