KNEO 300 User Manual V1.6.2

(For Admin User)

July 2024

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
1.6.1	Add chat history and company QA	V0.16.1	Oscar Law	2024/06/03
1.6.2	Support automatic startup	V0.16.2	Oscar Law	2024/07/19

Notice:

1. Kneron 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.

Contents

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, which is 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 a variety of 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 also 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.

2. EdgeGPT Server

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

- Product parameters

CPU	8-core A53, 2.0GHz
NPU	30 TOPS (INT8)
DRAM	16GB LPDDR4
eMMC	64GB
Power	DC12V, AC100-240V, 50-60HZ
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.

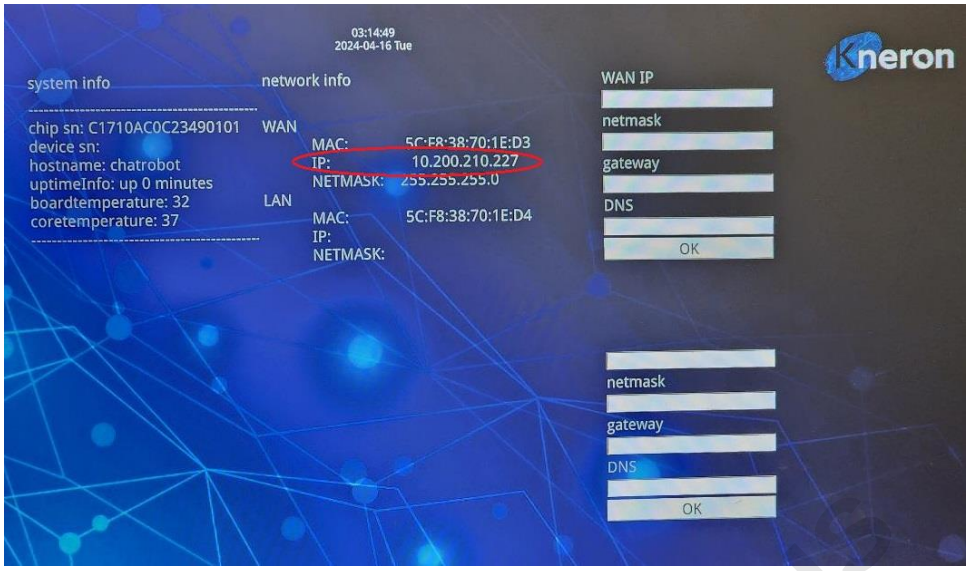

Figure 2-2 KNEO 300 IP Address

2.4 Network remote login

2.4.1 Remote Access

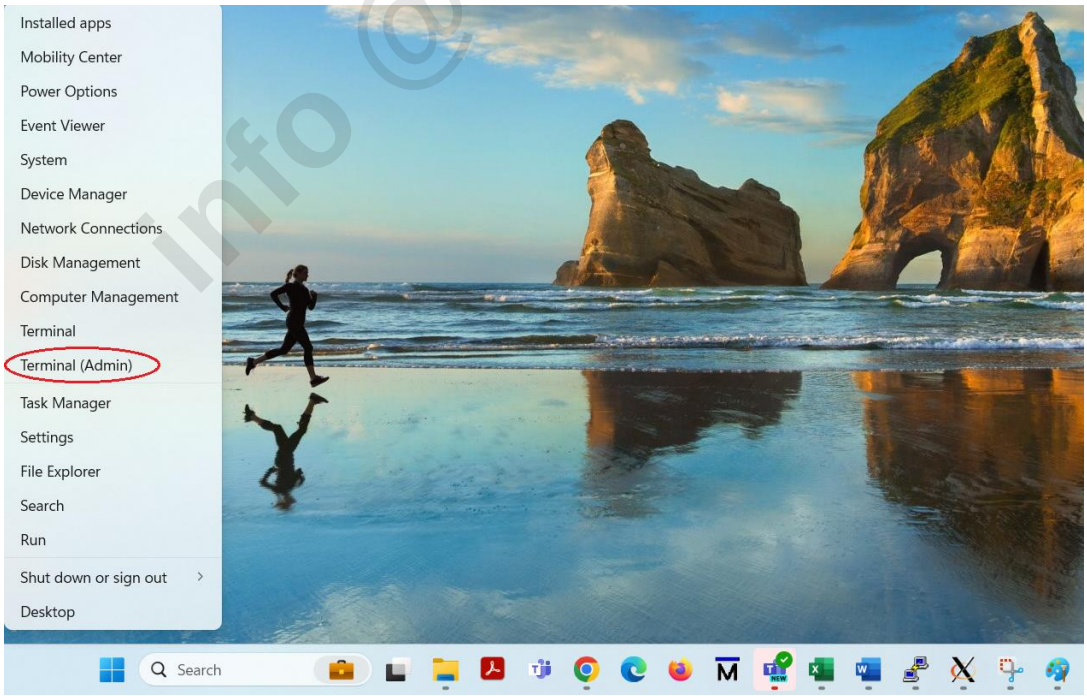

Figure 2-3 Windows PowerShell

Invoke the Windows PowerShell Terminal (Admin) to access the KNEO 300. First, click the lower left side Windows start icon with the right button, then select the Terminal (Admin) to create the terminal windows. The system administrator can use the command ping with the IP address (i.e. 10.200.210.227) to check machine accessibility. After that, it hits the CTRL-C to terminate the ping process and start to initialize the server using either ssh or putty approaches:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\oscar> ping 10.200.210.227

Pinging 10.200.210.227 with 32 bytes of data:
Reply from 10.200.210.227: bytes=32 time=14ms TTL=62
```

2.4.2 Automatic Startup

Kneron offers the automatic startup script during the v0.16.2 upgrade. If the administrator initializes the system_patch_autostart procedure during the installation, the administrator can ignore 2.4.3 ssh and 2.4.4 putty for the KNEO 300 activation, and then jump to 2.5 WEBUI Interface to start the inquiry.

2.4.3 ssh

To log in to the KNEO 300 with the command ssh. The username is linaro and the password is linaro

```
C:\Users\oscar > ssh linaro@10.200.210.227
linaro@10.200.210.227's password:
```

After the login, it displays the message as follows:

```
Welcome to Ubuntu 20.04 LTS (GNU/Linux 5.4.217-bm1684-g4758df7c6cfd-dirty aarch64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
   just raised the bar for easy, resilient and secure K8s cluster deployment.

   https://ubuntu.com/engage/secure-kubernetes-at-the-edge        overlay        /        overlay
```

```
rw,relatime,lowerdir = /media/root-ro,upperdir=/media/root-rw/overlay, workdir=/media/root-rw/overlay-workdir 0 0
/dev/mmcblk0p5 /media/root-rw ext4 rw,relatime 0 0
/dev/mmcblk0p4 /media/root-ro ext4 ro,relatime 0 0

Last login: Tue Apr 16 01:45:17 2024 from 10.200.211.128
linaro@chatrobot:~$
```

System background job is a critical step that must initialize KNEO 300, it allows the process to run on the server even if the system administrator logs out from the server. The administrator must invoke the utility screen to start the system background job in KNEO 300.

```
linaro@chatrobot:~$ screen
GNU Screen version 4.08.00 (GNU) 05-Feb-20

Copyright (c) 2018-2020 Alexander Naumov, Amadeusz Slawinski
Copyright (c) 2015-2017 Juergen Weigert, Alexander Naumov, Amadeusz Slawinski
Copyright (c) 2010-2014 Juergen Weigert, Sadrul Habib Chowdhury
Copyright (c) 2008-2009 Juergen Weigert, Michael Schroeder, Micah Cowan, Sadrul Habib Chowdhury
Copyright (c) 1993-2007 Juergen Weigert, Michael Schroeder
Copyright (c) 1987 Oliver Laumann

This program is free software; you can redistribute it and/or modify it under the terms of the
GNU General Public License as published by the Free Software Foundation; either version 3, or
(at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY;
without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See
the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program (see
the file COPYING); if not, see https://www.gnu.org/licenses/, or contact Free Software
Foundation, Inc., 51 Franklin Street, Fifth Floor,
Boston, MA 02111-1301 USA.

Send bugreports, fixes, enhancements, t-shirts, money, beer & pizza to screen-devel@gnu.org

Capabilities:
+copy +remote-detach +power-detach +multi-attach +multi-user +font +color-256 +utf8 +rxvt
+builtin-telnet

                                [Press Space or Return to end.]
```

Press the Space bar or hit the Return key to return to the terminal mode.

To activate the KNEO 300 software, first change to the subdirectory: Kneron_chatbot_prod, then apply the command chmod to change the shell scrip new_launch.sh to execution file, then invoke the software using the command: ./new_launch.sh

```
linaro@chatrobot:~$ cd kneron_chatbot_prod
linaro@chatrobot:~/kneron_chatbot_prod$ chmod +x new_launch.sh
linaro@chatrobot:~/kneron_chatbot_prod$ ./new_launch.sh
```

It displays the execution messages on the screen and takes 30 seconds to complete the setup process. Finally, it shows the message "webpack compiled successfully", which means that the software is initialized successfully.

After the server is initialized, just hold the CTRL-A + D at the same time to leave the screen mode. The software still runs on KNEO 300 even if the administrator turns off the host computer.

