

Openwebui User Manual

Part 1: Setup and Getting Started

Before starting Open-WebUI, you need to already start the LLM model and embedding model through middleware.

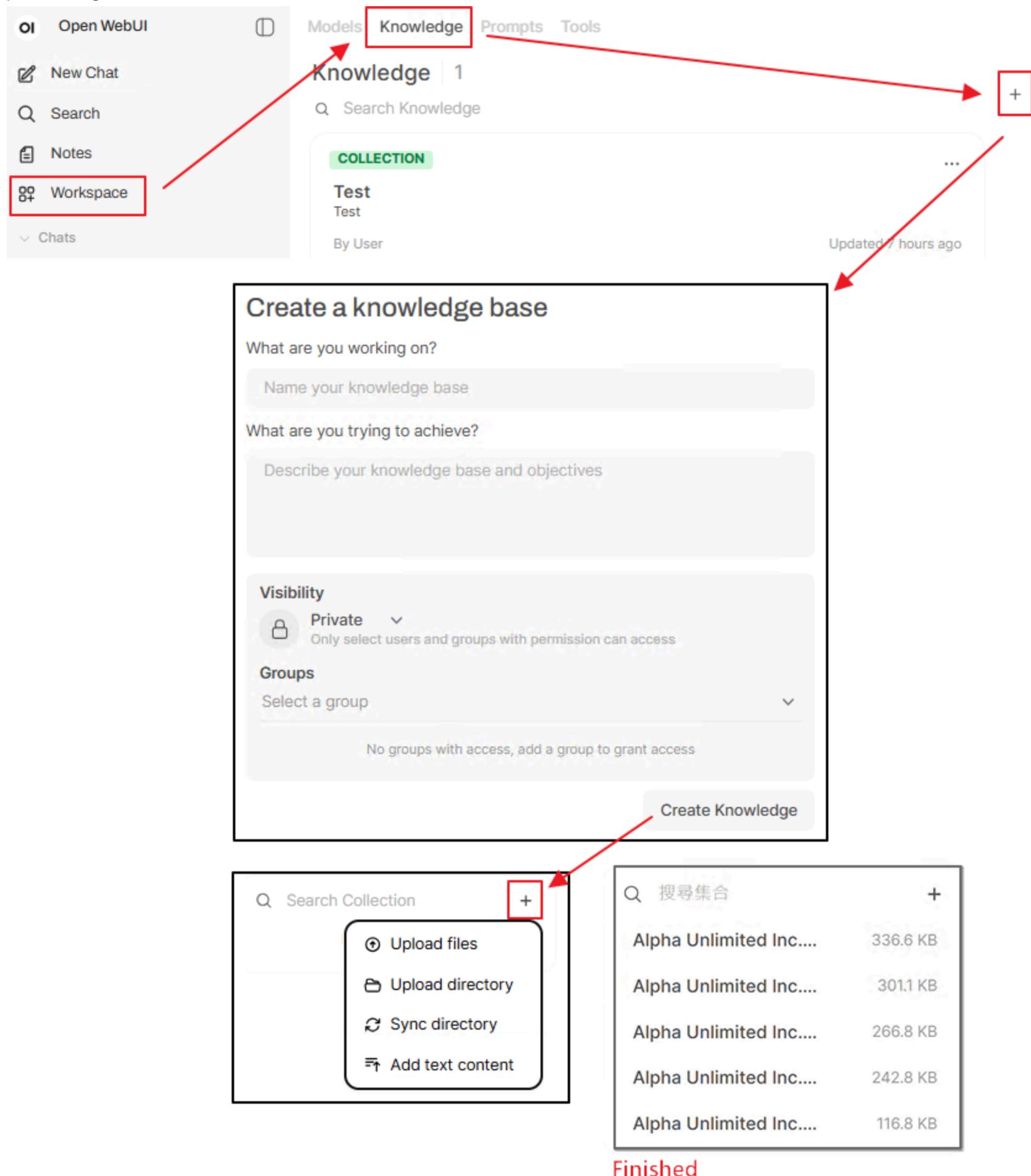
1. Please download `open-webui.zip` and `setup_and_run.bat` and place them in the same directory.

 `open-webui` 2025/11/28 下午 01:54 壓縮的 (zipped) 資... 92,571 KB
 `setup_and_run` 2025/11/28 下午 03:45 Windows 批次檔案 27 KB

2. Double-click `setup_and_run.bat` to launch Open-WebUI.

3. The first time you launch Open-WebUI, it will unzip `open-webui.zip`, install Python, and set up the environment. Then you can use Open-WebUI in your browser.

4. Upload files to Knowledge. We have prepared some sample questions in the `open-webui` folder. Please wait until the LLM finishes processing.



Part 2: Running Inference + RAG

Query Methods

1. **Normal Chat:** Directly enter your question in the chat box (will not perform collection retrieval)

2. Agent Chat:

- (1) Type <# hashtag symbol> in the chat box and click the collection you create.
- (2) Enter your question after the hashtag

COLLECTION Phison_Collection
aiDAPTIV RAG Collection: Phison_Collection (10 files)

#

+ Code Interpreter

Microphone icon

Verifying Execution Flow

Please confirm the following indicators:

1. **Time To First Token (TTFT) is between 2 ~ 8 seconds**
2. **RAG reference documents are displayed below the response**

Alpha Collection

How many days of full-pay sick leave does the company provide annually?

OI C:\Users\phison\Desktop\AgentBuilderClient\Installer_0.2.1\package\inference_model\Meta-Llama-3.1-8B-...

Time to first token: 2.32 s

The provided content does not mention the number of days of full-pay sick leave the company provides annually.
Total Time: 5.21 s

a7b45d63fd154d03b670bcb96cccf596_Alpha Unlimited Inc. ...

Edit, Copy, Share, Like, Dislike, Reply, Delete