

# Semantic understanding of instructions following

## 1. Course Content

After running the large model program, users can interact with the robot through voice conversation. User voice commands are first converted into text by the speech recognition large model. The text is then used by the large model and visual multimodality to accurately understand the user's commands and voice. Finally, the robot performs the specified actions according to the user's instructions and responds to the user.

## 2. Preparation

### 2.1 Content Description

This course uses the Jetson Orin NX as an example. For Raspberry Pi and Jetson Nano boards, you need to open a terminal on the host computer and enter the command to enter the Docker container. Once inside the Docker container, enter the commands mentioned in this course in the terminal. For instructions on entering the Docker container from the host computer, refer to the **[Configuration and Operation Guide] -- [Enter the Docker (Jetson Nano and Raspberry Pi 5 users see here)]** section of this product tutorial. For Orin and NX boards, simply open a terminal and enter the commands mentioned in this course.

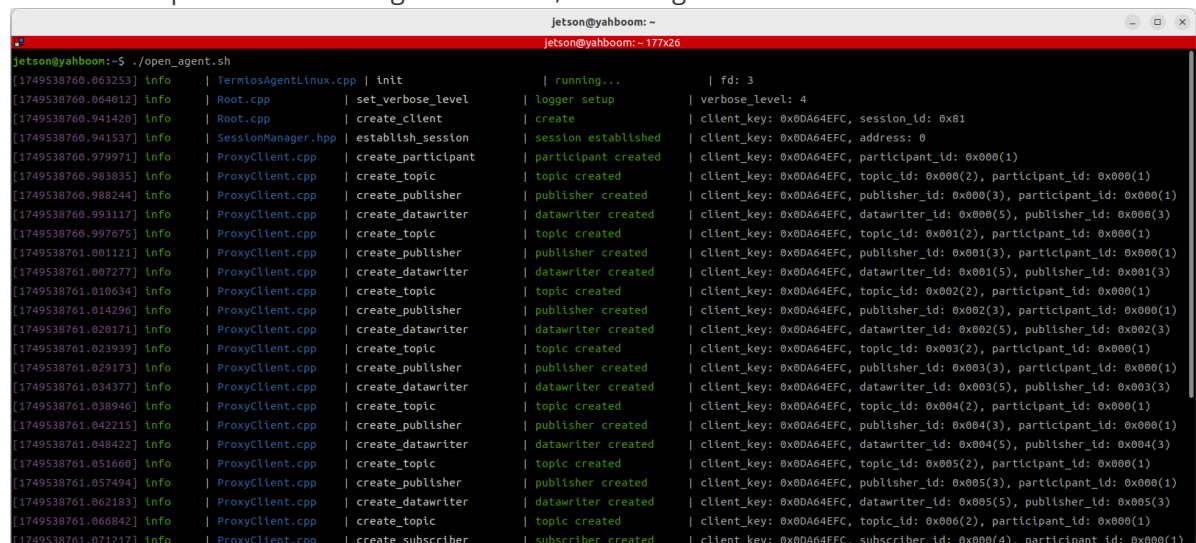
### 2.2 Start the agent

**Note: To test all cases, you must first start the docker agent. If it has already been started, you do not need to start it again**

Enter the command in the vehicle terminal:

```
sh start_agent.sh
```

The terminal prints the following information, indicating that the connection is successful



