

KNEO 300 User Manual V1.2

(For Admin User)

Mar, 2024

Revision History:

Doc Version	Description	Firmware Version	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



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.

2



Table of contents

KNEO 300 User Manual V1.21
Table of contents
1 Introduction4
2. Device operation tutorial4
2.1 Product Overview4
2.2 Accessories list5
2.3 Power on6
2.4 Network remote login7
2.5 File upload and download9
2.6 Shutdown
3. Chatbot software usage guide11
3.1 Start the WEBUI interface11
3.1.1 Start up11
3.1.2 Admin settings
3.1.3 Service settings13
3.1.4 WEBUI interface introduction15
3.2 Others20
4. System update22
4.1 Update chatbot software22
4.2 Update model24
4.3 Update Firmware
5.FAQs



1 Introduction

KNEO 300 is an NPU-based edge AI server, specially 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. Device operation tutorial

2.1 Product Overview

• KNEO 300 series AI box appearance







• 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. Down : USB2.0x2
- 7. HDMI 2.0
- 8. TF Card
- 9. DC 12V
- 10. 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 Temperature: -20 $^\circ$ C~60 $^\circ$ C;	
Operating Environment	Storage Temperature: -20 $^{\circ}$ C~70 $^{\circ}$ C;	
	Operating Humidity: 10%~90%RH;	
Ethernet	2*Gigabit Ethernet	
	2*USB3.0	
USD	2*USB2.0	
Connecting Darts	1*RS232	
connecting Ports	1*RS485	

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 following screen, and the IP address of the device will be displayed in the red box.

				* R		
	system info	netwo	15:05:3 2024-01-0 ork info	is i Thu	WAN IP	Gneron
	chip sn: E0711AC0C23520009 device sn: hostname: chatrobot uptimeInfo: up 0 minutes boardtemperature: 40 coretemperature: 43	WAN LAN	MAC: IP: NETMASK: MAC: IP: NETMASK ⁻	SC:F8:38:70:1E:A7 192:168.10.106 255:255.255.0 SC:F8:38:70:1E:A8	netmask gateway DNS	
					netmask	
H					gateway DNS OK	
				Dest 1		



2.4 Network remote login

If the PC can successfully ping the IP address displayed on the monitor after startup, you can then use ssh to log in remotely. The port number is 22 and the username and password are both *linaro*. Users can use Windows or Linux systems, but it is recommended to use putty or SecureCRT software on Windows systems for remote login.

Command line: *ssh linaro@ip_address*

2.4.1 windows system

The following uses putty software as an example.

- (1) Download and install putty.
- (2) Start putty, enter the device IP address displayed on the monitor, the port number is 22

Session	Basic options for your Pul	ITY session
· Terminal · Keyboard · Bell	Specify the destination you want to Host Name (or IP address) 192.168.10.150	connect to Port 22
Features Window Appearance	Connection type: Raw Telnet Rlogin	SSH OSerial
Behaviour Translation ⊡ Selection	Saved Sessions	
Connection Data	Default Settings	Load Save
Proxy Telnet Rlogin III SSH	0	Delete
Serial	Close window on exit: Always Never On	ly on clean exit

Click the "Open" button.

(3) In the next interface, enter the username and password, both are *linaro*



🛃 192.168.10.150 - PuTTY	-	×
P login as: linaro P linaro@192.168.10.150's password:		
		\sim

(4) After the verification is passed, it will be displayed as follows.



After connecting the device, you can start the chatbot software according to 3.1.

2.4.2 Linux system

Take the ubuntu system as an example.

Use the command: ssh linaro@192.168.10.150

192.168.10.150 is the IP address of the device, linaro is the username, and you will be prompted to enter a password later. The password is also linaro.



After connecting the device, you can start the chatbot software according to 3.1.



2.5 File upload and download

Users can upload and download files through the SCP protocol on Windows or Linux systems. The username and password are both *linaro*, and the port number is 22. We recommend using winscp for access on Windows systems.

2.5.1 windows system

- Download and install winscp software.
- Start the winscp software, fill in the IP address of the device in the Host name column, and click "Login"



• Enter the username and password on the next screen, both are *linaro*, and click "OK".

