KNEO 300 User Manual V1.5

(For Regular User)

Jun, 2024

info @ kneron.us

Revision History:

Doc Version	Description	Firmware Version	Author	Date
0.9	Initial version	-		2023/10/23
1.0	Add set-up guide	-		2024/01/08
1.1	Add custom settings	-		2024/01/17
1.2	Add admin user guide	V0.14.1		2024/03/07
1.3	Update product image	V0.14.1		2024/04/03
1.4	Rewrite the guide	V0.15.0		2024/04/15
1.5	Update product interface	V0.16.0	Oscar Law	2024/05/07
1.7	Update base on Oscar's version	V0.16.0	Warren Kuo	2024/6/20

Notice:

1. Kneron (Taiwan) Co., Ltd may make changes to any information in this document at any time without any prior notice. The information herein is subject to change without notice.

2. THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY OR CONDITION OF ANY KIND, EITHER EXPRESS, IMPLIED OR STATUTORY, INCLUDING, WITHOUT LIMITATION, ANY WARRANTY OR CONDITION WITH RESPECT TO MERCHANTABILITY, FITNESS FOR ANY PARTICULAR PURPOSE, OR NON-INFRINGEMENT .KNERON DOES NOT ASSUME ANY RESPONSIBILITY AND LIABILITY FOR ITS USE NOR FOR ANY INFRINGEMENT OF PATENTS OR OTHER RIGHTS OF THE THIRD PARTIES WHICH MAY RESULT FROM ITS USE.

3. Information in this document is provided in connection with Kneron products.

4. All referenced brands, product names, service names, and trademarks in this document are the property by their respective owners.

info @ kneron.us

Contents

KNEO 300 User Manual V1.5 1
1. Introduction 4
2. Device Tutorial 5
2.1 Product Overview 5
2.2 Accessories List 6
2.3 Power on 6
2.5 WEBUI Interface 7
2.5.1 Session Initialization 7
2.5.3 Free Chat Mode 10
2.5.4 Knowledge Base Mode 11

info @ kneron.us

1. Introduction

KNEO 300 is an NPU-based edge AI server, especially used to implement LLM applications, supporting 30TOPS AI computing power, equipped with an all-metal casing, fan cooling, and rich peripheral interfaces. Compared with traditional GPU LLM inference, it has the advantages of low cost, low power consumption, and high efficiency, and can be widely used in fields such as enterprise AIGC.

KNEO 300 has built-in Kneron self-developed edge chatbot software, mainly used to answer questions and provide information. Its function is similar to an advanced offline virtual assistant. Here are some of the key features and uses of this chat product:

1. Q&A: Ability to answer various questions covering a wide range of topics such as science, history, culture, technology, etc.
2. Language Understanding: Strong understanding of natural language and the ability to understand and respond to complex and abstract queries.
3. Text generation: In addition to answering questions, you can write articles, create stories, generate creative content, etc.
4. User interaction: Able to have smooth conversations with users and provide helpful answers and suggestions based on database and other information. Wide range of applications: education, customer support, HR, company training, IT support, etc.
5. Privacy and Security: This system adopts offline mode, which greatly protects the security of user information, data, and privacy.

info @ kneron.us

2. Device Tutorial

2.1 Product Overview

- KNEO 300 series AI box appearance

Figure 2-1 KNEO 300 Series AI Box

- KNE300 series AI box peripheral interfaces (from left to right)

1. UP : RS232
2. Down : RS485
3. UP : Ethernet (1000mbps)
4. Down : USB3.0x2
5. UP : Ethernet (1000mbps)
6. HDMI 2.0
7. TF Card
8. DC 12V
9. Power button

@ kneron.us

- Product parameters

CPU	8-core A53, 2.0GHz
NPU	30 TOPS (INT8)
DRAM	16GB LPDDR4
eMMC	64GB
Power	DC12V, AC100-240V, 50-60HZ
Power consumption	Typical=20W, Max=45W
Operating System	Ubuntu
Size	210mm*130mm*45mm
Operating Environment	Operating Temperature: -20℃~60℃; Storage Temperature: -20℃~70℃; Operating Humidity: 10%~90%RH;
Ethernet	2*Gigabit Ethernet
USB	2*USB3.0
Connecting Ports	1*RS232
	1*RS485

Table 2-1 KNEO 300 Product Specification

2.2 Accessories List

After receiving the device, check whether the accessories are complete:

- KNEO 300 AI box
- One 12V-5A power adapter
- One HDMI cable
- One Ethernet cable
- A pack of expansion screws

In addition, during use, you also need the following conditions:

- Display
  Monitor or TV with HDMI port.
- Network
  100M/1000M wired network.

