V1.0

Oct. 2025

ASUS AI SuperBuild

User Guide

Table of Contents

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 ASUS AIPC: NUC 14 Pro, NUC 14 Pro AI, NUC14 Pro AI+, and NUC 15 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.

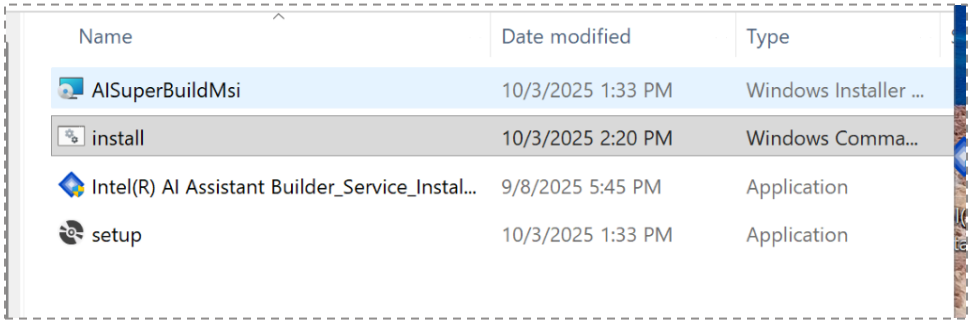
1.2 Installation

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.

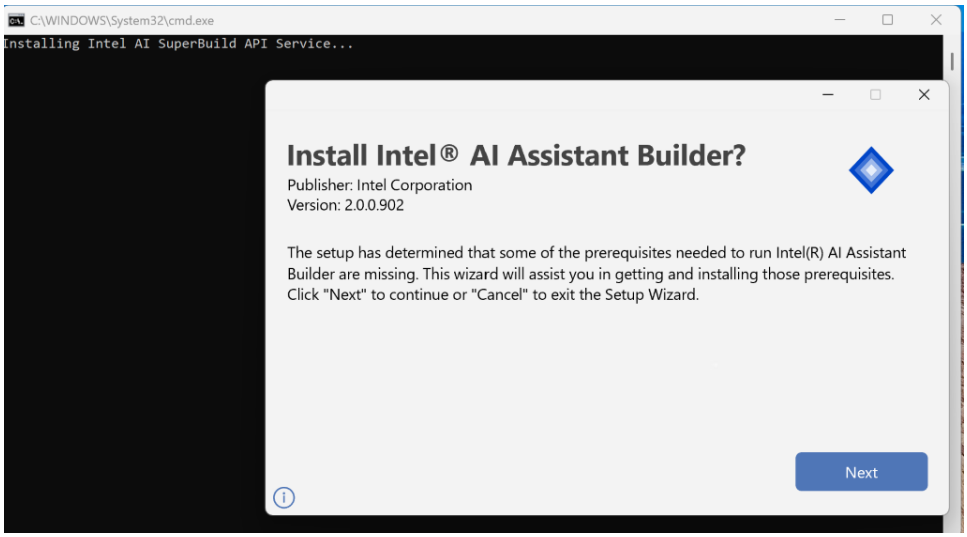
Installation Steps

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

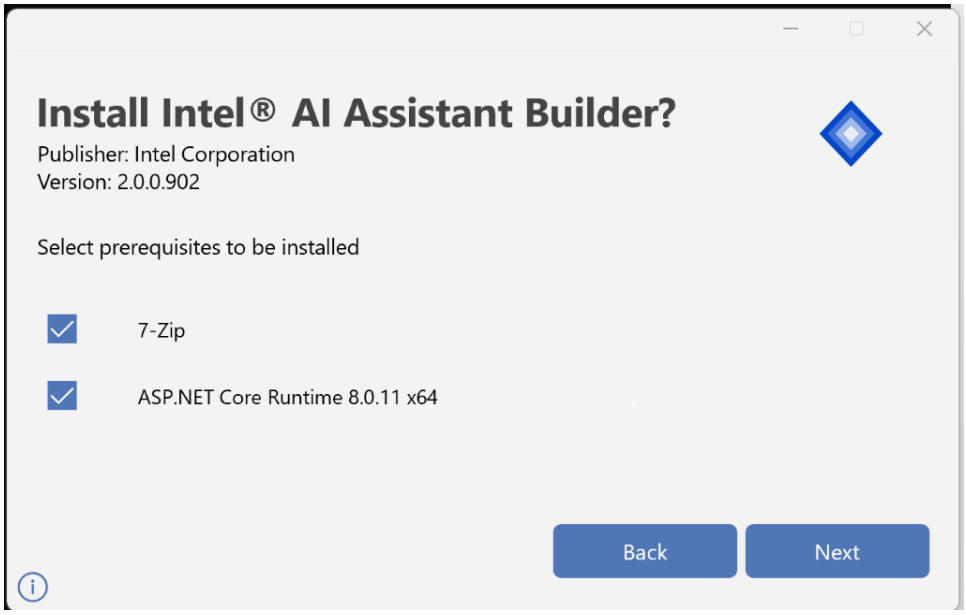
2. Run the Installer Script:
   Open the extracted folder and double-click the install.cmd file. This command script will initiate the entire installation process.



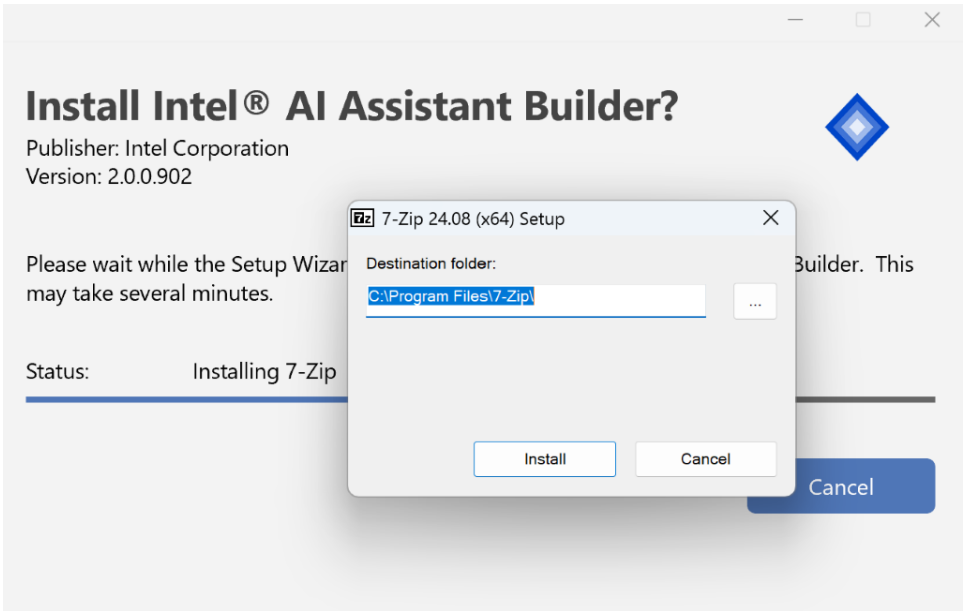
3. Follow the On-Screen Prompts:
   A series of installation wizards will appear. Follow the on-screen instructions, clicking Next, Install, and Accept as prompted to install the prerequisite software (Intel® AI Assistant Builder, 7-Zip, ASP.NET Core Runtime) and the main AI SuperBuild application.
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