192.168.10.150 - WHISCP				- 8 X
Local Mark Files Commands Tabs Options Remote Help				
🖽 🔠 Tyrchronize 📷 🥐 📰 🌘 🦪 Gunur - 🛛 Transfer Settings De	fault • 1 🖉 •			
📹 system_patch_2 Documents 🗙 👎 192.168.10.150 × 🖷 New Tab +				
R#XX# - 💋 - [] 🖬 🐄 🕲 🗞	My docu - 🞁 • 🝸 • 🛛 = 🛛 🖬 🐧	C C First Tiles B		
🎲 Upload • 📝 5/0 • X 📈 🗋 Properties - 🔛 New • 💽 🖂 🗑	Doveload • 2 Edt - X 🔜 Departer	- 12 Nov - 1 🗷 🖂 🗹		
Fijobi,AJ_Box\release\system_patch_20240105\				
Nere 16	here	erez - 1931415350 X generality for local Autoricating.	Sie Oargol * Riphs Ower	

• After the login, you can use the mouse to drag files to the box or download files from the box.

Initial 20240103 – 192108-10.150 – WINSCP							
Local Mark Files Commands Tabs Options R	iemote Help						
🖽 🕃 😂 Synchronize 🖬 🥏 🔯 🔞	Queue • Transfer Settings D	Jolault • 😹 •					
🛄 system_patch_2 Documents 🗙 💻 192.160	1.10.150 × 🐺 New Tab +						
- n \$108 - 😰 - 🝸 - 🖛 - 🛛 🖬 🕅	20 5	Inaro • 📅 • 🝸 • 👘 • 🗧 • 🔛 🖬 🌚 🖏 🖏 Find Files 🎭					
Upload + Z Edt + X A D Properties	- New + 1	118 Doueload + 2 Edk + X of D Properties + 10 New + 10 E					
Frijob/AI_Box/release\install_20240103\		(home/linearc)					
Name	Sz	Name	Size	Changed	Rights	Owner	
		K		2023/1/14 11:04:10	PHORT-DIT-IX	root	
M contert.zip	453,358 KI	bsp-debs		2024/1/2 12:51:42		root	
10 install.sh	2 KI	instal_20240103		2024/1/4 18:17:04	recenter-s	linaro	
kreron_chatbot_prod.zip	61,364 KI	Recon_chatbot_prod		2024/1/4 21:30:45		linaro	
kneronchatbot.so	5,269 KI	C epuinto	2 KB	2023/12/17 18:12:38	ne-rw-r	linaro	
ingine.conf	2 KJ	instal_20240103.zip	916,102 KB	2024/1/4 20:22:26	IN-IW-I	linaro	
- README.txt	1 KI	auto_launch_zh.log	1 KB	2024/1/5 20:27:04	nu-rw-r	linaro	
wheels.zip	402,447 KJ	auto_launch_en.log	1 KB	2024/1/5 20:31:11	PH-PW-1	linaro	



2.5.2 Linux system

Take the ubuntu system as an example.

• Upload file.

Take uploading the file install_20240103.zip to the 192.168.10.150 /home/linaro directory as an example. Use the command: *scp install_20240103.zip linaro@192.168.10.150:/home/linaro*



192.168.10.150 is the IP address of the device, /home/linaro is the address where the file is stored after uploading, and linaro is the username. The system will prompt you to enter the password, both of which are linaro.

• Download file

Take downloading the file install_20240103.zip from the /home/linaro directory of the device to the local as an example.

Use the command: scp linaro@192.168.10.150:/home/linaro/install 20240103.zip.



192.168.10.150 is the IP address of the device, /home/linaro/install_20240103.zip is the file that needs to be downloaded, and linaro is the user name. The system will prompt you to enter the password, which is also linaro. After the file is downloaded, it is stored in the current directory.

2.6 Shutdown

Please try not to disconnect the power directly when shutting down but run it first sudo poweroff before powering off to avoid damaging file system data.

In addition, if you have successfully entered the Linux system, you can also press and hold the power button. The system will detect and safely turn off the power of the system and development board.