2.3 Power on

- Connect the power cable to the 12V-5A power adapter.
- Connect the device and monitor with an HDMI cable.
- Plug the network cable into UP: Ethernet (1) and connect to the network.
- The device will automatically turn on after being powered on. The monitor will display the IP address (i.e. 10.200.210.227) on the screen.

info @ kneron.us

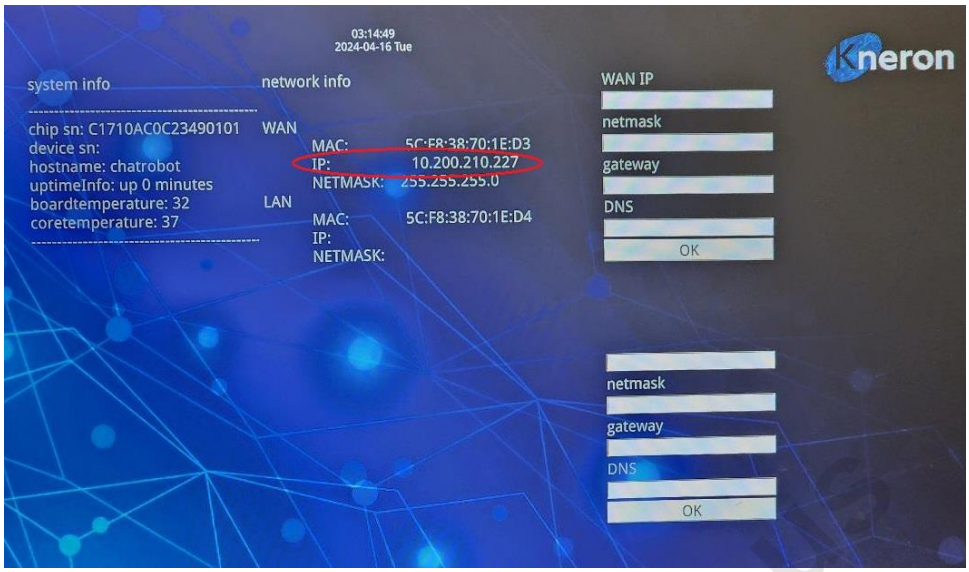

Figure 2-2 KNEO 300 IP Address

2.5 WEBUI Interface

2.5.1 Session Initialization

The user starts the web access using the WEBUI interface. For internal intranet access, it first types the IP address (i.e. 10.200.210.227) with port 3000 in the browser, and the web address becomes (10.200.210.227:3000). For external internet access, it types the website domain <domain.com> with port 3000 in the browser, and the website address becomes (<domain.com>:3000)

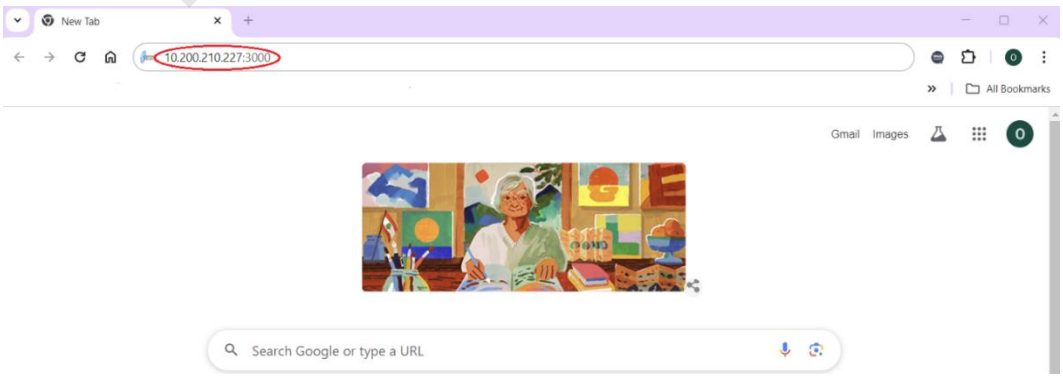

Figure 2-3 Browser Access

It invokes the WEBUI interface shown in Figure 2-6, the default interface is English, and the user can click

the icon and switch to Chinese.