```
linaro@chatrobot:~$ screen
[detached from 19219.pts-0.chatrobot]
linaro@chatrobot:~$ [detached from 19219.pts-0.chatrobot]
linaro@chatrobot:~$ screen -ls
There is a screen on:
        19219.pts-0.chatrobot   (04/16/24 03:32:22)     (Detached)
1 Socket in /run/screen/S-linaro.
```

The administrator/user can access KNEO 300 using the browser interface in Section 2.4.4

2.4.4 putty

It downloads the putty from the website (i.e. https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html) and chooses the 64-bit x86 package from the Windows Installer. After downloading the software, just double-click the binary execution file and follow the instructions to complete the installation process. Then start the putty shown in Figure 2-4, and enter the IP address (i.e. 10.200.210.27) and port number 22.

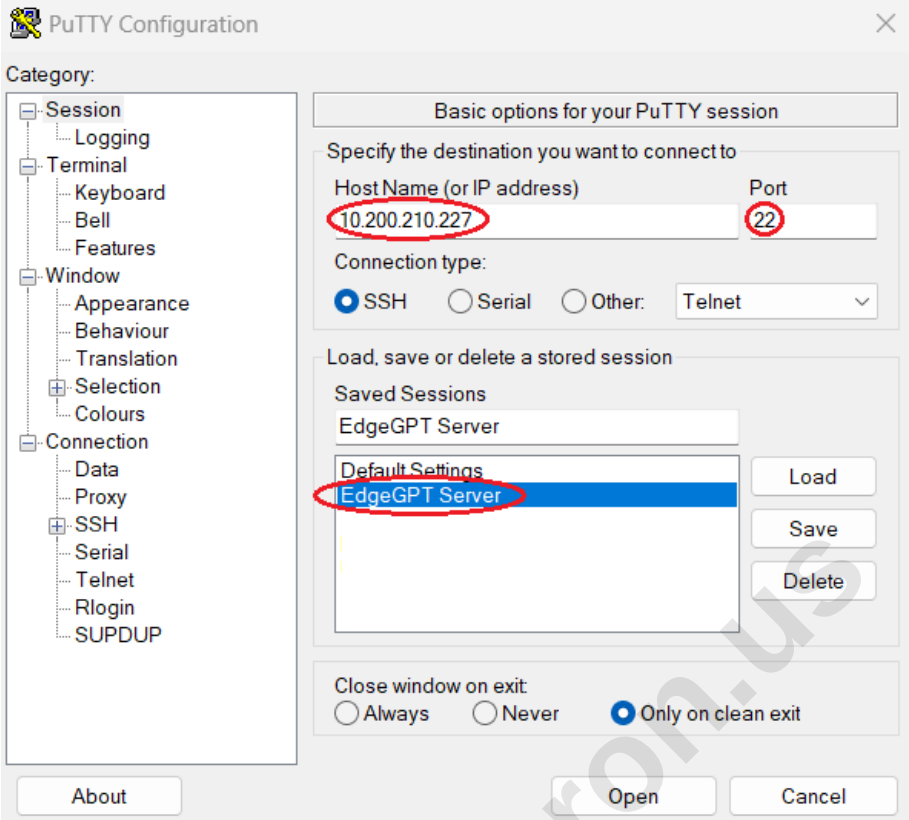

Figure 2-4 Putty home screen

The administrator and users can save the IP address and port number in Saved Sessions (i.e. EdgeGPT Server) and reload the session using the Load command to initialize KNEO 300 the next time. Currently, it hits the Open button to start the putty session with the username: linaro and password: linaro

