

大模型语音控制软件在环实验

1. 实验目的

探索实时语音指令识别（ASR）与大模型控制的协同工作流程。思考与实时语音大模型控制的区别。

2. 实验要求

- 软件要求：Windows 10及以上版本；RflySim工具链^[1]。
- 硬件要求：笔记本/台式电脑1台^[2]。

3. 实验地址

例程目录：[\[安装目录\]\RflySimAPIs\6.RflySimExtCtrl\3.CustExps\e1.LLM_CtrlUAVExps\3.VoiceCtrlUAVSim](#)

- `Main_OpenAI_api.py`：主控制程序：OpenAI 调用、指令解析、UDP 包构建与发送
- `OpenAI_audio.py`：通用的实时语音识别模块：音频采集、WebSocket 交互、文本返回；调用模型ID：`volc.bigasr.sauc.duration`
- `LLM_UAV.slx` / `LLM_UAV_init.m`：MATLAB/Simulink 在环仿真模型（如已集成）
- `RflyUdpMavlinkRealSim.bat`：PX4 SITL 启动脚本

4. 实验内容或步骤

4.1 步骤1：获取大模型API密钥与语音识别服务鉴权

本实验需要首先在 `\文档\0gre\.env` 文件中配置大模型 API 密钥和语音识别服务的鉴权参数。

4.1.1 获取大模型 API 密钥

完整配置说明可参考实验

`[RflySim 安装目录]\RflySimAPIs\6.RflySimExtCtrl\3.CustExps\e1.LLM_CtrlUAVExps\1.CommandCtrlUAVSim`

。

进入火山引擎官网 <https://www.volcengine.com>，在顶部工具栏的「大模型」菜单中点击「豆包大模型」，并使用个人账号登录。在左侧导航栏中点击「体验中心」→「语言模型」，将模型切换为 **DeepSeek-V3**，点击「API 接入」后按页面指引创建并记录 **API Key**、**接口地址**和 **model 参数值**。