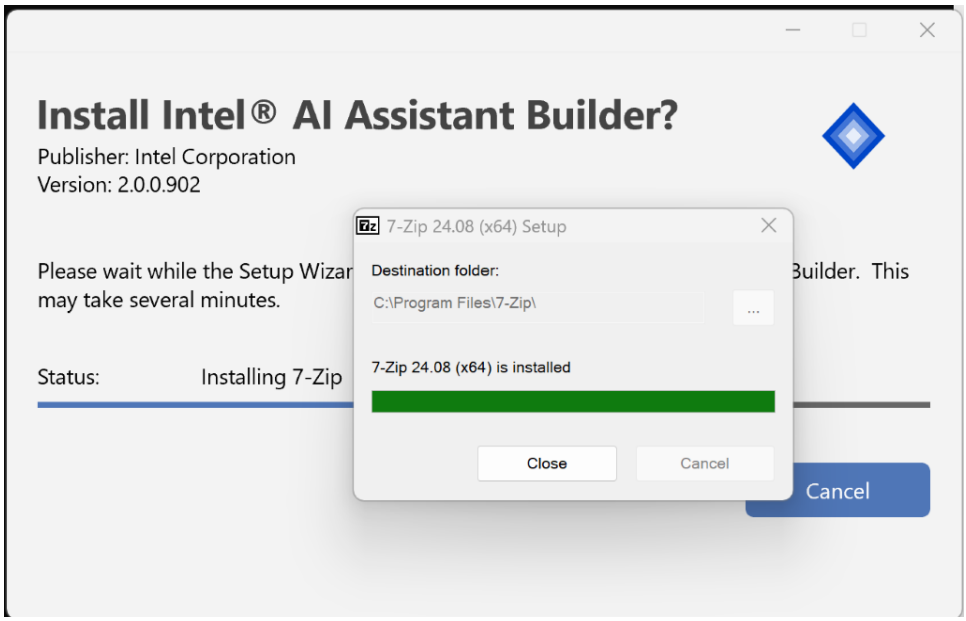
Select Next



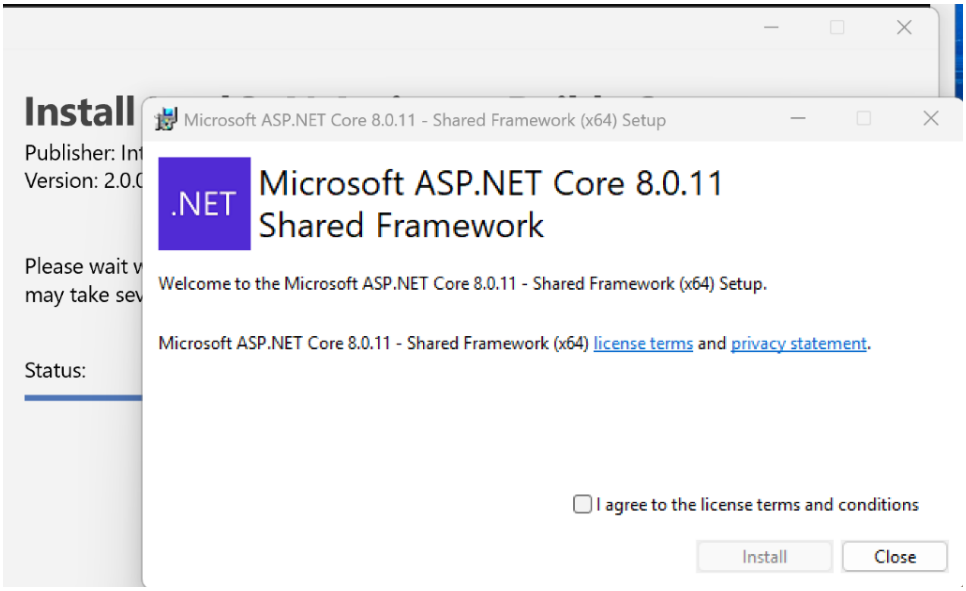
Select Install



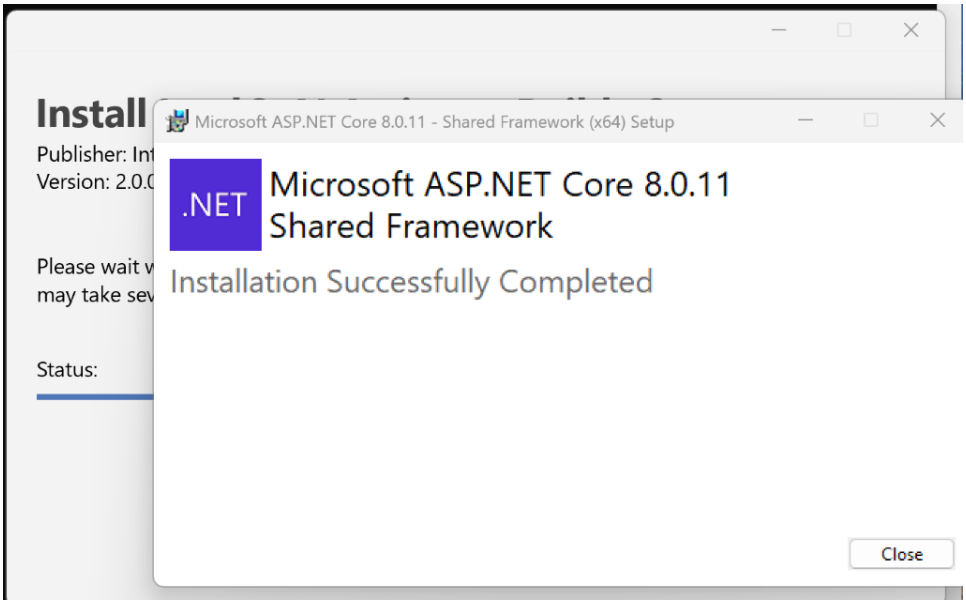
Select Close



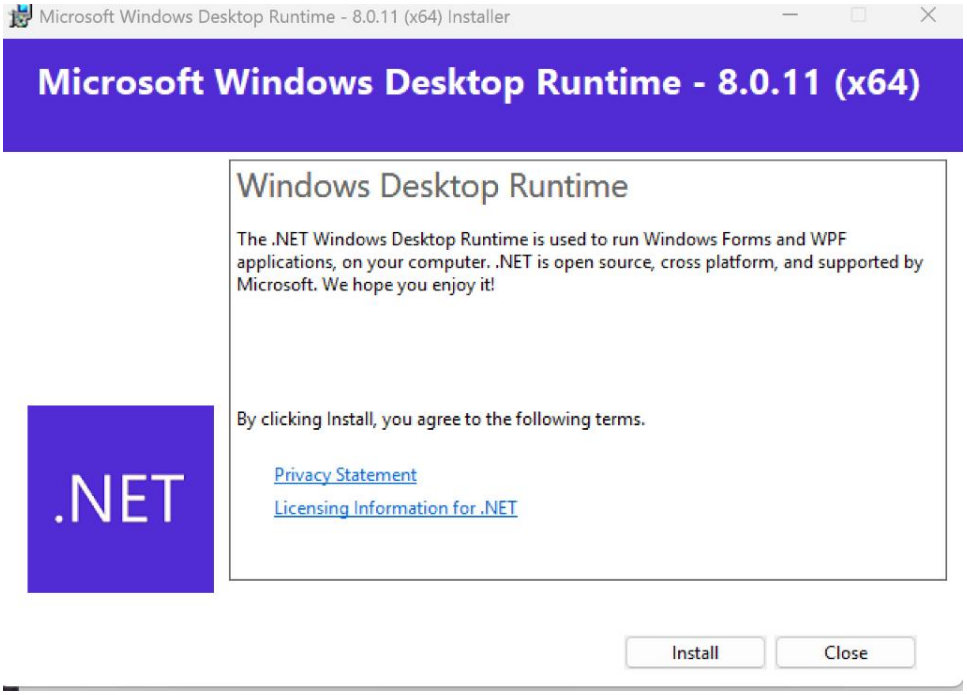
I agree, then install



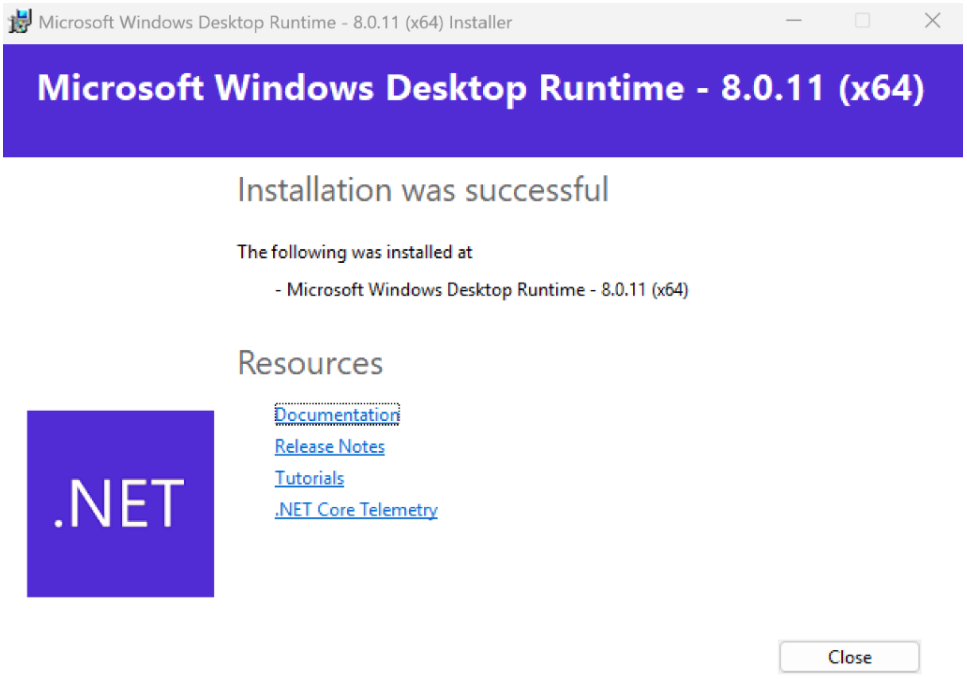