```
> login as: linaro
> linaro@10.200.210.227's password:
```


2.5 WEBUI Interface

2.5.1 Session Initialization

The administrator 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)

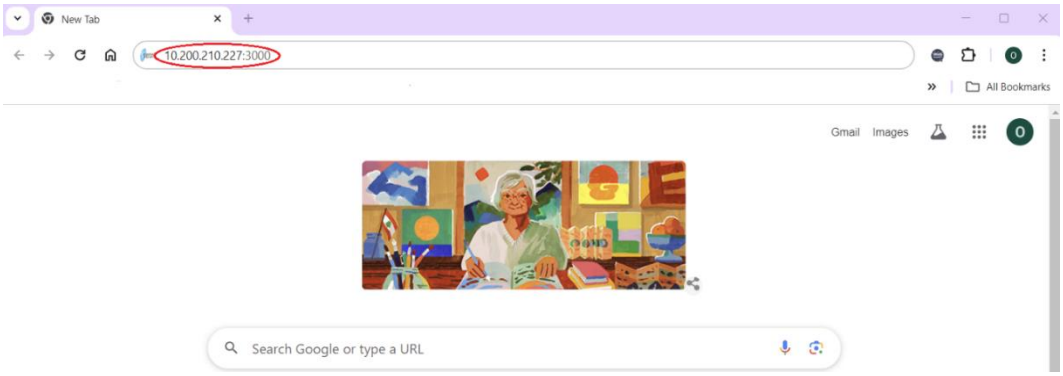

Figure 2-5 Browser Access

It invokes the WEBUI interface shown in Figure 2-6, the default interface is English, and the administrator can click the icon and switch to Traditional Chinese.

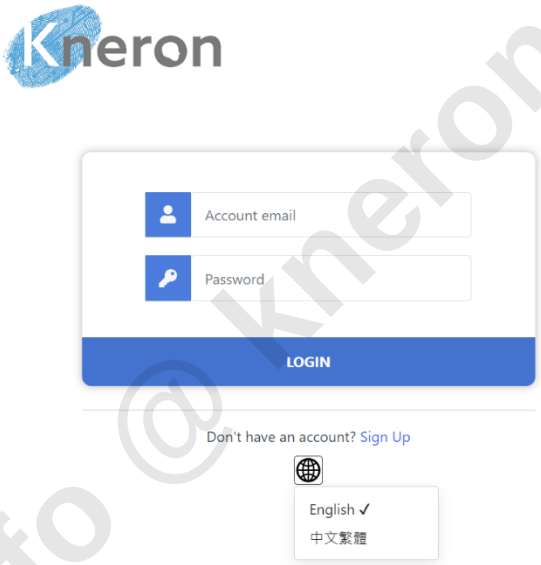

Figure 2-6 WEBUI Login Page (English)

The administrator can log in to the system using the default username: admin@kneronchatbot.com and Password: admin123 to start the WEBUI session.

Figure 2-7 WEBUI Login Page (Chinese)

The default WEBUI session is English, the administrator can click the icon to switch between English and Traditional Chinese. The username is shown in the bottom left-hand corner and the user clicks the button to log out from the system.

Figure 2-8 WEBUI Session (English)

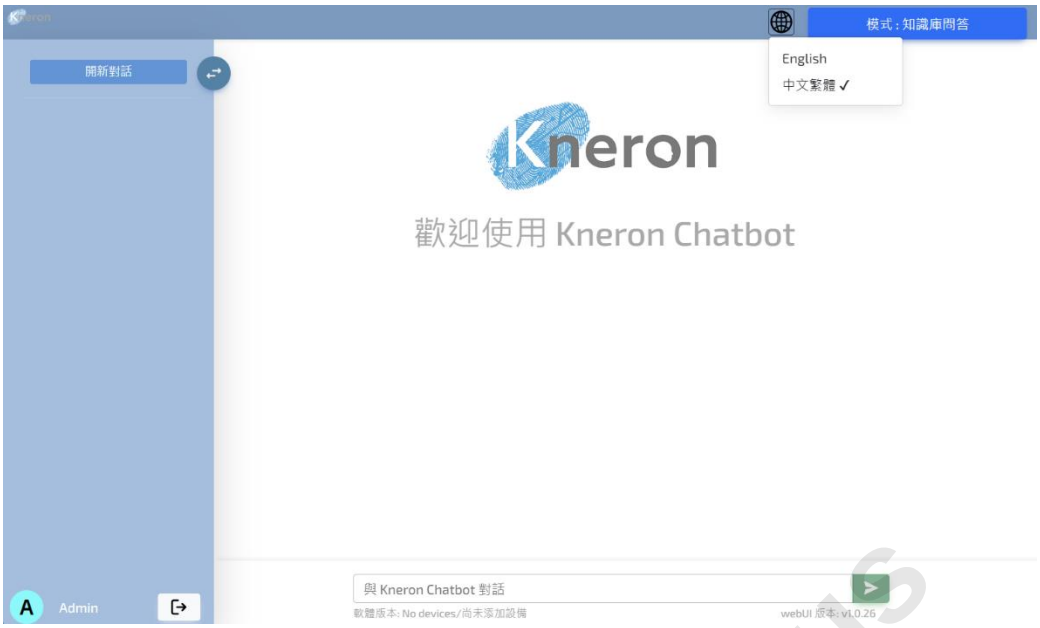

Figure 2-9 WEBUI Session (Chinese)

The administrator clicks the top right-hand corner menu button to invoke the Setting Menu, then toggles the language between Chinese and English. The KNEO 300 can support multiple servers but strongly recommend only running the session on the current server. The administrator first sets the Language box to English, then enters the device IP address (i.e. 10.200.210.227) in the Host box and presses the ADD DEVICE button. It takes a few minutes for KNEO 300 to initialize the system. After the initialization is completed, it displays the Device added in the Device List dialogue box. The administrator can click the chat history using the top left-hand corner history button to hide/unhide the chat history. ***If the automatic startup process is initialized, the administrator adds the host once. The host is stored and shown in the device list for future access without initialization again.***

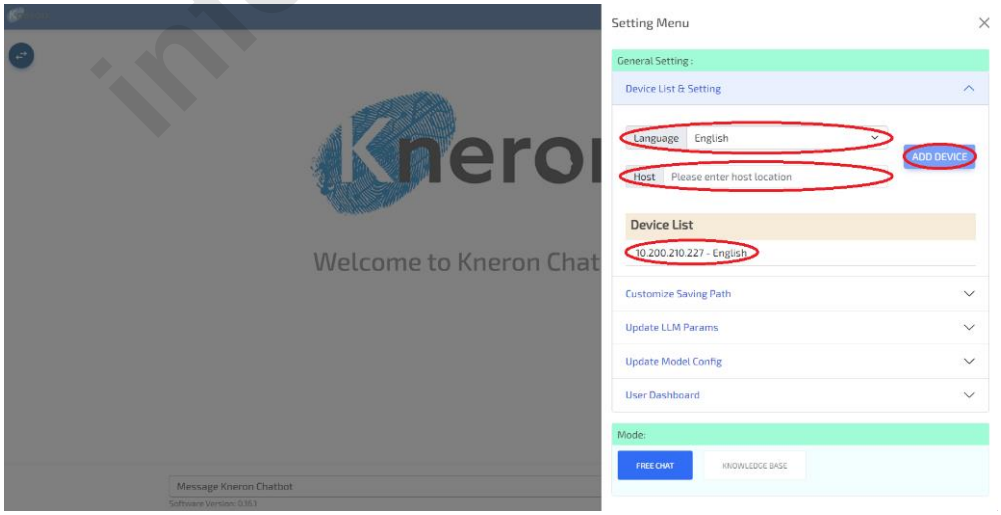

Figure 2-10 WEBUI Setting Menu

2.5.2 Public/Custom Database

All the databases are stored in the directory: /home/linaro/kneron_chatbot_prod/kneron_doc_chat/knowledge_base/content, which is further divided into the EN (English) and ZN (Chinese) subdirectories. They store the different language databases dependent on the system setting. All the databases are stored under the EN subdirectory if the language is set to English. ***The administrator is responsible for backing up all the databases during the system update***.

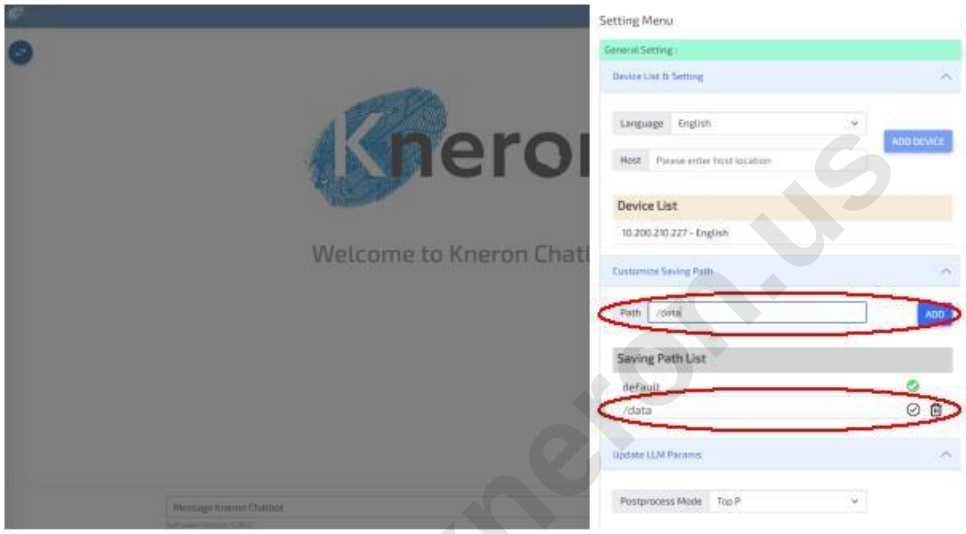

Figure 2-11 WEBUI Data Path

The administrator can set the custom database path using the Path box, followed by ADD button. It adds the custom data path to the Saving Path List. The administrator can select different data path using the check mark symbol ⊘ and remove the path using the garbage 🗑 symbol. it is useful to save the database to the other directory or mounted devices.

If the administrator wants to store the custom database in external storage (i.e. USB drive and NAS storage), please refer to Chapter 5 External Storage. It shows how to mount the external storage to the KNEO 300. After the external storage is successfully mounted, the administrator can enter the mounted directory (i.e. /mnt/U2 or /mnt/U3)[1] in the Path box with ADD button, and the custom path is shown up in the Saving Path List. The administrator clicks the symbol ⊘ to select the custom path.

[1] /mnt/U2 (USB drive) and /mnt/U3 (NAS storage) are the external storage mounted to the KNEO 300 in Chapter 5

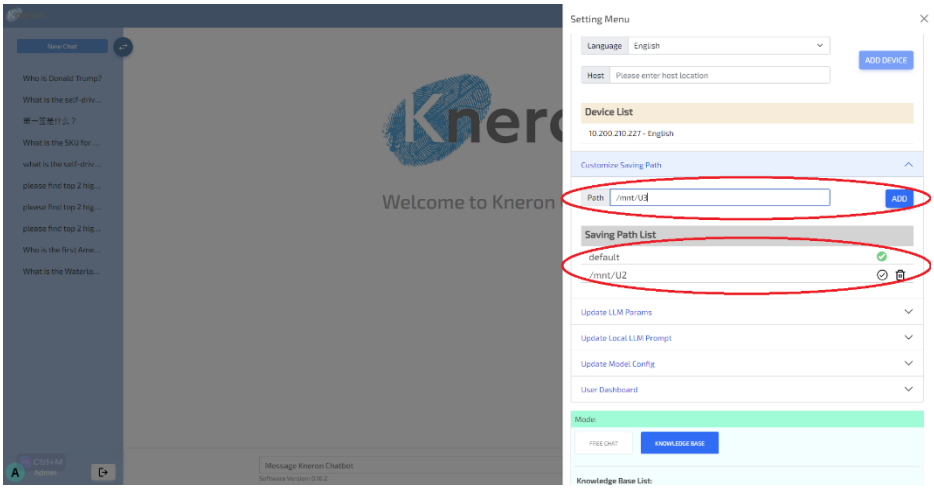

Figure 2-12 WEBUI Data Path List

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 bottom right-hand buttons.

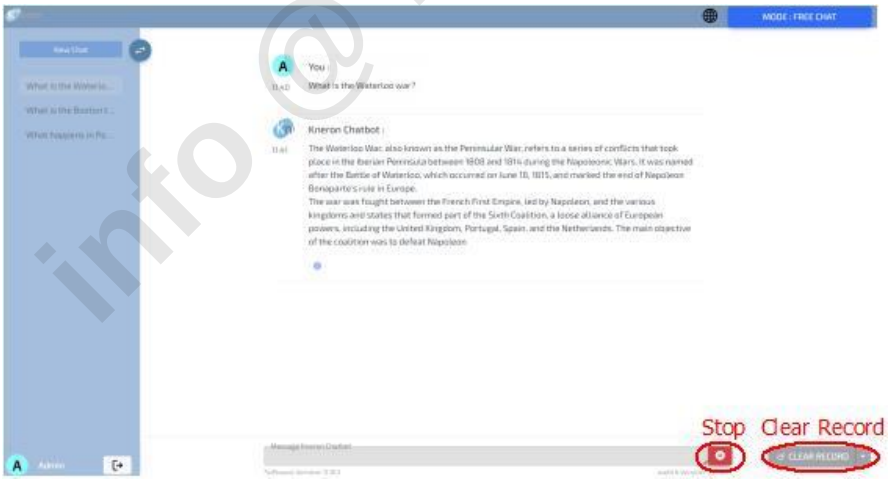

Figure 2-13 Free Chat Mode

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 chat history is displayed under the New Chat on the left-hand side. The user stops the chat using the red stop button or cleans the inquiry using the CLEAR RECORD button.

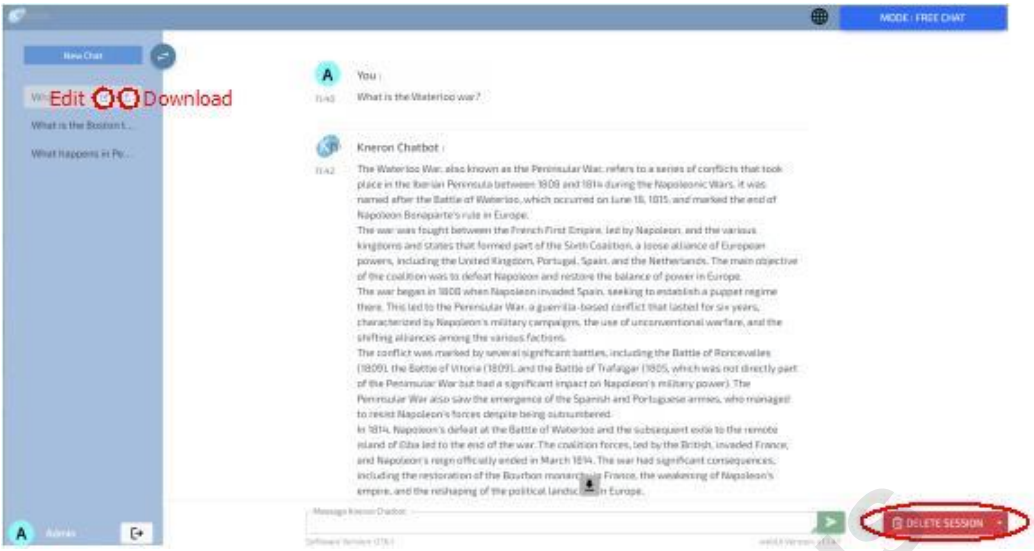

Figure 2-14 Chat Session

The user first clicks the chat session to re-invoke the dialogue, which allows the user to edit or download the chat. The user can modify the chat inquiry using the edit button and download the chat session using the download button. The KENO 300 automatically compresses the chat session in JSON format and saves it to the Download directory. Moreover, the user can select the chat session, and then click the DELETE SESSION button to delete the chat session. It displays the warning menu and asks the user to confirm removing the chat session from the history.

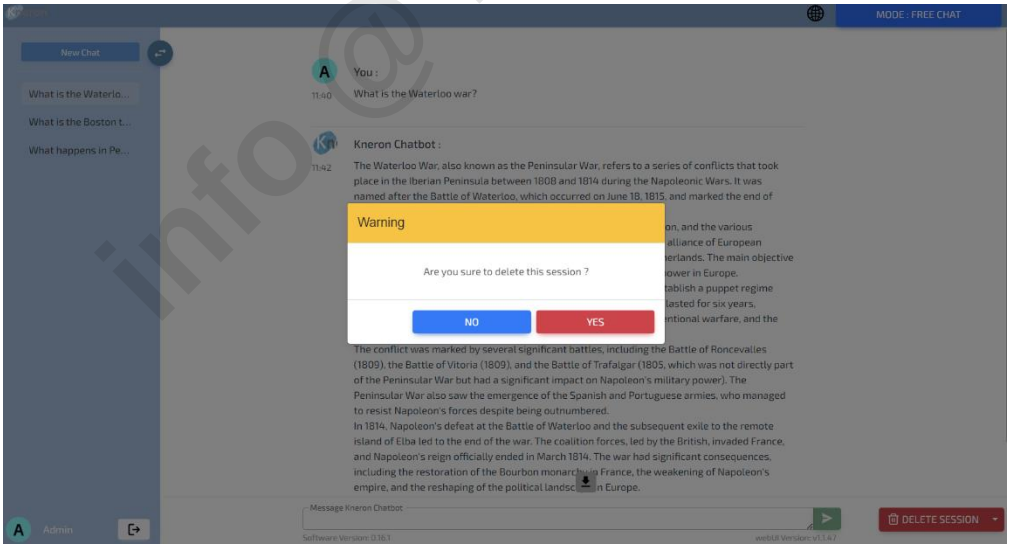

Figure 2-15 Chat Session Delete

The latest software provides additional control for inquiry results using Update LLM Params shown in Figure 2-16. 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 2. It recommends that users keep the default setting for general usage.

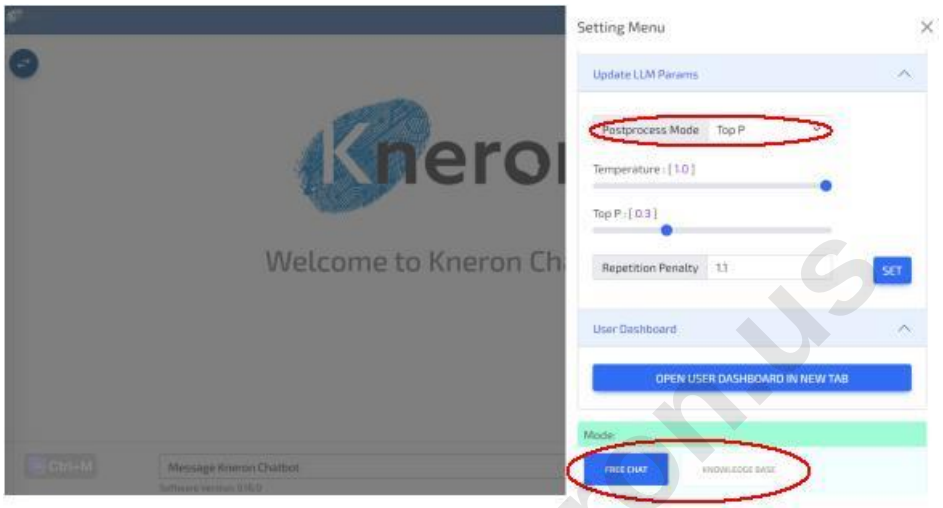

Figure 2-16 Chat Mode Setting

For the latest release, the KNEO 300 supports advanced Prompt Engineering using Update Local LLM Prompt, which offers the System Prompt and Question Prompt. The System Prompt refers to the role of the EdgeGPT server, which acts as the virtual assistant to answer the inquiry. The Question Prompt instructs the EdgeGPT server how to answer the questions using the relevant information without any false content. ***Since the inquiry accuracy is highly affected by the Prompt setting, please consult the Kneron FAE before any changes, otherwise, keep the default blank.***

