KNEO 300 User Manual V1.7

(For Admin 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 doc base on Oscar's version	V0.16.0	Warren Kuo	2024/6/20

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.

info @ kneron.us

Contents

KNEO 300 User Manual V1.7 1
1. Introduction 4
2. System Usage 5
2.1 Product Overview 5
2.2 Accessories List 6
2.3 Power on 7
2.4 Network remote login 7
2.4.1 Remote Access 7
2.4.2 ssh 8
2.4.3 putty 10
2.5 WEBUI Interface 11
2.5.1 Session Initialization 11
2.5.2 Public/Custom Database 14
2.5.3 Free Chat Mode 14
2.5.4 Knowledge Base Mode 16
2.6 Power Off 18
3. System Administration 19
3.1 User Registration 19
3.1.1 Account Setup 19
3.1.2. Access Permission 20
3.2 File Management 22
3.2.1 scp 22
3.2.2 winscp 22
4. System update 24
4.1 Account Setup 24
4.2 System Update 25

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. System Usage

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

info @ kneron.us

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.

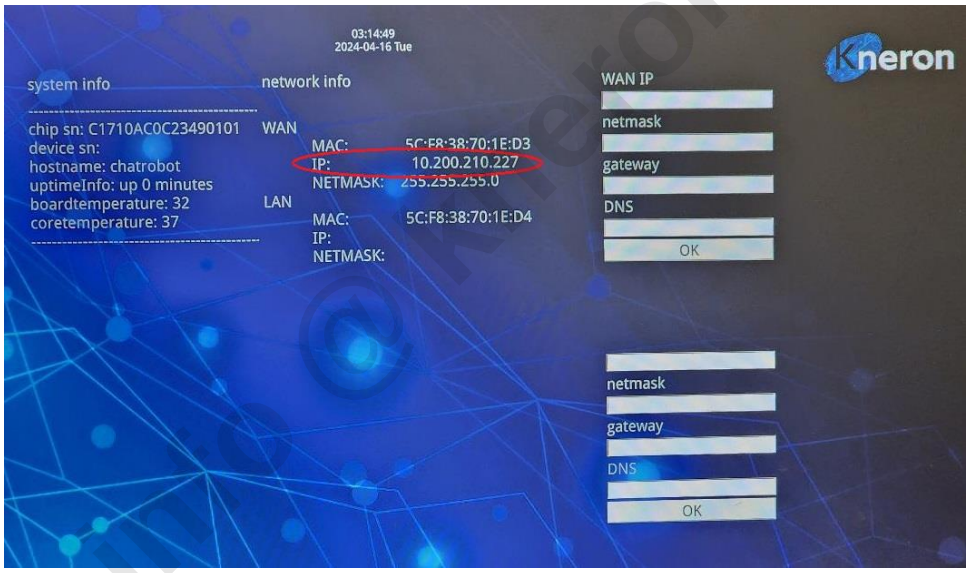

Figure 2-2 KNEO 300 IP Address

2.4 Network remote login

2.4.1 Remote Access

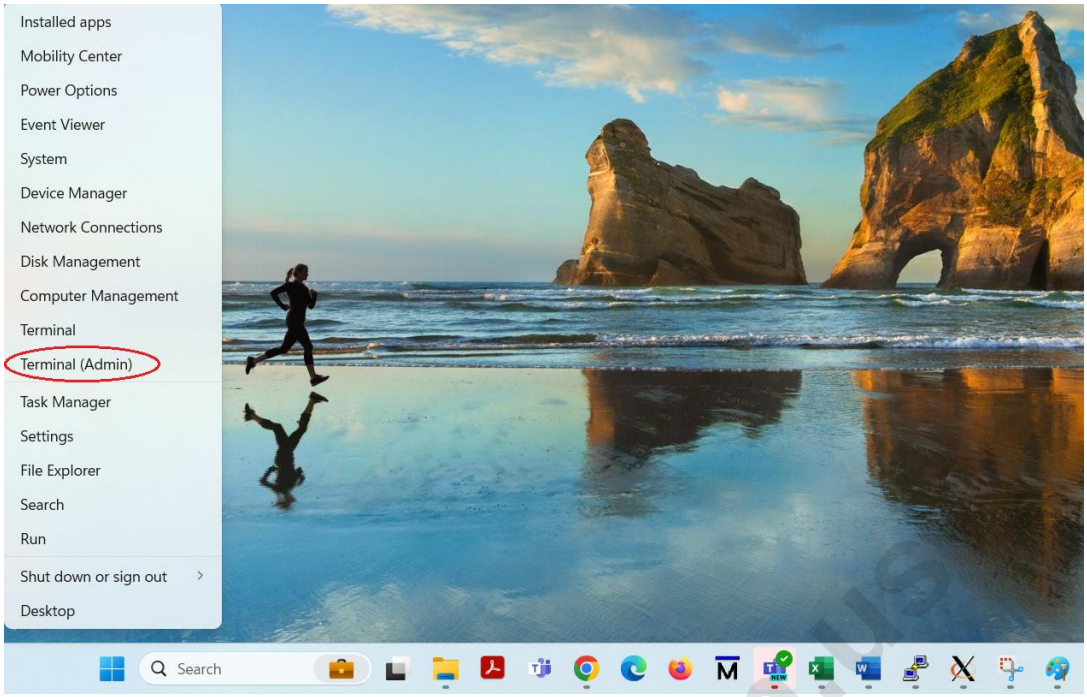

Figure 2-3 Windows PowerShell

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 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:
```

info @ kneron.us

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:~$
```