Select close



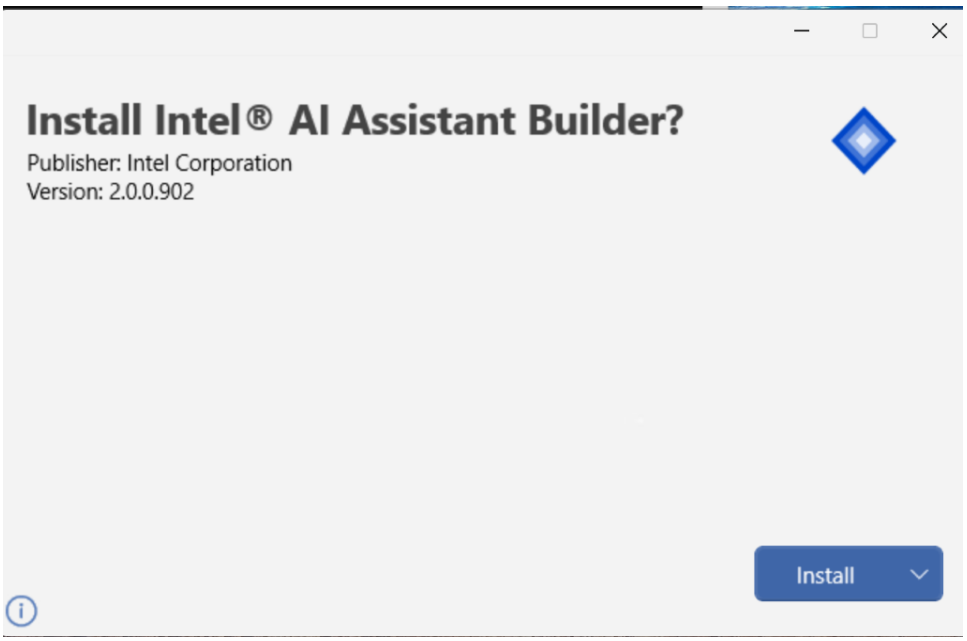
Install



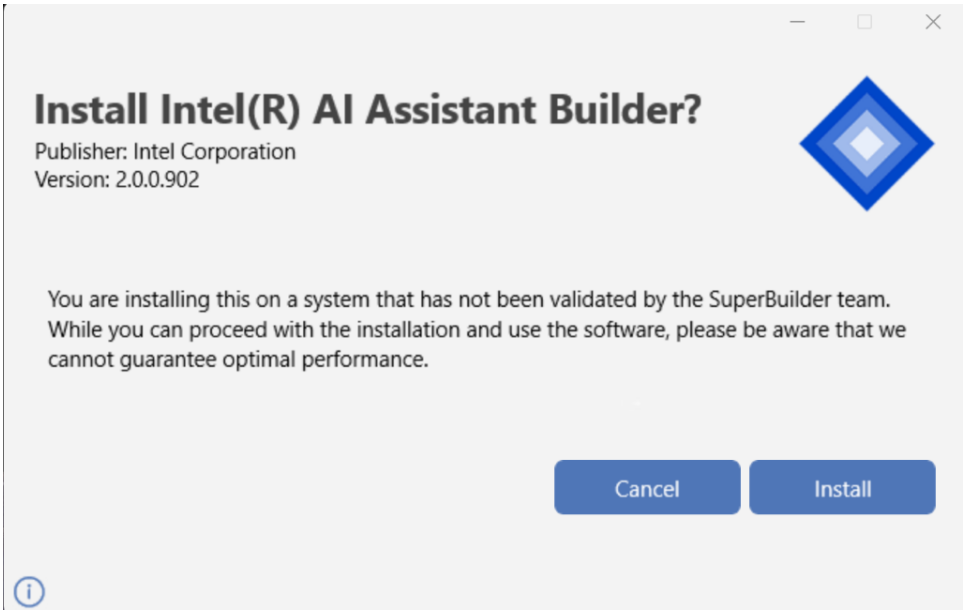
Close



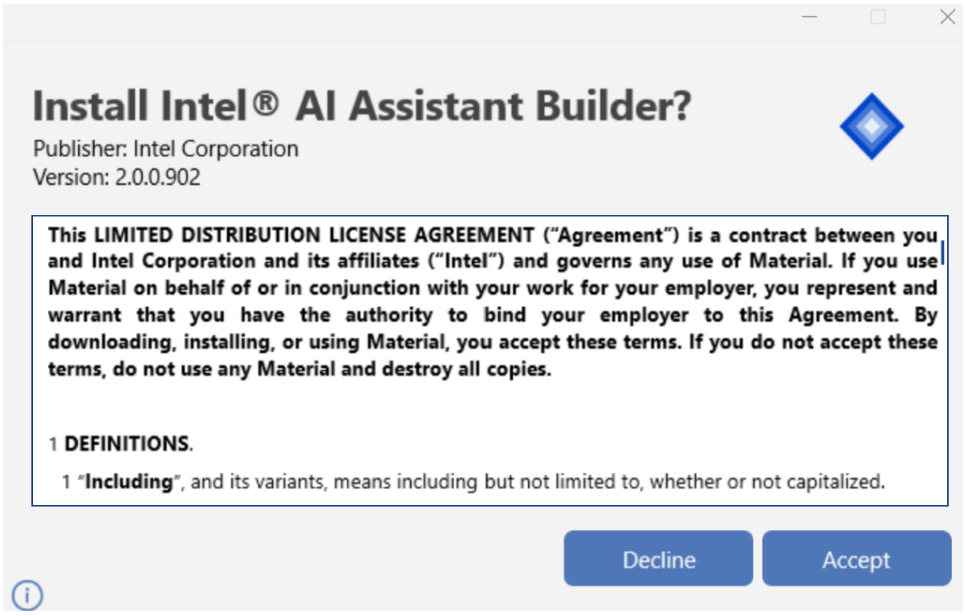
Install



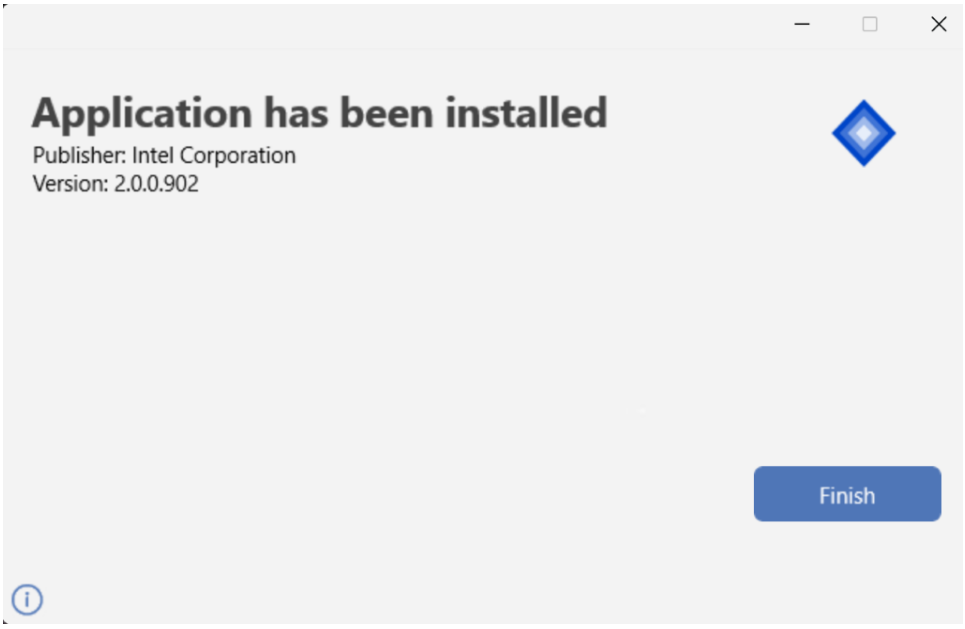
Install



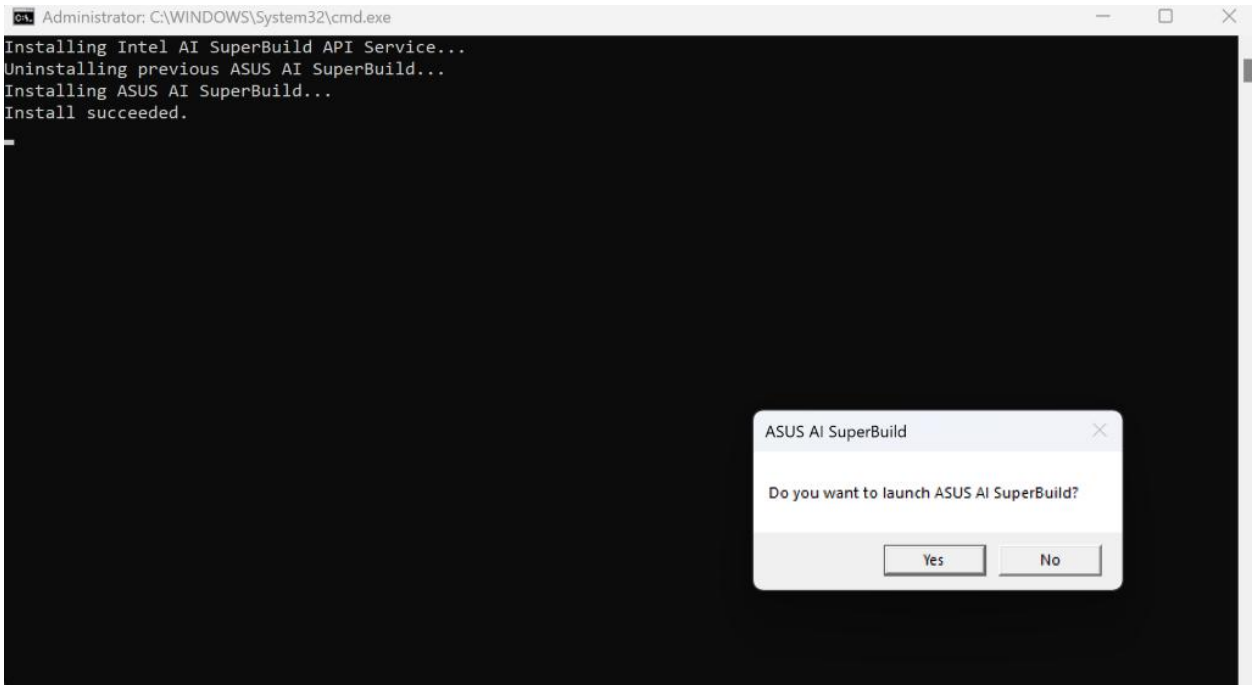
