Text-to-Speech

Convert input text into an audio file through the API interface.

Preparation

Before running the example program, you need to install the corresponding model package on the device. For the model package installation tutorial, please refer to the Model List section. For detailed model descriptions, please refer to the Model Introduction section.

> **Tip**
>
> AI Pyramid provides a dedicated text-to-speech solution with voice cloning. For details, please refer to the CosyVoice section.

Before running this example program, please ensure that the following preparations have been completed on the LLM device:

1. Use the apt package manager to install the `llm-model-melotts-en-us` model package.

```
apt install llm-model-melotts-en-us
```

2. Install the `ffmpeg` tool.

```
apt install ffmpeg
```

3. After installation, restart the OpenAI service to make the new model take effect.

```
systemctl restart llm-openai-api
```

Example Program

On the PC side, use the OpenAI API to pass text information to implement the text-to-speech function. Before running the example program, modify the IP part of `base_url` below to the actual IP address of the device.

```python
from pathlib import Path
from openai import OpenAI

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

speech_file_path = Path(__file__).parent / "speech.mp3"
with client.audio.speech.with_streaming_response.create(
  model="melotts-en-us",
  voice="alloy",
  input="The quick brown fox jumped over the lazy dog."
) as response:
  response.stream_to_file(speech_file_path)
```

Request Parameters

Parameter Name	Type	Required	Example Value	Description
input	string	Yes	"Hello, welcome to the system"	Text content to generate audio from, with a maximum length of 1024 characters
model	string	Yes	melotts-zh-cn	Available TTS models include `melotts-ja-jp`, `melotts-zh-cn`, `melotts-en-us`, etc.
voice	-	No	-	The MeloTTS model does not support voice style selection
response_format	string	No	mp3	Audio output format, supports `mp3`, `opus`, `aac`, `flac`, `wav`, `pcm`, etc.
speed	number	No	1.0	Speech generation speed, range 0.25 ~ 2.0, default value is 1.0

Response Example

- The generated audio file data will be stored in the `speech_file_path` path defined in the example program.