3. Chatbot software usage guide

3.1 Start the WEBUI interface

3.1.1 Start up

- a. Using 'screen', you can start a process that will continue to run on the server even if the SSH connection is disconnected, so that the website can continue to serve external parties.
 - i. Enter screen on the command line. After Figure 2 appears, press the space bar or Return key.



- b. Enter the following commands in order on the command line, as shown in the figure below
 - i. cd kneron_chatbot_prod
 - ii. chmod +x new_launch.sh
 - iii. ./ new_launch.sh (it takes about 30 seconds)

linar@chatrobotr/#ceineschatbot_prod linar@chatrobotr/#ceineschatbotroschatbotroschatbotroschatbotroschatbotroschatbotroschatbotroschatbotroschatbotr	
limaro@chatrobotr/knerom_chatbot_prod\$./launch_en.sh	
Enetopol January Sert Karoon Chatbat version: 0.11.0	
Load /data/kneron_chatbot_prod/kneron_doc_chat/kneron_models/chat/tokenizer.model Done! Device (&) loadim	
Model[/data/kneron_chatbot_prod/kneron_doc_chat/kneron_models/chat/kneron_llm_1024.bmodel] loading	
uone: /home/linaro/kneron_chatbot_prod/kneron_doc_chat/webui_edge_en.py:226: GradioUnusedKwargWarning: You have unused kwarg parameters in Row, please remove them: {'justify': 'center'}	
with gr.Row(jusify="center"): /home/jimar/knoron chathot prod/knoron doc chat/webui edne en ov:238: GradioBenrecationWarning: The `style' method is deprecated. Please set these arguments in the constructor instead	
chatbot = gr.Chatbot(
/nome/linato/kneron_chatoc_prod/kneron_doc_chat/webul_edge_en.py:242: bradioueprecationwarning: ine style method is deprecated. Please set these arguments in the constructor instead. query = gr.Textbox(
Running on local URL: http://0.0.0:19999	
To create a public link, set 'share=True' in 'launch()'.	
	_
DNI Erran version J. B. B. (DNI) AL-Exh-36	
Copright (5) 2015-201 Justgen Weiger, Alexador Mauroy, Amadeus: Slawinski	
Copyright (c) 2018-2013 Juergan weigert, Nainel Schwarder, Michel Schward, Sadrul Habib Chewdhury Copyright (c) 2018-2019 Juergan weigert, Michel Schwarder, Michel Schward, Sadrul Habib Chewdhury	
Copyright (c) 1993-2007 Juargan 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 WINKOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for nore 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 02211-1301 USA.	
Send bugreparts, fixes, enhancements, t-shirts, money, beer & pizza to screen-develögnu.org	
Canabilities:	
+copy +remote-detach +power-detach +multi-attach +multi-user +font +color-256 +utf8 +rxvt +builtin-telnet	

c. the WEBUI starts, press and hold the Ctrl, A and D keys on the keyboard at the same time,



and the following picture will appear. This step allows even if the remote network connection is disconnected, using screen The created session (session) will still be maintained



d. Enter "screen Is" to see the sessions created by screen

```
Inaro@chatrobot:~$ screen -ls
There is a screen on:
        2181.pts-2.chatrobot (02/15/24 03:53:43) (Detached)
1 Socket in /run/screen/S-linaro.
```

e. To retrieve the session created by screen, you need to enter "screen -r"

linaro@chatrobot:~\$ screen -r 2181.pts-2.chatrobot

f. After Launch is successful, it will display

Compiled successfully!

You can now view auth in the browser.