Accept



Finish

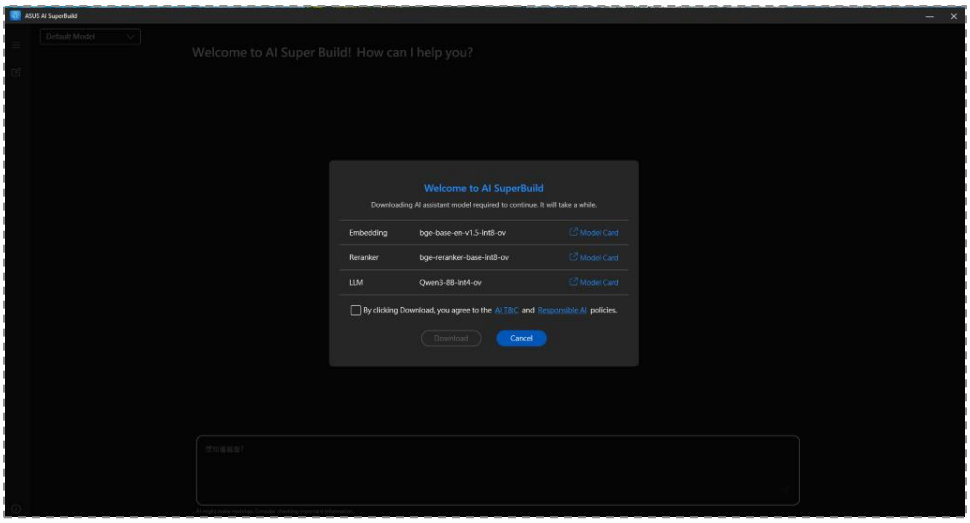


4. Launch the Application:
   Once the installation is complete, a prompt will ask if you want to launch ASUS AI SuperBuild. Click Yes. A new desktop shortcut will also be created for future access.

