**1.你计划将这个应用面向什么类型的用户?这些类型的用户他们面临什么样的痛点，你设想的用户故事是什么样呢?**

**语言学习者（樱木建二）**

**面对的用户类型：**

主要为各种语言学习者，如考研党、雅思党、有外语需求的工作党等。

**面对的用户痛点：**

* **时间被工作/学业切碎**：通勤+加班文化让持续学习难
* **“面子压力”导致不敢开口**：怕被同学/同事笑“中式口音”，宁愿沉默。
* **素材与练习割裂**：刷题/背单词多，**真实对话场景**少；纠错滞后、缺节奏。
* **升学/求职双轨压力**：CET-4/6、考研英语、求职面试/出国语言考试并行，**目标各异**且容易乱。
* **学习质量参差**：短视频/笔记平台信息碎片化，**缺系统化路径**与阶段性反馈。

**设想中的用户故事：**

1. 作为**一线城市白领（晚间经常加班）**，我希望在**地铁通勤的15分钟**内完成“听-跟-复述”一轮练习，并拿到**1-2个即时纠错点**，这样我**2周后**能在晨会汇报中自然表达。*验收：连续打卡≥10天，汇报时填自评+同事匿名评1次。*
2. 作为**应届生**，我需要一份“外企求职口语地道表达替换清单”**（按HR高频问题场景分类），**并能在模拟群面里试用，系统给出冗余词的标注。*验收：清单≥50条、覆盖5类场景；模拟面试得分较首轮↑20%。*
3. 作为**考研党**，我想把作文模板**口语化重写**并做**1分钟口语复述**，系统指出**语法/逻辑两类错误**并给替换建议。*验收：作文评分提升一个档位；复述错误率下降50%。*

**逻辑与思维训练者（福尔摩斯）**

**面对的用户类型：**

主要为各种需要对逻辑思维进行训练的人群，如考公/事业编群体、咨询顾问、战略分析师、产品经理等。

**面对的用户痛点：**

* **信息噪音大**：微信群/短视频/论坛信息量巨大，**事实与观点混杂**，常被“反转”带节奏。
* **工作场景“要结论、要方案”压力大**：缺“已知/未知/待验证”的结构化框架，讨论常走偏。
* **复盘文化薄弱**：做完就散，**缺证据库与假设验证记录**，难以沉淀方法论。

**更详细的用户故事：**

1. 作为**产品经理**，我要把一个舆情事件在**10分钟内**拆成“**事实/假设/证据需求**”三栏，并生成**5个追问**发给同事补全。*验收：团队在当天内完成事实闭环，结论与证据一一对应。*
2. 作为**本科生**，我要用“**反事实推演**”写一段作业说明：**如果X不成立会怎样**，并让系统**打回逻辑漏洞**。*验收：老师反馈“论证链完整”，扣分点从“逻辑不清”转为“资料不全”。*
3. 作为**运营同学**，我要把活动复盘按“**目标-动作-结果-偏差-下次假设**”自动成表，并沉淀为**下次模板**。*验收：模板被团队复用≥2次。*

**个人理财与小微经营者（范蠡）**

**面对的用户类型：**

范围广如有理财需求的人群等。

**面对的用户痛点：**

* **金融素养参差**：对风险、估值、现金流概念把握不足，易被热点裹挟。研究显示我国居民金融素养仍在补课阶段。
* **缺“退出机制”**：涨跌随情绪操作，**止盈/止损纪律**难以执行。
* **资产与负债错配**：房贷、车贷、教育/赡养负担下，**流动性安全垫不足**。
* **小微经营现金流脆弱**：账期长、回款慢，**回撤红线**与**再平衡**不清晰。

**更详细的用户故事：**

1. **月光白领小陈**工资 1.5w，房租 5k，追热点满仓，没应急金。跌了不敢卖，月底付房租才发现没现金。*验收：口袋分账→应急/房租/日常/投资；核心指数定投＋小卫星仓；设止盈止损、工资日自动转入。*
2. 作为**卤味店老板老周**，月流水 8w，平台 15 天结款，进货要现钱。看着赚钱却老缺钱，只能借高息贷。*验收：做 13 周现金流表；谈缩短结算或做保理；设“回款红线”；备付金=1 个月固定成本；月度再平衡。*
3. 作为**三口之家小杨**，房贷车贷压力大，把备用金全买三年理财。父亲住院急用钱，产品取不出，只能借亲戚。*验收：三桶法——流动桶（货基/定存阶梯）＞保障桶（医保+意外+重疾）＞增长桶（指数/债基）；每季复盘。*