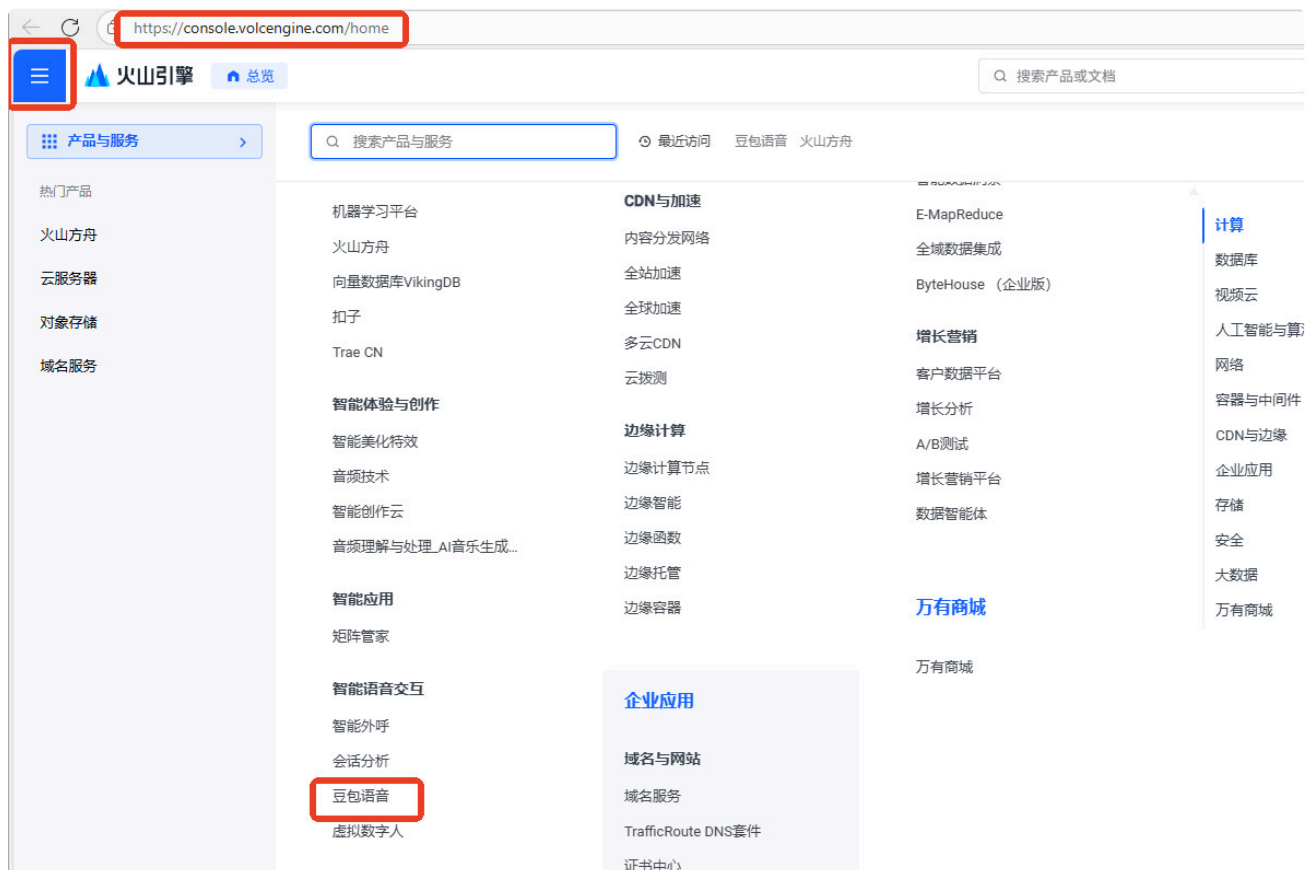
4.1.2 获取语音识别鉴权参数

具体请参考 <https://www.volcengine.com/docs/6561/163043#step1>: 创建应用

在火山引擎官网 <https://www.volcengine.com>，在顶部工具栏中点击「控制台」，并使用个人或单位账号登录。



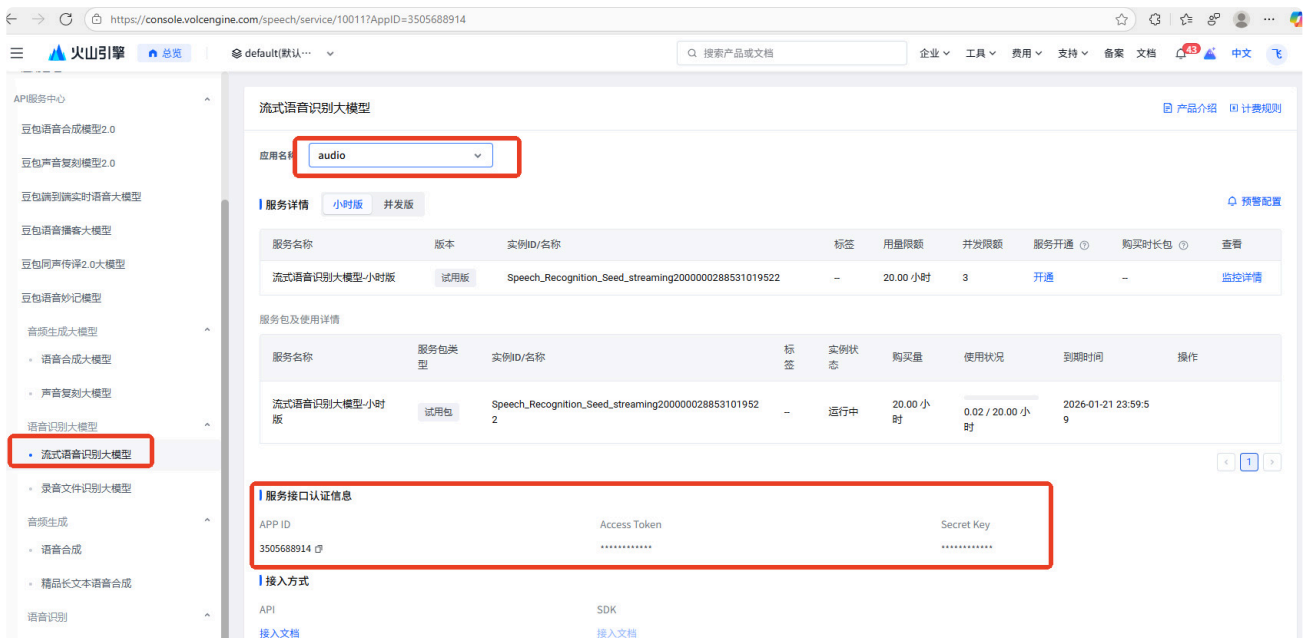
在火山引擎控制台 <https://console.volcengine.com/> 中，进入左侧产品与服务导航栏，选择智能语音服务中的豆包语音