Local: http://localhost:3000 On Your Network: http://10.200.210.181:3000

Note that the development build is not optimized. To create a production build, use npm run build.

webpack compiled successfully

3.1.2 Admin settings

a. After the service is started successfully, enter 10.200.210.181:3000 in the local computer browser (for example, the IP of this service device is 10.200.210.181)

olginin
Email abc@gmail.com 🤤
Password Password here
Sign In

Sign In

Don't have account Sign Up?

b. For first time use, select Sign_up ,

	10.200.210.181:3000 says Account created successfully	ed Diagram
Sign Up		
Email abc@gmail.com	Sign Up	
Password Password here 🔤 Confirm Password Please confirm passes Sign Up	Username/jernawu Emailjerna@kweron.us 2 Password Contirm Password Sign Up	
Already registered Sign In?	Already registered Sign In?	

c. As Admin, after logging in, please go to <u>http://10.200.210.181:5000/admin/user/</u> Set the mailbox permissions on the margin of the page. (Take device ip 10.200.210.181, <u>jenna@kneron.us</u> account as an example). The default user is a regular user after registration and login.

Dashboar	rd Home User		
Record was	successfully deleted.	0	
List (1)	Add Filter - With selected-	Search: username, email	
	Username	Email	Role
○(⁄)ੇ	jennawu	jenna@kneron.us	regular

d. Click the pen circled in red in the picture above, edit the jennawu account, change the regular in the role to admin, and click "save" to save.

Dashboard	Home	User	
List Edit			
	Email	jenna@kneron.us	
	Role	admin	~
		Save and Continue Editing Cancel	

3.1.3 Service settings

a. Enter 10.200.210.181:3000 in the local computer browser (for example, the IP of this service device is 10.200.210.181), and then log in with the updated Admin email password (wait 30 seconds). Note: Do not click the Sign-in button back and forth.



Sign In

Emailabc@gmail.com 🧰	
Password Password here	
Sign In	

Don't have account Sign Up?

b. After logging in, the following interface appears

Geron	Please Select Language ● English 中文 Confirm
	10.200.210.181 Add device
ialogus/對話 Kowledge Base Management/知識庫管理	
Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation with LLM model or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.	Please Select Mode of Use LLM Knowledge Base QEA
	Configure Knowledge Base
	Please Select a Knowledge Base to Load
	admin/kneron_doc_center -
	Add files to Knowledge Base
	Upload File/上傳文件
	Drop File Here - or - Click to Upload

- i. Step 1: You need to select the service device language. After selecting, please click the "Confirm" button.
- ii. Step 2: Fill in the IP address of the service device (take local as 10.200.210.181). Note that you can add multiple KNEO300 EdgeBOX IP addresses. To obtain the IP address, please refer to 2.3 Introduction. As shown in the figure, the device is added. Edge service is officially available to employees.

Greron	Please Select Language Conf
	10.200.210.181 // Add do
Dialogur/對話 Kowledge Base Management/知識場管理	
Welcome to Kneron Knowledge Base Chat! Please awitch modes on the right. Currently supports direct conversation with LLM model or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.	Please Select Mode of Use LLM O Knowledge Base Q&A
Device added.	Configure Knowledge Base Please Select a Knowledge Base to Load
	admin/kneron_doc_center
	Add files to Knowledge Base

Note: Do not switch between Chinese and English at will for the admin account. Once selected, other users are already using it. If you want to switch, please coordinate the service switching time with other regular users.



3.1.4 WEBUI interface introduction

3.1.4.1. WEBUI is a chat dialog box, where users can have interactive conversations. At the same time, the knowledge base management page performs operations such as deleting and merging existing knowledge bases.

Dialogue/對該 Kowledge Base Management/知識庫管理	
Welcome to Kneron Knowledge Base Chatl Please switch modes on the right. Cu or Q&A based on local knowledge base. For Q&A mode, after selecting a knowle	urrently supports direct conversation with LLM model dge base, you can start Q&A.
Device added.	
Please enter your question and press enter to submit	
neron	
	Please Select Mode of Use
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base to Load
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Create New Knowledge Base
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base to Load Create New Knowledge Base Enter New Knowledge Base Name
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Letter New Knowledge Base Add to Knowledge Base
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Letter New Knowledge Base Add to Knowledge Base Number of Matches
gue Welcome to Kneron Knowledge Base Chatl Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Lenter New Knowledge Base Add to Knowledge Base Number of Matches
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Enter New Knowledge Base Add to Knowledge Base Inter New Knowledge
gue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LM Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Create New Knowledge Base Enter New Knowledge Base Number of Matches Refresh knowledge Base list Reset
igue Welcome to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base Please Select a Knowledge Base Create New Knowledge Base Create New Knowledge Base Cherrent Answer Cherent Answer Cherrent Answer Cherrent Answer Cherrent Answer C
Reverse to Kneron Knowledge Base Chat! Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model
Velcome to Kneron Knowledge Base Chatl Please switch modes on the right. Currently supports direct conversation or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A. Model loaded successfully, you can start the conversation, or select a mode from the right side to begin.	with LLM model Please Select Mode of Use LLM Knowledge Base Q&A Configure Knowledge Base Please Select a Knowledge Base to Load Create New Knowledge Base Enter New Knowledge Base Add to Knowledge Base Number of Matches Refresh knowledge base list Reset Stop Current Answer

