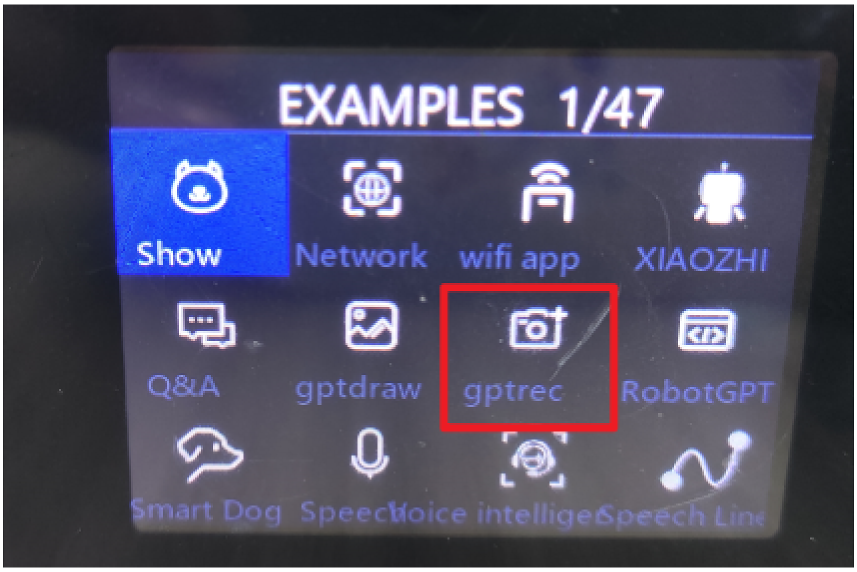
Picture to Text

Function Introduction

This case is about how to start the big program. You need to complete the configuration of the API-KEY related to the big model before you can use it normally.
This function is to describe the on-site picture.
The model platform used in the English version is: OpenRouter calls the free interface of meta
This function needs to be connected to the Internet to work properly