The administrator **must** invoke the utility screen to start the system background job in KNEO 300. **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**.

```
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
```

info @ kneron.us

```
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 and B at the same time to leave the screen mode. The software still runs on KNEO 300 even if the administrator logs out from 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

2.4.3 putty

The putty is a graphical remote login software, It first downloads the putty from the website (i.e.

info @ kneron.us

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.

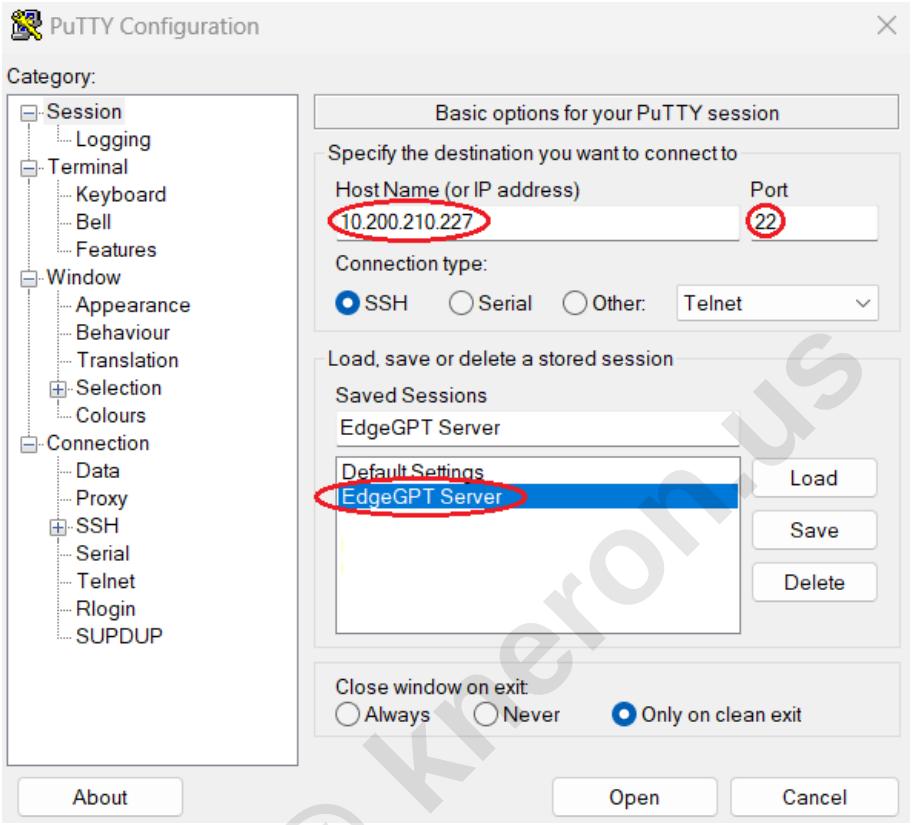

Figure 2-4 Putty home screen

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

info @ kneron.us

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)

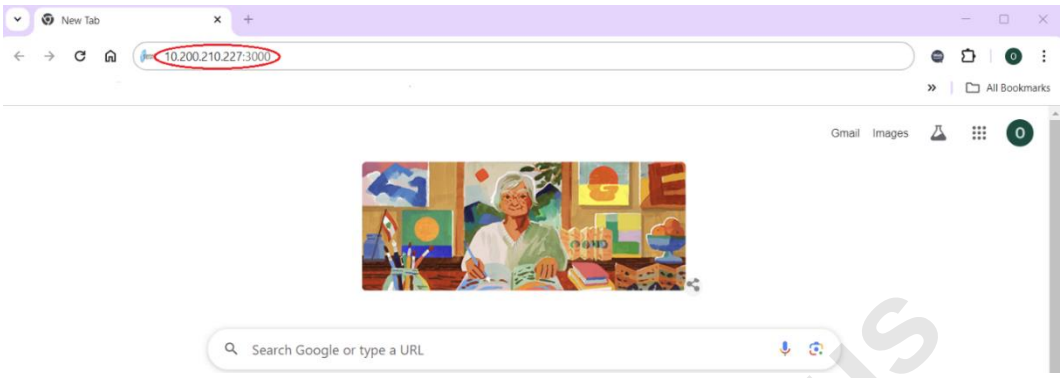

Figure 2-5 Browser Access

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 Chinese.

