KNEO 330 EdgeGPT Server
Quick User Manual

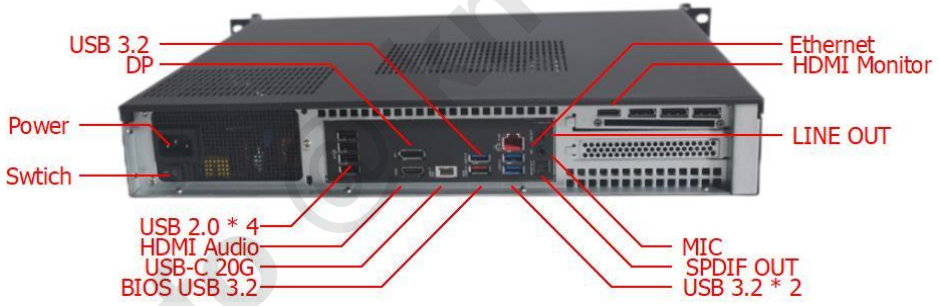
(v 0.20.9)

Mar 2025

info @ kneron.us

1 Introduction

The KNEO 330 is an EdgeGPT server powered by an NPU, tailored for Large Language Model (LLM) applications and offering 48 TOPS of AI computing performance. It features an all-metal body with fan-based cooling and boasts multiple peripheral interfaces for enhanced functionality. The KNEO 330 excels in cost-effectiveness, energy efficiency, and overall performance for Artificial Intelligence Generated Content (AIGC) applications. The system supports multiple-user access and functions like the chatbot with user-defined answers. It also supports embedded image (pdf) and video subtitles (srt) file formats. Finally, it allows the users to access the custom knowledge base through the private group.



The KNEO 330[1] is a standard equipment module for data centers. Please consult IT professionals about installation. The KNEO 330 adopts an air-cooling system, it is recommended to leave space at the top to ensure proper airflow for cooling. When the front panel ON/OFF button is activated (indicated by the blue `light), it initiates the system in active mode. When the button is turned off, it initiates a soft shutdown, placing the system in standby mode without cutting off the power supply. Pressing the button wakes up the system. The back panel switch directly connects/disconnects the power supply. When the switch is off, the system is completely shut down even if the front panel ON/OFF button is pressed. It recommends that the administrator turn off the ON/OFF button first, followed by the back panel switch, to move the system.

[1] Use only the top right HDMI port to connect to the monitor for display, while the other HDMI port is designated solely for audio connection only

CPU	Intel i5 10 cores 16 threads 4.6 GHz CPU
NPU	48 TOPS (INT8) equivalent
DRAM	32 Gb DDR4
Storage	2Tb SSD
Power	100-240V, 50/60Hz Avg 140W, Max 320W
Operating System	Ubuntu Linux
Size	428 x 350 x 66.6 mm (16.85 x 13.78 x 2.62 in)
Weight	7.3 kg (16.09 lb.)

1.1 Accessories List

- KNEO 330 EdgeGPT Server
- One AC power cord

1.2 Hardware Configuration

- Connect the power cable to the 100-240V 50/60Hz power cord.
- Connect the device and monitor with the HDMI cable.
- Plug the network cable into the Ethernet port and connect it to the network.
- Once powered on, the device will automatically start, and the default terminal is initialized. The administrator first logs in to the system with user ID: aiuser with password: aiuser, then types the command: ifconfig to display the server IP address shown in inet entry (i.e. 10.200.210.237) for web access

```
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.8.0-40-generic x86_64)

. . . .

Last login: Fri Sep 13 14:23:45 2024 from 10.200.211.96
aiuser@kneron330:~$ ifconfig
eno1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 10.200.210.237  netmask 255.255.255.0  broadcast 192.168.200.255
        inet6 fe80::2979:5613:2a22:d58  prefixlen 64  scopeid 0x20<link>
        ether 10:7c:61:74:cd:d0  txqueuelen 1000  (Ethernet)
```

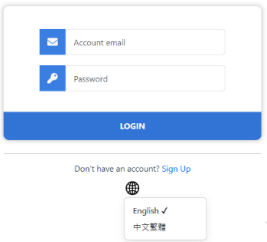
info @ kneron.us

2 WEBUI Interface

2.1 Session Initialization

The administrator initiates web access using the WEBUI interface. For intranet access, it must employ the prefix: Hypertext Transfer Protocol Secure (i.e. HTTPS), then enter the IP address (e.g., 10.200.210.237) followed by port 3000 in the browser, resulting in the web address (https://10.200.210.237:3000). For internet access, the domain name (e.g., <domain.com>) is entered with port 3000, resulting in the web address (https://<domain.com>:3000). The administrator must enter the IP address: (i.e. https://10.200.210.237:3000) to log in to the system.

