Intel® AI Super Builder

Product Overview and User Guide

Intel technologies may require enabled hardware, software or service activation.

No product or component can be absolutely secure.

Code names are used by Intel to identify products, technologies, or services that are in development and not publicly available. These are not "commercial" names and not intended to function as trademarks.

Customer is responsible for safety of the overall system, including compliance with applicable safety-related requirements or standards.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at intel.com.

Table of Contents

Revision history

Date	Description
January 2026	Name change; added GGUF backend for v2.7
September 2025	Major feature updates and revisions for v2.0
June 2025	Updated for v1.2.1 fixes and version references
May 2025	Added v1.2 features and updates
April 2025	Added v1.1 features and updates
March 2025	Added several new sections of information and additional clarity in response to community activity and user questions.
February 2025	Major revision for consistency, simplicity, and clarity.
January 2025	Added sections for Knowledge Base, Chat Sessions, Model Selection, Parameter Settings and Special Query Types, User Model Upload
December 2024	Added NPU support for Phi-3-4k model on Lunar Lake system, Model parameter tuning, Hugging Face model repo support
November 2024	Added instructions for proxy server configuration in section 1.2
November 2024	Initial release.

1.0 Introduction

Intel® AI Super Builder is Intel's Gen-AI reference design platform that enables the rapid creation of custom AI gents tailored to specific industry needs and proprietary data. These agents streamline everyday tasks and deliver intelligent solutions by leveraging your internal knowledge bases, **all while running entirely locally** on Intel®-based AI PCs. Your data and workflows remain private and secure, powered by cutting-edge large language models (LLMs), customizable agentic workflows, and performance-optimized processing.

1.1 Key Benefits

- **Simple & Accelerated Development**: Jumpstart your AI agents with a rich set of prebuilt **APIs**, reusable **templates**, and a user-friendly tooling environment designed for fast prototyping and deployment.
- **Flexible & Modular Architecture**: Use our turnkey front-end reference design for immediate rollout or integrate only the backend components you need to build a fully customized experience.
- **Secure & Local Execution**: Keep your proprietary data and IP safe with AI that runs directly on-device—no cloud required.
- **Customizable for Any Industry**: Tailor agents to meet the demands of your specific domain, whether in healthcare, finance, manufacturing, or beyond.
- **Scalable & Portable**: Package and deploy across a wide range of devices and use cases with ease.

1.2 Prerequisites

To download, install, and set up the Intel® AI Super Builder application for the first time, an internet connection is required. After initial setup, an internet connection is generally not needed unless you change the LLM models (which may require additional downloads). Note: Some MCP services in the Super Agent workflow may also require internet access.

1.2.1 Hardware Requirements

COMPONENT	MINIMUM REQUIREMENTS	RECOMMENDED REQUIREMENTS
Processor	Intel® Core™ Ultra processor Series 1 (Meteor Lake)	Intel® Core™ Ultra 200V series (Lunar Lake) and newer
Memory (RAM)	16GB	32GB
Storage	4GB for 1 Chat LLM	12GB for 3 Chat LLMs
Graphics	Integrated Intel® Graphics	Integrated Intel® Arc™ Graphics
Network	Broadband connection for LLMs and other components' download	

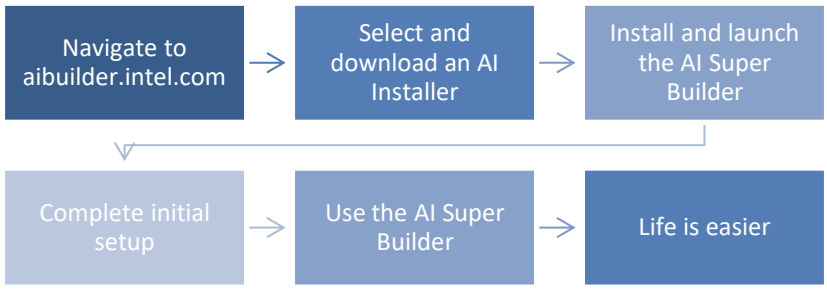
Note:

- Intel® AI Super Builder has been validated on limited Intel AIPC: MTL, LNL, ARL and PTL systems.
- Minimum Intel Graphics driver version is **32.0.101.8132**, and minimum NPU driver version is **32.0.100.4239**. (latest recommended)
- When using MCP services (Super Agent MCP):
  - Recommended AI Model is Qwen3-8B-int4 with 32GB RAM

1.2.2 Operating System (OS) Requirements

Intel® AI Super Builder has been validated for use on **Microsoft Windows 11 version 23H2 or newer**. During the installation process, Intel® AI Super Builder application may download and install required components.

1.3 Quick Start Guide: Easy steps to Get Up and Running



1.4 Download the Intel® AI Super Builder

Visit https://aibuilder.intel.com in your web browser to learn about and download the application.

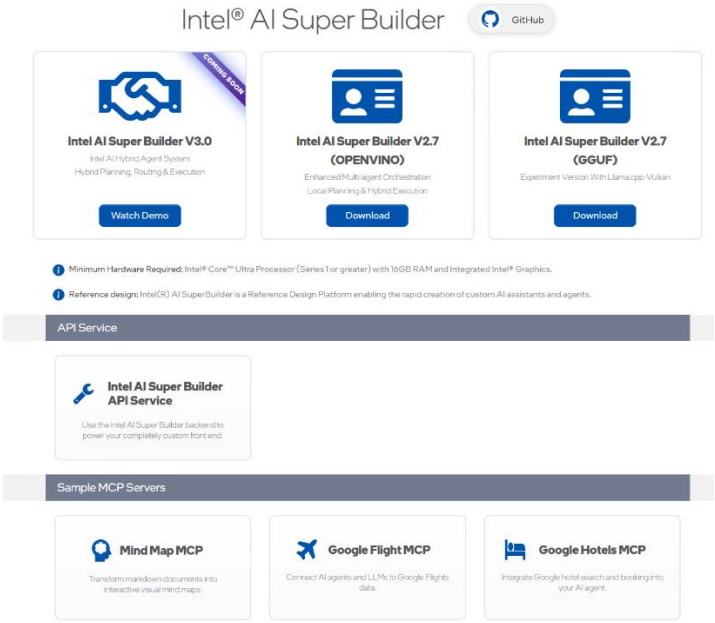
2.0 Getting started

Intel® AI Super Builder currently **supports ONLY one installer at a time** on your AI PC. If you want to switch to a different installer (for example, from OpenVINO to GGUF), you must completely uninstall the existing application, including removing the local database during uninstallation, before installing another one.

2.1 Select AI Installer Type and Install the Application

Step 1. **Go to the website:** Navigate to https://aibuilder.intel.com in your web browser.

Example Illustration of Intel® AI Super Builder Web Portal



Step 2. **Download the installer:** Select one of the available installers to start the download. For optimal performance on Intel® AI PCs, we recommend using the "OpenVINO" installer. To explore new models in GGUF format, use the "GGUF" installer.

Note: Going forward, this guide assumes you are using the "OpenVINO" installer. The experience with the "GGUF" installer is largely similar, but with different LLM model support.

Step 3. **Locate the downloaded file:** Once the installer is downloaded, locate it in your *Downloads* folder.

- The file name of the installer "Intel(R) AI Super Builder_Installer_**XX_Z.Z.Z.YYYY**.exe" has the following meaning:
  - **XX** defines the type of installer: IO for OpenVINO, and IC for GGUF (with llama.cpp backend)

- **Z.Z.Z** is the application version number.
  Note: **YYYY** is the application build date number.

Step 4. **Uninstall Previous Version (if previously installed):**

Go to Settings → Apps → Installed apps.

Search for Intel, then click the ⋮ (More options) next to the app and select Uninstall.

Remove Intel® AI Assistant Builder (application name before v2.7) or Intel® AI Super Builder (application name starting from v2.7).

When prompted, select "**Yes**" to remove local database and local models

Manually delete the following folders to ensure complete removal:
C:\ProgramData\IntelAIA\SqliteDB
C:\ProgramData\IntelAIA\local_models
C:\ProgramData\IntelAIA\vectordb_store

Step 5. **Start the installation:** Launch the installation wizard by double-clicking the downloaded file. The wizard will guide you through the required steps to successfully complete the installation.

Note: Various supporting components are required to install Intel® AI Super Builder. Including:

a) 7-zip
b) Microsoft Visual C++ Redistributable
c) ASP.NET Core Runtime 8.0
d) .Net 8.0 Runtime

If your system does not already have these components installed, they will be included in the Intel® AI Super Builder installation process. You must accept the respective End User License Agreements (EULA) to complete the installation.

If the installation of these required components fails for any reason, you may use the respective links above to manually install them and re-try the Intel® AI Super Builder installation.

Step 6. **Launch the Intel® AI Super Builder application:** If the "Launch when ready" checkbox was selected during installation, the Intel® AI Super Builder application will launch automatically. Otherwise, you can manually launch it from the Windows Start menu.

