

Model List

This endpoint is used to retrieve the list of model services currently installed on the device.

Model Installation and Viewing

1. Before use, please refer to the corresponding device's software update tutorial to add and update the M5Stack apt repository information and index.
 - [Module LLM Image & Software Upgrade Tutorial](#)
 - [LLM630 Compute Kit Image & Software Upgrade Tutorial](#)
2. View the list of available llm deb packages. Packages named in the `llm-model-name` format are model packages.

```
apt list | grep llm-model-
```

3. Use the `apt` command to install packages as needed, for example, install the `llm-model-qwen2.5-0.5b-p256-ax630c` package.
Note: models occupy significant space; it is recommended to install as needed.

```
apt install llm-model-qwen2.5-0.5b-p256-ax630c
```

Viewing Available Models

You can query the list of models currently available on the device from a PC via the OpenAI API. Before running the program, modify the IP part of the `base_url` below to the actual IP address of the device.

```
from openai import OpenAI
client = OpenAI(
    api_key="sk-",
    base_url="http://192.168.20.186:8000/v1"
)

client.models.list()
print(client.models.list())
```

Response Example

```
SyncPage[Model](data=[
Model(id='melotts_zh-cn', created=0, object='model', owned_by='user', permission=[], root=''),
Model(id='qwen2.5-0.5B-prefill-20e', created=0, object='model', owned_by='user', permission=[], root='')
Model(id='sherpa-ncnn-streaming-zipformer-20M-2023-02-17', created=0, object='model', owned_by='user',
Model(id='sherpa-ncnn-streaming-zipformer-zh-14M-2023-02-23', created=0, object='model', owned_by='user
Model(id='single_speaker_english_fast', created=0, object='model', owned_by='user', permission=[], root
Model(id='single_speaker_fast', created=0, object='model', owned_by='user', permission=[], root=''),
Model(id='qwen2.5-0.5B-p256-ax630c', created=0, object='model', owned_by='user', permission=[], root='')
],
object='list')
```