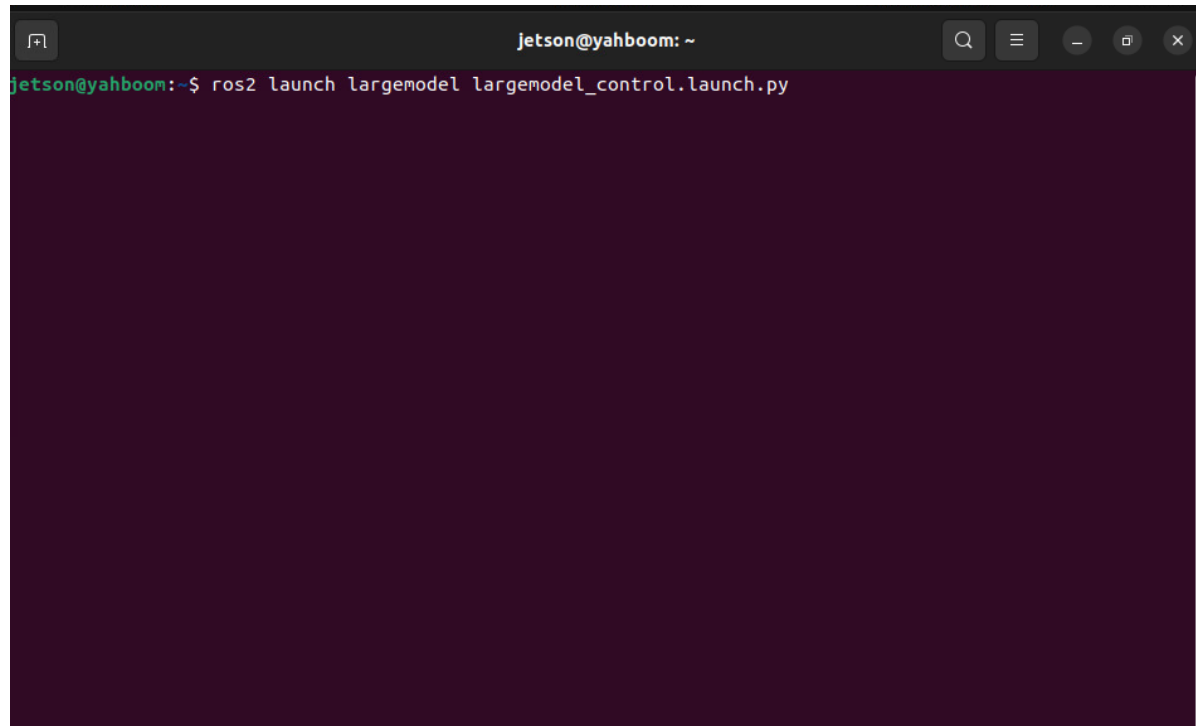
```
jetson@yahboom: ~  
jetson@yahboom: ~ 177x26  
jetson@yahboom:~$ ./open_agent.sh  
[1749538760.063253] Info | TermosAgentLinux.cpp | init | running... | fd: 3  
[1749538760.064012] Info | Root.cpp | set_verbose_level | verbose_level: 4  
[1749538760.941420] Info | Root.cpp | create_client | create | client_key: 0x0DA64EFC, session_id: 0x81  
[1749538760.941537] Info | SessionManager.hpp | establish_session | session established | client_key: 0x0DA64EFC, address: 0  
[1749538760.979971] Info | ProxyClient.cpp | create_participant | participant created | client_key: 0x0DA64EFC, participant_id: 0x000(1)  
[1749538760.983835] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x000(2), participant_id: 0x000(1)  
[1749538760.988244] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x000(3), participant_id: 0x000(1)  
[1749538760.993117] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x000(5), publisher_id: 0x000(3)  
[1749538760.997675] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x001(2), participant_id: 0x000(1)  
[1749538761.001121] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x001(3), participant_id: 0x000(1)  
[1749538761.007277] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x001(5), publisher_id: 0x001(3)  
[1749538761.010634] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x002(2), participant_id: 0x000(1)  
[1749538761.014296] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x002(3), participant_id: 0x000(1)  
[1749538761.020171] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x002(5), publisher_id: 0x002(3)  
[1749538761.023939] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x003(2), participant_id: 0x000(1)  
[1749538761.029173] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x003(3), participant_id: 0x000(1)  
[1749538761.034377] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x003(5), publisher_id: 0x003(3)  
[1749538761.038946] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x004(2), participant_id: 0x000(1)  
[1749538761.042215] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x004(3), participant_id: 0x000(1)  
[1749538761.048422] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x004(5), publisher_id: 0x004(3)  
[1749538761.051668] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x005(2), participant_id: 0x000(1)  
[1749538761.057494] Info | ProxyClient.cpp | create_publisher | publisher created | client_key: 0x0DA64EFC, publisher_id: 0x005(3), participant_id: 0x000(1)  
[1749538761.062183] Info | ProxyClient.cpp | create_datawriter | datawriter created | client_key: 0x0DA64EFC, datawriter_id: 0x005(5), publisher_id: 0x005(3)  
[1749538761.066842] Info | ProxyClient.cpp | create_topic | topic created | client_key: 0x0DA64EFC, topic_id: 0x006(2), participant_id: 0x000(1)  
[1749538761.071217] Info | ProxyClient.cpp | create_subscriber | subscriber created | client_key: 0x0DA64EFC, subscriber_id: 0x000(4), participant_id: 0x000(1)
```