进入「应用管理」，点击「创建应用」新建一个语音识别应用。



创建完成后，进入「流式语音识别大模型」，选择刚才创建的应用（例如这里是audio），激活模型的试用版后，在服务接口认证信息中记录以下两个密钥：



- **APP_KEY** (APP ID 应用密钥)：用于标识应用身份
- **ACCESS_KEY** (Access Token 访问密钥)：用于服务端鉴权校验

4.1.3 编辑配置文件

打开 `\文档\0gre\.env` 文件（如不存在请手动创建），按以下格式填写配置参数：

```

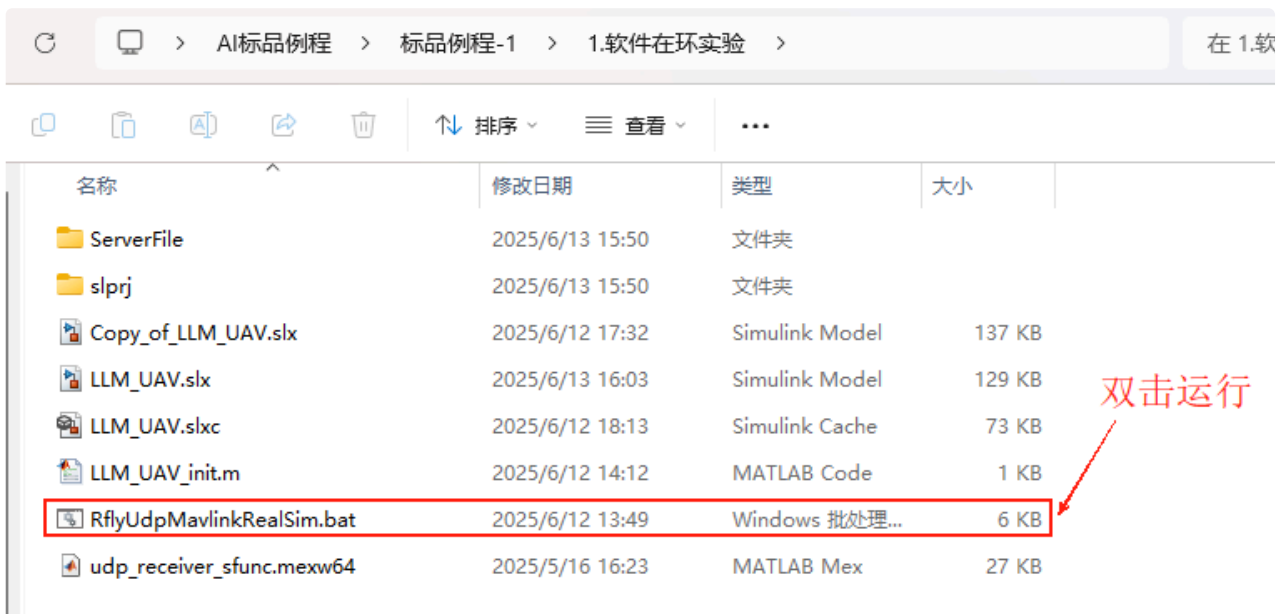
1 # 大模型 API 配置
2 VOLC_API_KEY=45f6ff5***** # 步骤 4.1.1 记录的 API Key
3 VOLC_NAME=deepseek-v3-250324 # 步骤 4.1.1 记录的 model 值
4 VOLC_PROVIDER=volcengine # 当前暂未使用, 可按示例填写
5 VOLC_ENDPOINT=https://ark.cn-beijing.volces.com/api/v3 # 步骤 4.1.1 记录的接口地址
6
7 # 语音识别 API 配置
8 VOLC_ASR_WS_URL=wss://openspeech.bytedance.com/api/v3/sauc/bigmodel
9 VOLC_ASR_APP_KEY=步骤 4.1.2 记录的应用密钥
10 VOLC_ASR_ACCESS_KEY=步骤 4.1.2 记录的访问密钥
11 VOLC_ASR_RESOURCE_ID=volc.bigasr.sauc.duration