3.1.4.2. WEBUI is the selection of chat mode. Currently we have launched two modes: free conversation and knowledge base conversation.



a. Free conversation mode

	Please Select I	Inde of Lice
Velcome to Kneron Knowledge Base Chatl Please switch modes on the right. Currently supports direct conversation with LLM model or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.		Knowledge Base Q&A
Device added.		Reset
here are 20 birds in the tree. 12 flies way, How many birds left in the tree		Stop Current Answer
here are 20 birds in the tree. 12 flies away means that the birds are not in the tree. So, there are 20 - 12 = <<20-12=8>>8 birds left in he tree.		
		5

Click the LLM check box to switch to free conversation mode. Users can enter questions they want to ask in the left chat box.

"Reset": Click this button to reset the dialog box.

"Stop current answer": Click this button to stop the current conversation.

- b. Knowledge base question and answer mode
 - i. Click the knowledge base Q&A mode, and users can conduct Q&A based on the knowledge base in this interface.
 - ii. Create new database.

	LLM O Knowledge Base Q&A
	Configure Knowledge Base
	Please Select a Knowledge Base to Load
Ť I	Create New Knowledge Base
J	Enter New Knowledge Base Name
	test
	Add to Knowledge Base
	Number of Matches
	Refresh knowledge base list
	Reset



- iii. Create a new knowledge base " in " Please select the knowledge base to load ".
- iv. exist Enter the name of the knowledge base to be created in "Enter new knowledge base name".
- v. Click " Add to knowledge base" button to complete the creation of the knowledge base.
- c. upload files
 - I. Currently supports documents in docx, pdf, txt, and md formats.
 - II. Drag the file to be uploaded to " Upload file " box.

Please Select Mode of Use			
C LLM Knowledge Base Q&A			
Configure Knowledge Base	•		
Please Select a Knowledge Base to Load			
admin/test	•		
Add files to Knowledge Base			
Upload File/上傳文件			
CMOS - Wikipedia.pdf 963.5 KB	Download		
Upload Files and Load Knowledg	e Base		
Knowledge Base File list (Total Files: 0)			
Number of Matches	1		
•			
Refresh knowledge base lis			
Reset			

- III. Click the " Upload file and load knowledge base " button to upload the document.
- IV. After the upload is complete, in the "Knowledge Base Document List" will display the names of the documents that have been uploaded and the number of documents currently in the knowledge base.

Welcome to Aneron Knowledge Base Chatt Please switch modes on the right. Currently supports direct conversation with LLM model or Q&A based on local knowledge base. For Q&A mode, after selecting a knowledge base, you can start Q&A.	Please Select Mode of Use LLM O Knowledge Base Q&A
Device added.	Configure Knowledge Base Please Select a Knowledge Base to Load
There are 20 birds in the tree. 12 files way, How many birds left in the tree	admin/test
There are 20 birds in the tree. 12 flies away means that the birds are not in the tree. So, there are 20 - 12 = <<20-12=8>8 birds left in the tree.	Add files to Knowledge Base Upload File/上時文件
New knowledge base"test", has been added, it will be stored after the files are uploaded and loaded successfully. Please complete file uploading before starting the conversation.	Drop File Here - or -
Files CMOS - Wikipedia.pdf have been uploaded to the knowledge base and are now loaded. Please start asking questions.	Click to Upload
	Upload Files and Load Knowledge Base
	Knowledge Base File list (Total Files: 1) CMOS - Wikipedia.pdf
se enter your question and press enter to submit	