The WEBUI interface is set to English by default, it can click the icon to switch to Traditional Chinese. The administrator can log in to the system using the username: admin@useradmin.com with the password: admin123.

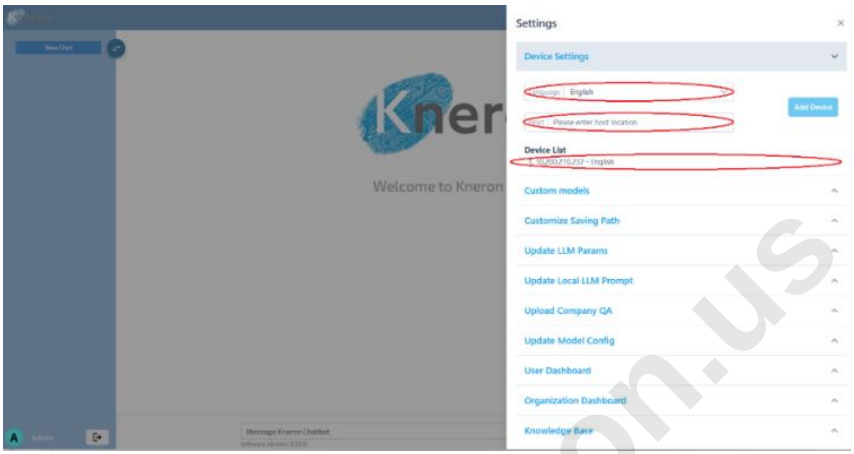


The menu in the top-left corner displays the chat history and includes a toggle button to switch it on or off. The username is shown at the bottom-left corner, along with a button for logging out. In the top-right corner, the administrator can find the Free Chat and Knowledge Base modes toggle button. The administrator clicks the setting button in the top-right corner menu button to open the Settings Menu and switches the language between English and Chinese using the language button. The software release is indicated in the version display.

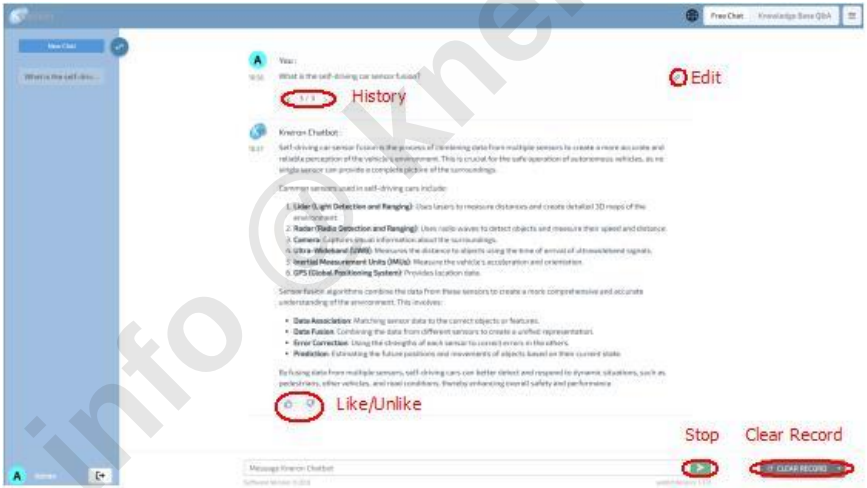
In the Setting menu, the administrator first sets the language to English in the Language box, enters the device IP address (e.g., 10.200.210.237) in the Host field, and clicks the ADD DEVICE button. It takes a few minutes for the KNEO 330 to initialize the system. Once initialization is completed, the device IP address appears in the Device List dialog box. The machine is linked with the system, it is no longer required to initialize the machine in the future. The device IP address will be shown in the Device List dialog box during the next login. The administrator can follow the same steps

info @ Kneron.us

to add additional machines for inquiries, it is strongly recommended to run only one machine during an inquiry session.