```

保存并关闭配置文件。程序运行时会自动从该文件读取配置并初始化大模型客户端与语音识别服务。

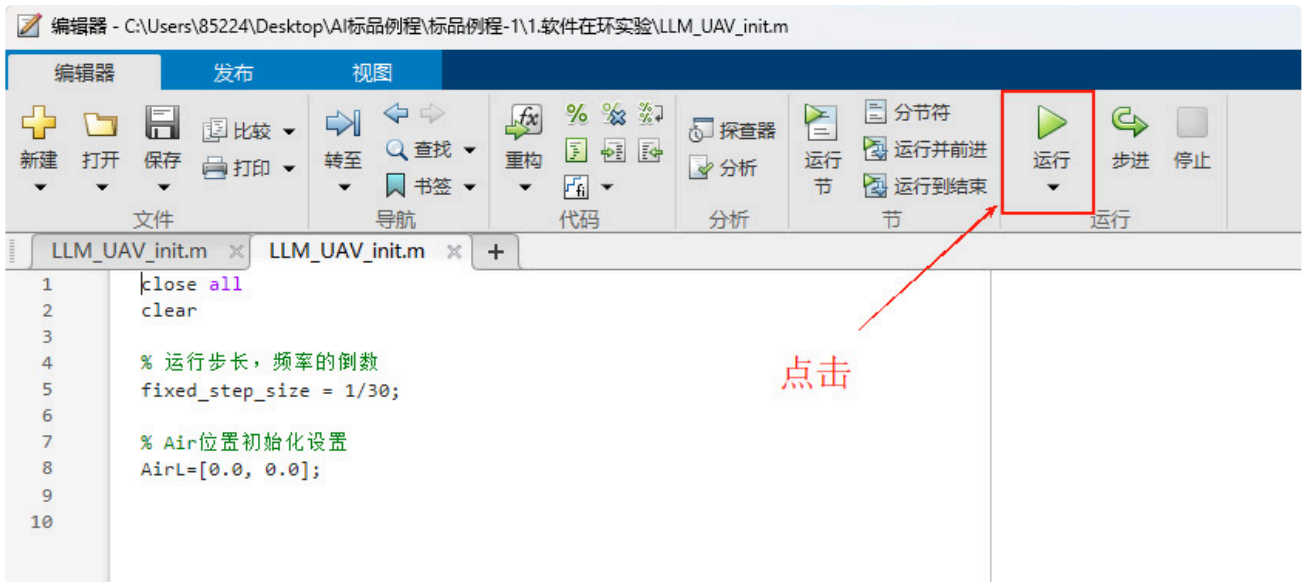
4.2 步骤2：启动飞控软件在环

在本实验文件夹下：双击运行批处理文件：[RflyUdpMavlinkRealSim.bat](#)