Hint:

- Please keep the file name specifications in the uploaded file name and do not include special characters, such as (), \$, {}, etc.
- Normally, it takes about 15 seconds to upload a file of size 25kb, just for reference. Upload speed will be affected by file size, type, format and current network environment.
- d. Question and answer based on knowledge base.
 - i. On the right side of the UI, there is a dropdown menu labeled "Please select the knowledge base to load" that allows you to select the desired database.
 - ii. Type your question in the dialog box (shown in below figure) on the left and click Enter to start the conversation.

Please enter your question and press enter to submit	

e. Download and view knowledge base files.

evice added.	Please Select Mode of Use
	C LLM O Knowledge Base Q&A
here are 20 birds in the tree. 12 flies way, How many birds left in the tree	
	Configure Knowledge Base
here are 20 birds in the tree. 12 flies away means that the birds are not in the tree. So, there are 20 - 12 = <<20-12=8>>8 birds left	Please Select a Knowledge Base to Load
n the tree.	admin/test
lew knowledge base"test", has been added, it will be stored after the files are uploaded and loaded successfully. Please complete	Add files to Knowledge Base
le uploading before starting the conversation.	Upload File/上傳文件
iles CMOS - Wikipedia.pdf have been uploaded to the knowledge base and are now loaded. Please start asking questions.	
	Drop File Here
that is the cmos-transistor?	- or -
	Click to Upload
MOS stands for Complementary Metal-Oxide-Semiconductor. It is a type of transistor used in digital electronics. CMOS	
ransistors are commonly used in integrated circuits, microprocessors, and other digital devices. They are known for their low	
ransistors to amplify and switch electronic signals. The N-channel transistor is made of a P-type semiconductor material, while	
he P-channel transistor is made of an N-type semiconductor material. The two transistors are connected in a complementary	Upload Files and Load Knowledge Base
ray to form a CMOS inverter, which is the basic building block of CMOS circuits. CMOS technology is widely used in digital	Knowledge Base File list (Total Files: 1)
Source [1] 10.200.210.181/content/EN/admin/test/test/original/CMOS - Wikipedia.pdf	CMOS - Wikipedia.pdf

- i. In the generated answer, you can view the source information and documents.
- ii. Open a new browser tab, copy the file path mentioned in the KNEO300's answer regarding the source, and paste the path into the tab.
- f. Other function introduction



Upload Files and Load Knowledge Base	
owledge Base File list (Total Files: 1) CMOS - Wikipedia.pdf	
iumber of Matches	1
Refresh knowledge base list	
Reset	
Stop Current Answer	

- I. "Number of Matches": The number of entries found in the given database that are related to or matching the query text. Currently the default is 1
- II. "Refresh knowledge base list ": Click this button to refresh the directory of the current knowledge base.
- III. "Reset "Click this button to reset the dialog box.
- IV. Stop current answer: Click this button to interrupt the current conversation.
- g. Web caching
 - If you have logged in as admin before, after the device firmware is updated, when you log in with the same interface, due to the existence of web page cache, as shown in the figure below, you need to click the mailbox->log out-> and then log in again.





3.2 Others

If the network you are connected to is a non-public IP, that is, it is used within a local network (such as a home, school, or corporate network) and is used for communication within the internal network (such as printers, smart devices, etc.), These IP addresses are usually automatically assigned by the local network's router. Please follow these steps:

3.2.1. on port 3000. On your local computer, please set up a local proxy. ssh - NfL 127.0.0.1:8080:127.0.0.1:3000 linaro@192.168.200.102 (This is an example of device ip address, please change it to your actual machine ip address) linaro@192.168.10.150's password:

3.2.2. Set up the proxy on the local machine, taking chrome as an example.a. Go to settings.



÷ -	O Chrome chrome://settings/system		
0	Settings	Q Search settings	
•	You and Google	System	
Ê	Autofill and passwords		
۲	Privacy and security	Use naroware acceleration when available	
Ø	Performance	Open your computer's proxy settings	Ø
۲	Appearance		
Q	Search engine		
	Default browser		
Ċ	On startup		
	Languages		
<u>*</u>	Downloads		
Ť	Accessibility		
٩	System		
Ð	Reset settings		
*	Extensions 🛛		
0	About Chrome		