## 3. Run the case

## 3.1 Start the program

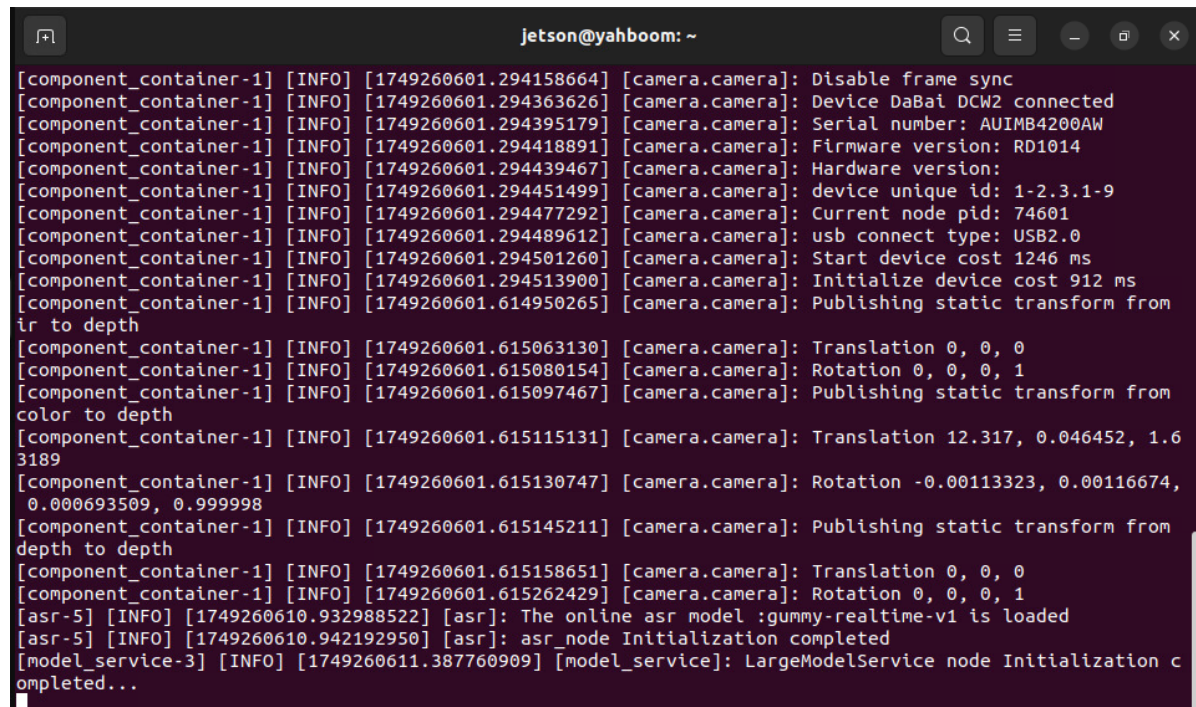
Open the terminal on the vehicle and enter the command:

```
ros2 launch largemodel largemodel_control.launch.py
```



A terminal window titled 'jetson@yahboom: ~' with a search icon, menu icon, and window controls. The command 'ros2 launch largemodel largemodel\_control.launch.py' has been entered and is being executed. The terminal background is dark purple.