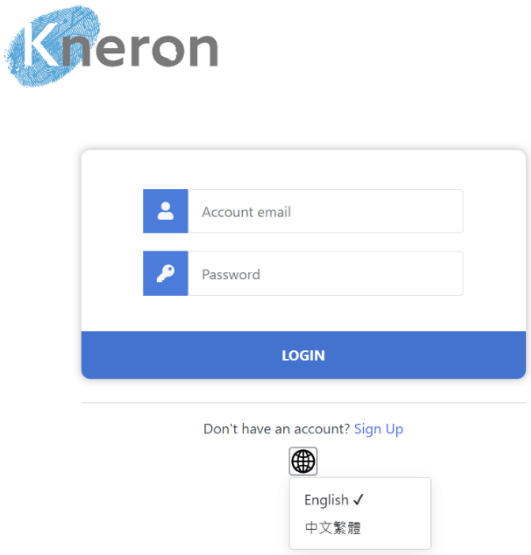

Figure 2-4 WEBUI Login Page

The default WEBUI session is English, the user can click the icon to switch between Chinese and English.

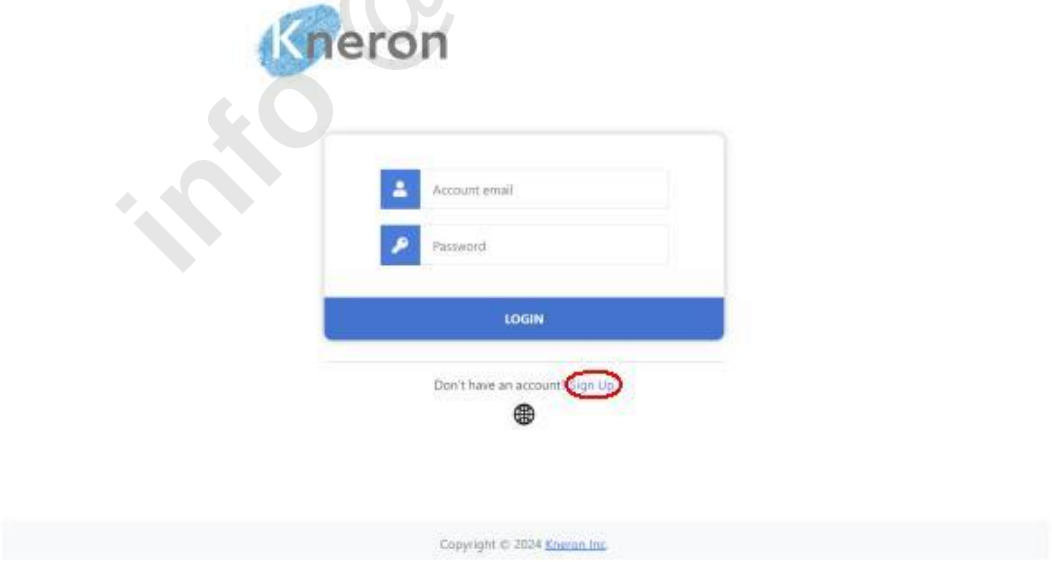

Figure 2-5 New User Sign Up

For new users, they must sign up with the new user account using the Sign Up button during the login. The new user fills up the new account information in the pop-up menu, including the username, e-mail

info@kneron.us

address, and password. Then, the new user can click the SIGN UP button to register the new account. It requires the e-mail rather than the username to log in to the KNEO 300. The username is displayed in the lower left-hand corner after the login.