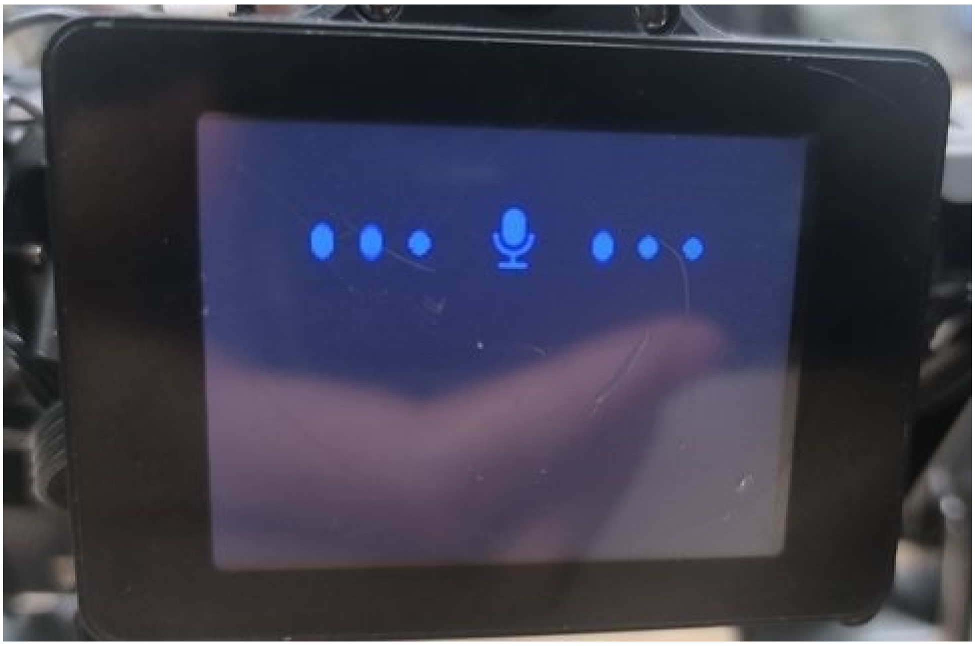
Function Experience

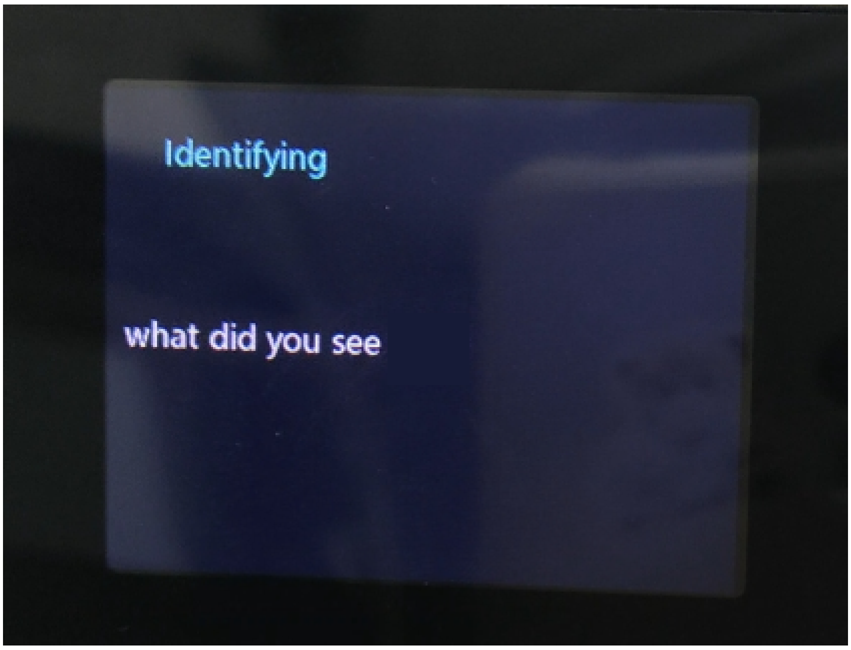
1. Turn on the robot dog first, press the button on the upper right of the "dog head" to enter the sample mode, and then select the function of image to text.



2. After entering the function of image to text, wake up by voice first, "lulu".

3. When you hear a ding, you can say the question you want to ask.

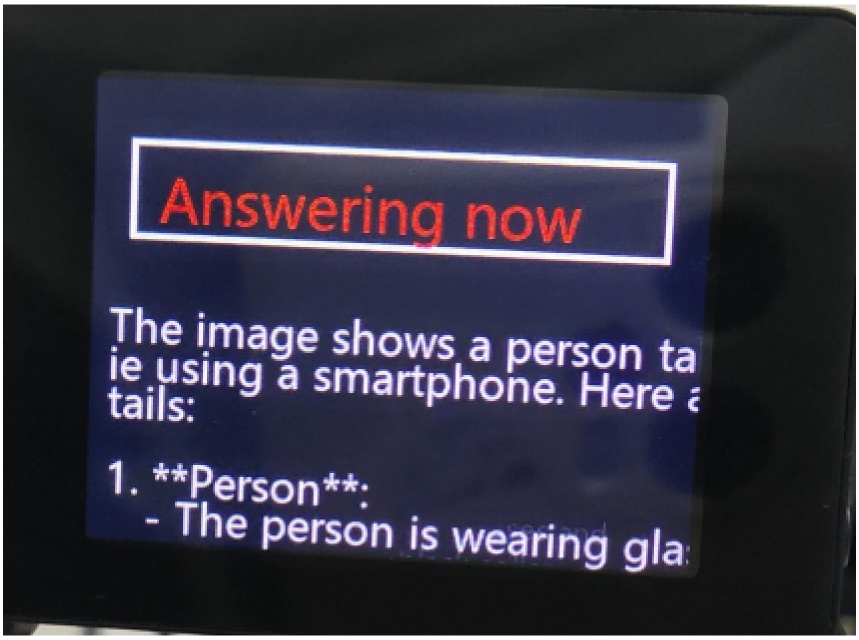




Identifying

what did you see

4. The robot dog will record a picture of the scene based on the question and then answer.



Answering now

The image shows a person ta
ie using a smartphone. Here a
tails:

1. **Person**:
 - The person is wearing gla

Program source code

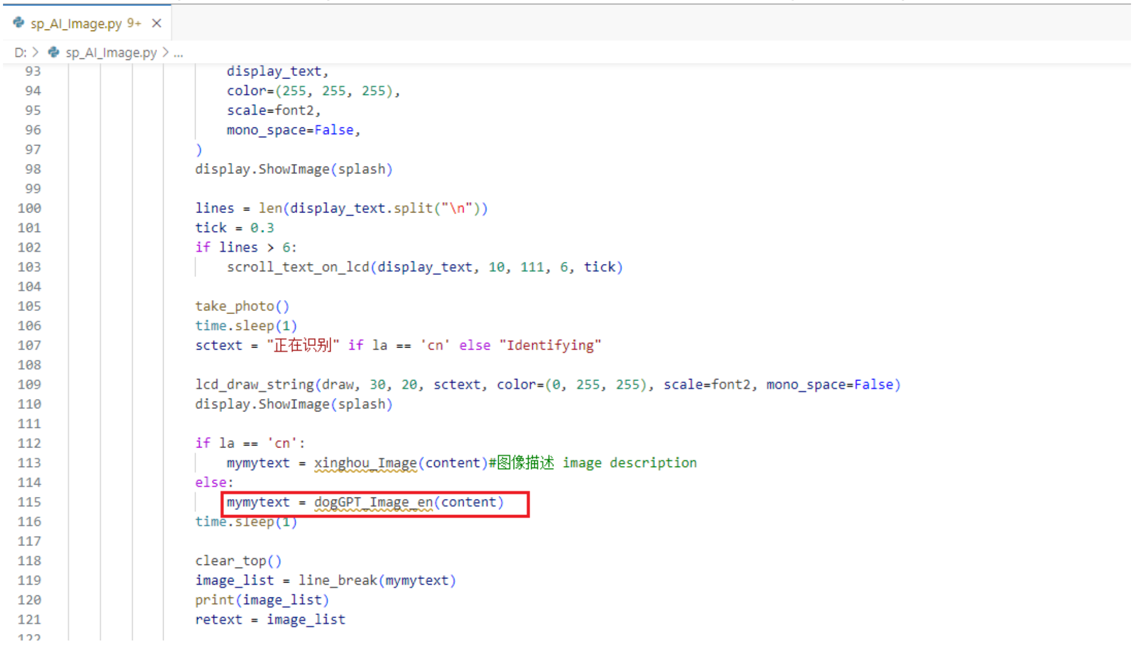
1. First, log in to the robot dog system through VNC
2. Then enter the terminal

```
cd /home/pi/RaspberryPi-CM4-main/demos/pic_comprehension/
tree
```

3. Directory structure description
 ├── anspeech_imageswer.mp3 #Reply audio
 ├── audio.py #Recording file
 ├── auto_platform.py #Recording file dependency environment
 ├── language_recognize.py #Speech recognition
 ├── libnyumaya.py #Speech wake-up
 ├── rec.jpg #Record the picture of the scene
 ├── sp_AI_Image.py #Main program file
 ├── xinghou_ImageAPI.py #Graphics file interface
 └── xinghou_tts.py #Speech synthesis file interface
4. If you want to replace the model interface you want

- You can first use the python version of the platform corresponding to the interface program, and fill in the necessary information according to the platform interface and instructions.
- Then encapsulate the executable file into a function. You can refer to the method of "**sp_AI_Image.py**" and put it in the path /home/pi/RaspberryPi-CM4-main/demos/pic_comprehension. For example, the added file name is "mychatgpt.py"
- Open the file **sp_AI_Image.py**, and replace from xinghou_ImageAPI import * with the newly added from mychatgpt import * in the header

