Creating Images with Text

Experiment Objective

Understand and master a basic function of the AI model. Interact with the smart car through verbal dialogue, enabling the smart car to generate and display random images based on semantic meaning.

Experiment Steps

1. Observe the IP address of the OLED screen and log in to the remote desktop via VNC.
2. According to the prerequisite configuration tutorial, both the Chinese and English versions require the Tongyi Qianwen key.
3. Open a new terminal and run the following command:

```
cd /home/pi/project_demo/09.AI_Big_Model/
```

```
#Startup command for Chinese version
python3 Image_create/Image_main.py
```

```
#Startup command for English version
python3 Image_create/Image_main_en.py
```

4. Wake up the car using the wake-up phrase "Hi, Yahboom" (for international users).
5. After waking up successfully, the car will respond with a honking sound. After waiting for about half a second, you can then speak the desired image you want the car to generate.
6. After the robot recognizes your voice, wait a few seconds for the generated image to appear on the terminal interface.
7. This completes the conversation. To continue, repeat steps 4-6.

Experimental Results

1. Wait for the robot to wake up

```
pi@yahboom:~/project_demo/09.AI_Big_Model $ python3 Image_create/Image_main.py
serial /dev/myspeech open
Waiting for keyword...
```

2. After the robot wakes up successfully, a green box with an image will appear. You can then speak the image you want to generate.

```
JackShmReadWritePtr::~JackShmReadWritePtr - Init not done for -1, skipping unlock
Current volume: 111279.0, boot threshold: 3000, End threshold: 1500
start recording
3000 111279.0
Current volume: 111884.0, boot threshold: 3000, End threshold: 1500
3000 111884.0
Current volume: 80749.0, boot threshold: 3000, End threshold: 1500
3000 80749.0
Current volume: 65331.0, boot threshold: 3000, End threshold: 1500
3000 65331.0
Current volume: 68820.0, boot threshold: 3000, End threshold: 1500
```

3. After a while, the terminal will display the image generated by the large model.

```
###speak iat closed ###
Q: A picture of a little girl.
```

Experiment main source code analysis

```
def main():
    while True:
        if detect_keyword():

            start_recording()
```

```
            if TTS_IAT_Tongyi:
                content = rec_wav_music_Tongyi()
            else:
                content = rec_wav_music()

            if content != "":

                print("Q:"+content)
                tongyi_image_creat(content)

            if content == 0:
               break

        time.sleep(0.1)
```

detect_keyword: Wake-up function handler
start_recording: Recording function handler
tongyi_image_creat: Visual model image generation interface

Chinese version-specific options:
rec_wav_music_Tongyi: Tongyi Qianwen voice recognition
rec_wav_music: iFlytek Spark voice recognition
You can choose between two voice recognition options. You can enable or disable them in the **API_KEY.py** file. When TTS_IAT_Tongyi=True, either Tongyi Qianwen or iFlytek Spark voice recognition is enabled.

English Version
Speech recognition and synthesis use the iFlytek Spark API by default. You don't need to specify the iFlytek Spark API in API_KEY.py; simply fill in the **openAI_KEY** key.

Modifying the recording duration, start threshold, and end threshold

1. In the terminal, enter:

```
cd /home/pi/project_demo/09.AI_Big_Model/
nano audio.py
```

2. Find the source code shown below.

```
quitmark = 0
automark = True
def start_recording(timel = 3,save_file=SAVE_FILE):
    global automark,quitmark
    start_threshold = 3000 #30000
    end_threshold = 1500 #20000
    endlast = 15
    max_record_time = 5
```

- start_threshold: The threshold for starting recording when sound is detected (reduced to 5000 in quiet environments and increased to 150000+ in noisy environments).
- end_threshold: The threshold for stopping recording when sound is detected. A recommended value is 30-50% of start_threshold.
- endlast: The number of times to stop recording. Here, 15 is used. For example, Recording will automatically terminate if 15 consecutive sound values meet the stop threshold.
- max_record_time: Recording duration, set to 5 here.
  Note: start_threshold > end_threshold. This is a required rule, and its value can be determined by the environment.

3. Directory Structure of the Main Files for this Experiment

```
├── creat_img_tongyi.py #Tongyi Qianwen Image Generation API
├── Image_main_en.py #English Version Main Program Entry
├── Image_main.py #Chinese Version Main Program Entry
├── tongyi_speak_iat.py #Tongyi Qianwen Speech Recognition
└── xinghou_speak_iat.py #iFlytek Spark Speech Recognition
```

Overall Flowchart