2.2 First Use and Initial Setup

2.2.1 Download Required Files

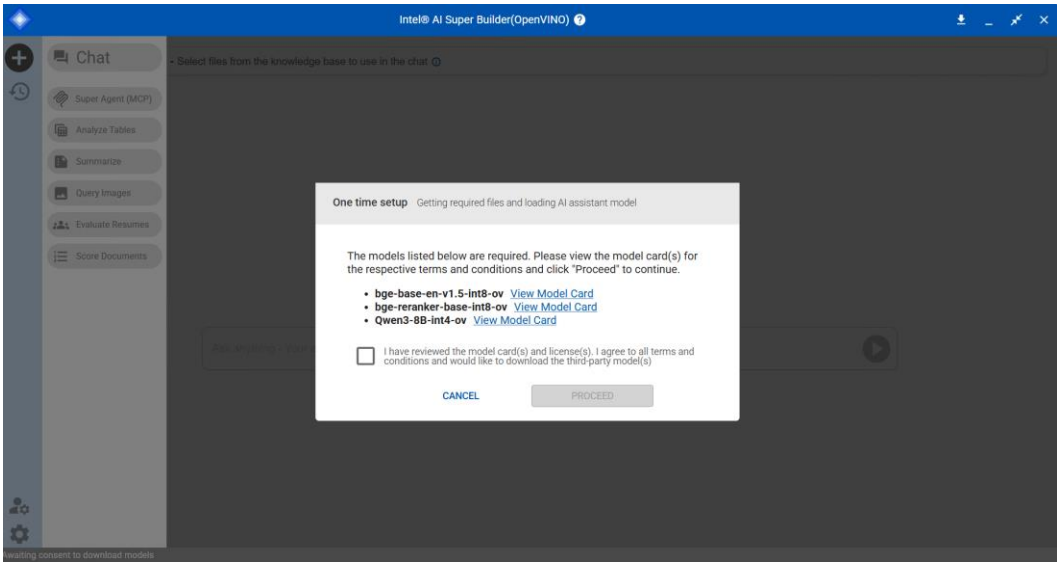
The first time Intel® AI Super Builder runs; you will be prompted to download the required model files. Please follow these steps:

Step 1. Check the box to accept the terms and conditions of use.

Step 2. Click the “Proceed” button to start the download.

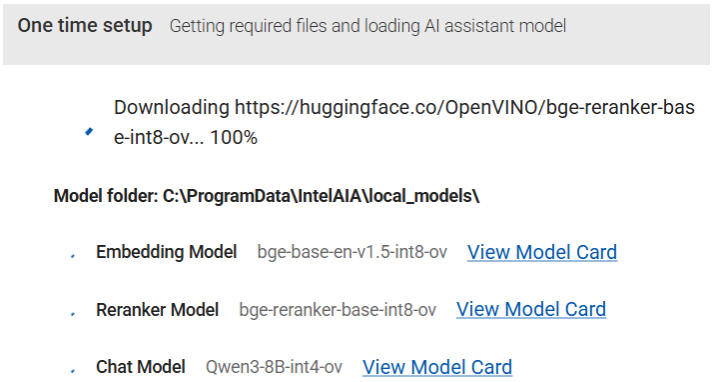
Step 3. Monitor application status—the status bar at the bottom will show progress and disappear once the download is complete.

Example Illustration of Intel® AI Super Builder Application Main Screen at the First Launch

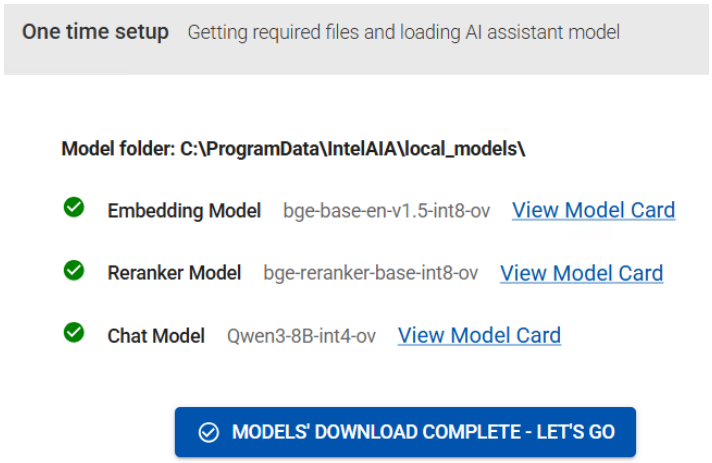


Note: Intel® AI Super Builder cannot be used until all required files are downloaded and activated. Many features will remain disabled until this process is complete. Initial setup typically takes a few minutes, but may take longer depending on your network and hardware.

Example Illustration of LLM, RAG, and Other Components' Download in Progress Dialogue Window



Example Illustration of LLM, RAG, and Other Components' Download Complete Dialogue Window



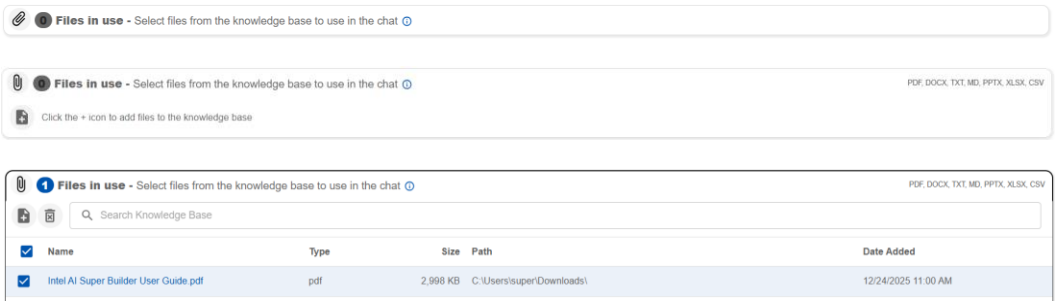
2.2.2 Add Documents to Knowledge Base

The agent uses the LLM model to answer your questions, but its capabilities can be significantly enhanced by leveraging your custom documentation. You can "train" your agent by adding relevant documents (PDF, DOCX, TXT, MD, PPTX, XLSX and CSV formats) improve its effectiveness. To add documents, simply choose one of the following methods:

a) **Drag and Drop** file(s) to the application.

b) **Click the "Add Files" paperclip icon** and browse to add one or more files.

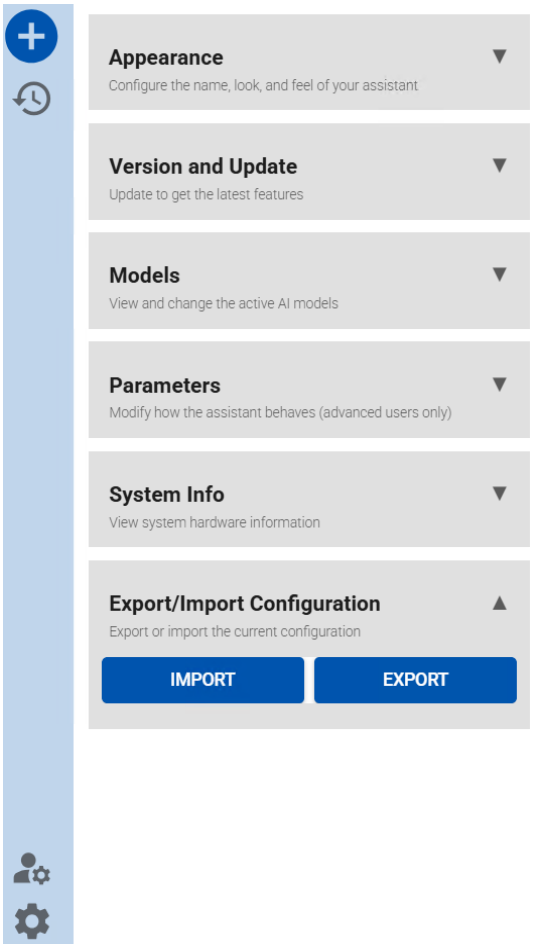
Example Illustration for adding files



For more information on managing these files, please refer to section 3.0 Knowledge Base / File Library.

2.2.3 Customize the Application – Settings

Click the settings icon  the left sidebar to customize your AI Agent.



Through the settings menu, you can:

- **Personalize the app's look and language.**
  - Set a custom name, logo, or color theme.
  - Change UI language.
- **Switch the LLM and RAG models** for your AI Agent.
  **Only use this advanced feature if you understand its effects*
  - Select a model from the drop-down.
  - Upload your own model.
  - Convert text models from HuggingFace or a local folder.
  - Change the model download source.

  Note: When you change the model, it will start downloading if required. The application is unavailable during this process. You can track download process via UI.
- **Adjust LLM and RAG parameters**.
  **Only use this advanced feature if you understand its effects.*
- **View the hardware** details of your AI PC.
- **Export** the application's configuration or **Import** a saved one.
  Save your current theme and setup, then apply it later to quickly switch between specialized or themed agents.

2.2.4 Customize the Application – Admin Mode

To enable or disable “Admin mode”, click on the  located in the left panel.