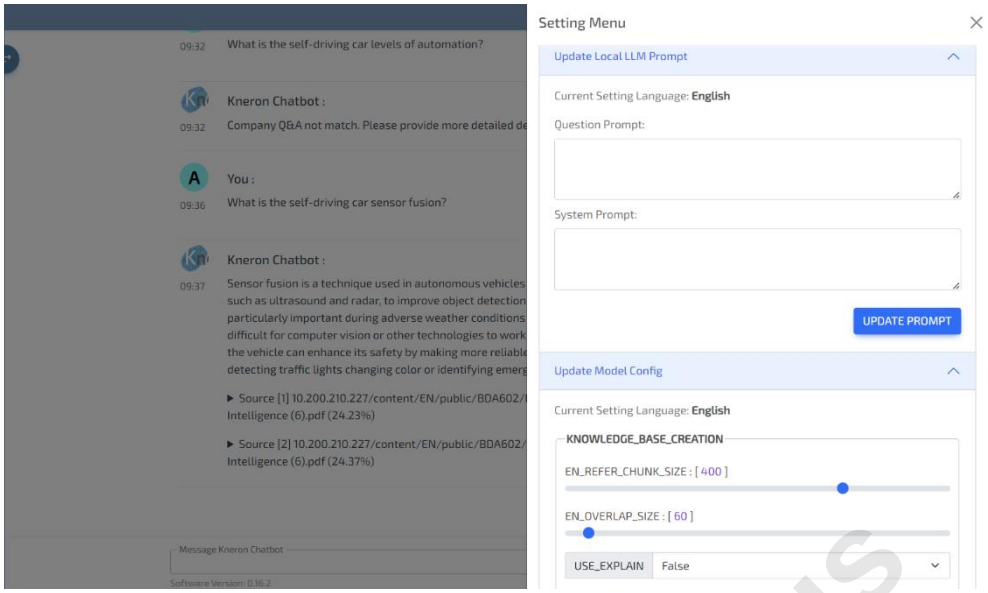

Figure 2-17 Update Local LLM Prompt

2.5.4 Knowledge Base Mode

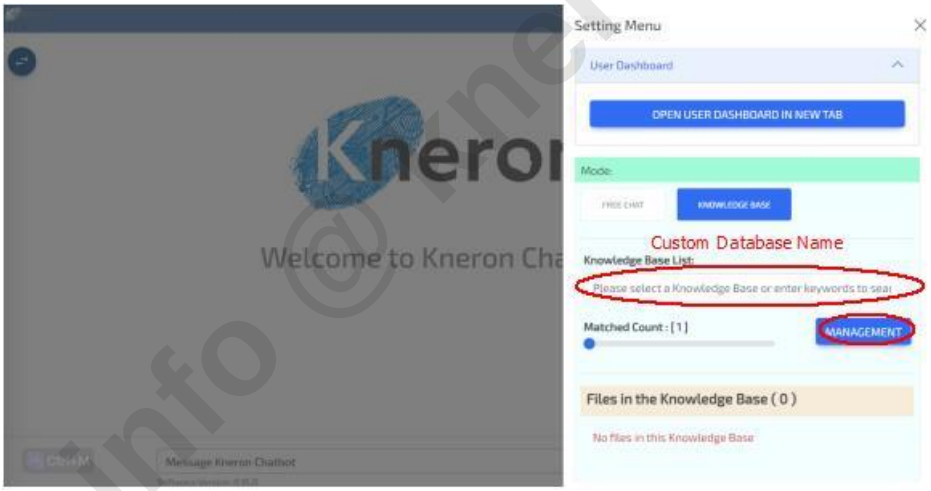

Figure 2-18 Knowledge Base Management

For knowledge base inquiry, the administrator first creates the custom database using the MANAGEMENT button. After pressing the button, the management pop-up menu shows up. The administrator enters the database name <user>/<database> in the Knowledge Base List box (i.e. public/bda602), then clicks the MANAGEMENT button and uploads the files in the Drop files box with the UPLOAD 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 administrator can create the database for the public or user directories. However, the general user can create the database under its user directory only.

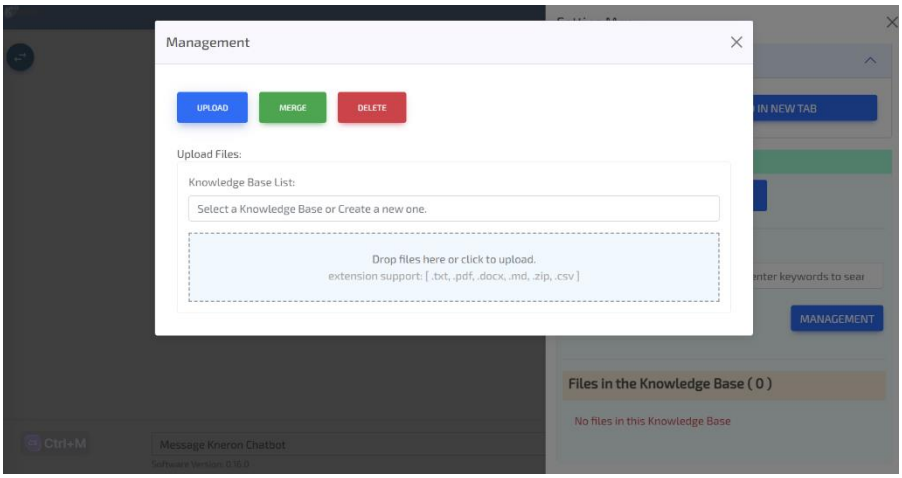

Figure 2-19 Custom Base Creation

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 administrator 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 2-20 Custom Database Inquiry

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.

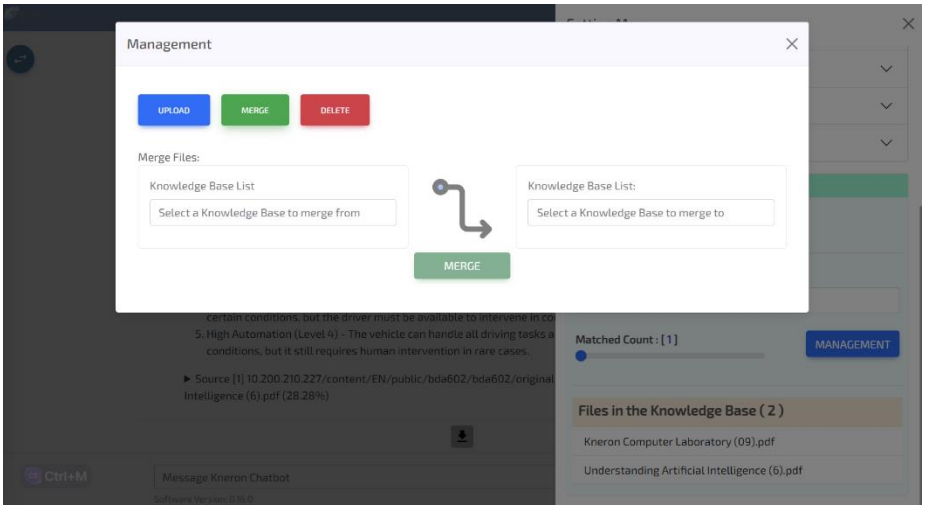

Figure 2-21 Merge Custom Database

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.

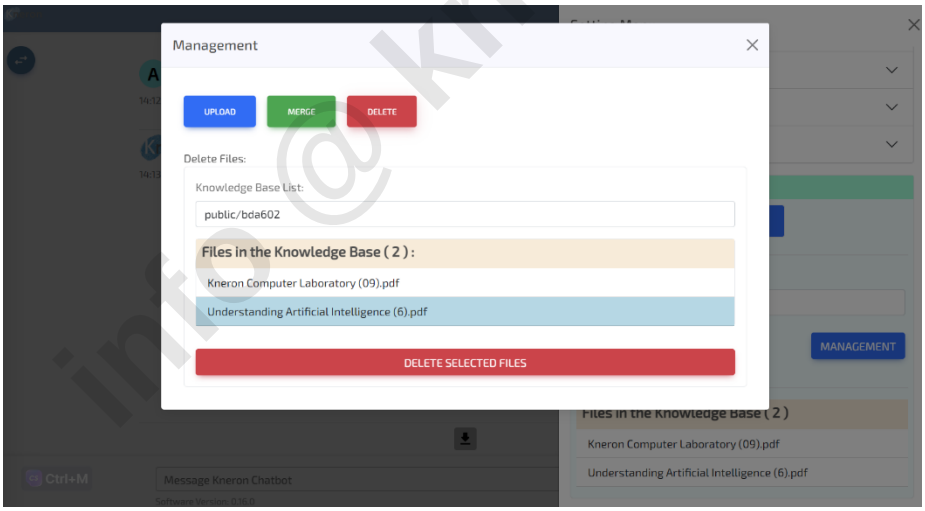

Figure 2-22 Delete Single/Multiple Files or Custom Database

The administrator can configure the model in the knowledge base mode in the submenu Update Model Config, which is further divided into the KNOWLEDGE_BASE_CREATION[2] and KNOWLEDGE_BASE_QA[3]

[2] USE_EXPLAIN is not supported in KNEO 300 yet

[3] REPHASE_QUERY is not supported in KNEO 300 yet

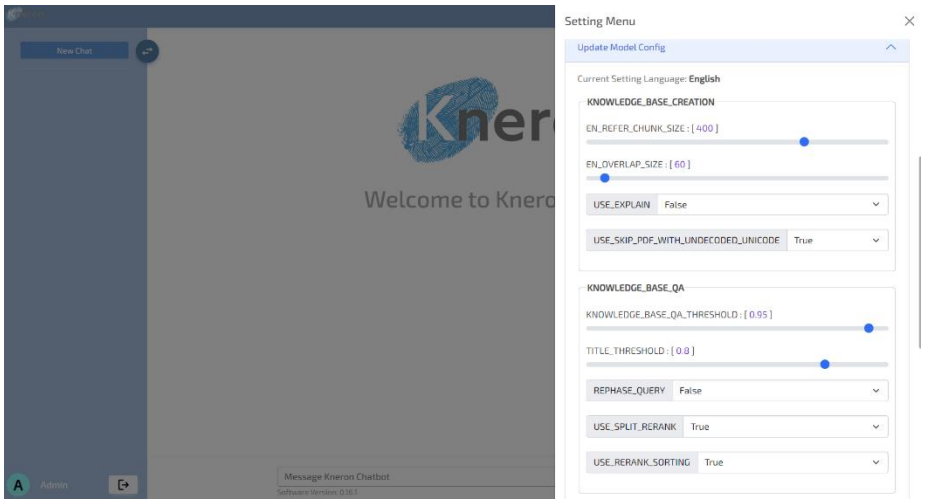

Figure 2-23 Knowledge Base Creation/Inquiry

The administrator can optimize the performance to adjust the model processing chunk size (number of characters) using EN_REFER_CHUNK_SIZE (0 to 512). The accuracy is improved when the chunk size is reduced with the drawback of longer processing time. The processing chunk can be linked together through the chunk overlapping using EN_OVERLAP_SIZE, which MUST BE less than 50% of the chunk size. USE_SKIP_PDF_WITH_UNDECODED_UNICODE is set to TURE, which ignores the pdf document with an undefined character (i.e. undecoded Unicode), and the undefined character crashes the system.

The administrator adjusts the QA threshold to improve the matching between the inquiry and source using KNOWLEDGE_BASE_QA_THRESHOLD. It is easy to match the inquiry to the knowledge base source using the higher threshold, but the drawback is less accuracy. Moreover, the administrator can adjust the TITLE_THRESHOLD to improve the matching between the inquiry and source using the knowledge base uploaded file name.

The administrator can further set USE_SPLIT_RERANK to TRUE, which further splits and ranks the source to improve the accuracy within the same document. Alternatively, the administrator also sets USE_RERANK_SORTING to TRUE, then sorts and ranks the different sources to retrieve the data with the longer processing time. The major difference between USE_SPLIT_RERANK and USE_RERANK_SORTING, the first one targets a single source (same document), and the latest one refers to different sources (various documents).

If the options, USE_SPLIT_RERANK and USE_RERANK_SORTING are turned on, it slows down the operations. The administrator can set the options to TRUE or FALSE with the trade-off between the result accuracy and inquiry speed.

2.5.5 Company Q&A

	A	B
1	Questions	Answers
2	Who is the first American president?	George Washington
3	Who is the longest British ruler?	Queen Elizabeth II
4	When is the Boston tea party?	December 16, 1773.
5	Which countries are the axis in the World War 2?	Germany, Italy and Japan
6	Who first landed on the moon?	Neil Armstrong
7	Why the first industrial revolution was so important?	The First Industrial Revolution (1760-1840) revolutionized manufacturing, transportation, and communication, laying the groundwork for modern industrial society.
8	What is the nobel prize ?	The Nobel Prize is a prestigious international award recognizing outstanding contributions in Physics, Chemistry, Medicine, Literature, Peace, and Economic
9	Who is the first black American president?	Barack Obama
10	What were the four important inventions in the ancient China?	Paper, Printing, Gun Powder, and the Compass
11	What was the most famous Pyramid?	The Great Pyramid of Giza, also known as the Pyramid of Khufu

Figure 2-24 Company Q&A Spreadsheet

Currently, KNEO 300 offers the conventional one-to-one chatbot feature to set the questions with predefined answers using the spreadsheet in the Knowledge Base Mode[4]. The system supports a single spreadsheet only, which can be updated and reloaded to overwrite the previous one. The sample chatbot.csv is defined in Figure 2-24. The first row shows the comment: col A – Questions and col B – Answers. The questions and answers start in the second row with the questions and their corresponding answers.

- The comments are defined in the first row (i.e. row 1), which is ignored during the processing
- The questions are defined in the first column (i.e. col A)
- The answers are defined in the second column (i.e. col B)

Then, the administrator applies scp[5] command to transfer the spreadsheet from Windows to the KNEO 300 directory: /home/linaro/kneron_chatbot_prod/kneron_doc_chat