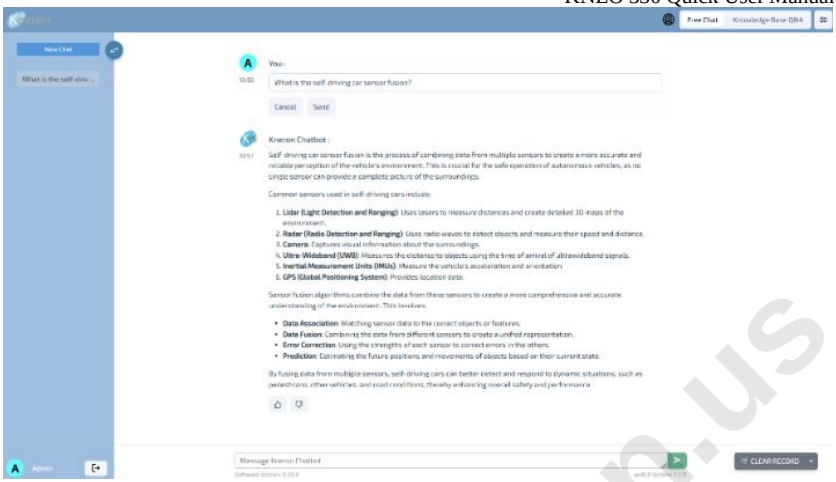
2.2 Free Chat Mode



For the default Free Chat mode, the user enters the inquiry into the Message Kneron Chatbot box and presses the green arrow key, the response appears in the Dialogue Box. The chat history is shown under the New Chat section. The user edits the prompt using the edit icon and iterates the history with the history icon (< n/n >) where n indicates the times of modification, The user can stop the chat using the red stop button or clear the inquiry using the CLEAR RECORD button.

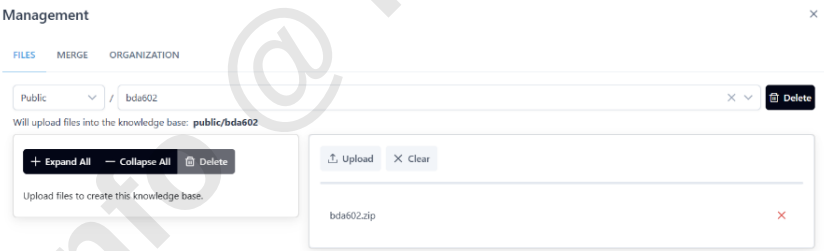
The like/unlike icon is used for system statistical analysis. By clicking the edit icon, the administrator opens a window where the prompt can be modified. The Send button evaluates the prompt, and the Cancel button dismisses the inquiry.

info@kneron.us



The user can click on the chat tag to reopen a chat session, and the chat tag can be modified by using the edit button. To download a chat session, the user clicks the download button; the session will be saved in JSON format and compressed into the local machine Download directory. Additionally, the user can select a chat session and click the DELETE SESSION button to remove it. A warning message will appear, prompting the user to confirm the deletion of the chat session from history.

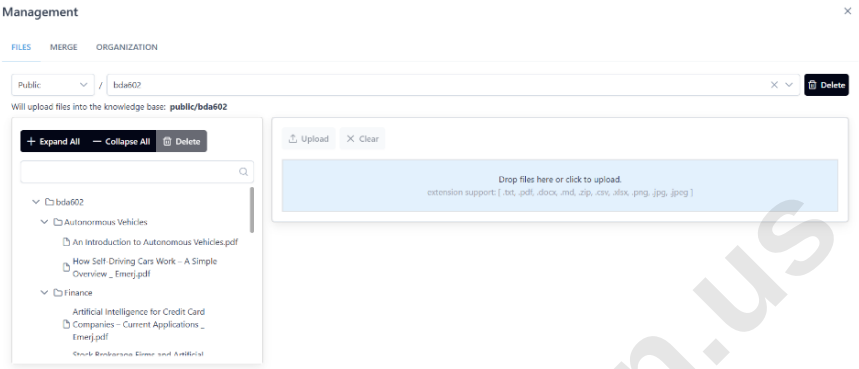
2.3 Knowledge Base Mode



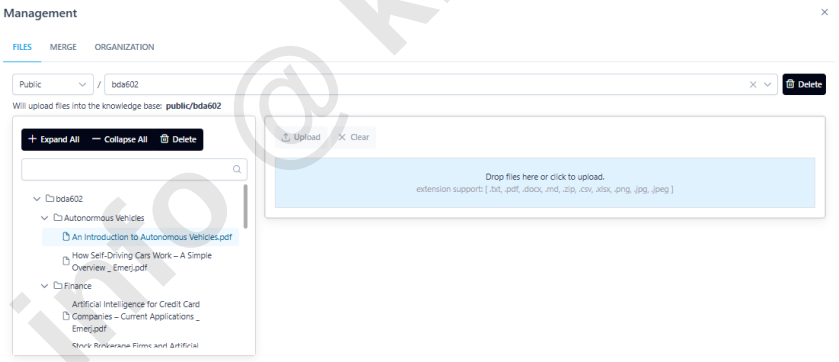
The administrator can create a knowledge base to handle all the custom inquiries, the inputs can be either a single file or multiple files stored in a directory. The administrator clicks the MANAGEMENT button to pop up the management menu. The administrator then enters the name <user>/<database> (e.g., public/bda602) in the Knowledge Base List box for knowledge base creation. There are two types of knowledge bases: public and personal. The public knowledge base is accessible to all users, while the personal knowledge base is shared only with invited users within the Company Organization. Next, the administrator clicks the Management button again and uploads the data[2] using the Drop files box and the Upload button. The administrator clicks the Confirm button on the Confirm Settings page, then the system automatically uploads the files to

[2] Multiple files are stored in a directory and compressed into a zip format before being uploaded to the system.

the system, For uploading multiple files, the administrator can click + Expand All to display the files within the subdirectory and use - Collapse All to hide the list.