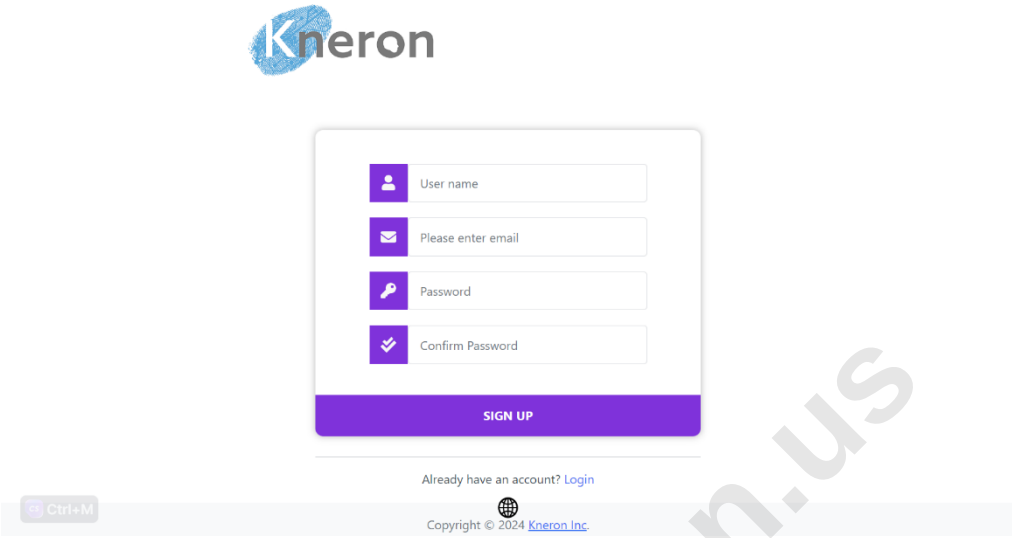

Figure 2-6 New User Registration

Figure 2-7 WEBUI Session

The user can click the chat history using the top left-hand corner history button to hide/unhide the chat history. The user clicks the top right-hand Setting button to go to the setup menu and toggle between different chat modes.

info @ kneron.us

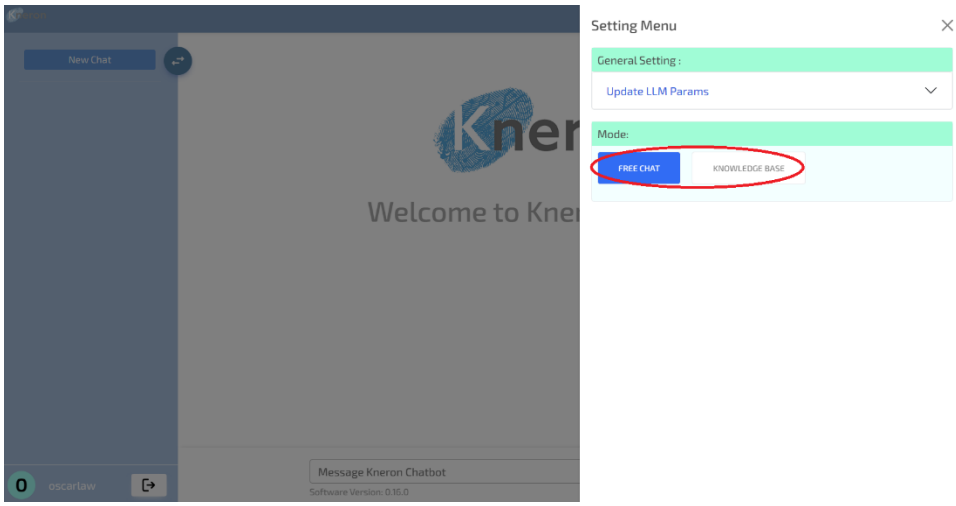

Figure 2-8 WEBUI Setting Menu

2.5.3 Free Chat Mode

Currently, the KNEO 300 supports two chat modes: Free Chat and Knowledge Base, which can be selected using the chat mode button: FREE CHAT and KNOWLEDGE BASE. Free Chat is the default chat mode, which answers the general inquiry, and the Knowledge Base is the custom database created by the users. The administrator can switch between Free Chat and Knowledge Base mode using the mode buttons.
For the default free chat inquiry, the user enters the inquiry in the Message Kneron Chatbot box and hits the green arrow key, then the result is shown in the Dialogue Box. The users can stop or clean the inquiry using the CLEAR RECORD button.