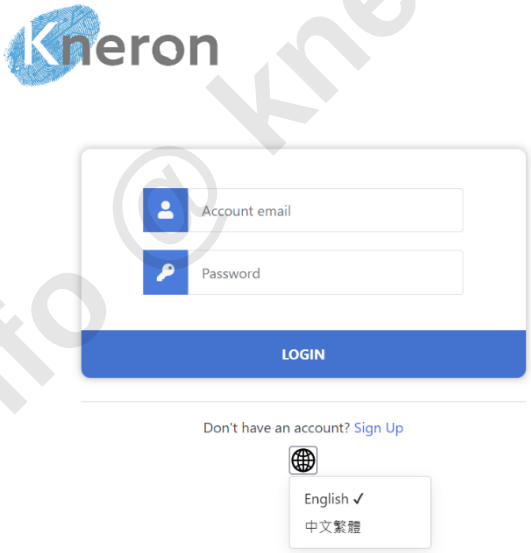

Figure 2-6 WEBUI Login Page

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

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

info @ kneron.us

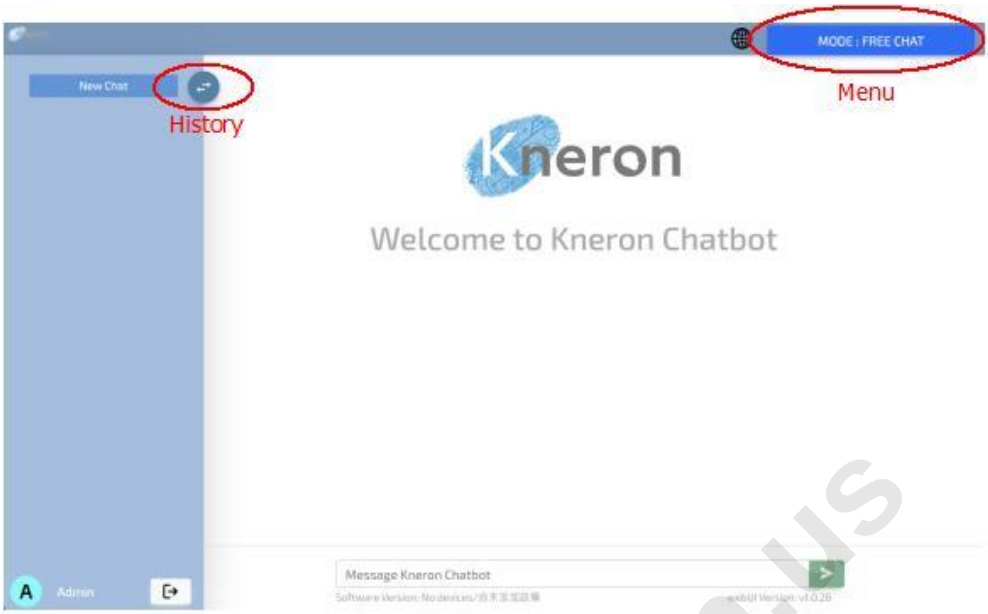

Figure 2-7 WEBUI Session

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 running the session on the current server only. 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.

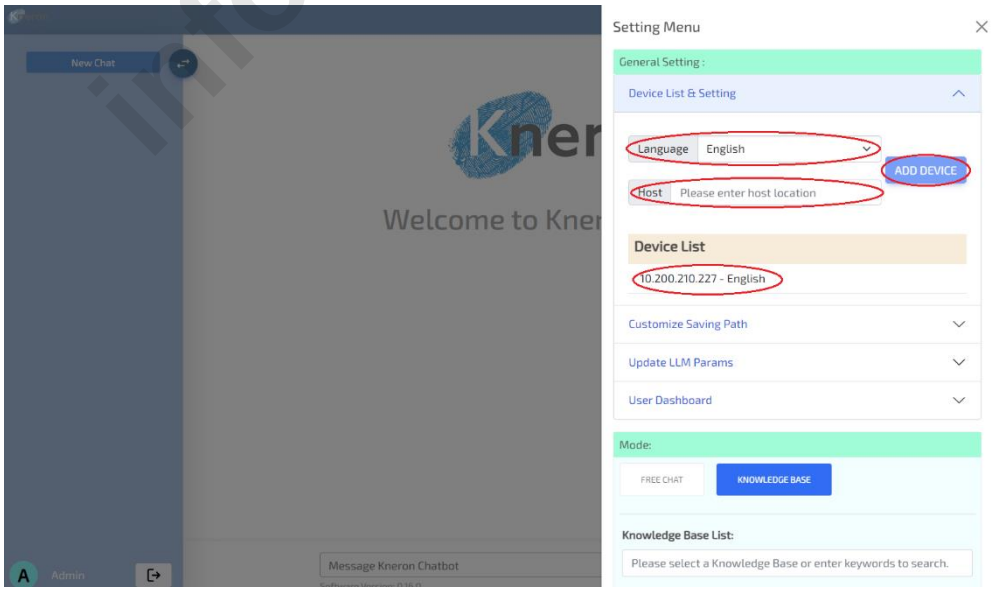

Figure 2-8 WEBUI Setting Menu

info @ kneron.us

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. If the language is set to English, all the databases are stored under the EN subdirectory. After the administrator initializes the servers using the language English. All the databases are stored under the subdirectory (EN), the users can access those databases only. The administrator can re-initialize the system using the language Chinese, which allows the users to access those databases under the subdirectory (ZN). **The administrator is responsible for backing up all the databases during the system update**.