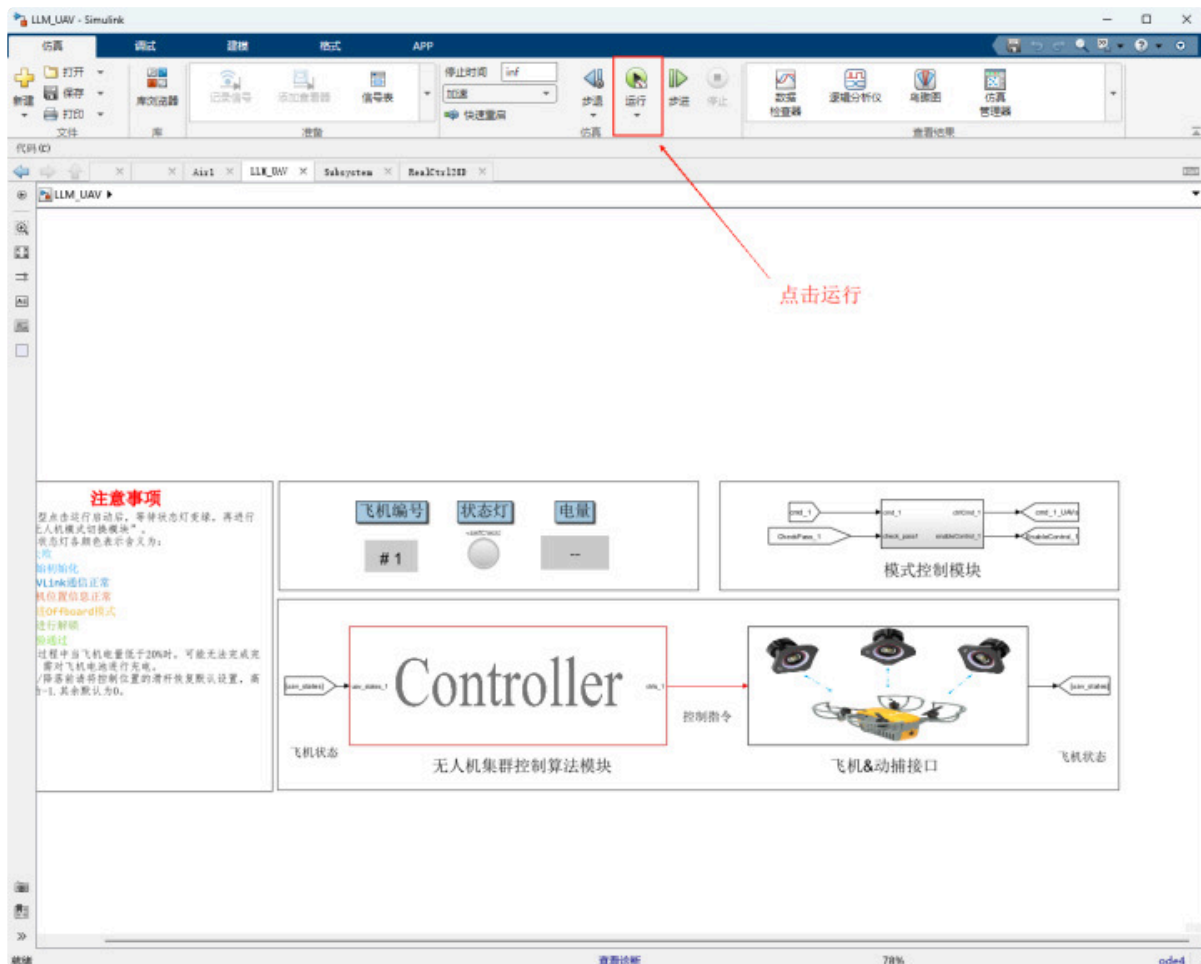


4.3 步骤3：运行外部控制程序

在同文件夹下，打开 [LLM_UAV_init.m](#) 文件，并在matlab中运行，进行参数初始化



在同文件夹下，打开LLM_UAV.slx文件，并在matlab中点击运行

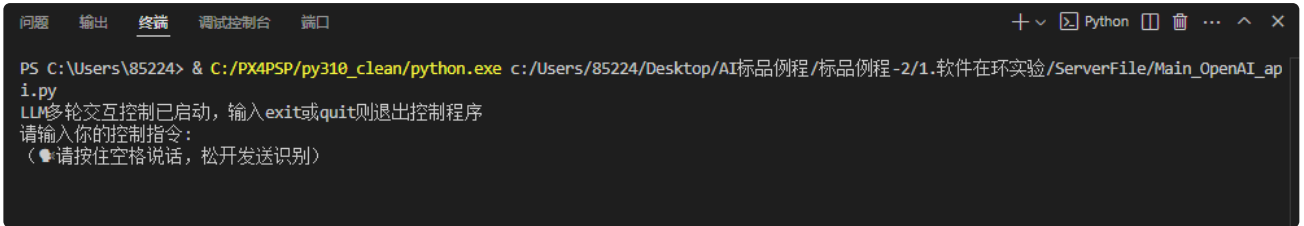


4.4 步骤4：启动大模型服务端并发送指令

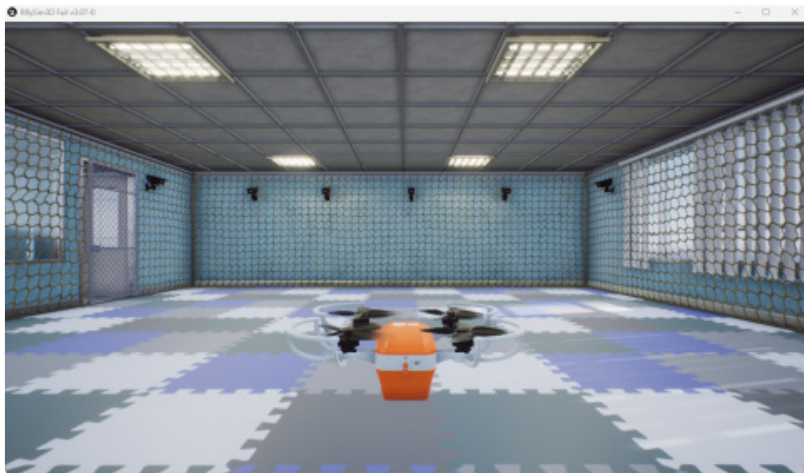
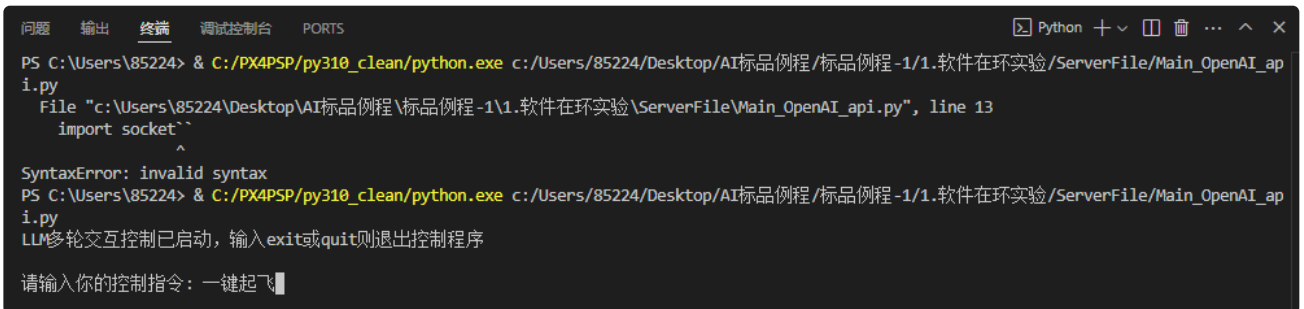
使用原先配置好python环境的Visual Studio Code打开程序 \ServerFile\Main_OpenAI_api.py; 并运行代码



按住空格，说话输入语音指令（请确保电脑有可用的麦克风）。输入的指令如：



- 输入：****一键起飞****。程序将会依次启动：开启Offboard模式、开启解锁模式、开启起飞模式；最后飞机将解锁并起飞升空
- 输入：**向左飞行1米**。程序将启动Left_Right_Fly，并按要求向左飞行1米
- 输入：****降落****。程序将启动Land_mode模式，飞机将降落到地面
- 输入：**锁定**。程序将启动Lock_mode模式，飞机电机将停止转动
- 输入：**解锁**。程序将启动UnLock_mode模式，飞机电机将解锁。