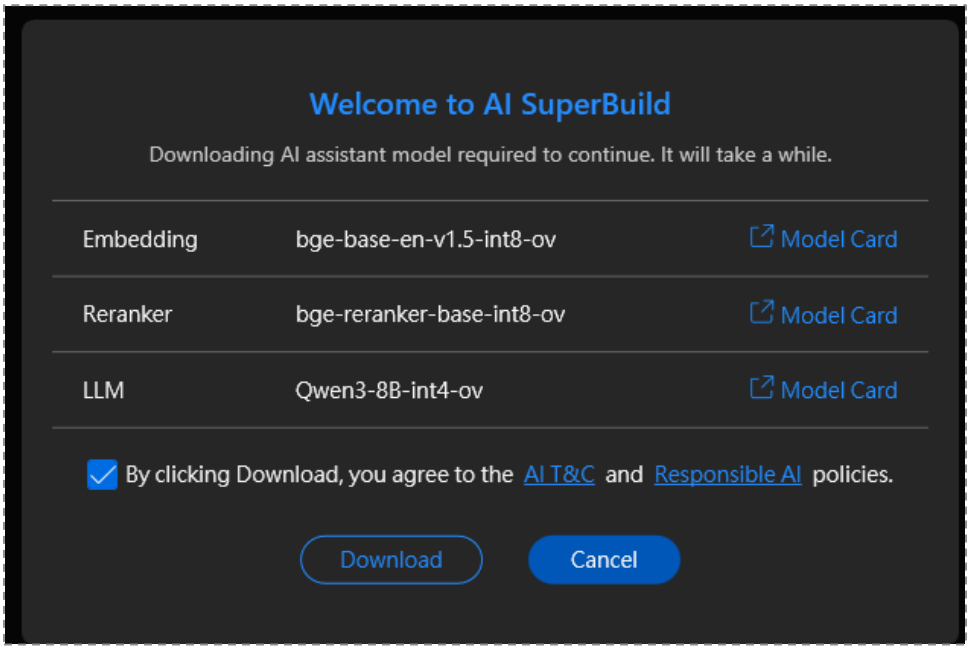


First-Time Setup: Model Download

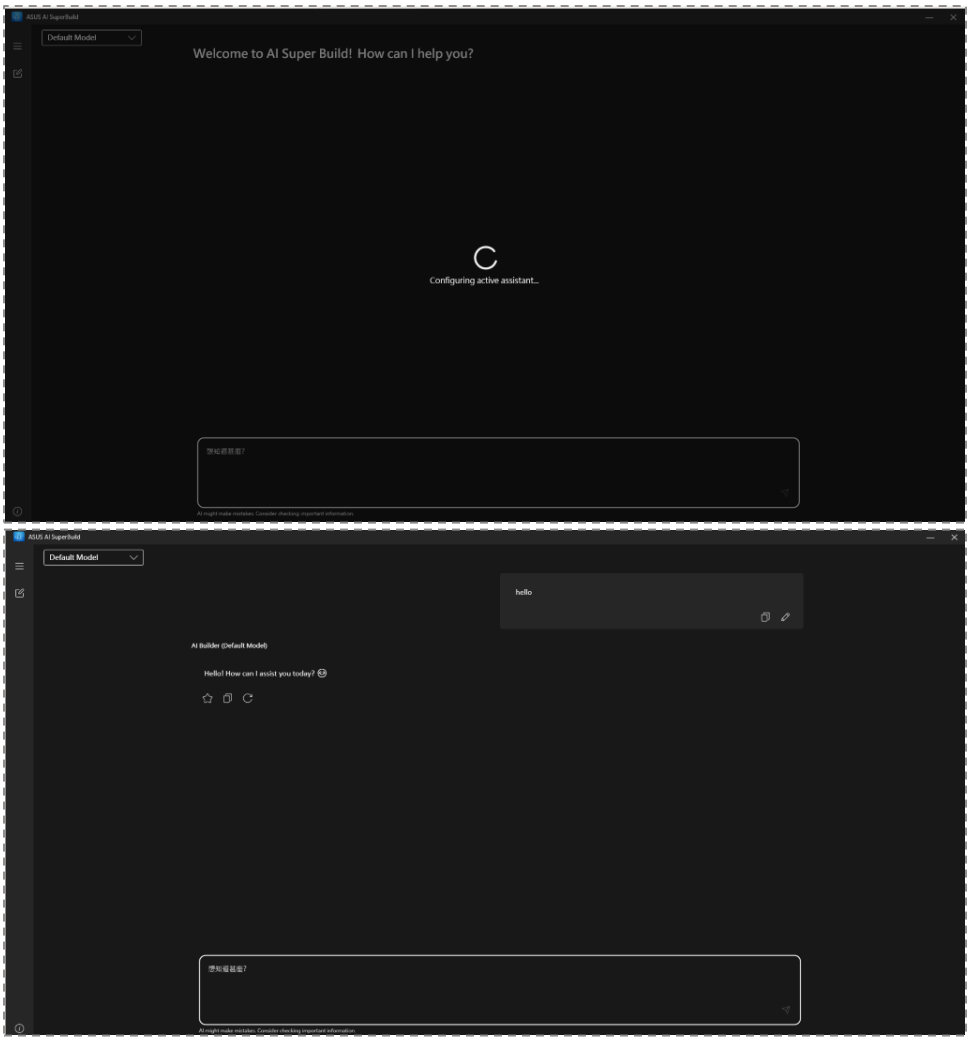
1. Initial Model Download:
   Upon the first launch, a "Welcome to AI SuperBuild" window will appear, prompting you to download the required AI assistant models (Embedding, Reranker, and LLM).



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

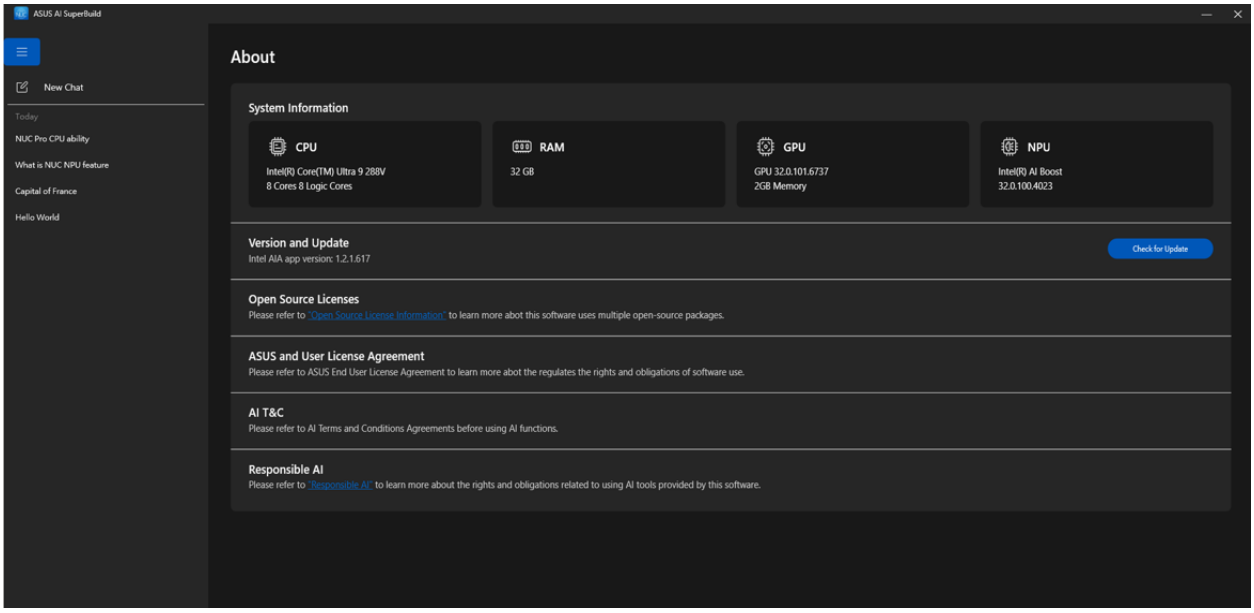
- When the software is used for the first time, it may take approximately 35 to 50 seconds to respond to the initial prompt.
- Text generation speed may be slow in the current version. This is slated for improvement in a future release.

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

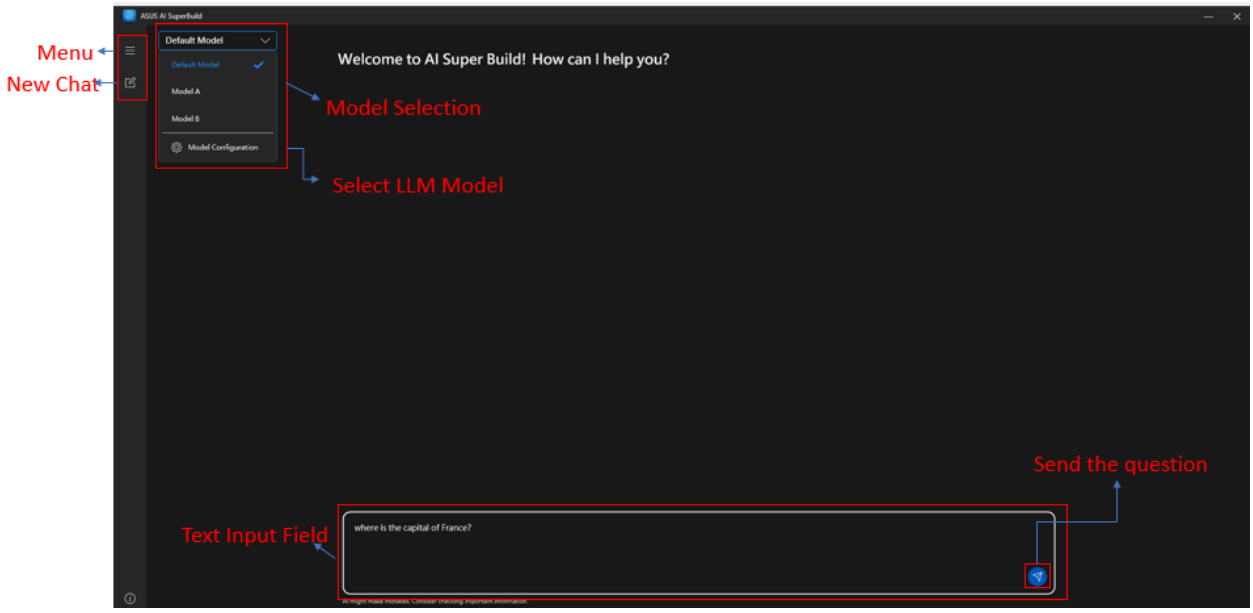
	and driver version, which is used for accelerating AI tasks (e.g., Intel(R) AI Boost 32.0.100.4023).
Version and Update	This section shows the currently installed application version of AI SuperBuild (e.g., 1.2.1.617).
Check for Update	Clicks to check for and download the latest version of the application.
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.