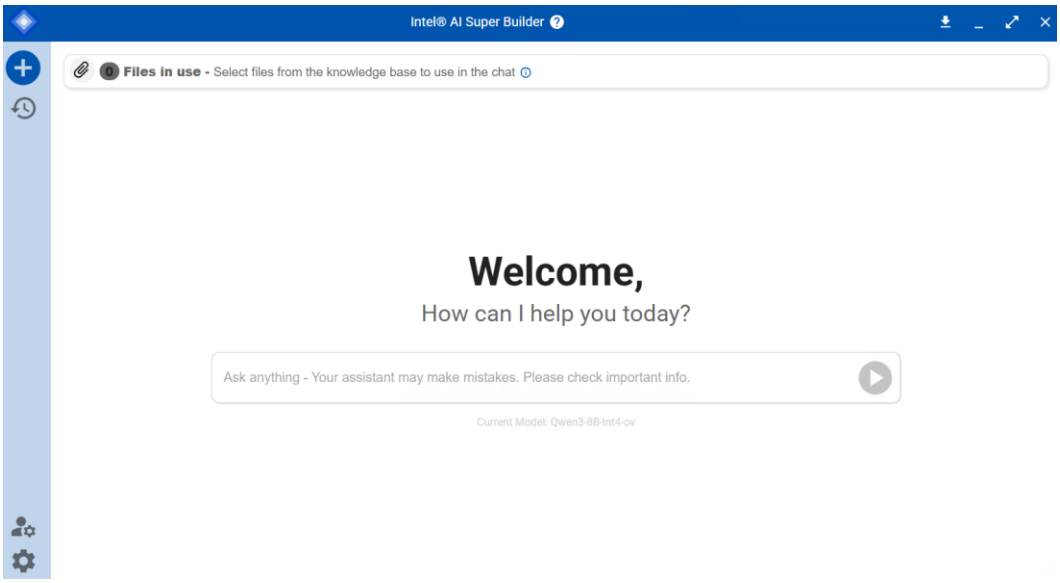
When Admin mode is enabled, the following functions are available:

- Adjust settings (appearance, models, parameters, etc.)
- Modify the knowledge base (add/remove documents used by the AI agent as references to answer questions)
- Use of “Special Queries”

When Admin mode is disabled, access to the above noted functions is removed from the interface and cannot be adjusted.

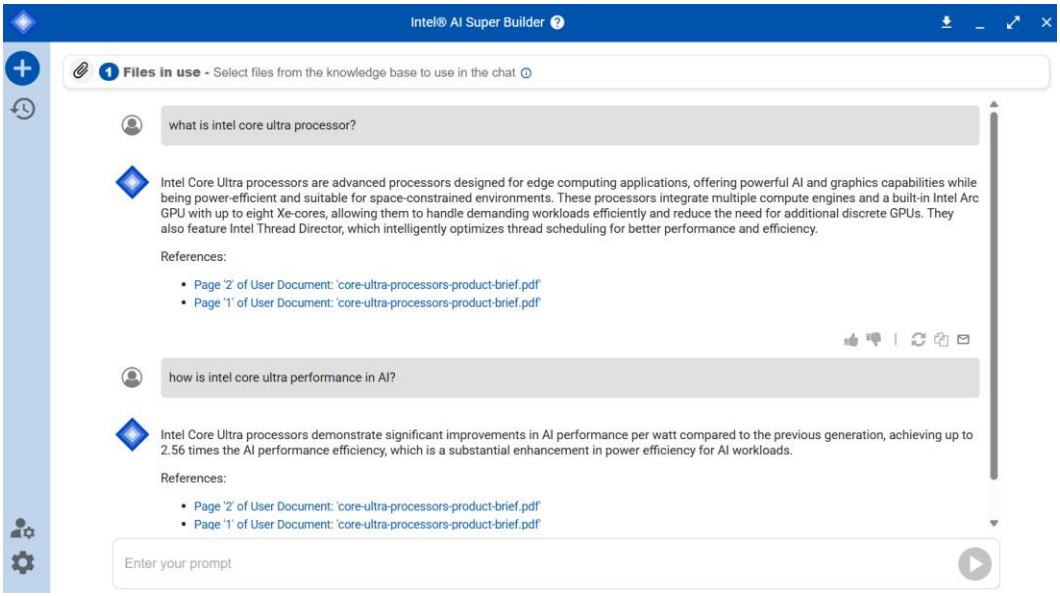
2.2.5 Interact with the AI Agent

After completing the initial setup and any optional customizations, you are ready to start chatting with your AI Agent.



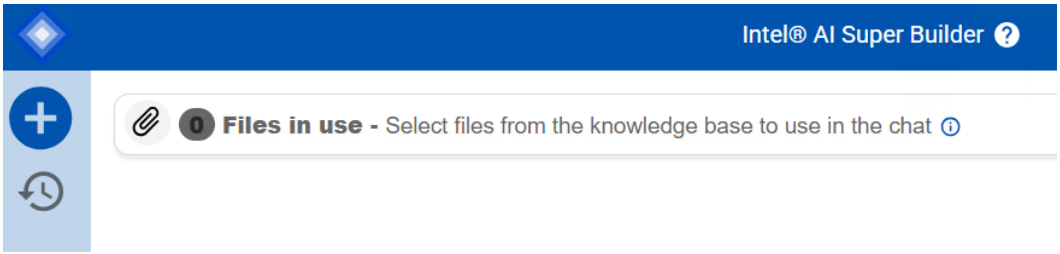
To improve your agent’s responses, add documents related to the topics you want it to address. This provides context and relevant references for your questions. The next section explains how to do this.

Example Illustration of Interaction with the AI Agent



2.2.6 Start a New Chat and Access Chat History

To begin a new conversation, simply click the “Plus” icon at the top of the left sidebar. If you want to revisit or continue a previous chat, just select the “History” icon found beneath the “Plus” icon.



3.0 Knowledge Base / File Library

You can improve your agent's performance by adding files and documents to the shared "Knowledge Base," which helps the agent find information when answering your questions. The agent will also show you which document it used for its response.
The knowledge base is available across all chat sessions. When starting a new session or using certain features, you'll need to choose which files to use within that session. If you prefer, you can adjust the settings so that every session automatically uses all available files in the knowledge base:

- Go to Settings > Parameters > Other > "Use All Files".

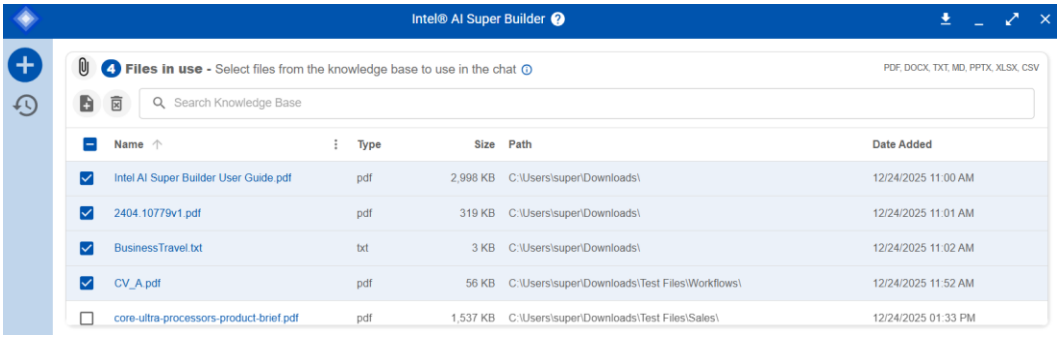
Tip: For the most accurate responses, only select files **relevant** to your current discussion. If you want to switch topics, begin a new chat and pick files related to that subject.

File formats supported: PDF, DOCX, TXT, MD, PPTX, XLSX, and CSV.

File management includes:

- Select files for use with a given chat or "feature" session
- Add/Remove files to/from the knowledge base
- Search for a given document
- Sort the documents by various attributes

Example Illustration of Knowledge Base Feature



Note: When documents are incorporated into or deleted from the knowledge base, the original source files remain in their initial location and are not altered. Only the embedding files generated from these files are retained within the knowledge base.