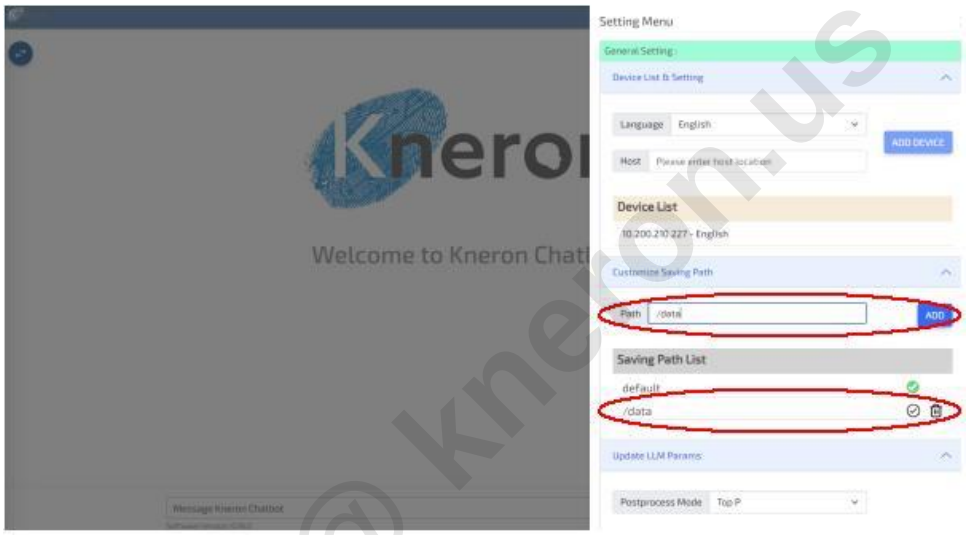

Figure 2-9 WEBUI Data Path

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. It is useful to store the databases in the external mount devices.

The administrator can select different data paths 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 mount devices.

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.

info @ kneron.us

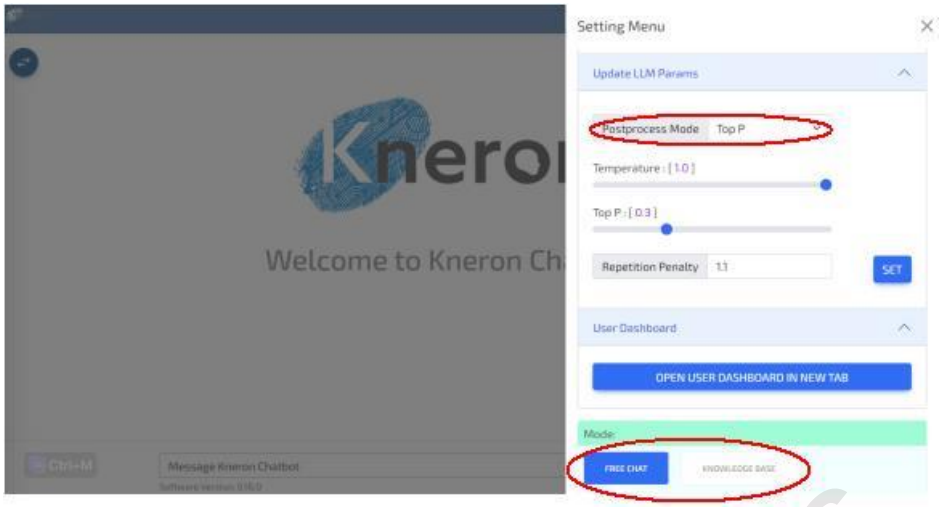

Figure 2-10 Chat Mode

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 users can stop or clean the inquiry using the CLEAR RECORD button.

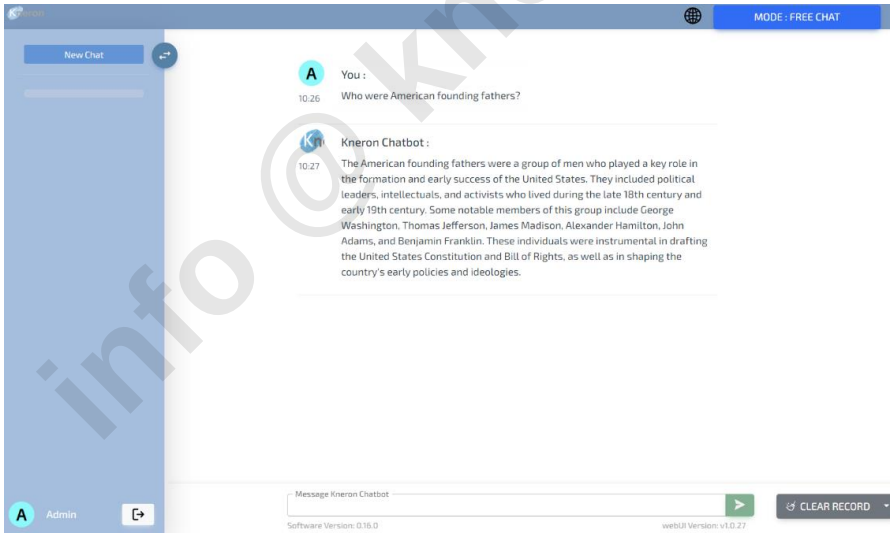

Figure 2-11 Free Chat Mode

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

info @ kneron.us

create the same output with a higher penalty. The penalty is set between 1 and 10.

2.5.4 Knowledge Base Mode

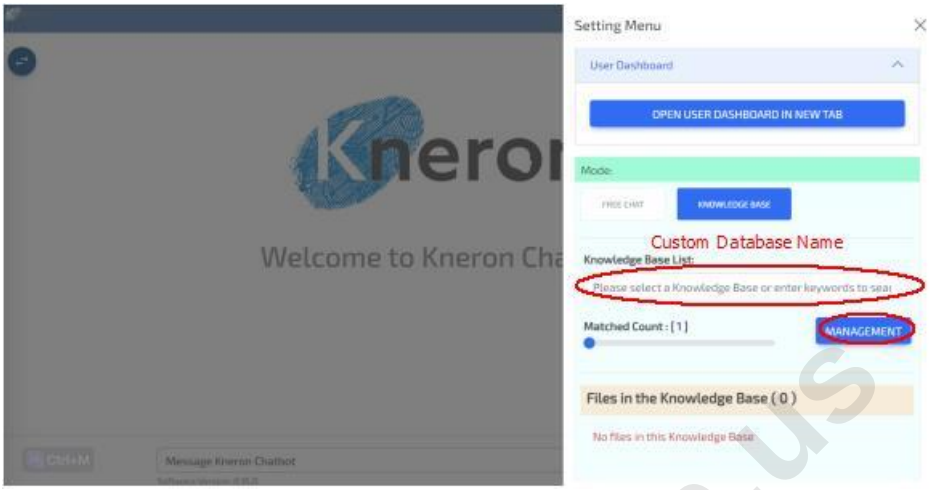

Figure 2-12 Knowledge Base Management