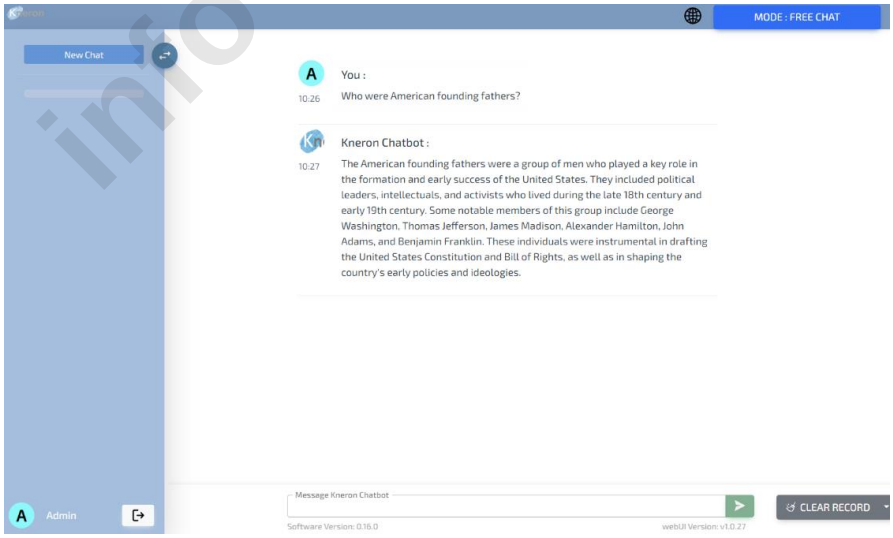

Figure 2-9 Free Chat Mode

The latest software provides additional control for inquiry results using Update LLM Params shown in

info @ kneron.us

Figure 2-10. It offers two postprocessing modes: Top P and Greedy, Top P creates random outputs based on the [Top P] sliding bar setting between 0 and 1. The output becomes more random for larger values. Temperature controls the diversity of the outputs, whereas a lower temperature produces more conservative and predictable outputs. On the other hand, a higher temperature leads to more diverse and creative outputs. The temperature ranges from 0 to 1. Greedy normally produces the same outputs due to the highest matching probability. Finally, the Repetition Penalty controls the output repetition, it will not create the same output with a higher penalty. The penalty is set between 1 and 10.