After initialization is complete, the following content will be displayed



A terminal window titled 'jetson@yahboom: ~' showing the output of the launch command. The output consists of multiple lines of log messages from different components. The messages include information about camera initialization, such as disabling frame sync, connecting the DaBai DCW2 camera, and publishing static transforms. It also shows the ASR model 'gummy-realtime-v1' being loaded and initialized. The terminal background is dark purple.

```
[component_container-1] [INFO] [1749260601.294158664] [camera.camera]: Disable frame sync
[component_container-1] [INFO] [1749260601.294363626] [camera.camera]: Device DaBai DCW2 connected
[component_container-1] [INFO] [1749260601.294395179] [camera.camera]: Serial number: AUIMB4200AW
[component_container-1] [INFO] [1749260601.294418891] [camera.camera]: Firmware version: RD1014
[component_container-1] [INFO] [1749260601.294439467] [camera.camera]: Hardware version:
[component_container-1] [INFO] [1749260601.294451499] [camera.camera]: device unique id: 1-2.3.1-9
[component_container-1] [INFO] [1749260601.294477292] [camera.camera]: Current node pid: 74601
[component_container-1] [INFO] [1749260601.294489612] [camera.camera]: usb connect type: USB2.0
[component_container-1] [INFO] [1749260601.294501260] [camera.camera]: Start device cost 1246 ms
[component_container-1] [INFO] [1749260601.294513900] [camera.camera]: Initialize device cost 912 ms
[component_container-1] [INFO] [1749260601.614950265] [camera.camera]: Publishing static transform from
ir to depth
[component_container-1] [INFO] [1749260601.615063130] [camera.camera]: Translation 0, 0, 0
[component_container-1] [INFO] [1749260601.615080154] [camera.camera]: Rotation 0, 0, 0, 1
[component_container-1] [INFO] [1749260601.615097467] [camera.camera]: Publishing static transform from
color to depth
[component_container-1] [INFO] [1749260601.615115131] [camera.camera]: Translation 12.317, 0.046452, 1.6
3189
[component_container-1] [INFO] [1749260601.615130747] [camera.camera]: Rotation -0.00113323, 0.00116674,
0.000693509, 0.999998
[component_container-1] [INFO] [1749260601.615145211] [camera.camera]: Publishing static transform from
depth to depth
[component_container-1] [INFO] [1749260601.615158651] [camera.camera]: Translation 0, 0, 0
[component_container-1] [INFO] [1749260601.615262429] [camera.camera]: Rotation 0, 0, 0, 1
[asr-5] [INFO] [1749260610.932988522] [asr]: The online asr model :gummy-realtime-v1 is loaded
[asr-5] [INFO] [1749260610.942192950] [asr]: asr_node Initialization completed
[model_service-3] [INFO] [1749260611.387760909] [model_service]: LargeModelService node Initialization c
ompleted...
```

## 3.2 Test case

Here are two reference test cases. Users can compile their own test instructions

- Please move forward quickly for 1 meter, then slowly back away 0.5 meters like a turtle. Then, turn left 30 degrees, turn right 90 degrees, then move horizontally 0.5 meters to the left, and then move horizontally 10 centimeters to the right.
- Please perform a dance first, then tell me a joke about a kitten and a puppy.