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

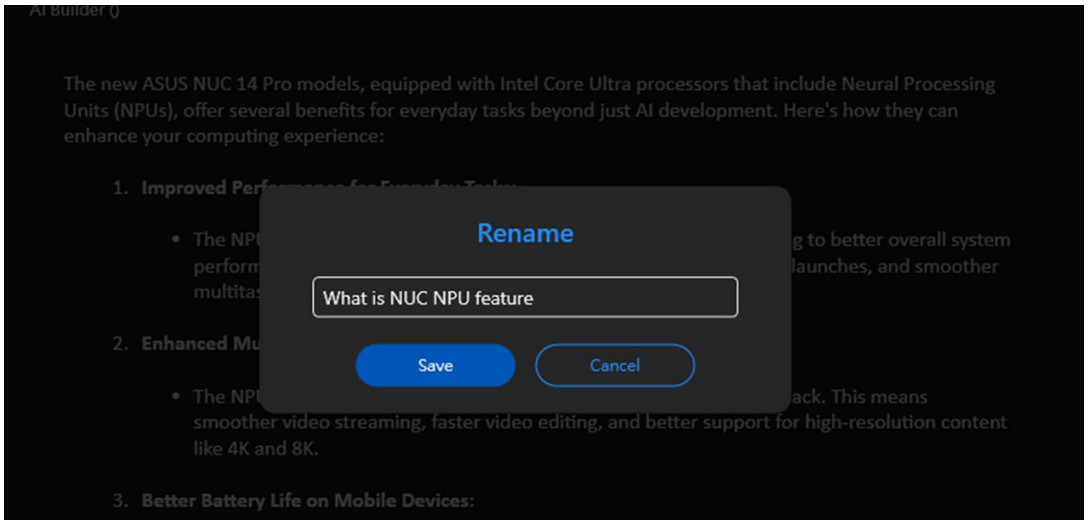
The main interface is your central hub for interacting with the AI SuperBuild assistant. The layout is designed to be intuitive, with a navigation panel on the left for managing chats and accessing settings, the main conversation window in the center, and your input tools at the bottom.

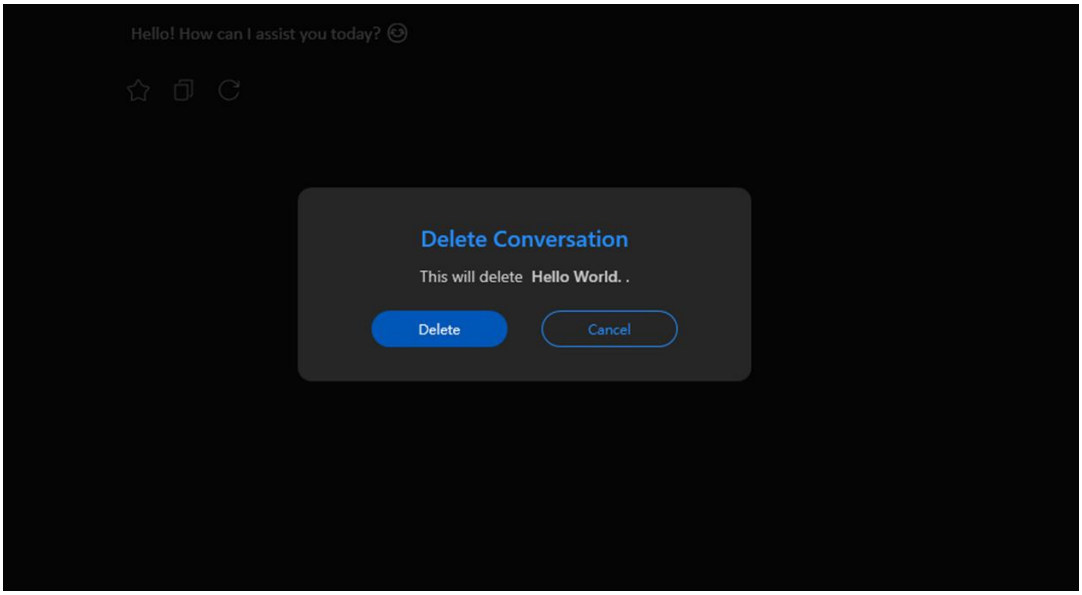


UI Element	Description
Menu (☰)	Toggles the visibility of the left-hand navigation panel.
New Chat	Click to start a new, blank conversation session.
Model Selection	Allows you to select the active AI model for your session from a list of available models like "Default Model", "Model A", etc.
Model Configuration	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.

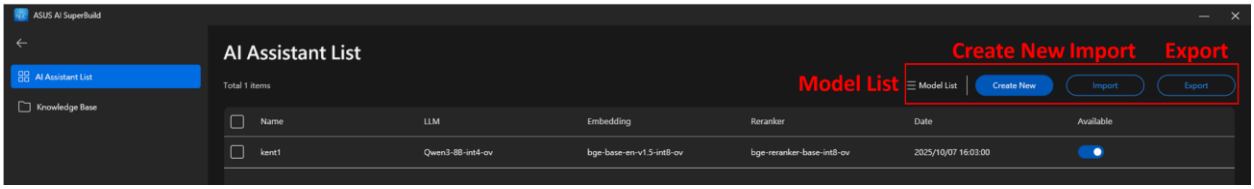
Delete a Chat	Right-click on a chat and select "Delete" to permanently remove the conversation.

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.

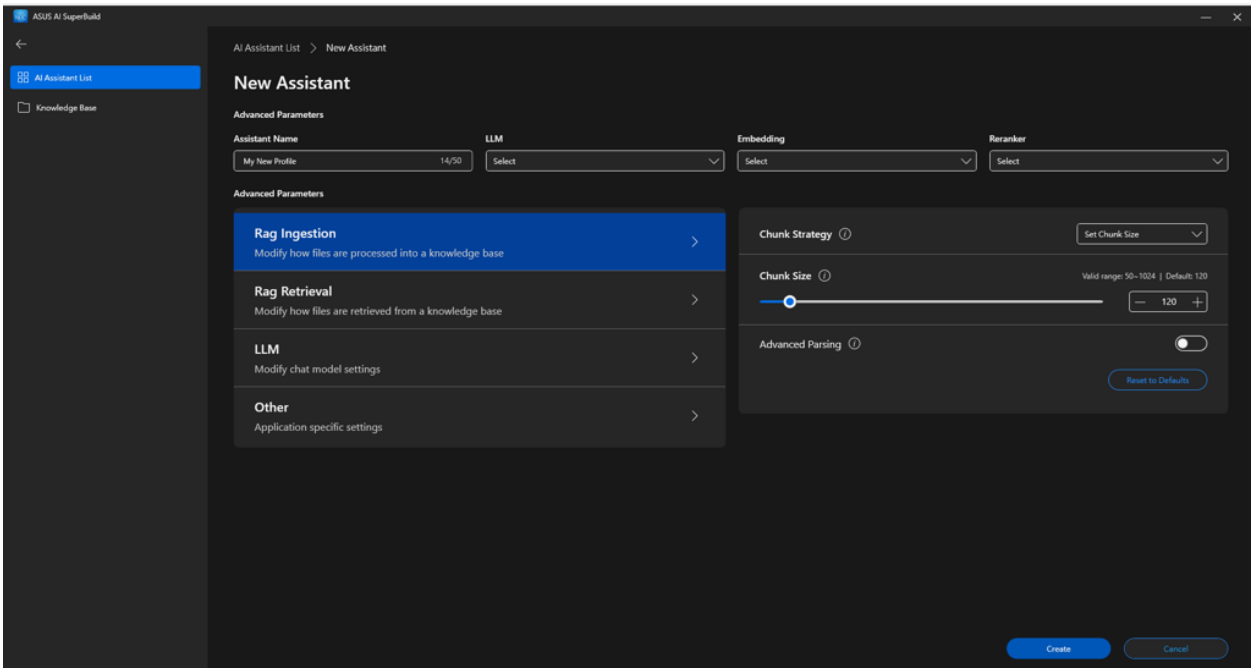


Button/UI Element	Description
Create New	Opens the "New Assistant" wizard to create a new, customized AI assistant profile from scratch.
Import	Allows you to import a previously exported assistant configuration file, making it easy to share or restore settings.
Export	Allows you to export the selected 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)	Enables or disables an assistant, controlling its visibility in the main chat interface's model selection dropdown.