For knowledge base inquiry, the administrator 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 administrator enters the database name <user>/<database> in the Knowledge Base List box (i.e. public/bda602), then clicks the UPLOAD FILES 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.

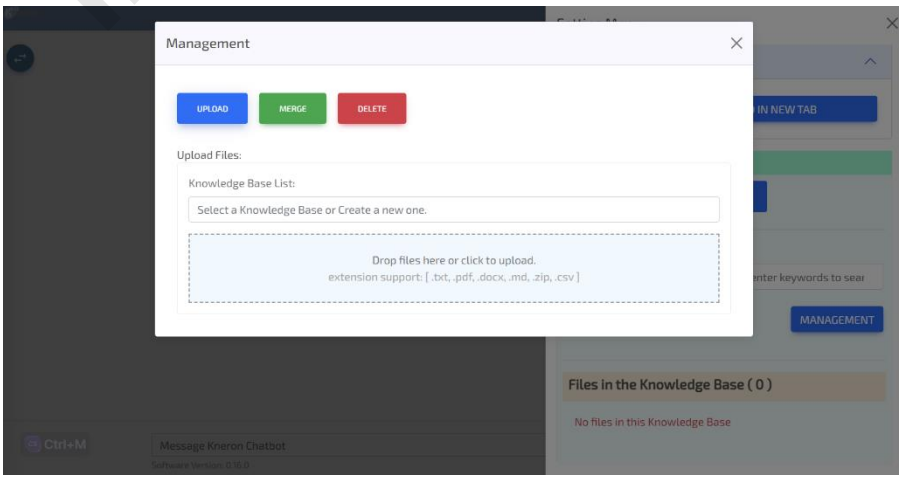

Figure 2-13 Custom Base Creation

info@kneron.us

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.

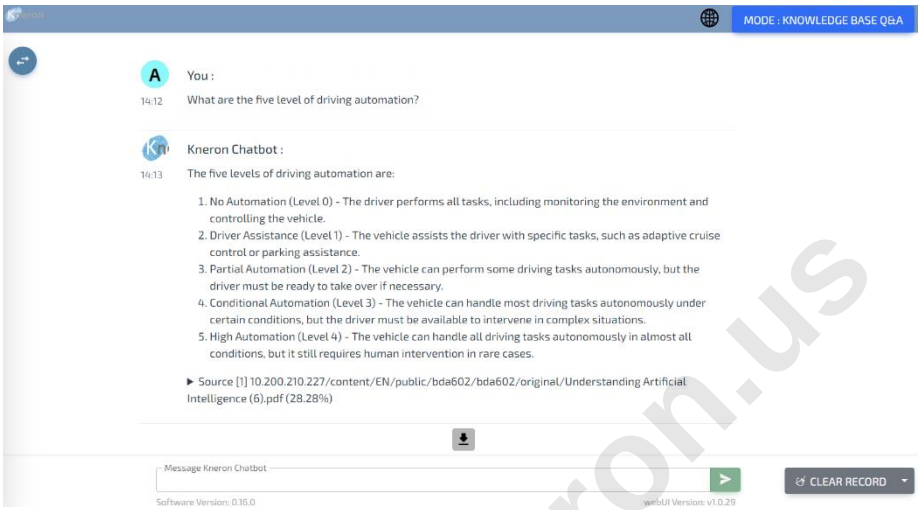

Figure 2-14 Custom Database Inquiry

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.

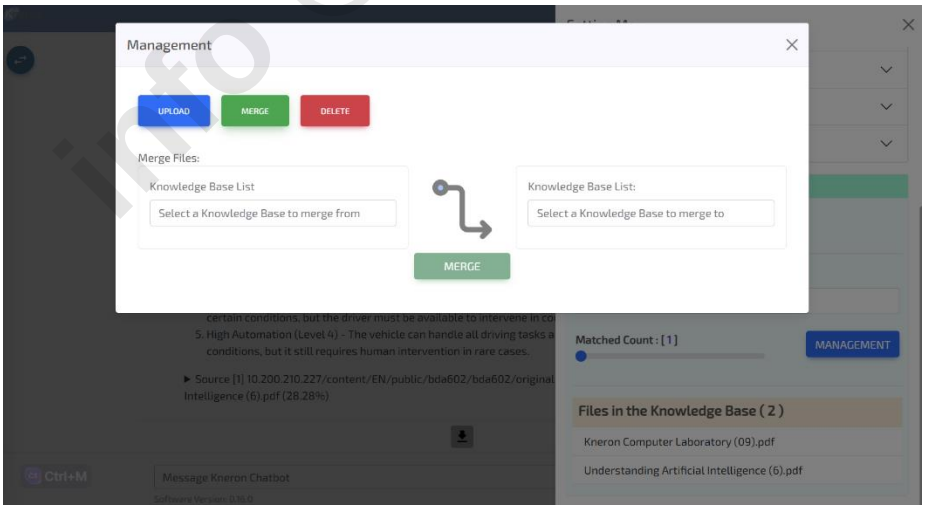

Figure 2-15 Merge Custom Database

The administrator clicks the DELETE button to delete the files or databases, the administrator can highlight

info @ kneron.us

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.

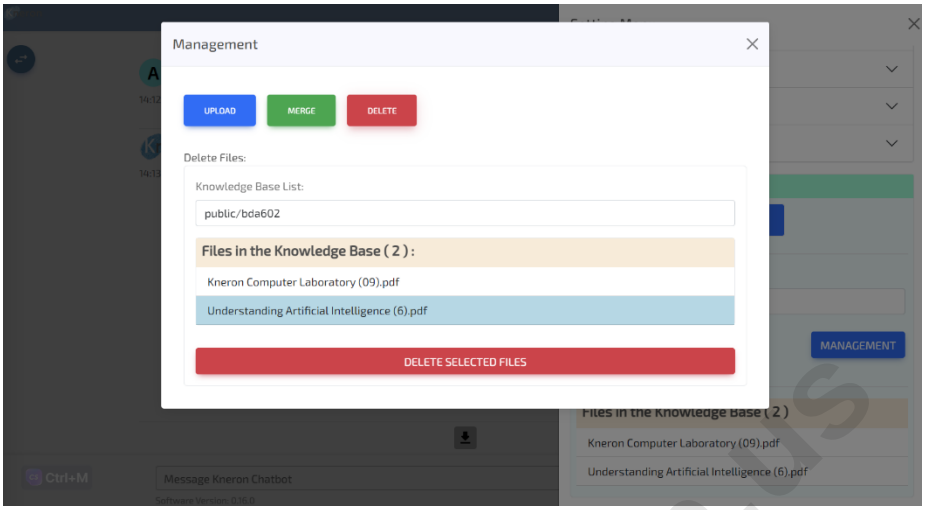

Figure 2-16 Delete Single/Multiple Files or Custom Database

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
```