First, wake the robot using "Hi, yahboom." The robot responds: "I'm here, please." After the robot responds, the buzzer beeps briefly (beep-). The user can then speak. The robot will detect sound activity, printing 1 if there is sound activity and 0 if there is no sound activity. When the speech ends, it detects the end of the voice, and stops recording if there is silence for more than 1 second.

```
jseton@yahboom: ~  
[asr-5] [INFO] [1749262297.654733341] [asr]: 1  
[asr-5] [INFO] [1749262297.753996196] [asr]: 1  
[asr-5] [INFO] [1749262297.845296499] [asr]: 1  
[asr-5] [INFO] [1749262297.935830158] [asr]: 1  
[asr-5] [INFO] [1749262297.997093005] [asr]: 1  
[asr-5] [INFO] [1749262298.097002049] [asr]: 1  
[asr-5] [INFO] [1749262298.186356406] [asr]: 1  
[asr-5] [INFO] [1749262298.277009071] [asr]: 1  
[asr-5] [INFO] [1749262298.367562404] [asr]: 1  
[asr-5] [INFO] [1749262298.465003189] [asr]: 1  
[asr-5] [INFO] [1749262298.554886744] [asr]: 1  
[asr-5] [INFO] [1749262298.645609043] [asr]: 1  
[asr-5] [INFO] [1749262298.736733464] [asr]: 1  
[asr-5] [INFO] [1749262298.827381264] [asr]: 1  
[asr-5] [INFO] [1749262298.918001576] [asr]: 1  
[asr-5] [INFO] [1749262299.008533007] [asr]: 1  
[asr-5] [INFO] [1749262299.098226614] [asr]: 1  
[asr-5] [INFO] [1749262299.190759036] [asr]: 0  
[asr-5] [INFO] [1749262299.280469046] [asr]: 1  
[asr-5] [INFO] [1749262299.371976416] [asr]: 1  
[asr-5] [INFO] [1749262299.461162572] [asr]: 1  
[asr-5] [INFO] [1749262299.551932322] [asr]: 1  
[asr-5] [INFO] [1749262299.641563578] [asr]: 1  
[asr-5] [INFO] [1749262299.733164071] [asr]: 1  
[asr-5] [INFO] [1749262299.824791477] [asr]: 1  
[asr-5] [INFO] [1749262299.915240322] [asr]: 1  
[asr-5] [INFO] [1749262300.005843444] [asr]: 0  
[asr-5] [INFO] [1749262300.065963947] [asr]: 0  
[asr-5] [INFO] [1749262300.156661434] [asr]: 0
```

```

[ast-5] [INFO] [1755156719.099792058] [ast]: -
[ast-5] [INFO] [1755156719.760003194] [ast]: -
[ast-5] [INFO] [1755156719.820762066] [ast]: -
[ast-5] [INFO] [1755156719.881593937] [ast]: -
[ast-5] [INFO] [1755156719.941811730] [ast]: -
[ast-5] [INFO] [1755156720.000970695] [ast]: -
[ast-5] [INFO] [1755156720.061617049] [ast]: -
[ast-5] [INFO] [1755156720.093336087] [ast]: -
[ast-5] [INFO] [1755156720.152681918] [ast]: -
[ast-5] [INFO] [1755156720.212754381] [ast]: -
[ast-5] [INFO] [1755156722.936148518] [ast]: Please move forward quickly for one meter.
[ast-5] [INFO] [1755156722.938242922] [ast]: @okay, let me think for a moment...
[model_service-3] [INFO] [1755156730.318148362] [model_service]: Decision making AI planning: Move forward quickly for one meter.
[model_service-3] [INFO] [1755156732.685838517] [model_service]: "action": ['set_cmdvel(0.8, 0, 0, 1.25)'], "response": "Okay, I'm moving forward o
quickly for one meter! Watch me go!
[action_service-4] [INFO] [1755156741.161007099] [action_service]: Published message: Robot feedback: Execute set_cmdvel(0.8,0.0,0.0,1.25) complet
ed
[model_service-3] [INFO] [1755156743.316941871] [model_service]: "action": ['finishtask()'], "response": "I've moved forward one meter quickly! Mis
sion accomplished!

```

- **Decision-making layer large model output:** 1. Quickly move forward 1 meter, slowly retreat 0.5 meters, turn left 30 degrees, turn right 90 degrees, move 0.5 meters to the left, and move 0.1 meters to the right.
- **Decision-making model output:** `action": ['set_cmdvel(0.5, 0, 0, 2)', 'set_cmdvel(-0.1, 0, 0, 5)', 'move_left(30, 1.5)', 'move_right(90, 1.5)', 'set_cmdvel(0, 0.5, 0, 1)', 'set_cmdvel(0, -0.1, 0, 1)'], "response": Okay, I'll start executing your instructions now. I'll move forward quickly, then slowly move back, and then spin around like I'm doing a complicated dance., action": ['finishtask()'], "response": I've completed all the actions. It feels like I've completed a wonderful dance performance! If you have other tasks you need my help with, feel free to let me know!`

The action list contains **finishtask()**, indicating that the execution layer model has determined that the robot has completed the user's command and entered the **waiting state**. At this point, you can wake Xiaoya up again to end the current task:

```
[asr-5] [INFO] [1755157804.85588514] [asr]: -
[asr-5] [INFO] [1755157804.916058818] [asr]: -
[asr-5] [INFO] [1755157804.977469353] [asr]: -
[asr-5] [INFO] [1755157805.036861951] [asr]: -
[asr-5] [INFO] [1755157806.548471736] [asr]: Finish this task.
[asr-5] [INFO] [1755157806.549680063] [asr]: @ okay, let me think for a moment...
[model_service-3] [INFO] [1755157809.584289821] [model_service]: Decision making AI planning:Could you please provide more details about the task you want me to complete?
[model_service-3] [INFO] [1755157812.603812689] [model_service]: "action": ['finishtask()'], "response": I'm ready to help, but I need a bit more info on what task you'd like me to finish! Let me know the details and I'll get started right away.
```