5. 关键知识点

5.1 关键知识点1：火山引擎实时语音识别（ASR）WebSocket 协议：

包括鉴权机制、二进制消息格式、音频流式传输与响应解析

5.2 关键知识点2：PyAudio 实时音频采集：

基于 PortAudio 实现 16kHz 单声道麦克风采样与 PCM 编码

5.3 关键知识点3：WebSockets 异步通信：

利用 `asyncio` 实现非阻塞音频推送与识别结果接收

5.4 关键知识点4：Pynput 键盘事件监听：

通过空格按下/抬起控制录音起止，实现 Push-to-Talk 交互模式

5.5 关键知识点5：多线程与事件同步：

守护线程运行异步事件循环，主线程通过 `Event` 阻塞等待识别结果

6. 参考资料

1. [PX4 官方文档（Offboard 控制与 MAVLink 通信）](#)
2. [MAVLink 协议说明文档](#)
3. [MATLAB/Simulink UAV Toolbox 文档（飞行控制与在环仿真相关示例）](#)
4. [OpenAI API 官方文档（大模型接口调用参考）](#)
5. [火山引擎豆包大模型产品文档与 API 说明](#)
6. [火山引擎语音识别鉴权方法](#)
7. [火山引擎大模型流式语音识别 API](#)
8. [PyAudio 音频采集库文档](#)
9. [WebSockets Python 异步通信库](#)