info @ kneron.us

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 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.

info @ kneron.us

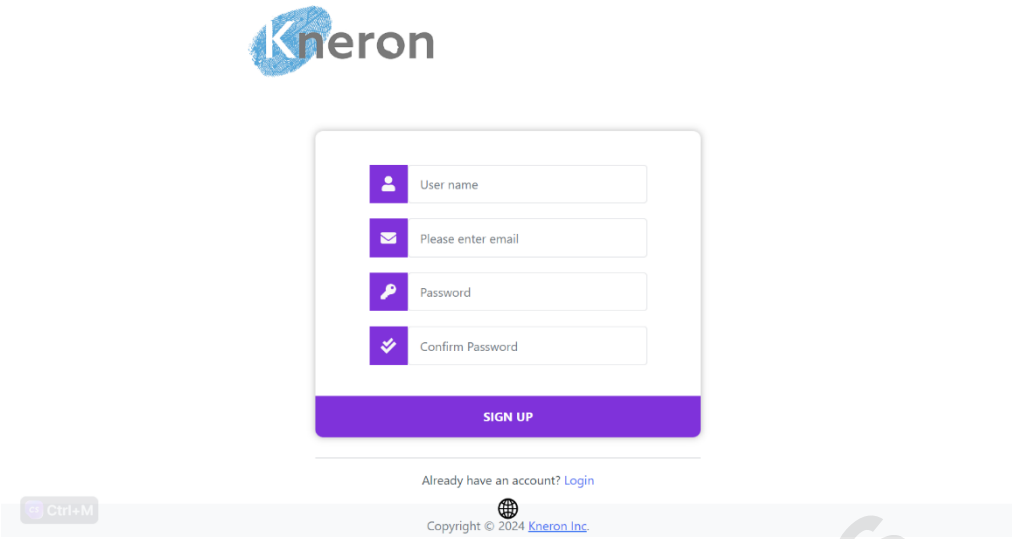

Figure 3-2 New User Registration

3.1.2. Access Permission

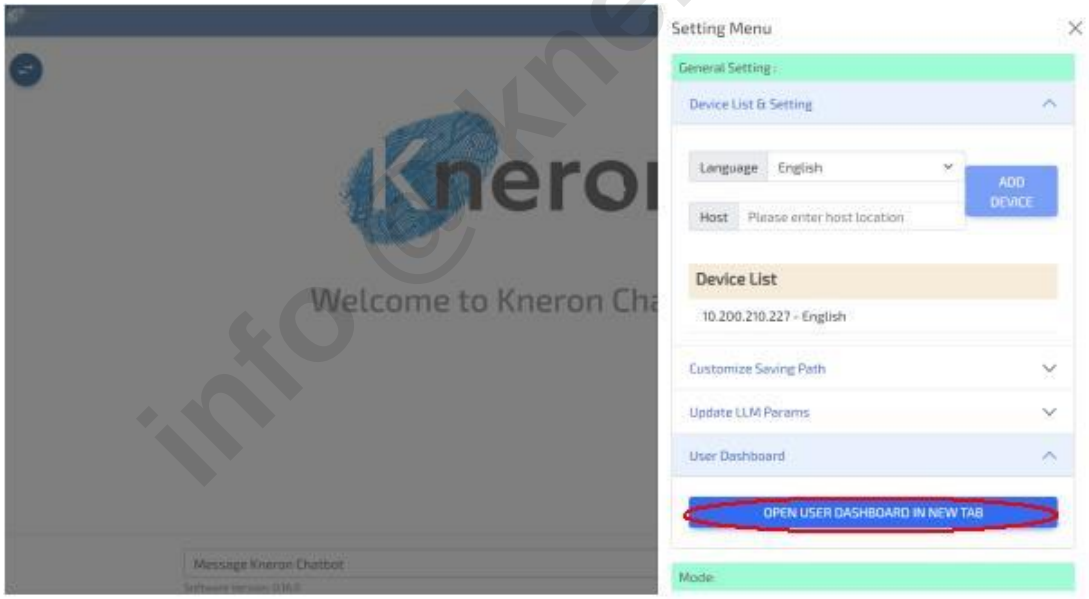

Figure 3-3 User Dashboard

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.