3.2.3. Find the system in the left column, click Open your computer's proxy settings, and then set the SOCKs proxy server (127.0.0.1) and port (8080)



0 😐	< > Wi-Fi	
2 Search	🛜 Wi-Fi	
HUAZHU-JiHotel	Automatic proxy configuration	
Connected	Web proxy (HTTP)	
DNS	Secure web proxy (HTTPS)	
WINS 802.1X	SOCKS proxy	
Proxies	Server	127.0.0.1
Hardware	Port	8080
	Proxy server requires password Your credentials may be sent unencrypted.	<u>6</u> °
	Exclude simple hostnames	
	Forget This Network	Cancel OK
Control Center	nopasswd	A 3
🔄 Siri & Spotlight		
Privacy & Security	Topway-0422	ê 🦷

b. In your local computer browser, enter: <u>http://localhost:8080/</u>

4. System update

4.1 Update chatbot software

Please follow the steps below to update the chatbot software, taking firmware install_20240103 as an example.

Download the update package (zip compressed format) install_20240103.zip from Kneron development center <u>Developers | Kneron – Full Stack Edge AI</u>, you need to apply the account first, then provide the email address that you registered to Kneron sales team, or the contact window that you purchased, to get the authority of the web access. Bellowing figure shows the snapshot of the Kneron development center that you can download from.

Note: The official website will also include an update guide. The update process may differ in the future, so please refer to the documentation on the official website for the most accurate information.



Kneron AI chat robot

Document name	Version	Latest modified	EIP No.	
C KNEO300				Open folder

• Copy the update package to the device. The recommended storage address is /home/linaro/ Please refer to 2.5 for how to upload files.

• Unzip the update package and run the command "unzip install_20240103.zip"



• Enter the update package directory and run the command "chmod +x install.sh" to set the running permissions



• Execute the installation script and run the command "./install.sh". After the installation is completed, the device will automatically restart.

🚰 linaro@chatrobot: ~/inst	il_20240103
linaro@chatrobot:~/ir	stall 20240103\$./install.sh
Archive: kneron chat	bot prod.zip
inflating: /data/kr	eron_chatbot_prod/README.md
inflating: /data/km	eron chatbot prod/README Kneron APIs.md
inflating: /data/kr	eron_chatbot_prod/README_internal.md
inflating: /data/kr	eron chatbot prod/Release notes.md
inflating: /data/kr	leron_chatbot_prod/envsetting.sh
inflating: /data/kr	eron chatbot prod/install.sh
inflating: /data/km	eron_chatbot_prod/kneron_doc_chat/LICENSE
extracting: /data/km	eron chatbot prod/kneron doc chat/kneron models/chat/README.md
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/chat/tokenizer.model
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en/tokenizer.model
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/config.json
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/config_sentence_transformers.json
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/modules.json
extracting: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/sentence_bert_config.json
inflating: /data/kr	eron chatbot prod/kneron doc chat/kneron models/npu models/small/en tokenizer/special tokens map.json
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/tokenizer.json
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/tokenizer_config.json
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/en_tokenizer/vocab.txt
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/zh/tokenizer.model
inflating: /data/kr	eron_chatbot_prod/kneron_doc_chat/kneron_models/npu_models/small/zh_tokenizer/bge-small-zh-v1.5/README.md
inflating: /data/km	eron chatbot prod/kneron doc chat/kneron models/npu models/small/zh tokenizer/bge-small-zh-vl.5/config.json

• Re-run the chatbot software and confirm whether the update is successful by checking the software version (shown in the red box)





4.2 Update model

The steps for updating the model are the same as updating the Chatbot. The only difference is the download address of the model update package (please contact the supplier or Kneron technical support to obtain the download address).

4.3 Update Firmware

Please contact the supplier or Kneron technical support for firmware update

5.FAQs

O known by