### 3.2.2 Case 2

Similar to the test in Case 1, first wake the robot with "Hello Xiaoya". After the robot responds, the buzzer beeps briefly (beep). The user can then speak. After the speech is complete, the robot responds and dances according to the user's instructions.

```
[asr-5] [INFO] [1755157280.787749548] [asr]: -
[asr-5] [INFO] [1755157280.847145806] [asr]: -
[asr-5] [INFO] [1755157280.908246162] [asr]: -
[asr-5] [INFO] [1755157280.968440567] [asr]: -
[asr-5] [INFO] [1755157281.027254669] [asr]: -
[asr-5] [INFO] [1755157281.087533833] [asr]: -
[asr-5] [INFO] [1755157281.148294612] [asr]: -
[asr-5] [INFO] [1755157281.207741999] [asr]: -
[asr-5] [INFO] [1755157281.268495898] [asr]: -
[asr-5] [INFO] [1755157283.634476198] [asr]: A dance first then tell me a joke about a kitten and a puppy.
[asr-5] [INFO] [1755157283.635895931] [asr]: @ okay, let me think for a moment...
[model_service-3] [INFO] [1755157287.628281862] [model_service]: "action": ['dance()'], "response": Alright, let's get the party started! *dances with flair* Now, here's a sweet joke for you: Why did the kitten and the puppy sit together on the porch? Because they were both waiting for their 'paw-ty' to begin!
[action_service-4] [INFO] [1755157318.918941121] [action_service]: Published message: Robot feedback: Execute dance() completed
[model_service-3] [INFO] [1755157321.135486255] [model_service]: "action": ['finishtask()'], "response": Hope you enjoyed the dance and the cute joke! Kittens and puppies are just too adorable together!
```

## 4. Code Analysis

This lesson uses the basic programming framework for the basic AI embodied intelligence gameplay. For code analysis, refer to the [2. AI Large Model Basics - 4. Embodied Intelligence Gameplay Core Source Code Interpretation] section.