3.2 Creating a New Assistant

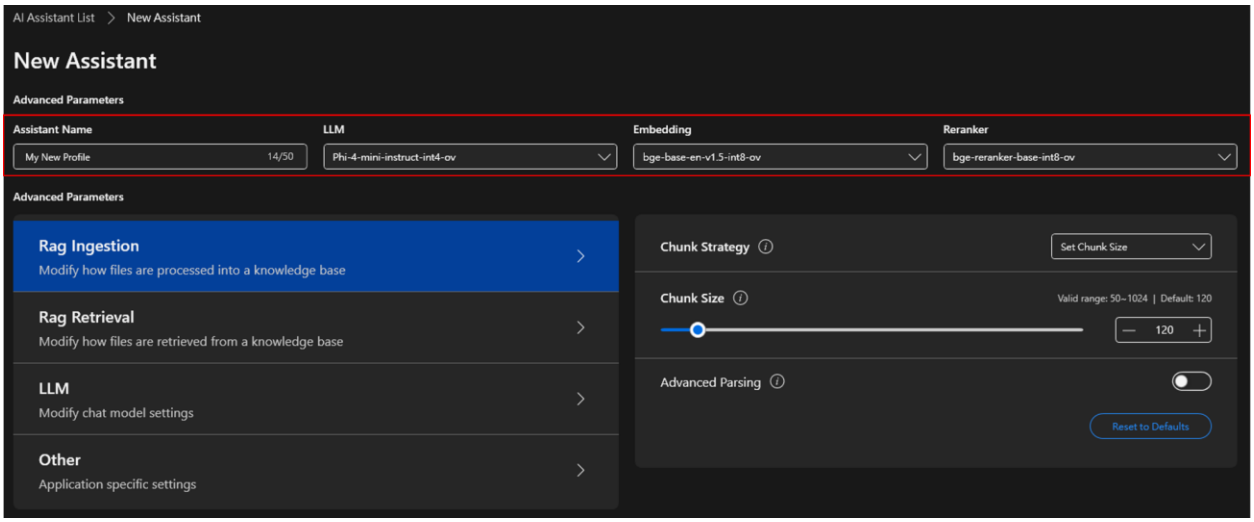
When you click "Create New," you are taken to the New Assistant configuration page. This is where you define the core components and behaviors of your new AI assistant, from the base

models to the fine-grained parameters that control its performance.



3.2.1 Basic Settings

These are the fundamental components of your AI assistant.



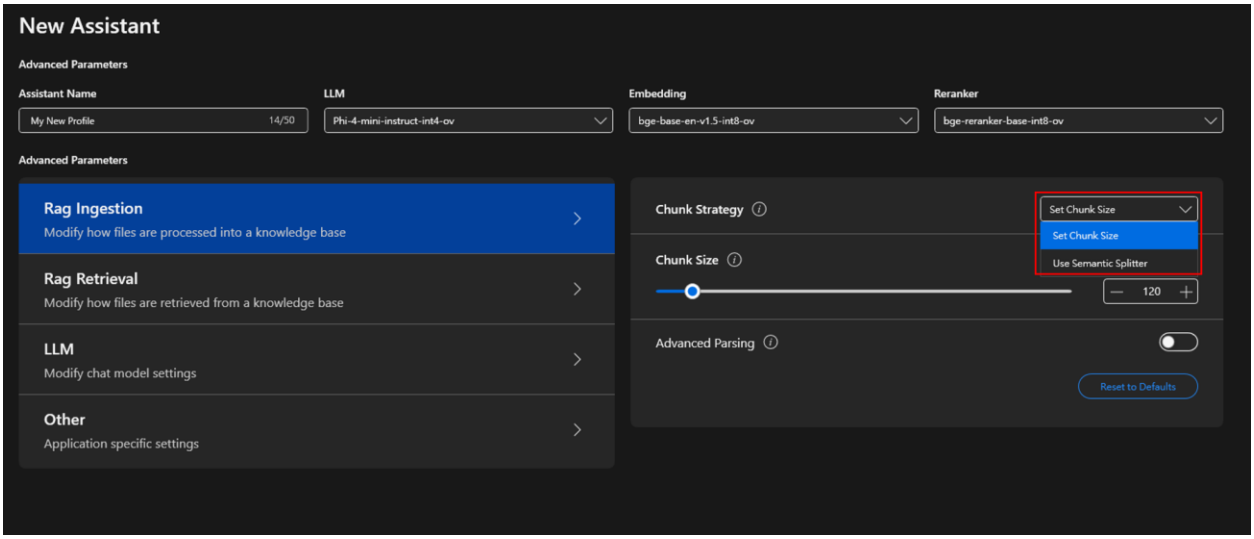
Parameter	Description
Assistant Name	A custom name for your new assistant profile. This name will appear in the Model Selection dropdown on the main chat screen.

LLM (Large Language Model)	The core AI engine that generates human-like text to answer questions, summarize documents, and engage in conversation. From the dropdown, select the base language model that will power this assistant.
Embedding	The model 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.

3.2.2 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.

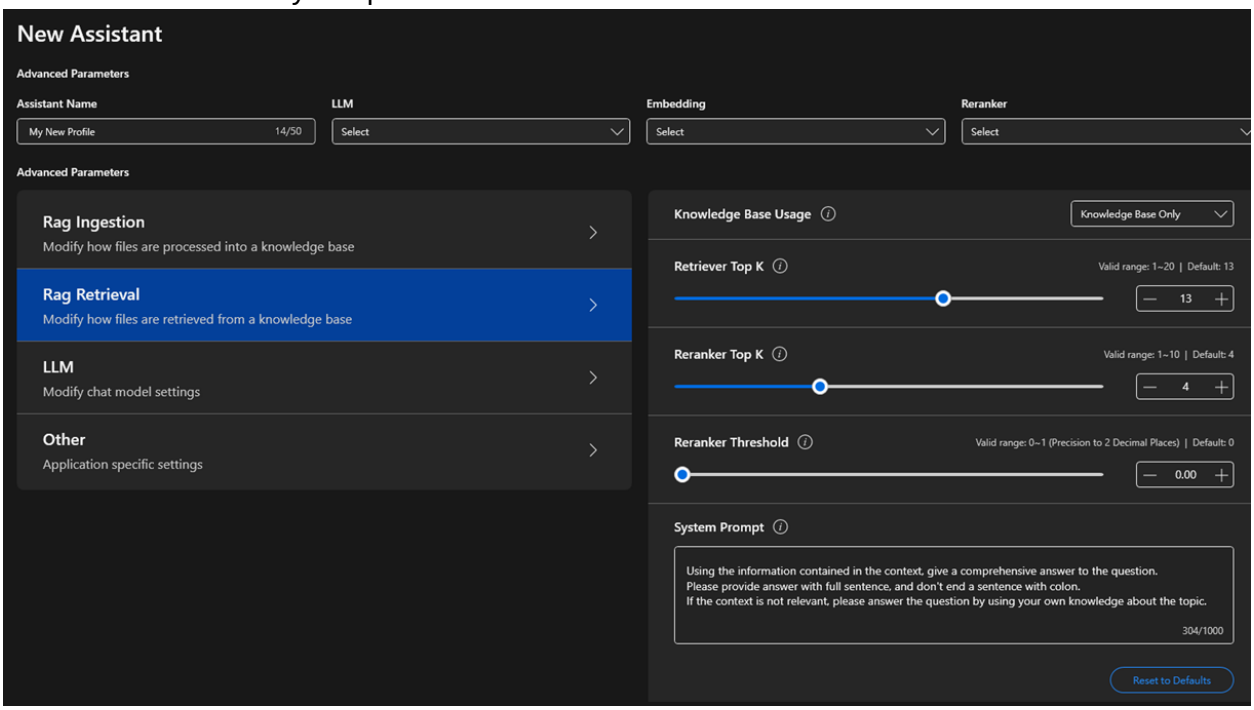


Parameter	Description
Chunk Strategy	Defines how large documents are split into smaller pieces ("chunks"). This is critical for how the material is indexed and searched. Options include "Set Chunk Size" or "Use Semantic Splitter".
Chunk Size	Sets the size of each text chunk. A smaller size may yield 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.