```
C:\Users\oscar\>scp  chatbot.csv  linaro@10.200.210.227:/home/linaro/kneron_chatbot_prod/kneron_doc_chat
linaro@10.200.210.227's password:
chatbot.csv                                                   100%  976   190.8KB/s   00:00
```

The administrator logs in to the KNEO 300 using ssh[6], then changes to the directory: /home/linaro/

[4] In the Free Chat mode, the system provides the general answer rather than the pre-defined one for the same inquiry.

[5] scp usage is defined in Section 3.2.1

[6] ssh usage is defined in Section 2.4.3

kneron_chatbot_prod/kneron_doc_chat. It initializes the environment using the Shell command: source ../envsetting.sh, then invokes the Python command: python3 company_qa_db_creator.py <spreadsheet> <language> to integrate the spreadsheet into the system, where <language> is defined as either English or Chinese.

```
C:\Users\oscar> ssh linaro@10.200.210.227
linaro@10.200.210.227's password:
Welcome to Ubuntu 20.04 LTS (GNU/Linux 5.4.217-bm1684-g4758df7c6cfd-dirty aarch64)

 * Documentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

 * Strictly confined Kubernetes makes edge and IoT secure. Learn how MicroK8s
   just raised the bar for easy, resilient and secure K8s cluster deployment.

   https://ubuntu.com/engage/secure-kubernetes-at-the-edge
overlay/overlay                         rw,relatime,lowerdir=/media/root-ro,upperdir=/media/root-
rw/overlay,workdir=/media/root-rw/overlay-workdir 0 0
/dev/mmcblk0p5 /media/root-rw ext4 rw,relatime 0 0
/dev/mmcblk0p4 /media/root-ro ext4 ro,relatime 0 0