3.1 How does the AI agent answer questions?

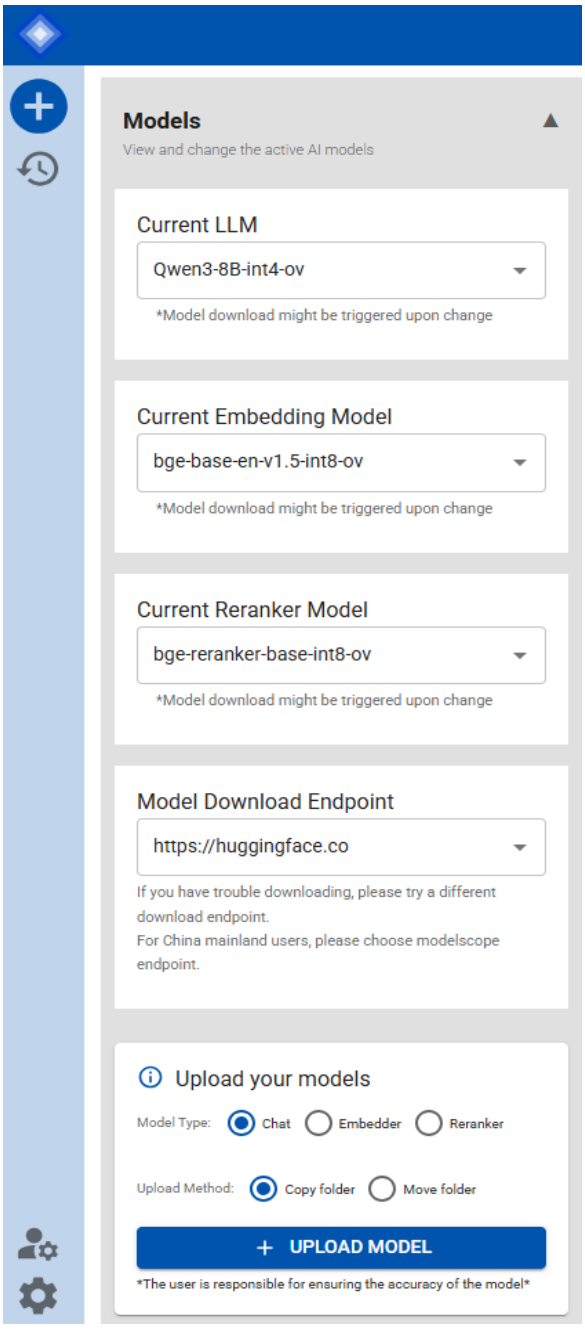
Your AI agent draws from three primary sources to generate responses:

- **LLM baseline knowledge**
  The Large Language Model (LLM) provides broad, public-domain understanding (general facts, language, reasoning). It **does not** inherently know your private policies, internal processes, or personal data.
- **Grounding with your Knowledge Base**
  When you add documents into **Knowledge Base**, the agent retrieves relevant passages ("chunks") and uses them to ground its response. Properly prepared, well-tagged documents dramatically improve precision.
- **Tools & Agents (optional)**
  If configured, the agent may call tools (e.g., MCP agents) to search for information or perform tasks. This extends capabilities beyond static files (e.g., querying an approved data source)

Troubleshooting Unexpected Responses:

When your agent provides unsatisfactory answers, it's usually because the LLM lacks baseline knowledge about your specialized topic (technical subjects, recent developments, industry-specific processes), or your uploaded files aren't easily searchable due to poor document structure, unclear terminology, missing keywords, or fragmented information across multiple files. To fix this, review your documents for clarity, add keyword-rich summaries, and ensure information is well-organized and accessible.

4.0 Model selection



You can select different models from the dropdown list

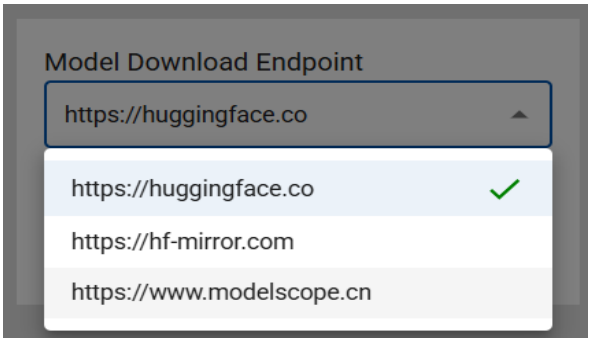
- Go to Settings → Models.
- Use the model dropdown to choose a supported model.
  If it isn't installed, it will download automatically and then become active.

Models can be selected, downloaded, and/or removed from your local system from the dropdown list.

You can also upload custom models to enhance your AI Agent's capabilities, including different chat models, embedder models, and reranker models.

Click the "i" icon next to "Upload your models" for detailed instructions and requirements for uploading custom models.

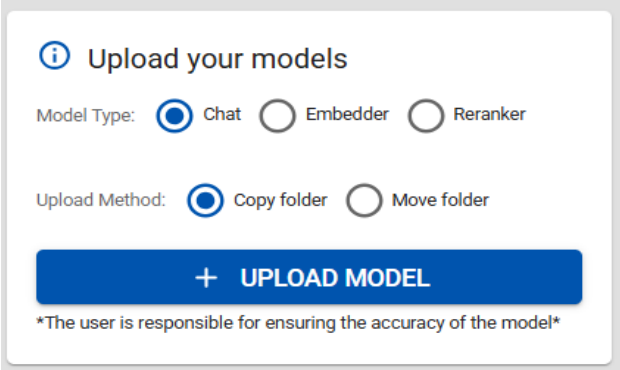
4.1 Change the Model Download Endpoint



Mainland China users should select the "ModelScope" endpoint from the dropdown for quicker and more reliable access to models in the People's Republic of China.

4.2 Upload a Model

You can find more intel OpenVINO compatible models in OpenVINO Huggingface: https://huggingface.co/OpenVINO/models



Please note the procedure provided may not work with all models, and users are responsible for verifying the feasibility of using any given model. Please feel free to contact us if you would like to receive personal recommendations.

Important Notes:

- **Chat models** must be compatible with Inte OpenVINO format to function properly
- **Embedding and reranker models** must be compatible with their vector size
- **Before using new embedding or reranker models**, you must remove all files from your Knowledge Base, as existing embeddings will be incompatible with the new models
- Uploading models is an advanced feature that should only be performed by those who understand the impact of using custom models
- You are responsible for ensuring compatibility and accuracy of any uploaded models

4.3 Convert a Model

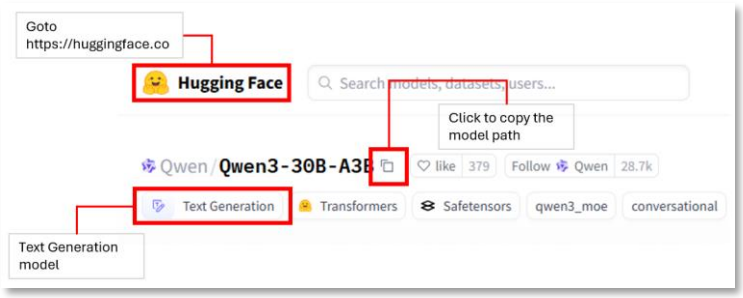
The model conversion feature allows you to convert most* "Text Generation" and "Image-Text-to-Text" models from huggingface.co or a model on your local machine, to an OpenVINO-compatible format for use in the application, all in a few easy steps.

Note: *Not all models will work. You can find supported architecture and models here: https://openvinotoolkit.github.io/openvino.genai/docs/supported-models/
It is important to verify the accuracy of converted models.

Step 1. Locate the model you wish to convert.

- If the model is on your **local machine**, note its path.
- If the model is on **"Hugging Face"**, copy the Hugging Face path.

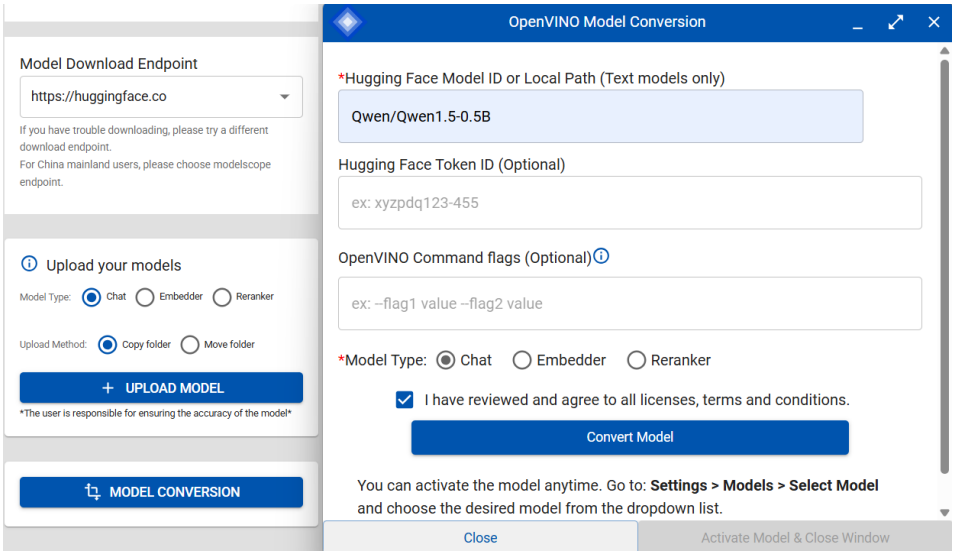
Example Illustration of Getting Model Path from Hugging Face



Step 2. Enter the Model ID path, and add the optional Token ID (if required for Hugging Face sign-in) and OpenVINO command flags as needed or optional.

Step 3. Accept the terms and conditions for the target model and proceed with the model conversion.