- Then find this place and replace it with the API function interface you encapsulated.



```
                display_text,
                color=(255, 255, 255),
                scale=font2,
                mono_space=False,
            )
            display.ShowImage(splash)

            lines = len(display_text.split("\n"))
            tick = 0.3
            if lines > 6:
                scroll_text_on_lcd(display_text, 10, 111, 6, tick)

            take_photo()
            time.sleep(1)
            sctext = "正在识别" if la == 'cn' else "Identifying"

            lcd_draw_string(draw, 30, 20, sctext, color=(0, 255, 255), scale=font2, mono_space=False)
            display.ShowImage(splash)

            if la == 'cn':
                mymytext = xinghou_Image(content)#图像描述 image description
            else:
                mymytext = dogGPT_Image_en(content)
            time.sleep(1)

            clear_top()
            image_list = line_break(mymytext)
            print(image_list)
            retext = image_list
```

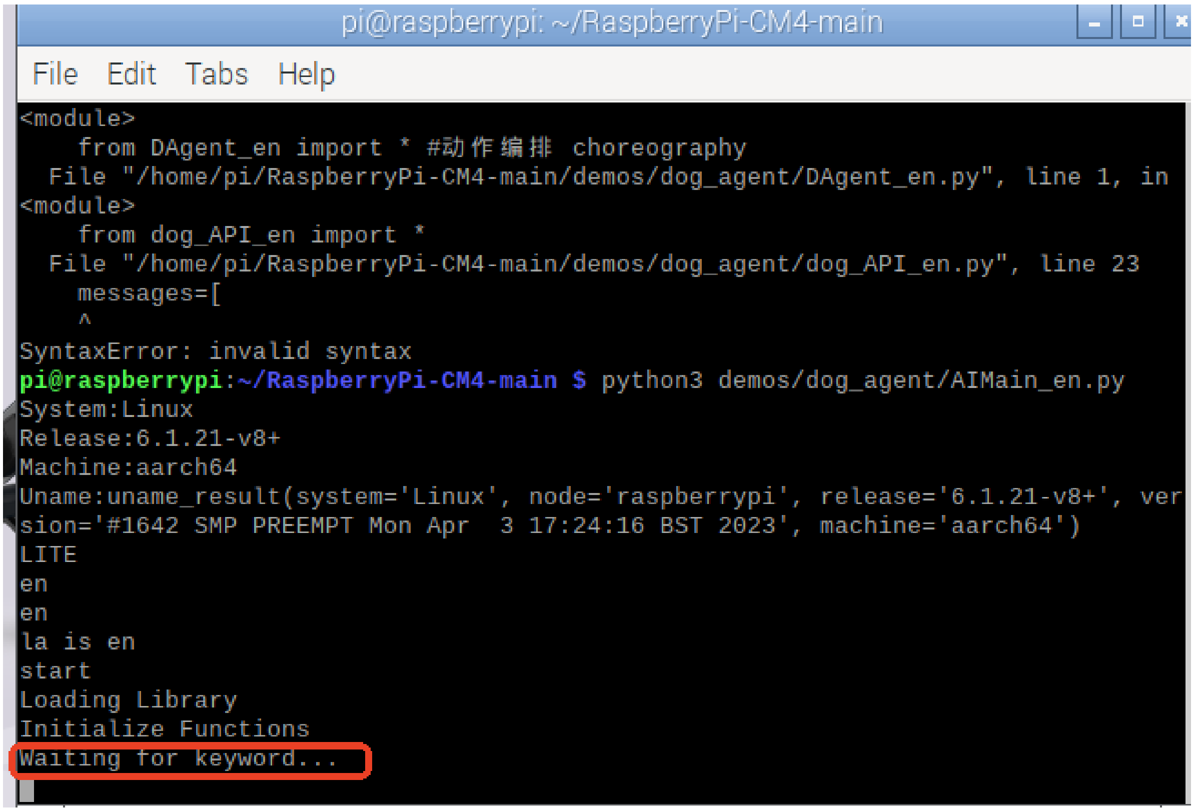
- Then restart the car and enter this function again, you can run the model platform you replaced. If it cannot run, it means there is an error. You need to check the syntax and logic of the newly added file yourself.