info@kneron.us

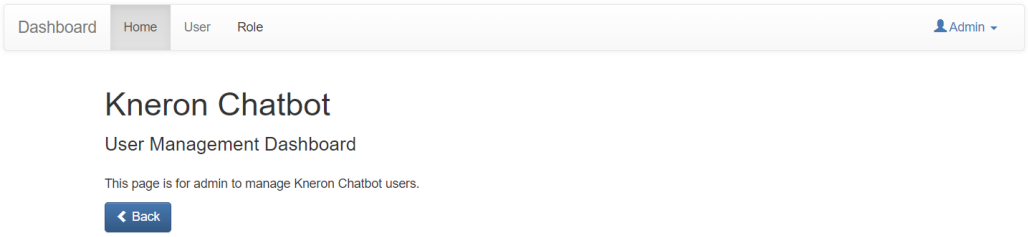

Figure 3-4 Dashboard Home Menu

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

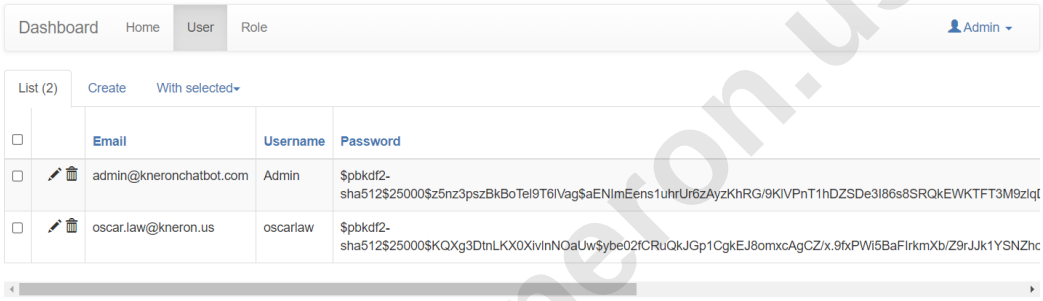

Figure 3-5 Dashboard User Menu

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.

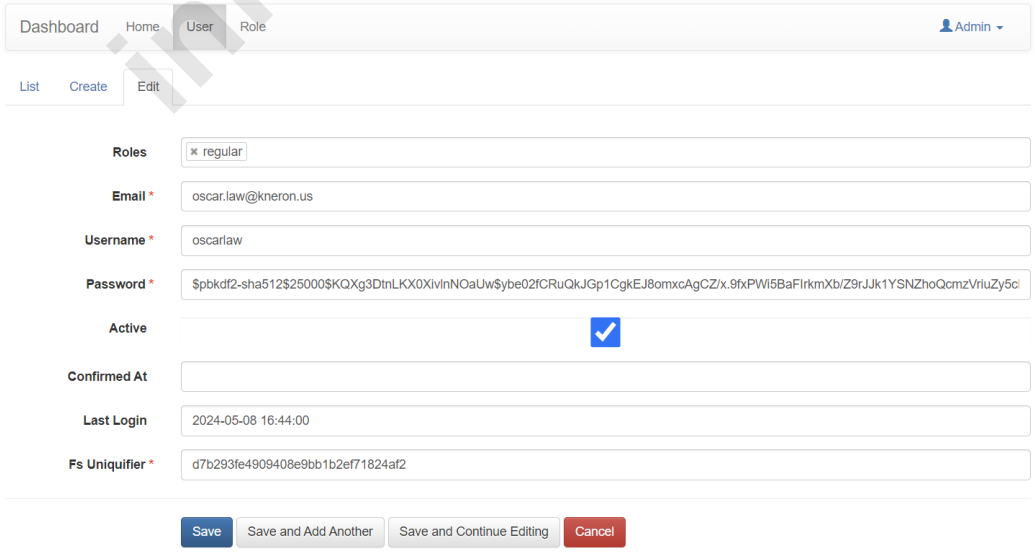

Figure 3-6 User Role Modification

info @ kneron.us

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>. If the administrator can copy the entire directory from Windows to KNEO 300 with an additional option: scp -r <directory> <user_name>@<ip address>:<user directory>. It is useful to back up the entire database 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 alternative file transfer graphical interface: winscp, the administrator downloads the software from the website (https://winscp.net/eng/download.php) and follows the instructions to install the software.

info @ kneron.us

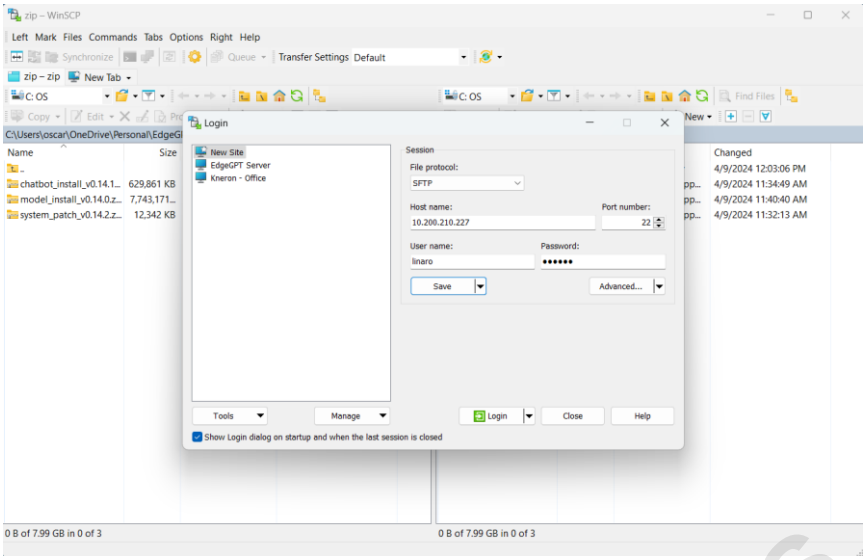

Figure 3-7 Winscp Remote Login

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 to 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 traverses different directories using the buttons above each panel and drags the files from one panel to another. The administrator can drag the entire directory between Windows and the KNEO 300.