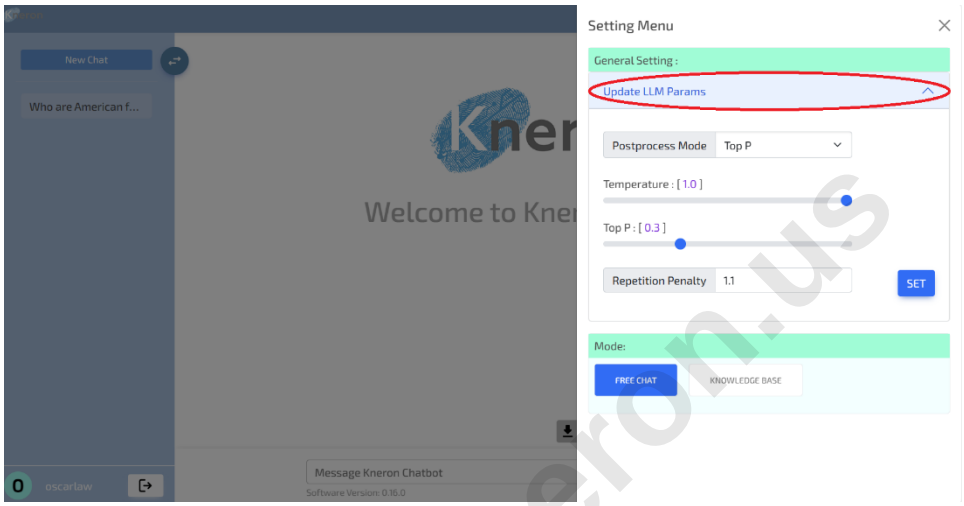

Figure 2-10 Chat Mode Setting

2.5.4 Knowledge Base Mode

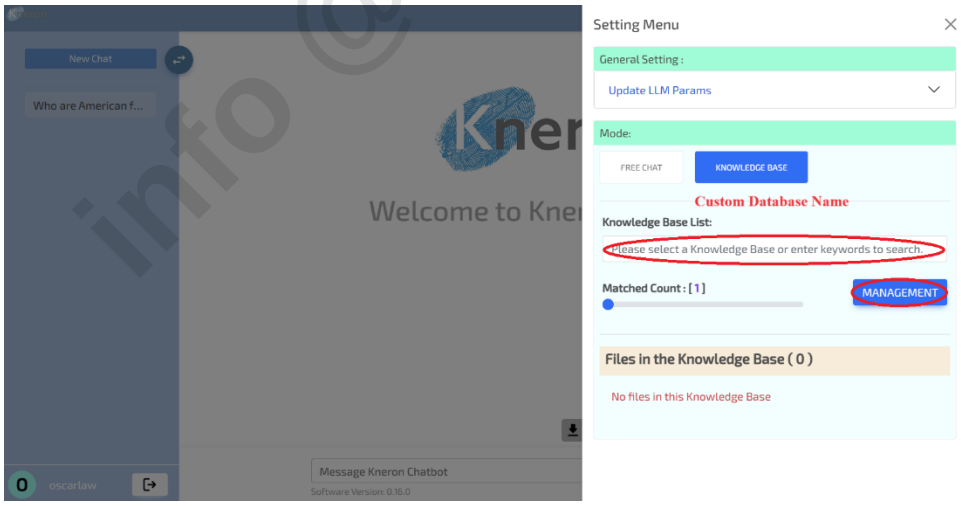

Figure 2-11 Knowledge Base Management

For knowledge base inquiry, the user first presses the KNOWLEDGE BASE button to switch to the custom knowledge base mode and creates the custom database using the MANAGEMENT button. After pressing the button, the management pop-up menu shows up. The user enters the database name <database> in the Knowledge Base List box (i.e. bda602), then clicks the UPLOAD button and uploads the files in the

info @ kneron.us

Drop files box with the UPLOAD FILES button. KNEO 300 supports multiple file formats .txt (text), .pdf (portable document format), .docx (Microsoft Words), .csv (Microsoft Excel) .md (markdown-formatted text), and .zip (compressed file): The file name must not contain any special characters (i.e. (), {}, [] etc.). It takes a few minutes or more to upload the file depending on the file size. The user can create the database under the user directories only.

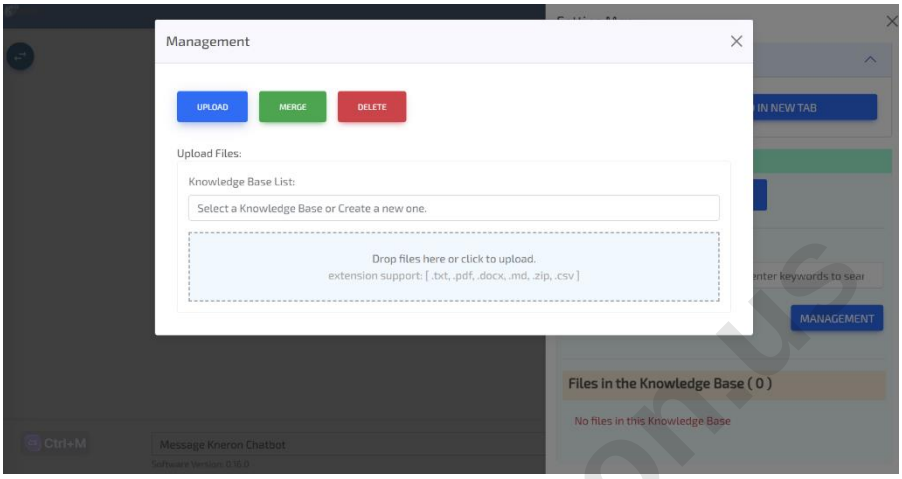

Figure 2-12 Custom Base Creation

After the custom database is created, it clicks the top right-hand corner cross symbol and returns to the knowledge base inquiry page. The custom database name is shown in the Knowledge Base List box and the database content is displayed in the Files in the Knowledge Base box. The user can choose different databases in the Files in the Knowledge Base box for inquiry, then, type in the prompt in the Message Kneron Chatbot. The administrator can increase the matching results using the Matched Count: [n] sliding bar where n is 1, 2, 3 to access more matching results.