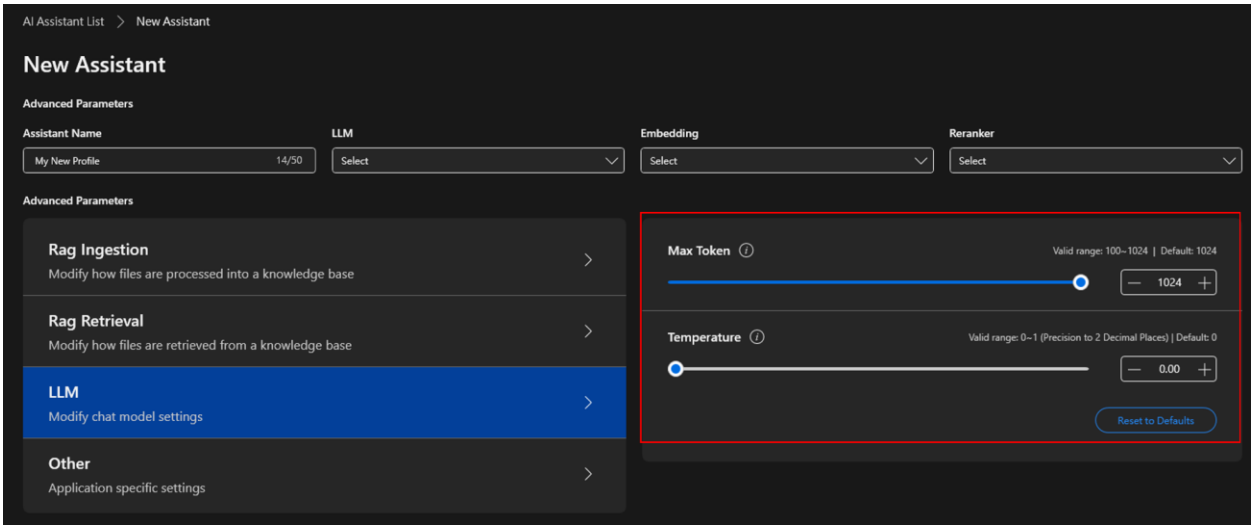


Parameter	Description
Retriever Top K	Controls how many of the most relevant documents ("chunks") are initially fetched from the knowledge base. A higher number provides more context but may increase processing time.
Reranker Top K	Determines how many of the initially retrieved documents are passed to 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

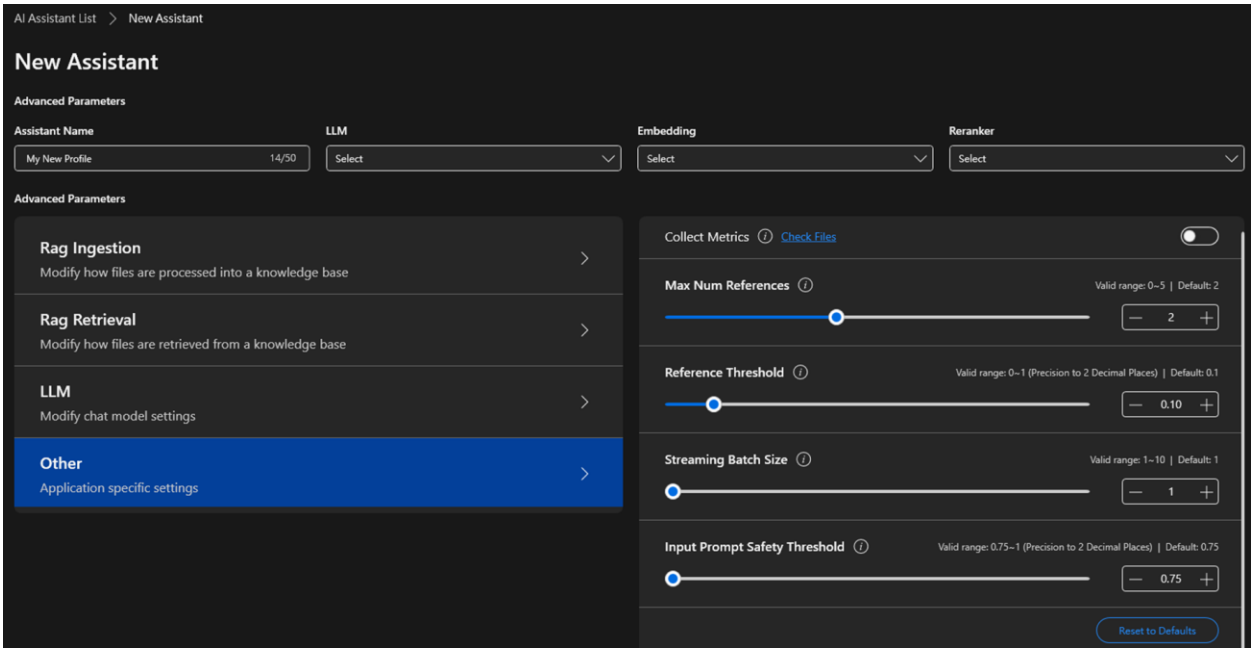
Here you can modify the chat model's settings to control the style and length of the AI's responses.



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.



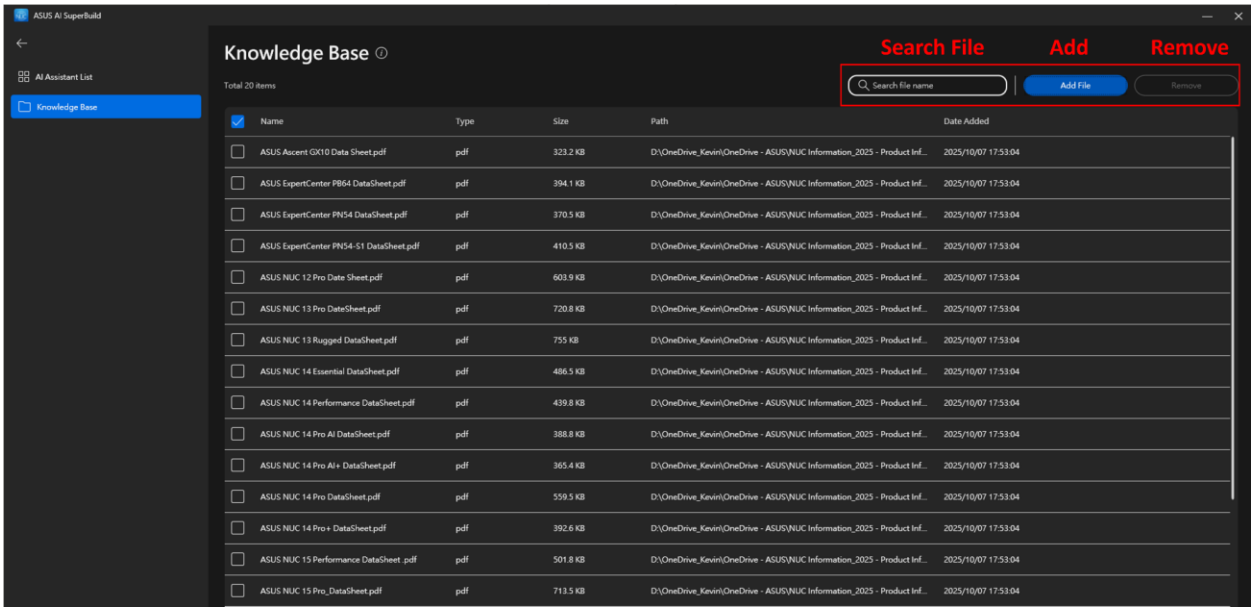
Parameter	Description
Collect Metrics	Enables or disables the collection of performance and usage data for analysis.
Max Num References	Sets the maximum number of source documents from the knowledge base that can be cited in a single response.
Reference Threshold	Defines the minimum relevance score a 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.

3.3 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.

3.4 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.