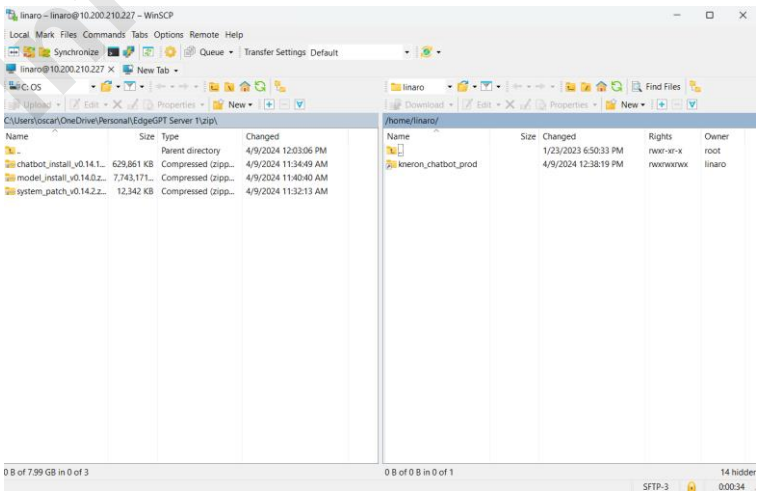

Figure 3-8 Winscp Remote Access

info @ kneron.us

4. System update

4.1 Account Setup

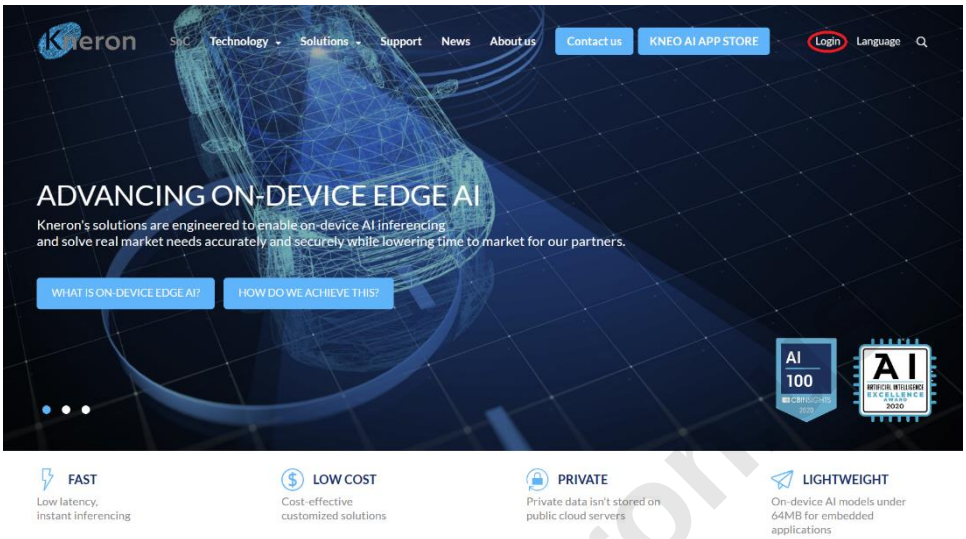

Figure 4-1 Kneron Home Page

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

info @ kneron

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.**

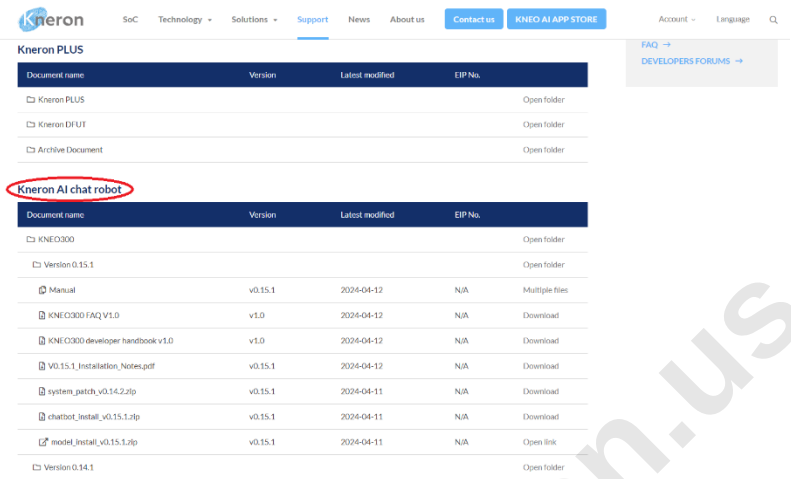

Figure 4-3 KENO 300 Document

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, and model_install_<release>.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[1]

5. Uncompress the .zip files using the command: unzip <installation file>.zip, which creates the

[1] 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.

info @ kneron.us

subdirectory with the same name <installation files>[2]

```
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.

info @ kneron.us

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

info @ kneron.us