7. 常见问题

Q1：UDP 端口占用错误

A1：现象：若运行 Python 程序时出现如下报错：

```
1 | Traceback (most recent call last):
2 |   File "...\Main_OpenAI_api.py", line 301, in <module>
3 |     chat_api = OpenAI_APIs()
4 |   File "...\Main_OpenAI_api.py", line 129, in __init__
5 |     self.UDPConnect = UDPSendRece()
6 |   File "...\Main_OpenAI_api.py", line 20, in __init__
7 |     self.sock.bind(("", 5005))
8 | OSError: [WinError 10048] Only one usage of each socket address (protocol/network address/port) is normally permitted
```

原因：5005 端口被其他程序占用。

解决方法：

1. 在终端输入 `netstat -ano | findstr "5005"` 查找占用该端口的进程 PID
2. 在任务管理器的「详细信息」面板中根据 PID 找到对应进程并结束
3. 或者修改代码中的端口号为其他未占用端口（需同步修改 Simulink 模型中的目标端口）

Q2：语音识别无响应或识别失败

A2：现象：按住空格录音后松开，长时间无识别结果或返回空文本。

可能原因及解决方法：

1. **麦克风权限未开启：**检查 Windows 隐私设置 → 麦克风，确保允许桌面应用访问麦克风
2. **鉴权参数错误：**检查 `.env` 文件中 `VOLC_ASR_APP_KEY` 和 `VOLC_ASR_ACCESS_KEY` 是否正确，确认已在火山引擎控制台激活流式语音识别服务
3. **网络连接问题：**检查网络连接，确保能访问 `wss://openspeech.bytedance.com`
4. **音频质量问题：**在安静环境下测试，确保说话清晰且麦克风音量合适
5. **PyAudio 未正确安装：**重新安装 PyAudio 或使用 `pip install pyaudio` 确认安装成功

Q3：大模型 API 调用失败

A3：现象：语音识别成功但大模型无响应，或返回鉴权错误。

解决方法：

1. 检查 `.env` 文件中 `VOLC_API_KEY`、`VOLC_NAME` 和 `VOLC_ENDPOINT` 是否正确
2. 确认火山引擎控制台中该 API Key 对应的模型（DeepSeek-V3）已开通且有剩余额度
3. 检查 `VOLC_ENDPOINT` 格式是否完整（需包含 `https://` 前缀和完整路径）
4. 尝试在火山引擎控制台的「API 接入」页面重新测试接口连通性

Q4：MATLAB/Simulink 连接失败

A4：现象：Simulink 模型运行后无法与 Python 程序通信，无法控制无人机。

解决方法：

1. 确认 PX4 SITL 已成功启动（Step 4.2）
2. 检查 `LLM_UAV_init.m` 中的 UDP 端口配置与 Python 程序一致（默认 5005）
3. 检查防火墙是否阻止了 UDP 通信，必要时添加例外规则
4. 确认 Simulink 模型中的 `udp_receiver_sfunc.mexw64` 文件存在且版本匹配

Q5: 录音保存但无法识别

A5: 现象: 可以看到 `last_record.wav` 文件生成, 但识别结果为空或错误。

解决方法:

1. 用音频播放器打开 `last_record.wav` 检查录音质量
 2. 确认录音格式为 16kHz 单声道 16bit PCM (代码中已设置)
 3. 检查录音时长是否过短 (建议至少 1 秒) 或过长 (建议不超过 60 秒)
 4. 尝试在代码中增加录音前的延迟, 避免音频开头被截断
-

1. <https://rflysim.com/> ↩

2. 推荐配置请见: <https://rflysim.com/doc/zh/HowToInstall.pdf> ↩