Last login: Thu Jun  6 02:11:39 2024 from 10.200.211.96
linaro@chatrobot:~$ cd kneron_chatbot_prod/kneron_doc_chat/
linaro@chatrobot:~/kneron_chatbot_prod/kneron_doc_chat$ source ../envsetting.sh
linaro@chatrobot:~/kneron_chatbot_prod/kneron_doc_chat$  python3  company_qa_db_creator.py  chatbo
t.csv English
English
get 1 items.
get 2 items.
get 3 items.
get 4 items.
get 5 items.
get 6 items.
get 7 items.
get 8 items.
get 9 items.
get 10 items.
```


To initialize the Company QA, the administrator first sets the USE_COMPANY_QA to TRUE in the Setting Menu, then adjusts the COMPANY_QA_THRESHOLD to the higher threshold to match with the answers in the spreadsheet. If the ONLY_COMPANY_QA is set to TRUE, the system only searches for the answers

from the spreadsheet.

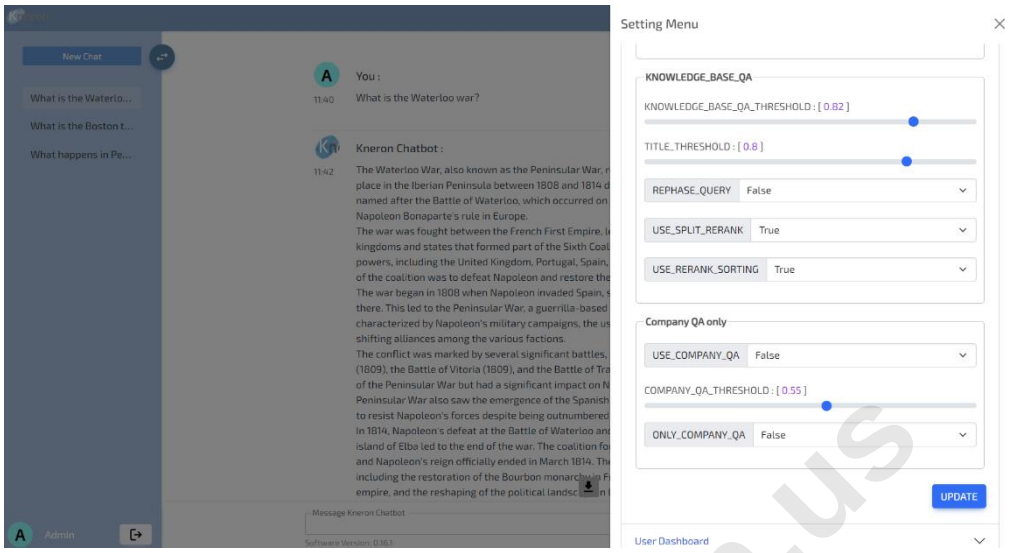

Figure 2-25 Company Q&A Setting

After the administrator clicks the UPDATE button, the system updates all the settings for the Company QA questions and answers.

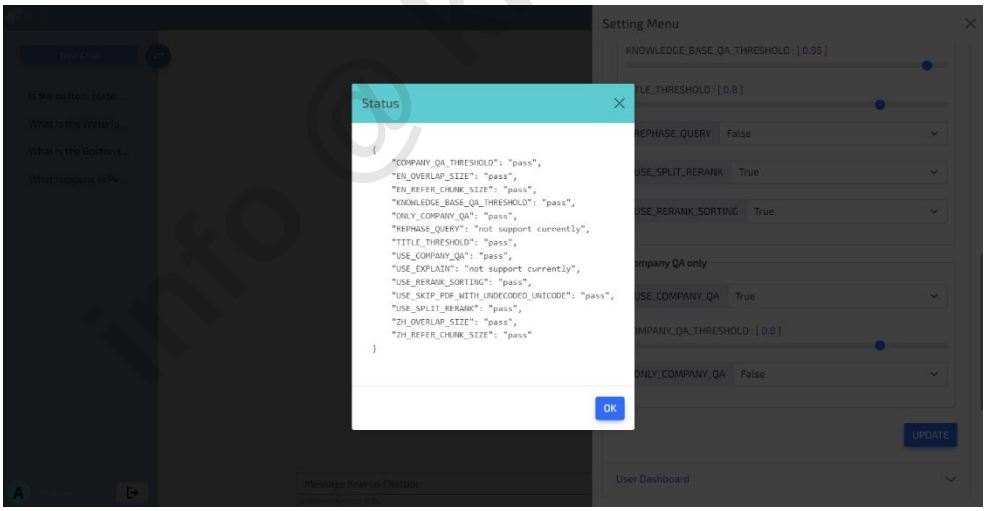

Figure 2-26 Company Q&A Update

The administrator enters the questions in the Dialogue box and the system returns the predefined answer. If the ONLY_COMPANY_QA is set to TRUE, the system can't find the answers in the Company Q&A, it shows the message "Company Q&A does not match. Please provide a more detailed description.", which means that the question is not defined in the Company Q&A spreadsheet yet.

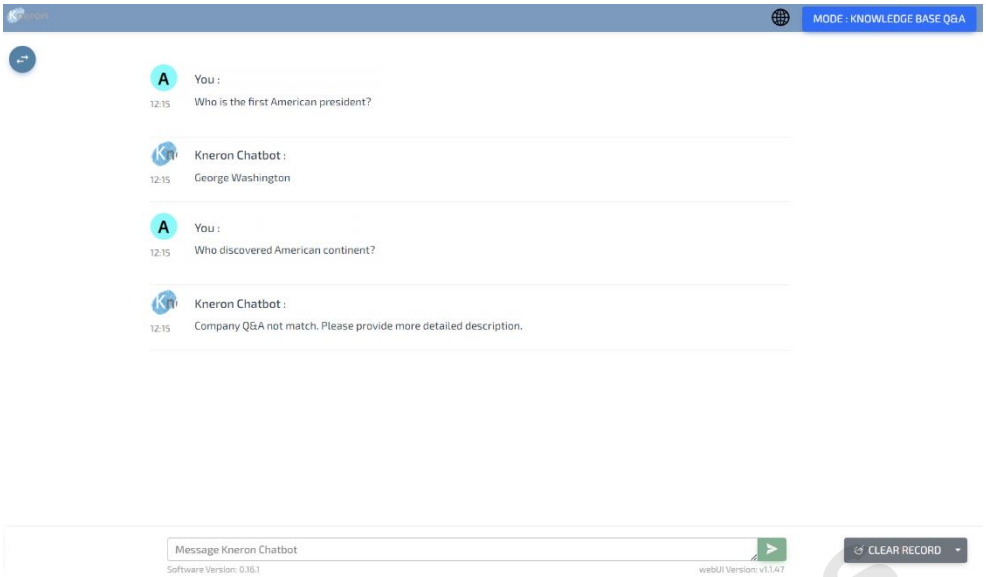

Figure 2-27 Company Q&A Inquiry

If the ONLY_COMPANY_QA is turned off and set to False, the system first searches the Company Q&A spreadsheet for the answer followed by the custom knowledge base. ***The ONLY_COMPANY_QA is recommended to be turned off for the general inquiry.***

2.6 Power Off

To power off the system, it recommends all the users log out of the system first, then the administrator initializes the command to shut down the system and allows the data to be properly stored in the database. **Please don't disconnect the power before the software shuts down, it may damage the file system.**