How to run this case in the terminal

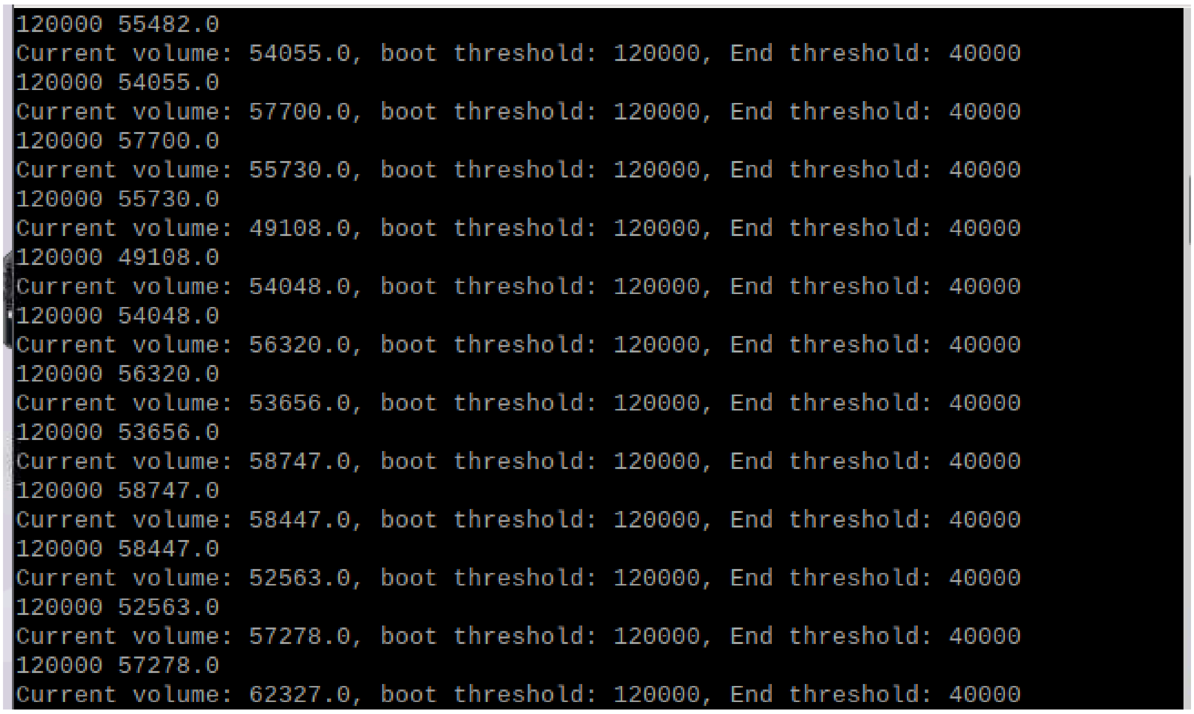
1. End the big program first to prevent the screen from being distorted. For how to end it, please go to the tutorial of ending the big program in Chapter 1. I will not describe it here.
2. Enter the following command in the terminal

```
cd /home/pi/RaspberryPi-CM4-main
sudo python3 demos/pic_comprehension/sp_AI_Image.py
```

3. When waiting for keyword appears, wake up the robot dog with "lu lu". After a few seconds, a scrolling prompt "current volume, boot threshold, end threshold" will appear. You can then input voice to the robot dog.