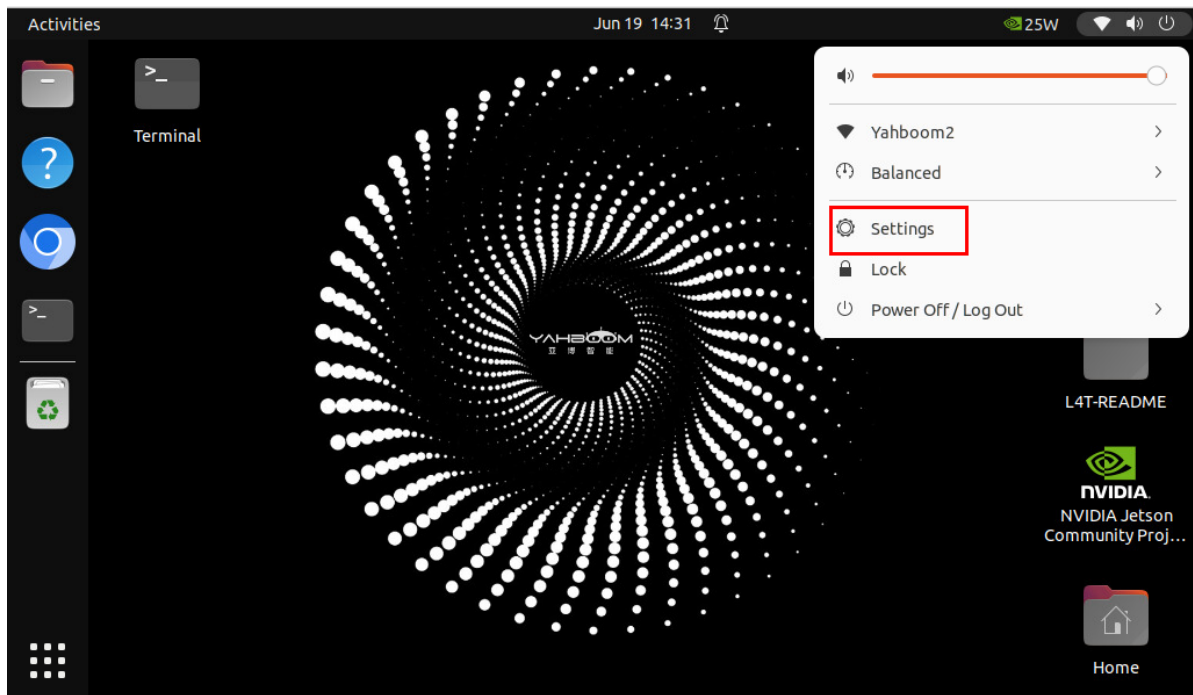
## 5. Common Problem Solutions

### 5.1 Microphone Recording Too Sensitive

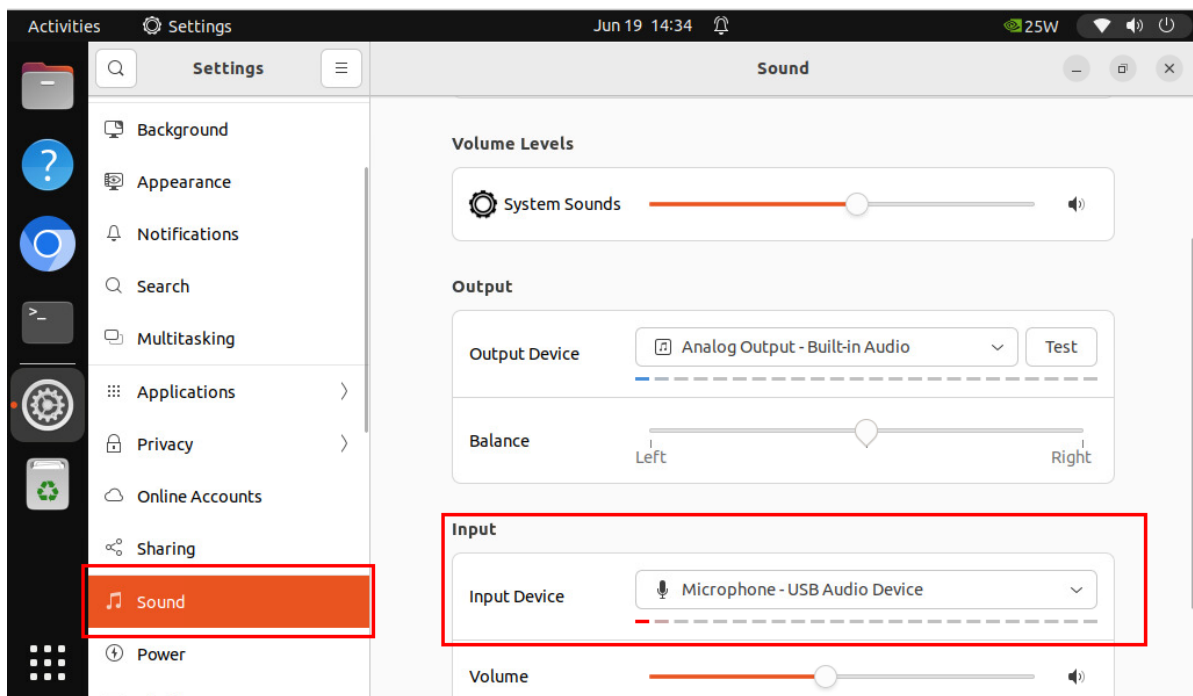
If you find that the VAD Voice Activity Detection indicator remains "1" after the speaking phase and you cannot stop recording, this indicates that the microphone is set too sensitively and there is constant voice activity. You can try reducing the microphone sensitivity.

First, connect to the robot car screen via VNC. Click the options bar in the upper right corner to find the Settings option.





Scroll down to the Settings list on the left and find the Sound option. On the Sound page, find the Input Device. Drag the Volume button below to adjust the sensitivity. Try adjusting it to a comfortable value while recording.



## 5.2 Microphone Recording Insensitivity

If the speaker is far away from the robot, the VAD voice activity detection function may not detect voice activity, causing the recording to end prematurely before the speaker finishes speaking. In this case, refer to the steps in **5.1 Microphone Recording Excessively Sensitive** to increase the microphone sensitivity.

### Note:

- If the microphone sensitivity is set too high, it may misinterpret ambient noise as speech activity.

## 5.3 Incomplete Speech Recognition

Different speech recognition models may have different recognition performance for the same audio. We recommend using the default Paraformer series model or the local SenseVoiceSmall model. (The local speech recognition model is currently only available on the Jetson Orin Nano and Jetson Orin NX series.) For instructions on switching speech models, refer to [2. AI Large Model Basics - 5. Configuring the AI Large Model].