```
C:\Users\oscar> sudo poweroff
```

3. System Administration

3.1 User Registration

3.1.1 Account Setup

Figure 3-1 New User Sign Up

For new users, they can 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 address, and password. Then, the new user can click the SIGN UP button to register the new account. It requires 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.

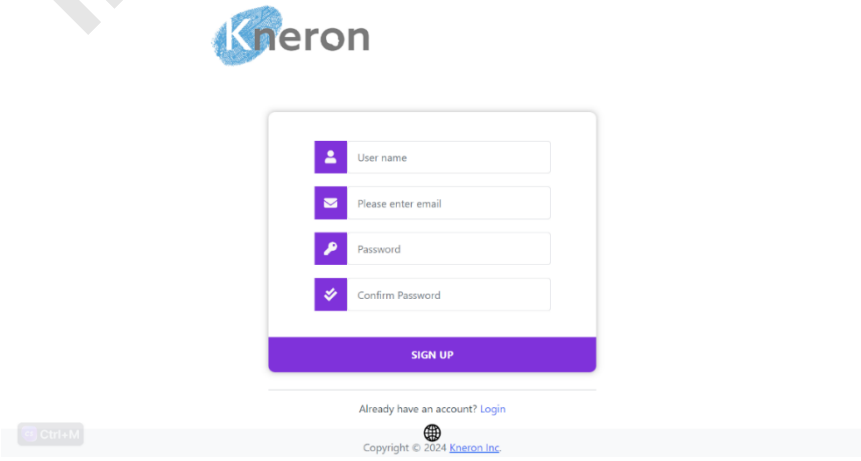

Figure 3-2 New User Registration

3.1.2 User Password Change

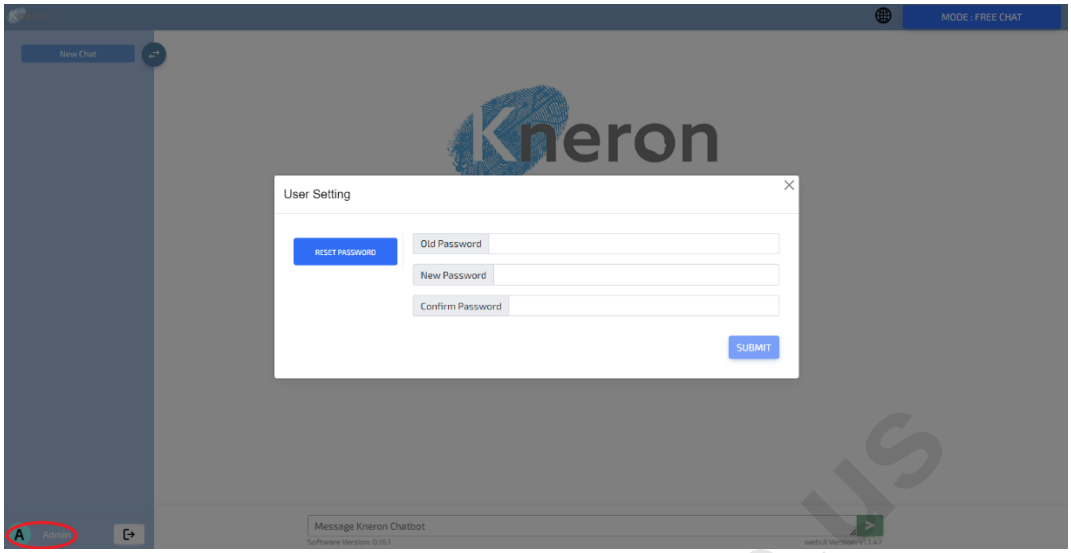

Figure 3-3 User Password Change

The user clicks the bottom left-hand corner username to reset the password, which pops up in the User Setting menu. The user enters the old password followed by the new one, then presses the SUBMIT button to reset the password.

3.1.3 Access Permission

Figure 3-4 User Dashboard

The administrator has the right to set the access permissions, it first clicks the USER DASHBOARD IN NEW TAB in the Setting menu, and it opens the user dashboard in another tab. The dashboard consists of three menus, Home, User, and Role. The administrator can change the user access permissions in the User menu.

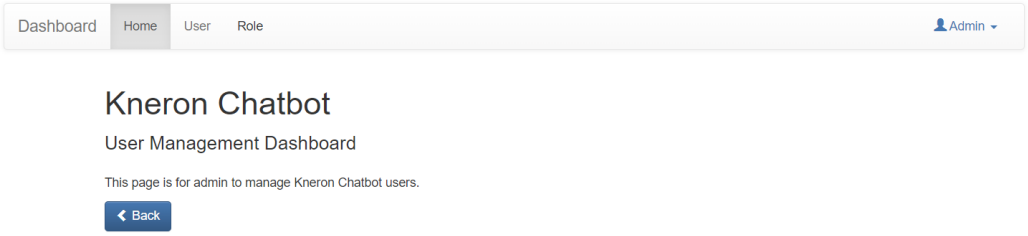

Figure 3-5 Dashboard Home Menu

After the administrator changes to the user dashboard, which shows all the users registered in the KNEO 300.

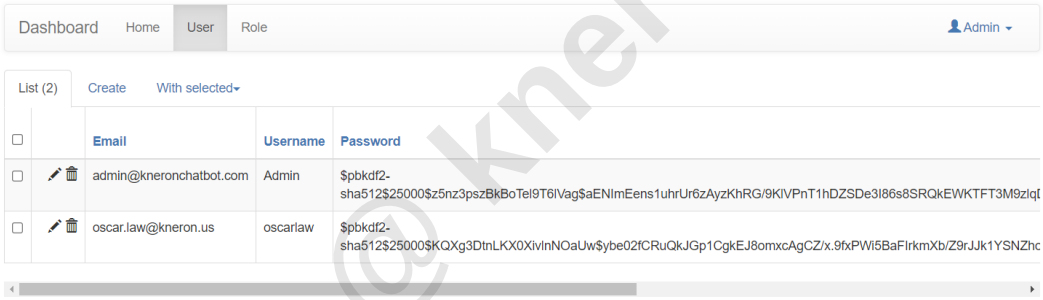

Figure 3-6 Dashboard User Menu

The default role of all users is set to regular. The administrator can click the pen icon to edit the user role in the Roles box and include the additional role admin access permission, then press the Save button to save the changes. The administrator can view the role in the Role menu.

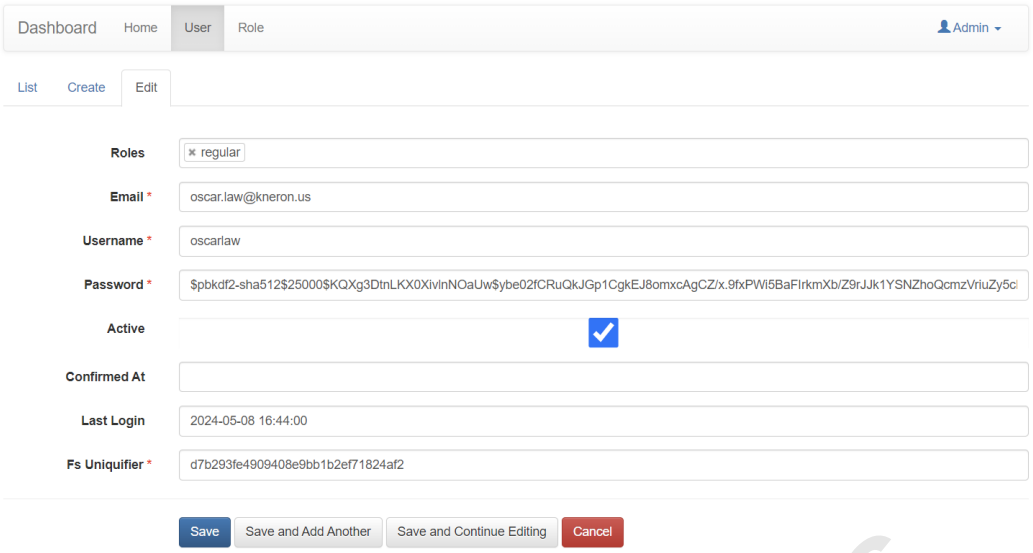

Figure 3-7 User Role Modification

3.1.4 User Password Reset

Due to the security, the administrator has no right to reset the user password. If the user forgets the password, the user must contact the administrator to delete the user account from the Dashboard User Manual in Figure 3-6, then register the same username in the Login Page (Figure 3-1). The new username MUST BE SAME as the old one, otherwise, the user can't access the previous custom database.

3.2 File Management

The administrator can transfer the files between Windows and KNEO 300 using the SCP protocol for file management. It is useful to back up the custom databases and install the new file release. The username and password are set to linaro with port number 22 for the KNEO 300.

3.2.1 scp

To follow the same instructions in 2.4.1 to invoke the Windows PowerShell Terminal (Admin) first, then apply the built-in command scp to transfer the files between Windows and KNEO 300. The administrator can copy the README.txt file from Windows to KNEO 300 /home/linaro directory with the command: scp <file> <user name>@<ip address>:<user directory>

```
C:\Users\oscar>scp README.txt linaro@10.200.210.227:/home/linaro
linaro@10.200.210.227's password:
README.txt                                    100%    8     0.5KB/s   00:00
```

The administrator can copy the README.txt from KNEO 300 to Windows using the similar command: scp <user name>@<ip address>:<user directory>/<file name> <.> where "." means the current directory

```
C:\Users\oscar>scp linaro@10.200.210.227:/home/linaro/README.txt .
linaro@10.200.210.227's password:
README.txt                                    100%    8     0.3KB/s   00:00
```

3.2.2 winscp

Currently, it provides a file transfer alternative graphical interface: winscp, the administrator downloads the software (https://winscp.net/eng/download.php) , and follows the instructions to install the software.

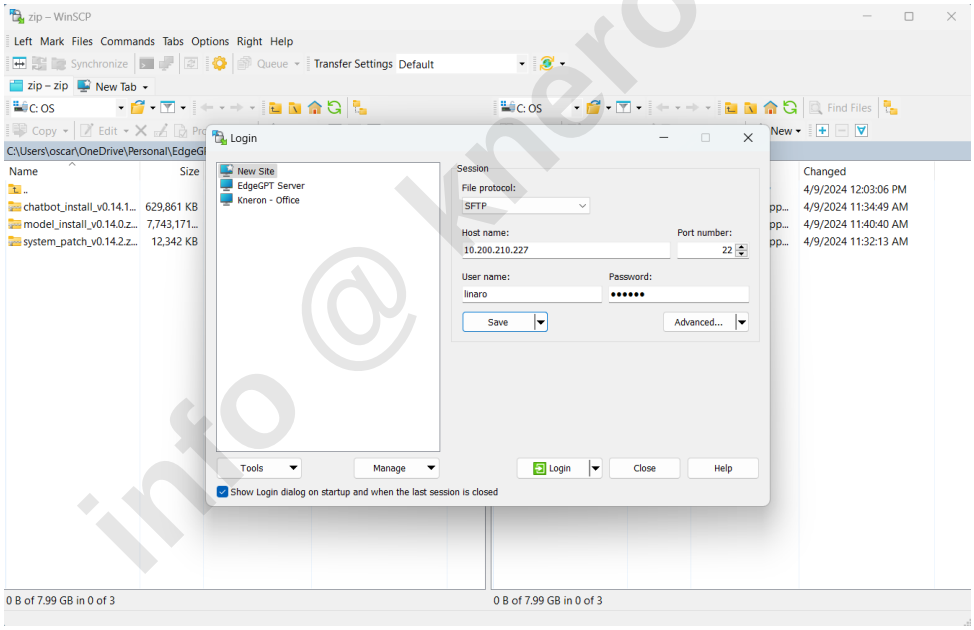

Figure 3-8 Winscp Remote Login

To remote access the KNEO 300, it first invokes winscp, then fills in the login information: including IP address (i.e. 10.200.210.227), username (i.e. linaro), and password (i.e. linaro). The administrator clicks the Save button and sets the current setting for the default login session. After that click the Login button to log in the KNEO 300.

The winscp interface is divided into the left and right panels, the left panel refers to the Windows directory, and the right one sets to the KNEO 300 home directory. The administrator can traverse different

directories using the buttons above each panel and drag the files from one panel to another one.

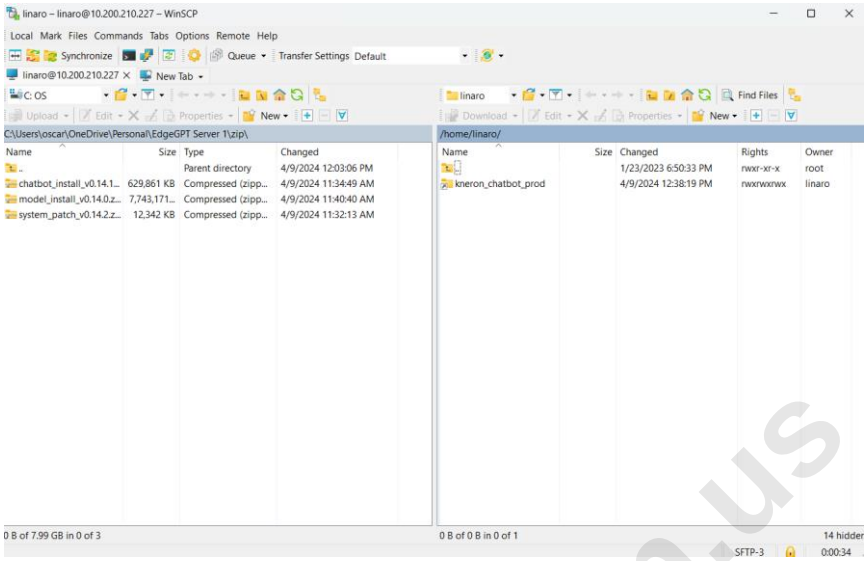

Figure 3-9 Winscp Remote Access

4. System update

4.1 Account Setup

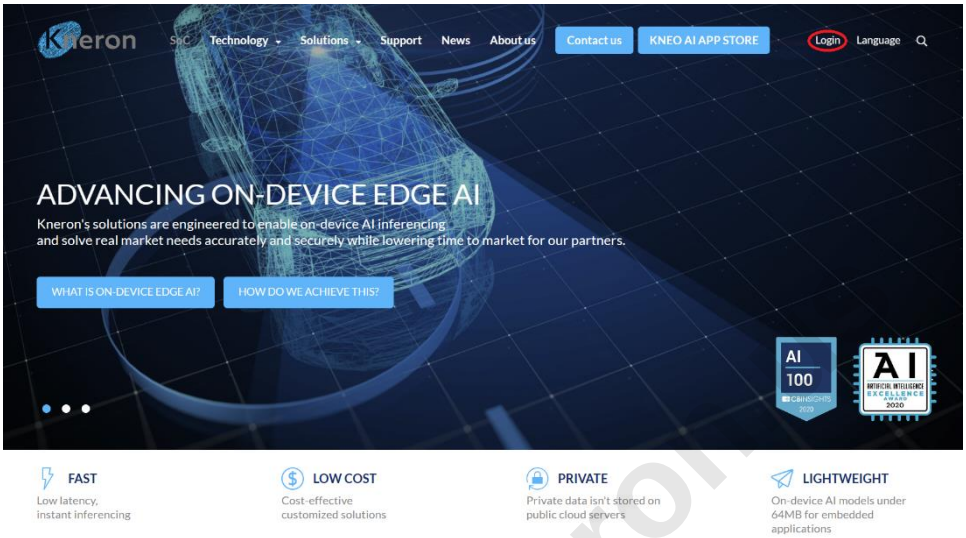

Figure 4-1 Kneron Home Page

For system updates, the administrator must register the user account on the Kneron home page. The administrator clicks the top right-hand Login button, which displays the login page. The administrator clicks the Create an account button and then follows the instructions to set up the user account. After the account setup, please provide the login e-mail to the Kneron salesperson, who shall grant permission to access the KNEO 300 documents and updated firmware.

Figure 4-2 Kneron User Login

After the permission is granted, the administrator can access the KNEO 300 document from the Kneron developer site (https://www.kneron.com/support/developers/), then click the KNEO300 under the entry Kneron AI chat robot to access different releases. **If the entry hasn't shown up, please follow up with the Kneron salesperson for permission access.**

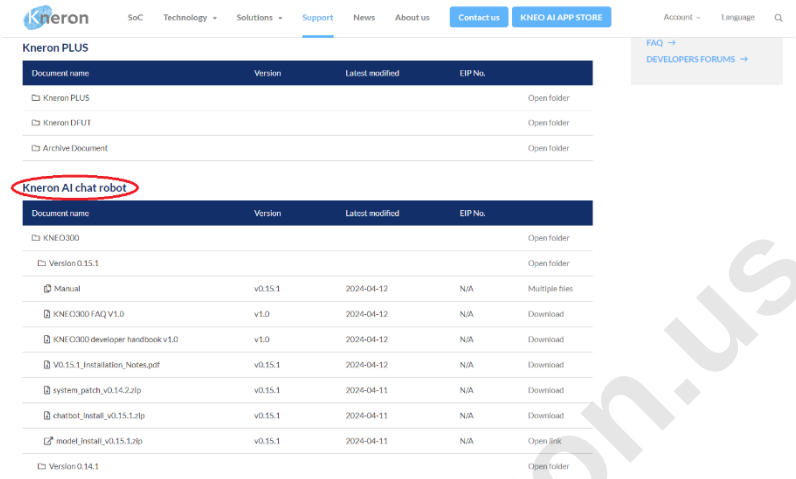

Figure 4-3 KENO 300 Document

The administrator can access both the administrator and regular user manual under the Manual subdirectory for every release. For system updates, please read the installation note in detail (<release>_Installation_Notes.pdf), the installation files include system_patch_<release>.zip, chatbot _install_<release>.zip, model_install_<release>.zip and the system_patch_autostart.zip. The administrator clicks the Download button and downloads them to the PC Downloads directory.

4.2 System Update

To update the system, please follow the following instructions:

1. Ask all the users to log out of the KNEO 300
2. Backup both the public and user databases in the directory: /home/linaro/kneron_chatbot_prod/kneron_doc_chat/knowledge_base/content
3. Reboot the system using the command: sudo reboot

