AI 角色扮演语音聊天网页

# 1. 这网页给谁用？他们卡在哪？我想象的用法是什么？

• 主要人群：

- 想练口语的人：需要有人“陪聊”和纠错。

- 喜欢某个角色/IP 的粉丝：想和“哈利·波特”“福尔摩斯”过把瘾。

- 老师/学生：课堂上做角色扮演、辩论练习。

• 常见痛点：

- 现有工具多是打字，不够沉浸；语音不稳定或太贵。

- 人设容易跑偏，说着说着就不像那个角色了。

- 初次使用门槛高，需要安装或注册。

• 我想象的几个小故事：

- 故事A（口语）：我点“苏格拉底导师”，说一段观点，3秒左右就听到语音回复，还顺带给我更地道的说法。

- 故事B（粉丝）：我选“哈利·波特”，问“我该进哪个学院？”，他用书里的语气回答我。

- 故事C（课堂）：老师让同学和“苏格拉底辩手”对话，机器人不断追问，让同学把论点讲清楚。

# 2. 需要哪些功能？先做哪个？这次我会做哪些？

• 必做（这次就做）：

1) 语音输入：浏览器录音→前端转16k单声道WAV→后端ASR→文字。

2) 语音输出：把AI文字用TTS读出来。

3) 角色搜索/切换：内置几张“角色卡”（口吻+设定），可搜可点。

4) 对话：流式显示回复（SSE），体验更顺滑。

5) 会话管理：清空/新建聊天，保留最近几轮上下文。

• 应该有（下一步）：

6) 纠错面板：把我刚说错的词句列出来，给替代表达。

7) 风格开关：正式/幽默/简洁。

• 可选：导出记录、选音色、移动端更友好。

# 3. 用哪家模型？为什么？

• 备选我看了这几个：GPT‑4o‑mini、通义Qwen 2.5、GLM。

• 我选 GPT‑4o‑mini，原因很直接：

- 反应快、价格不离谱，语音场景更顺手。

- 中英文都行，角色口吻比较稳定。

- 流式接口好用，Spring Boot 对接简单。

# 4. 除了“能聊天”，AI 角色还要会什么？（至少3项技能）

• 技能S1：人设不跑偏

- 做法：在系统提示里固定口吻、禁区、常用语；每轮让模型自检“像不像这个角色”。

• 技能S2：苏格拉底式追问

- 做法：当角色=苏格拉底时，自动生成1~2个追问（比如“你的假设是什么？”“有没有反例？”）。

• 技能S3：语言纠错+替代表达

- 做法：把我刚说的话再丢给模型，让它列出错误点+更自然的表达，放在侧栏。

• 技能S4：风格/语气控制（可选）

- 做法：前端加下拉框（正式/幽默/温柔；简要/详细），参数随请求一起发给模型。

# 实现简述（一点点技术说明）

• 前端：Vue3 + MediaRecorder 录音，OfflineAudioContext 转成 16k 单声道WAV，fetch 传给 /api/asr。

• 后端：Spring Boot 提供 /api/asr（连七牛ASR）、/api/chat/stream（连 LLM，SSE 流式返回）。

• 回复播放：前端用 Web Speech API 朗读；需要统一音色时再接后端TTS。

• 角色卡：一个简单的 JSON，包含“背景、口吻、口头禅、禁区、技能开关”。