Example Illustration of Utilizing the "Model Conversion" Feature

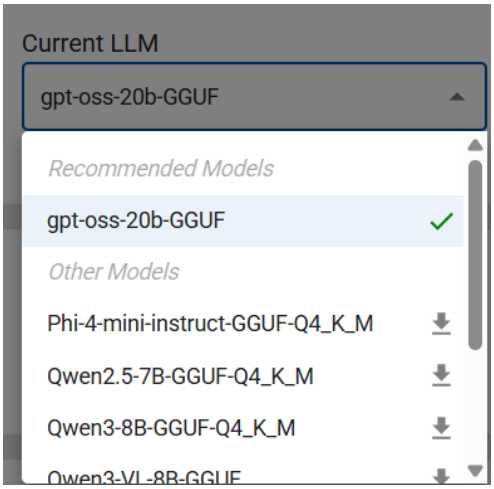


4.4 GGUF installer

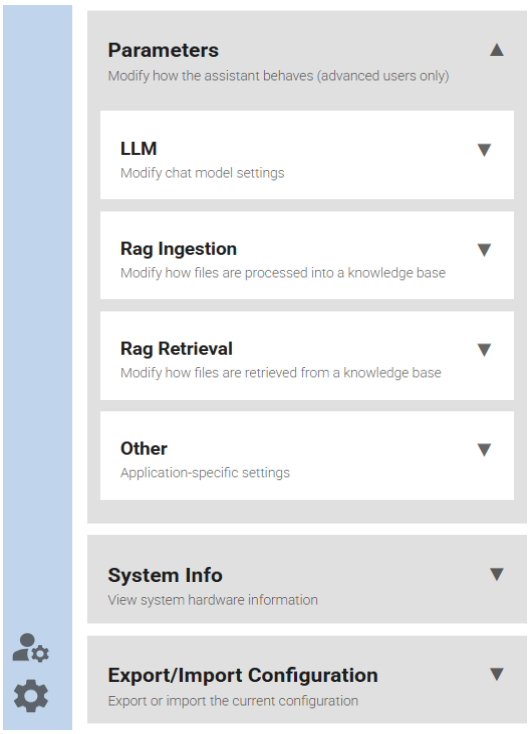
The GGUF installer needs LLM, embedder, and reranker models in the GGUF format, rather than the Intel OpenVINO format. The models you download are located at C:\ProgramData\IntelAIA\local_models and are different depending on the format. Before

installing the GGUF version, it's best to uninstall the OpenVINO version by deleting its local database and models.

Example Illustration of GGUF Models



5.0 Advanced Parameters Configuration



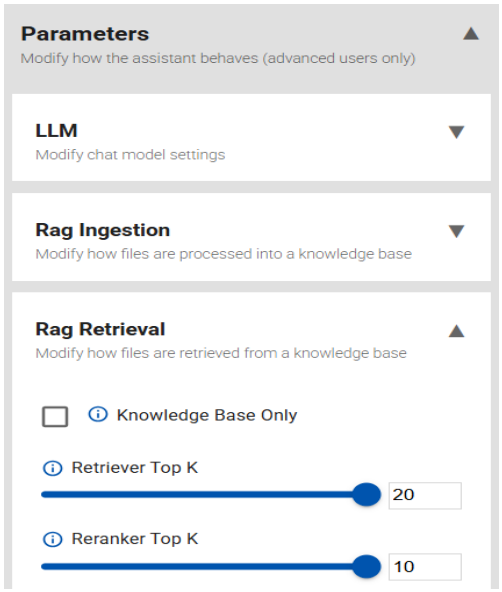
This feature is designed for advanced users who are familiar with the impact of changing application settings.

To adjust the default parameters for LLM, RAG, or other processing options, simply click the “Settings” icon and go to the “Parameters” section.

Pro Tip: Click the “i” icon next to each parameter for detailed explanations and helpful guidance.

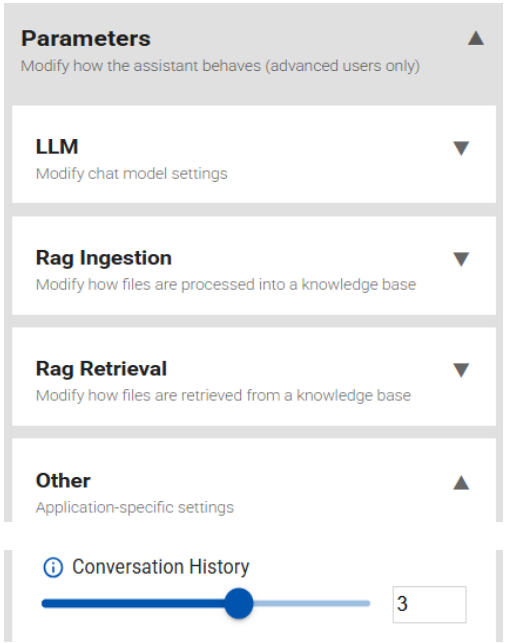
5.1 Practical Examples of Parameter Configuration to Enhance Agent Response

5.1.1 Retrieve More Context from Knowledge Base



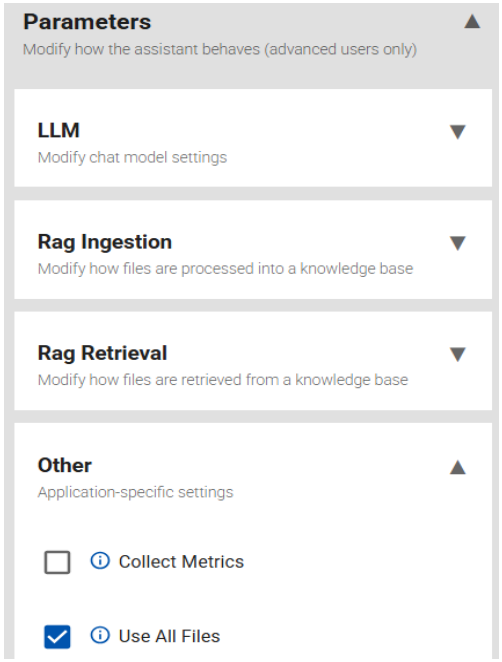
If you need more context from the Knowledge Base, try raising the Retriever Top K and Reranker Top K values.

5.1.2 Include Previous Chat Messages in the Context for the subsequent Responses



If you want the agent to remember earlier chat messages for future responses, try changing the "Conversation History" setting. By default, this history is not included, which allows for faster response times. Increasing the value will help the agent recall more details from earlier exchanges, enhancing the relevance and continuity of its answers.

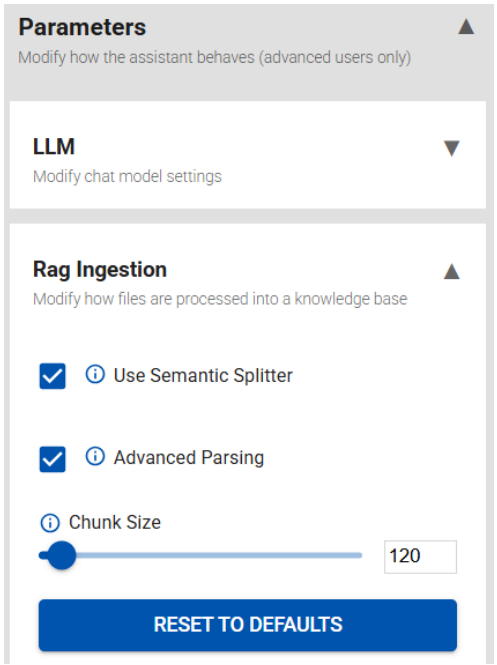
5.1.3 Use All Files from the Knowledge Base by default



If you prefer to have all uploaded files in the Knowledge Base used by default, enable the "Use All Files" setting.

Please note: When many files are included, overall performance may decrease, and the agent's responses could become less precise due to increased information noise.

5.1.4 Enhance Response Accuracy

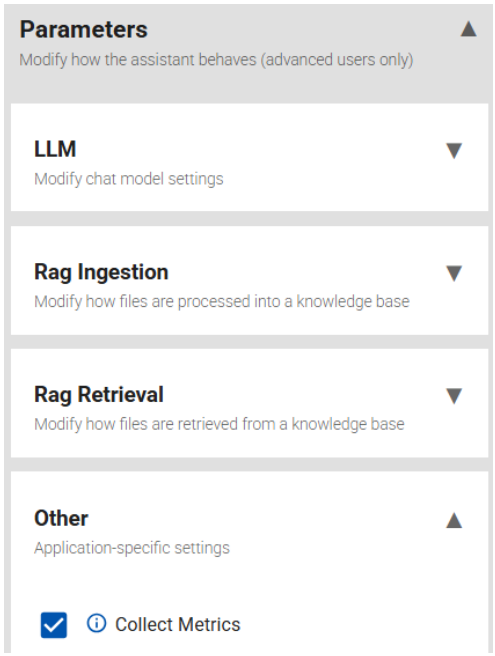


For improved accuracy, enable "Advanced Parsing" to extract tables from PDFs, and "Use Semantic Splitter" to group text by meaning instead of size.

PRO Tips:

- **Advanced Parsing**: Extracts tables from PDFs for better data accuracy. Uploads may take longer if there are many tables.
- **Use Semantic Splitter:** Groups text by meaning instead of size, which can slow down file ingestion but improve retrieval accuracy. The "Chunk Size" setting is ignored when this is enabled.

5.2 Collect Performance Metrics

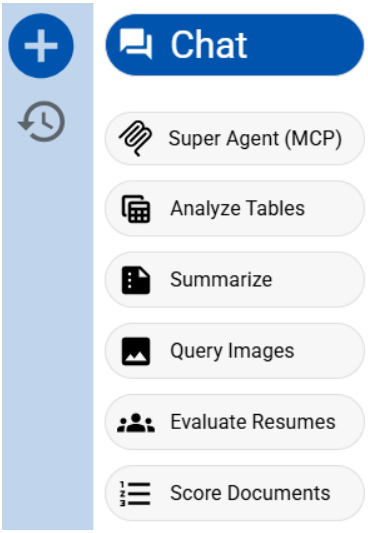


To track how well the app handles file uploads and generates responses, turn on "Collect Metrics" in the Parameters > Other section.