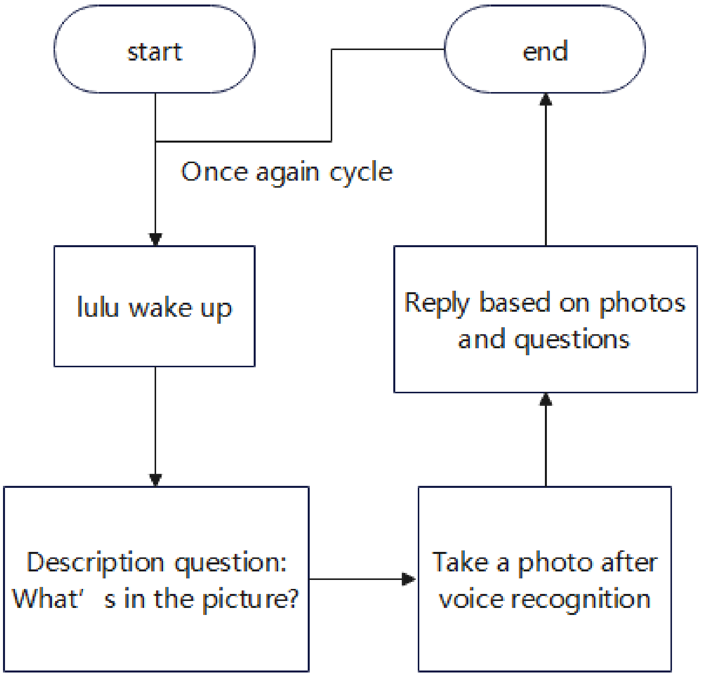
```
<module>
    from DAgent_en import * #动作编排 choreography
  File "/home/pi/RaspberryPi-CM4-main/demos/dog_agent/DAgent_en.py", line 1, in
<module>
    from dog_API_en import *
  File "/home/pi/RaspberryPi-CM4-main/demos/dog_agent/dog_API_en.py", line 23
    messages=[
    ^
SyntaxError: invalid syntax
pi@raspberrypi:~/RaspberryPi-CM4-main $ python3 demos/dog_agent/AIMain_en.py
System:Linux
Release:6.1.21-v8+
Machine:aarch64
Uname:uname_result(system='Linux', node='raspberrypi', release='6.1.21-v8+', ver
sion='#1642 SMP PREEMPT Mon Apr  3 17:24:16 BST 2023', machine='aarch64')
LITE
en
en
la is en
start
Loading Library
Initialize Functions
Waiting for keyword...
```



```
120000 55482.0
Current volume: 54055.0, boot threshold: 120000, End threshold: 40000
120000 54055.0
Current volume: 57700.0, boot threshold: 120000, End threshold: 40000
120000 57700.0
Current volume: 55730.0, boot threshold: 120000, End threshold: 40000
120000 55730.0
Current volume: 49108.0, boot threshold: 120000, End threshold: 40000
120000 49108.0
Current volume: 54048.0, boot threshold: 120000, End threshold: 40000
120000 54048.0
Current volume: 56320.0, boot threshold: 120000, End threshold: 40000
120000 56320.0
Current volume: 53656.0, boot threshold: 120000, End threshold: 40000
120000 53656.0
Current volume: 58747.0, boot threshold: 120000, End threshold: 40000
120000 58747.0
Current volume: 58447.0, boot threshold: 120000, End threshold: 40000
120000 58447.0
Current volume: 52563.0, boot threshold: 120000, End threshold: 40000
120000 52563.0
Current volume: 57278.0, boot threshold: 120000, End threshold: 40000
120000 57278.0
Current volume: 62327.0, boot threshold: 120000, End threshold: 40000
```

Functional principle

The specific flow chart is as follows:


start
end
Once again cycle
lulu wake up
Reply based on photos and questions
Description question: What' s in the picture?
Take a photo after voice recognition