The KNEO 330 supports multiple file formats, including .txt (text), .pdf (portable document format), .docx (Microsoft Word Document), .xlsx (Microsoft Excel Worksheet), .csv (Microsoft Excel Comma Separated Value File), .md (Markdown-formatted text), .jpg/.jpeg (Joint Photographic Experts Group), .png (Portable Network Graphic) and .zip (compressed files). File names should not contain special characters such as space, (), {}, or []. Depending on the file size, uploading may take a few minutes or longer. The administrator can create databases in both public and user directories, whereas general users can create databases only within their user directory.



To delete the database and files, the administrator selects the knowledge base and then clicks the button Delete to remove the database from the file system. The administrator can also highlight an individual file from the directory list to delete.

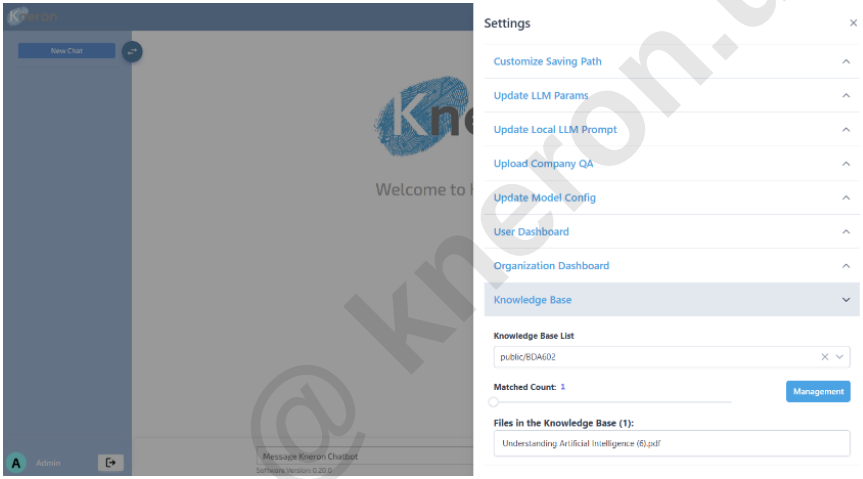
The user can share a custom knowledge base with others in a private group[3] during the database creation process. However, only the administrator can create a private group and invite the users to join the group. The administrator grants the modulator privilege to the user to invite another

[3] Please refer to KNEO 330 EdgeGPT Server Administrator Manual: Company Organization section for detailed

info@Kneron.us

user to join the group. While all users can access and share the knowledge base, only the creator can modify it. Once the database is created, the user clicks the ORGANIZATION button, selects the desired knowledge base from the Knowledge Base List, chooses the appropriate organization from the Selected Organizations menu, and presses the ADD button. This allows the knowledge base to be shared with private group users. Additionally, the user can remove the knowledge base from the group by using the REMOVE KNOWLEDGE BASE FROM ORGANIZATIONS option. A symbol (+ or x) is associated with the organization name. During removal, the user should ensure the symbol is set to +.

Once the custom knowledge base is created, the administrator clicks the cross symbol in the top right corner and returns to the knowledge base inquiry page. The knowledge base name appears in the Knowledge Base List box, and loaded files are displayed in the Files in the Knowledge Base box. **The Knowledge Base must be selected before inquiries.**



The administrator can select different databases from the Files in the Knowledge Base box for inquiry, enter a prompt into the Message Kneron Chatbot, and adjust the Matched Count: [n] sliding bar (where n is 1, 2, 3, 4, or 5) to access more matching results. Based on the matched count, the system displays the related document with its position in percentage and arranges them according to the match relevance. Additionally, the administrator clicks the source button, which shows the page source from the uploaded files.

3 System Administration

The administrator registers a user account (https://www.kneron.us) and follows the instructions to complete the registration process. Once the account is set up, the administrator sends a request to the email address (info@kneron.us), which includes the first name, last name, email address, machine model, and serial number. Kneron reviews the request and sends out the e-mail for the permission grant. Then, the administrator visits the online support (https://www.kneron.com/support/developers) and downloads the documents and software from the directory Kneron AI chat robot (KNEO 330). For any inquiries, please contact info@kneron.us

info@kneron.us

info @ kneron.us