Important notes:

- This feature is designed for use in Chat workflows only; it may interfere with Super Agent and other workflows.
- When enabled, performance data is saved as CSV files in C:/temp/IntelAia/<yyyy-mm-dd>.*.metrics.csv.
- Open these CSV files only when the agent is not running—otherwise, some metrics may not be recorded.

6.0 Special Features (Workflows)



Special features or workflows allow you to guide the agent to "focus" on specific documents, potentially yielding better, more accurate results and enabling it to answer more sophisticated questions based on the context of the specified document(s).

Currently there are 6 types of special workflows:

To use the "Special Features" click the "Plus" icon in the left side bar.

1. **Super Agent (MCP)*:** Super Agent routes queries through MCP services providing unlimited capability using any number of MCP servers and configured MCP agents.

   Note: Super Agent (MCP) is unique among the special feature workflows. Please see section 7 for detailed instructions on how to use and configure this feature.

2. **Analyze Tables:** Analyze files that have a table format where the first row contains column headers, and the rest contains data.

3. **Summarize:** Summarize files and ask follow-up questions based on the summary.

4. **Query Images*:** Query images for text extraction, captioning, and other tasks.

   Note : This requires a vision model. If necessary, you will be prompted to download the recommended model.

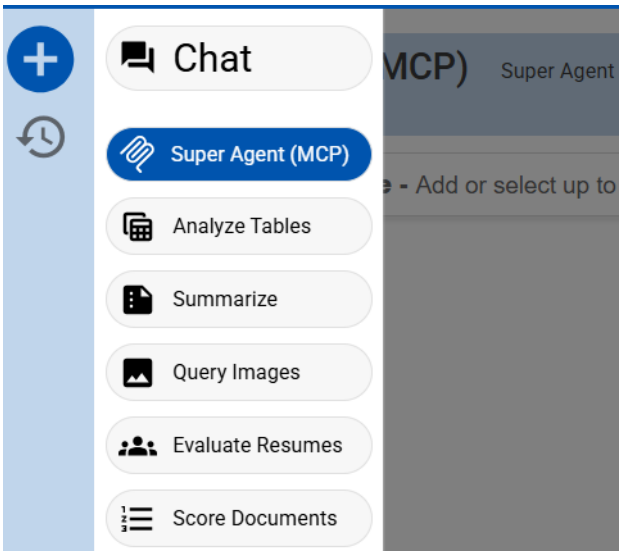
5. **Evaluate Resumes:** Evaluate resumes based on how well they match a specific job description. After the evaluation, you can ask follow-up questions.

6. **Score Documents:** Score unstructured documents against given scoring criteria.

PRO Tip: For the best, most accurate results, use small files that contain only information relevant to the topic.

7.0 Super Agent (MCP)

7.1 What is Super Agent (MCP)



Super Agent (MCP) enables the use of Model Context Protocol (MCP) services to connect AI models to different data sources and tools. MCP enables you to build agents and complex workflows on top of LLMs.

MCP enables nearly limitless capability by providing:

- A growing list of pre-built integrations that your LLM can directly plug into
- A standardized way to build custom integrations for AI applications
- An open protocol that everyone is free to implement and use

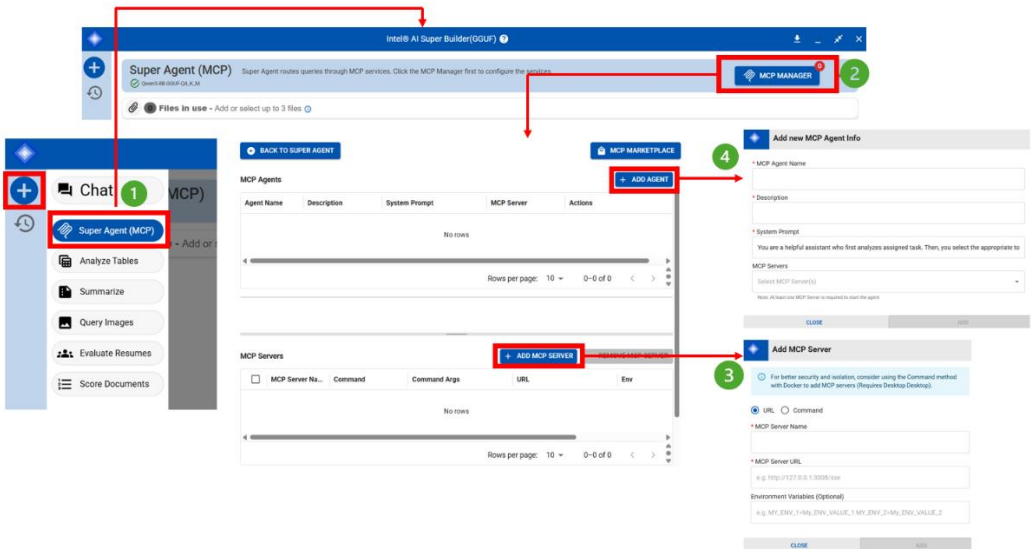
For more detail on MCP services please see the official Model Context Protocol (MCP) page.

7.2 How to configure and use MCP servers and agents

To set up Super Agent MCP, start by selecting the Super Agent (MCP) workflow and then choose “MCP Manager.” Here, you can add and configure both MCP servers and agents.

You can create custom MCP servers using our code generation framework, or you may use open source MCP servers. Be sure to configure MCP agents for the servers and start them before using Super Agent.

For further information, please see the screenshot below.



7.2.1 Deploying Custom-Generated MCP Servers

Custom MCP servers can be generated automatically using our codegen framework. These servers are specifically designed for Intel® AI Super Builder, ensuring optimal performance and smooth integration with the platform's features.

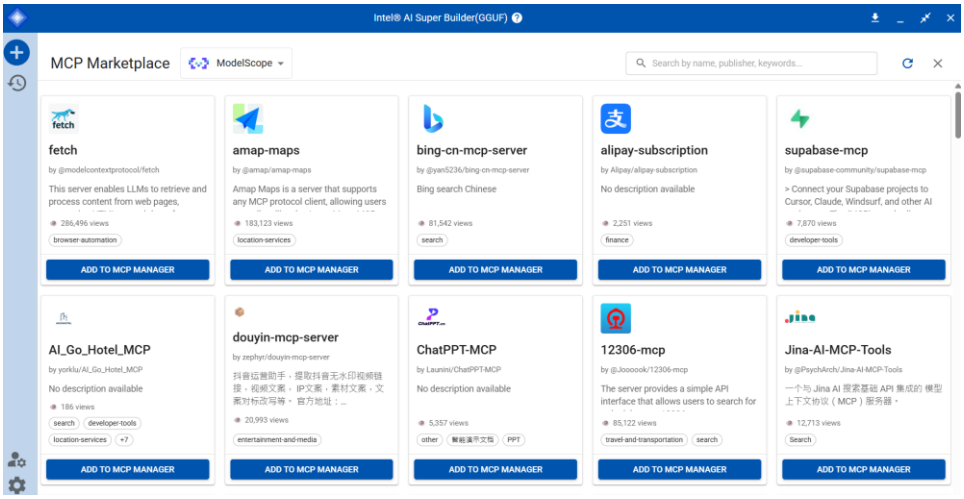
We also provide several examples of MCP server usage in Intel® AI Super Builder. Reviewing these examples can be valuable when configuring your own MCP servers and agents.

You can find instructions for setting up customized MCP servers by visiting our GitHub link.

7.2.2 Integrating Open Source MCP Servers

Intel® AI Super Builder supports community-maintained MCP servers from the official Model Context Protocol repository . These servers provide access to popular tools and services. Our MCP manager also includes the ModelScope MCP marketplace, allowing you to explore various MCP servers.

Illustration of ModelScope MCP Marketplace:



To download MCP servers, please ensure that all required libraries are installed. For detailed guidance, refer to our instructions provided on GitHub. We recommend using Docker for enhanced isolation and security; installation of Docker Desktop is necessary. Additionally, comprehensive instructions for installing MCP servers using pip, npx, and uvx are available on our GitHub page.