```
linaro@chatrobot: sudo reboot
```

4. After the system is rebooted, please follow the instructions in Session 3.2 to transfer the files from Windows to the KNEO 300 directory /data[7]

[7] Due to the disk space limit, please don't transfer the files to KNEO 300 home directory: /home/linaro for system updates, it results in the system crash.

5. Uncompress the .zip files using the command: unzip <installation file>.zip, which creates the subdirectory with the same name <installation files>[8]

```
linaro@chatrobot: unzip <installation file>.zip
```

6. After the uncompressing, please remove the installation file using the command: rm <installation file>.zip and change to the subdirectory using the command: cd <installation file>. The administrator must follow README.TXT to update the system.

```
linaro@chatrobot: rm <installation file>.zip
linaro@chatrobot: cd <installation file>
```

7. After the installation is complete, please apply the following commands to clean up the disk space.

```
linaro@chatrobot: cd /data
linaro@chatrobot: rm -r <installation file>
```

8. Repeat the above procedure until update all the installation files.

[8] Please install all the installation files one by one and not uncompress all the installation files once, it results in the disk space issues.

5. External Storage

KNEO 300 can store the knowledge base using the USB drive or Network Attach Storage (NAS). The USB drive is directly connected to KNEO 300 using the USB port and NAS storage is accessed through the internet.

5.1 USB Drive

Currently, KNEO 300 supports the USB drive with the exfat format rather than the vfat one. Please refer to the following instructions for USB drive setup:

It first checks the USB drive format using the command sudo fdisk -l, the USB drive is labeled as /dev/sda1 or /dev/sdb1. The format is also shown as HPFS/NTFS/exFAT

```
linaro@chatrobot:~$ sudo fdisk -l
Disk /dev/ram0: 64 MiB, 67108864 bytes, 131072 sectors
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 4096 bytes
I/O size (minimum/optimal): 4096 bytes / 4096 bytes

. . . .

Disk /dev/sdb: 114.61 GiB, 123048296448 bytes, 240328704 sectors
Disk model:  SanDisk 3.2Gen1
Units: sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disklabel type: dos
Disk identifier: 0x66123c14

Device     Boot Start       End   Sectors   Size Id Type
/dev/sdb1          32 240328703 240328672 114.6G  7 HPFS/NTFS/exFAT
```

The administrator can check the disk format using the command lsblk -f, which shows the sdb1 as exfat format[9]

[9] If the format is not exfat one, please download the utility and reformat the USB drive.

sudo apt-get install exfat-fuse exfat-utils
sudo apt-get install ntfs-3g
sudo mkfs.exfat /dev/sdb1

```
linaro@chatrobot:~$ lsblk -f
NAME     FSTYPE LABEL UUID                                 FSAVAIL FSUSE% MOUNTPOINT
sdb
└─sdb1   exfat        04EE-70EF                            106.6G     7% /mnt/U2
mmcblk0
├─mmcblk0p1
│        vfat         F2B0-5E69                               36M    72% /boot
├─mmcblk0p2
│        ext4         f54fdcb4-8192-4894-9e51-0d44a28b82d3    2.7G     2% /recovery
├─mmcblk0p3
│
├─mmcblk0p4
│        ext4         f709bf25-bf4c-435c-8c44-d9284aa63203  311.7M    81% /media/roo
├─mmcblk0p5
│        ext4         e5885da8-ffa3-43f7-a8a8-bf51627a65b0    2.2G    59% /media/roo
├─mmcblk0p6
│        ext4         aa24e5f0-26b2-4b27-9917-40e53e8e31aa    1.8G     2% /opt
└─mmcblk0p7
         ext4         af947db5-d030-4483-af60-fb81abe282ab   18.9G    53% /data
mmcblk0boot0

mmcblk0boot1
```

The administrator can create the mount volume /mnt/U2[10] using “sudo mkdir -p /mnt/U2”

```
linaro@chatrobot:/$ sudo mkdir -p /mnt/U2
linaro@chatrobot:/$ ls /mnt
U  U2
```

and link the volume with /dev/sdb1 using “sudo mount -o rw /dev/sdb1 /mnt/U2” and turn on the read/write permission using “sudo chmod 777 /mnt/U2”

```
linaro@chatrobot:/$ sudo mount -o rw /dev/sdb1 /mnt/U2
FUSE exfat 1.3.0
linaro@chatrobot:/$ sudo chmod 777 /mnt/U2
```

[10] The administrator must check the /mnt subdirectory using the command “ls /mnt”, which makes sure not to duplicate the subdirectory under /mnt.

5.2 NAS Storage

The administrator can mount the external NAS storage[11] on KNEO 300. The mounted volume /mnt/U3 is first created using “sudo mkdir -p /mnt/U3”

```
linaro@chatrobot:/$ sudo mkdir -p /mnt/U3
linaro@chatrobot:/$ ls /mnt
U  U2  U3
```

Then, the administrator installs the external NFS module using “sudo apt install nfs-common” and links the volume using “sudo mount -t nfs <ip address>:<directory> <mount volume>”. For example, the external NFS storage is referred to as the drive /mnt/kds/data_feed in the machine 10.200.100.60, the link command becomes “sudo mount -t nfs 10.200.100.60:/mnt/kds/data_feed /mnt/U3”

```
linaro@chatrobot:/$ sudo apt install nfs-common
Reading package lists... Done
Building dependency tree
Reading state information... Done
. . . .
Not creating home directory `/var/lib/nfs'.
Created       symlink       /etc/systemd/system/multi-user.target.wants/nfs-client.target       →
/lib/systemd/system/nfs-client.target.
Created       symlink       /etc/systemd/system/remote-fs.target.wants/nfs-client.target       →
/lib/systemd/system/nfs-client.target.
nfs-utils.service is a disabled or a static unit, not starting it.
Processing triggers for systemd (245.4-4ubuntu3) ...
Processing triggers for libc-bin (2.31-0ubuntu9) ...
linaro@chatrobot:/$ sudo mount -t nfs 10.200.100.60:/mnt/kds/data_feed /mnt/U3
linaro@chatrobot:/$ ls /mnt/U3
China  From_NAS  Taiwan  ftp  huggingface  sandiego  shenzhen  taibei
```


[11] Please contact the local Linux IT support to install the external NAS storage