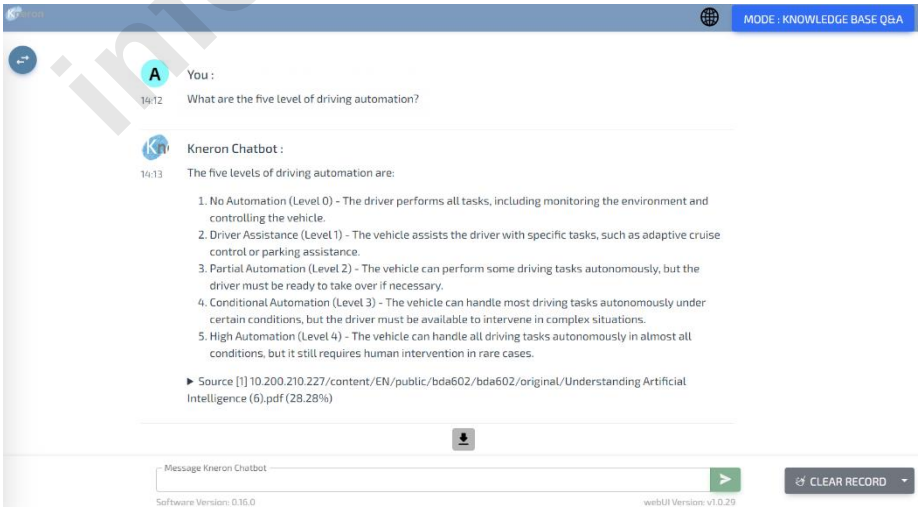

Figure 2-13 Custom Database Inquiry

info @ kneron.us

The pop-up menu offers two additional functions: merge and delete. To merge two databases, it first clicks the MERGE button, a new pop-up menu shows how to merge from one database to another.

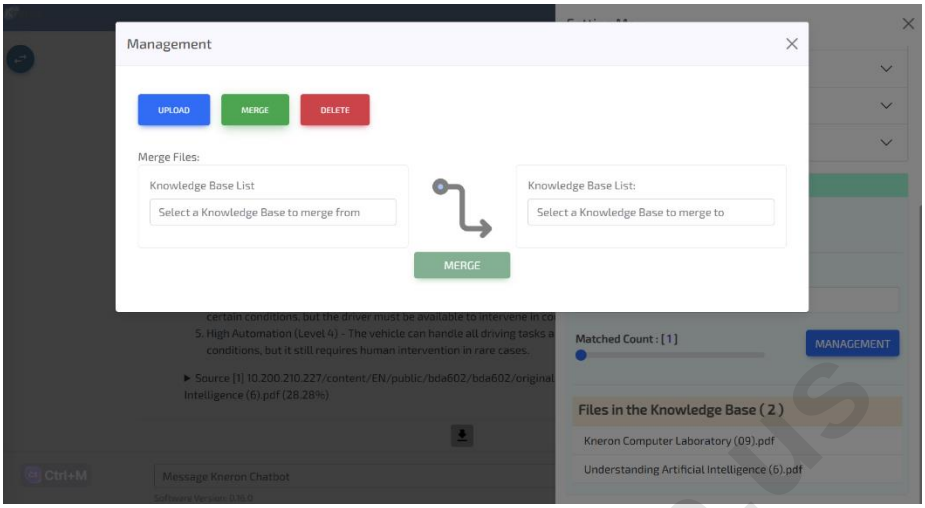

Figure 2-14 Merge Custom Database

To delete the files or databases, it clicks the DELETE button, the administrator can highlight single/multiple files to delete using the DELETE SELECTED FILES button. If there are no highlighted files, it deletes the entire custom database using the DELETE KNOWLEDGE BASE button and removes the custom database from the file system.

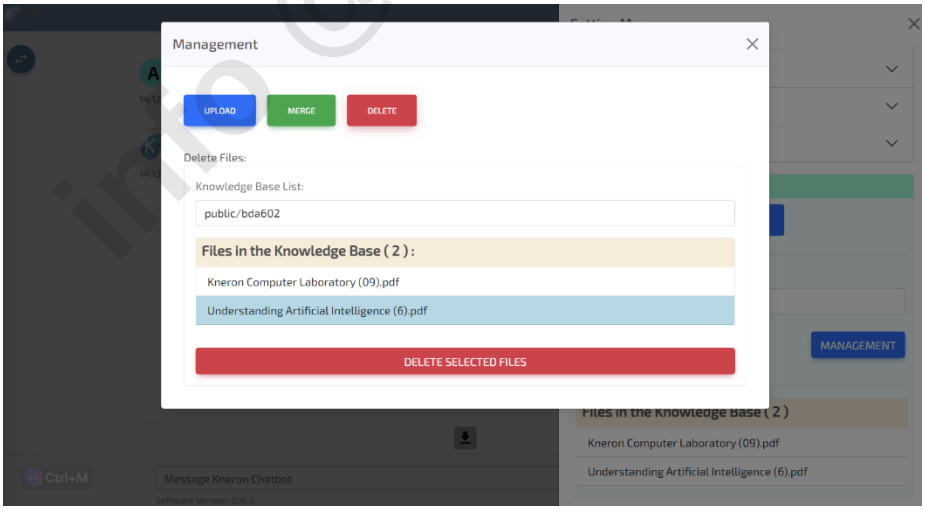

Figure 2-15 Delete Single/Multiple Files or Custom Database

info@kneron.us

info @ kneron.us