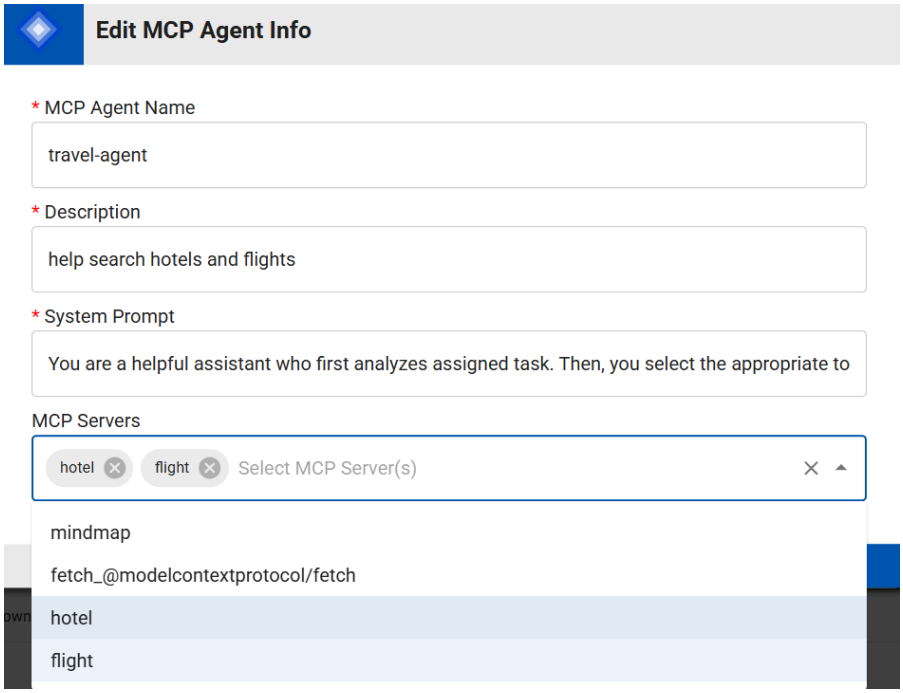
7.3 Configuring MCP Agents in Intel® AI Super Builder

Intel® AI Super Builder functions as a multi-agent orchestration framework, enabling users to coordinate and manage multiple AI agents efficiently. To utilize this system effectively, it is essential to configure MCP agents so they can interact with MCP servers.

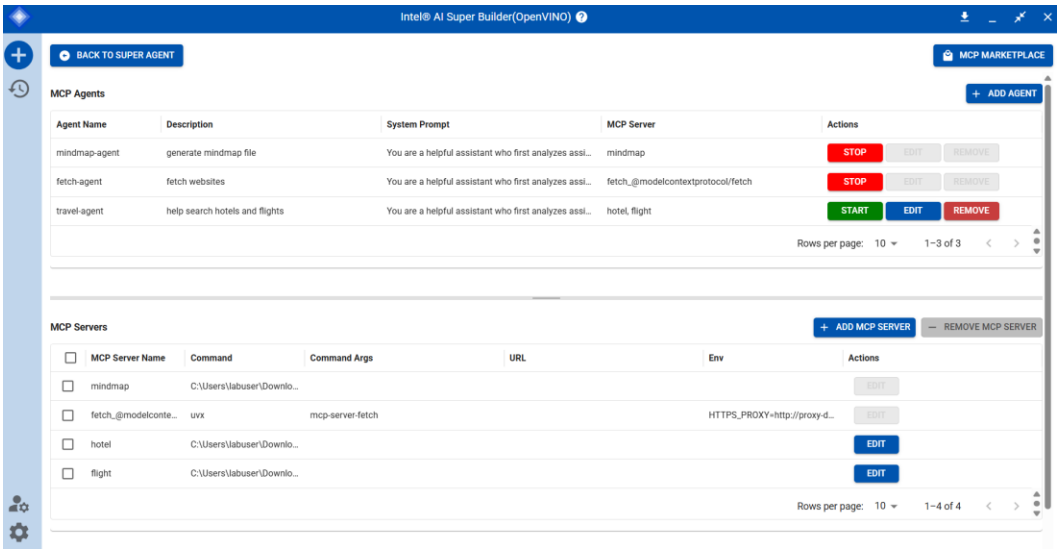
You have the flexibility to assign one or more MCP servers to a single MCP agent. By leveraging multiple MCP servers, an MCP agent can facilitate complex, multi-task workflows. This approach significantly enhances the capabilities and responsiveness of your AI agent, allowing it to handle a broader range of tasks and provide more comprehensive solutions.

See the screenshots below for examples of how to configure an MCP Agent. Also see the Business Travel Agent demo video as an example of how MCP servers can combine to provide multi-task responses within the Super Agent workflow.

How to configure an MCP Agent using one or more MCP Servers:



Multiple MCP Agents can be set up and started/stopped as needed:



7.4 Common problems / Tips & Tricks / Demo Videos

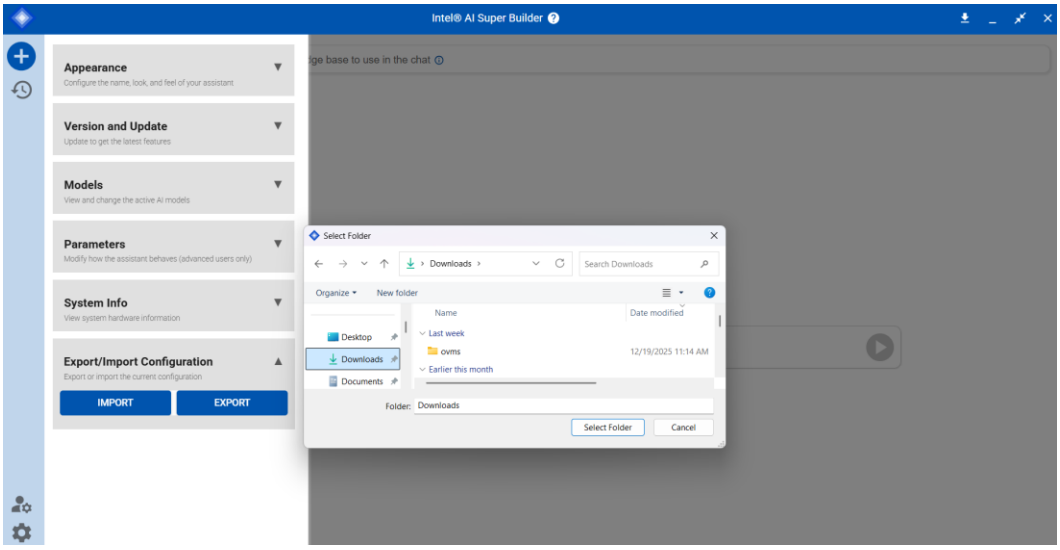
1. **System environment variables:** Intel AI Super Builder runs as a Windows Service, which means it uses System Environment Variables instead of User Environment Variables. If your system is behind a corporate firewall, also add HTTP_PROXY, HTTPS_PROXY, and NO_PROXY to the System Environment.
2. **Connection closed errors:** For detailed instructions and examples, see: Fixing MCP Error 32000 Connection Closed
3. **Demonstration videos:** See the following video demonstrations to see MCP servers in action.

Business Travel Agent	Time Zone Agent	Website Fetch Agent	File System Agent
Flight search and hotel booking with Google Flight and Google Hotel MCP servers	Get current time across different time zones worldwide	Fetch and process website content using the fetch MCP server	Access and manage your local file system

8.0 Import / Export an Agent

8.1 Export an Agent Configuration

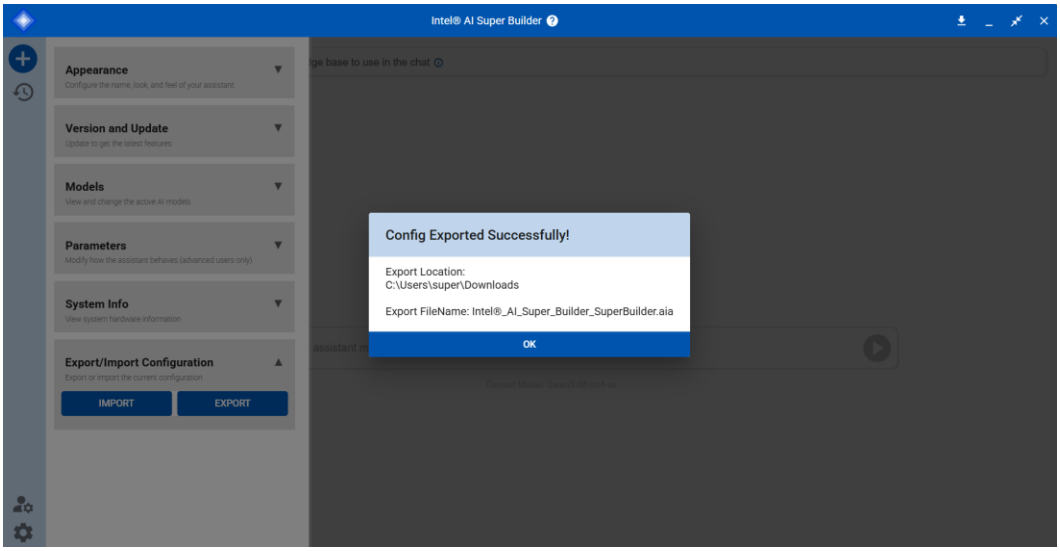
After customizing an AI agent to meet your needs, such as adding documents to the knowledge base, customizing its appearance, tuning parameters, and/or modifying the model(s) used, you can export the configured agent. To export an AI agent configuration, please follow the steps below:



Step 1. Click App Settings > Export/Import Configuration > EXPORT.

Step 2. Select a location to save the exported configuration.

Step 3. The agent configuration will be exported, and a confirmation message will be displayed.

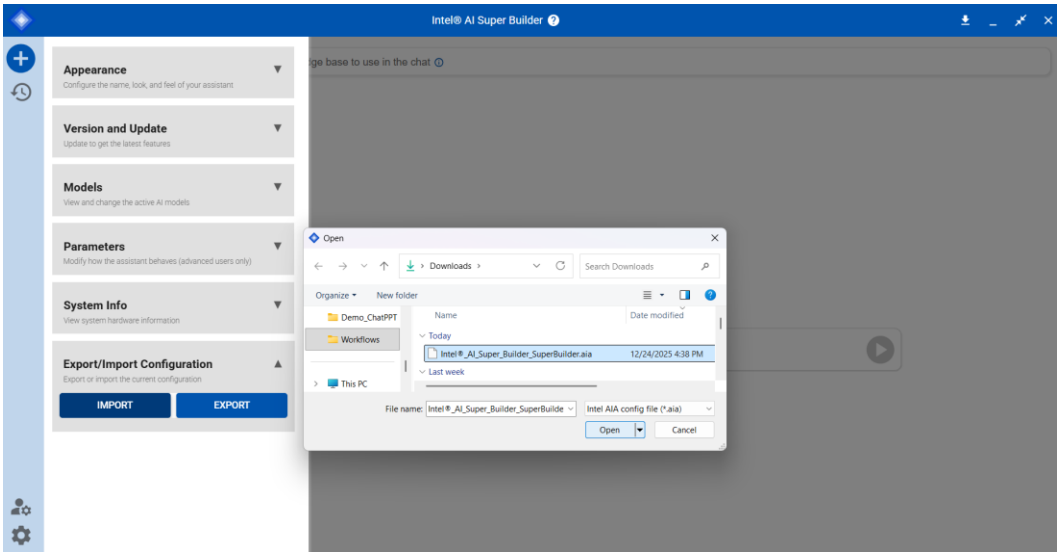


Once exported, the configuration can be imported later, or when distributing the configuration to another AI PC.

Note: OpenVINO installer configurations cannot be used with GGUF installer, and vice versa, due to differences between the models.

8.2 Import an Agent Configuration

To import an AI agent with all its configuration, please follow the steps below:



Step 1. Click App Settings > Export/Import Configuration > IMPORT.

Step 2. Select the. aia configuration file you want to import.

Step 3. All the associated configurations will automatically be applied to your application including:

- **Appearance**: Customizations related to look and feel.
- **Knowledge base**: Includes any feedback given to train or correct the agent
- **Parameters tuning:** Adjustments made to optimize performance.
- **Custom Models:** Custom models available in the Models drop down list.

Note: OpenVINO installer configurations cannot be used with GGUF installer, and vice versa, due to differences between the models.

9.0 Troubleshooting and Known Issues

9.1 Installation Issues

Some antivirus software such as **McAfee** Antivirus software is known to interfere with the installation process of Intel® AI Super Builder on Windows systems. If you encounter installation problems and have antivirus software installed, please stop the real-time scanning feature and then reinstall Intel® AI Super Builder. Once the installation is done and the models are loaded, you can re-enable it. Users might experience performance impact when antivirus real-time scanning is running.

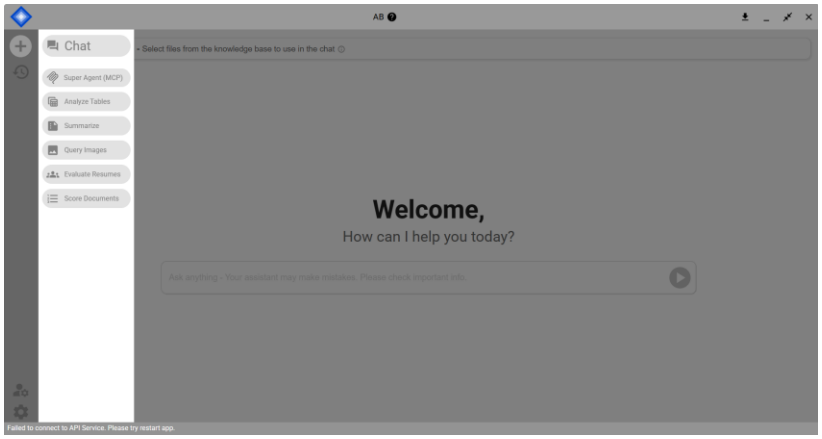
Important Note: If you have Intel AI Assistant Builder installed (before v2.7), please uninstall it by removing all the database and local models before installing version 2.7.

1. Press **Win + I** (Settings)⟶ **Apps** ⟶ **Installed apps** (Windows 11) or **Apps & features** (Windows 10).
2. Find Intel AI Assistant Builder app ⟶ click …⟶ **Uninstall** ⟶ confirm.
3. Click "Yes" to remove local database and local models.

9.2 Service Failure Issue

If you see failure in API Service or Chat error, please try to restart our service in Windows.

1. Click **Start** and type **Services**, then open the **Services** app.
   *(Alternatively, press **Win + R**, type services.msc, and press **Enter**.)*
2. In the list, find **Intel® AI Super Builder**.
3. **Right-click** it and choose **Start** or **Restart**.
   If Restart isn't available, click Stop, wait a few seconds, then click Start.



If this issue still exists, we recommend uninstall our app with removing local database and models, and try installing again.

9.3 Upgrade Errors

Since we change our application name in v2.7, please upgrade with removing your local database. We recommend uninstalling the application using Window's built-in *Add or Remove Programs* and click “Yes” to remove local database and models, then installing the updated version. **Also, once the application is launched, make sure to delete all the files from the knowledge base.**

9.4 Model Download Errors

This issue could be due to a few reasons:

- **HTTP/HTTPS Proxy for Enterprise Environments**: If you are an enterprise user and your organization uses HTTP or HTTPS proxy, you need to configure it on your AI PC to enable the Intel® AI Super Builder to download LLMs, RAG, and other necessary components. Please consult your organization’s IT department to determine which proxy server(s) are in use and how to configure them up on your device.

How to Set Proxy Environment Variables

1. Open Windows Settings (Win + I), go to System > About, and click Advanced system settings.
2. Click Environment Variables at the bottom of the System Properties window.
3. Under **System** variables, click New:
   - For HTTP proxy: Set name to HTTP_PROXY and value to your proxy address.
   - For HTTPS proxy: Set name to HTTPS_PROXY and value to your proxy address.
4. Click OK to save your changes and close the windows.
5. Restart your computer to apply the settings.

- **Model Download Endpoint**: Check your model download endpoint and consider selecting a different download server.
  - **For users in the PRC**, ***Model Scope*** is recommended.
  - **For users in other regions**, please choose the ***Hugging Face***.

9.5 Initial Load Time / Unresponsive

When the AI Super Builder is started, the API service and models must be initialized before the application can be used. During this time, the chat text entry field will be disabled, and a status message at the bottom of the window will indicate what is happening. When the application is ready, the status bar at the bottom of the window will be removed and the chat text entry will be enabled. Please note that the actual waiting time varies.

9.6 Model Loading Errors

If a “model loading error” occurs, please make sure to update the GPU and NPU drivers to the latest version. The NPU model requires NPU Driver **32.0.100.4239 at a minimum**. Please also try to restart the Intel AI Super Builder service in Windows.

9.7 Backend Not Ready

During the first-time Intel® AI Super Builder start-up, errors may occur due to backend not ready condition, especially if the application is running for the first time on a system with a slow network connection. You can right-click in app and then refresh the UI.

9.8 Export Config Issues

For configurations involving a large set of documents, only a partial set of text chunks is currently exported. Support for exporting a full set is currently under development.

9.9 Model Conversion Error

The model conversion tool within Intel® AI Super Builder only supports models compatible with the **Intel® OpenVINO** platform. You can find support models here: https://openvinotoolkit.github.io/openvino.genai/docs/supported-models/.

Here is the link to download already converted models OpenVINO huggingface: https://huggingface.co/OpenVINO/models.

9.10 Conversation History - Reset to Defaults Issue

Although the **Reset to Defaults** button sets the **Conversation History to 0**, the conversation history is still utilized in the context. To resolve this issue, please manually adjust the value using the slider.

9.11 White Title Bar Issue After Upgrade Issue

When upgrading from an older version to a newer one, you might see a white title bar at the top. This is a known issue. You can use the Appearance Reset button in the settings menu to fix the style. If you want to keep your existing style, make sure you export and back up your existing style before you reset. We recommend removing local database when you upgrade to v2.7.

9.12 Intermittent Qwen 2.5 Model on MTL Issue

The Qwen2.5 model has a known intermittent issue on MTL where they occasionally generate unwanted responses with exclamation marks. This behavior is not consistently reproducible, but retrying the query typically resolves the problem.

9.13 Phi-4-mini Model Generates Negative Scoring Issue

The Phi-4-mini model produces negative scores in the Document Scoring and Resume Evaluation workflows. It is recommended to use Qwen3 or Qwen2.5 models for these workflows instead.

10.0 Feedback and Support

For technical questions and feature requests, please use GitHub Issues.

We would love to hear about your experience. Please contact us at support.aibuilder@intel.com.