**法律风险意识用户（成步堂龙）**

**面对的用户类型：**

有需求但是缺乏相关知识的打工人等。

**面对的用户痛点：**

* **合同“先签后看”**：模板拼凑，关键条款（违约、争议解决、适用法/管辖）常被忽视。
* **证据意识薄弱**：聊天记录/发票/收据保存不规范，**证明难**。
* **路径不清**：**协商-调解-仲裁-诉讼**的成本/周期差异不了解；中国实践中**调解/和解应用广**。
* **劳动与合同纠纷常见**：加班、薪酬、工程/PPP等合同争议数量可观；“996”典型案例曾被点名。

**更详细的用户故事**：

1. **装修“先签后看”的王阿姨**，模板合同，增项多、拖工期，条款含糊。证据散、维权难。*验收：按时间线整理合同/收据/照片/录音；补充“里程碑验收+逾期扣款”；先市监调解，不成再仲裁/诉讼。*
2. 作为**996 要仲裁的小李**，口头承诺调休，实际无加班费；离职被压绩效。证据零散，差点错过时效。*验收：收集打卡/工时/任务邮件/群聊；列主张清单；先律师函促谈判，时效内提仲裁，备好替代诉求。*
3. **网店主小冯的图片侵权**，用供货商图被平台投诉，链接下架。没授权书，慌乱处理又被二次投诉。*验收：立即下架+留证；补签授权并提交来源说明；建立“合规素材库”；先平台调解，评估再和解或应诉。*

**健身入门者（安西光义）**

**面对的用户类型：**

有健身需求但是缺乏相关知识的人群、只能用短暂的时间运动的上班族、工作忙长期吃外卖或者饮食不均衡但是需要健身良好饮食的人等。

**面对的用户痛点：**

* **久坐+加班**：我国成年人的**身体活动不足**比例不容忽视，计划容易被工作打断。
* **动作不标准/怕受伤**：缺教练或“无脸出错”的安全感。
* **饮食难执行**：外卖/应酬多，**宏量配比**难落地。
* **中断复发**：出差/生病后难回正轨，**最低维持方案**缺失。

**更详细的用户故事：**

1. 作为**新手**，我需要**4周分级训练**（每周3次）+ **RPE强度提示**，并为**肩/膝不适**提供替代动作。*验收：周计划与替代动作单独成卡；RPE自评与训练日志同步。*
2. 作为**忙碌打工人**，我要一套“公司/家里可做”的20分钟无器械训练，配合**拉伸与放松**流程。*验收：可直接跟练；异常疼痛触发“减量/停练”提示。*
3. 作为**经常出差**的人，我要**最低健身维持方案**（每周2次、每次15分钟），避免“从零再来”。*验收：出差周不掉线，心率/主观疲劳可记录。*

**情绪支持与复原用户（艾洛叔叔）**

**面对的用户类型：**

工作、学习、备考、求职压力大的情绪需要排解的人群等。

**面对的用户痛点：**

* **“自己扛”的文化**与**求助污名**：公众心理健康素养不足、污名仍在，专科资源相对有限且集中。
* **家庭关系与界限**：与长辈期望冲突常引发内疚/愧疚，**界限话术**缺失。
* **工作/考试/婚恋多重压力叠加**：情绪命名能力弱，**只想要被倾听**而非大道理。
* **缺“微仪式”**：欠缺可日常执行的**呼吸/放松/睡前**流程。

**更详细的用户故事**

1. 作为**压力较大的职场人**，我需要一个**5分钟“情绪—身体—想法”三步梳理法**+ **呼吸练习**，得到**今天可执行的1个小目标**。*验收：日终自评情绪分较早上提升；小目标完成打勾。*
2. 作为**与父母沟通紧张的年轻人**，我要**界限话术库**（温和/坚定两版），并配**可能的父母反应**与**二次回应**。*验收：对话脚本三轮可演练，用户选择最顺嘴版本。*
3. 作为**失眠人群**，我想要**睡前微仪式**（热茶/呼吸/书写），并**一键记录睡眠质量**。*验收：连续7天完成≥5天，入睡时长缩短为目标。*

**2.你认为这个 APP 需要哪些功能?这些功能各自的优先级是什么?你计划本次开发哪些功能?**

对于App需要哪些功能我们采用了 **MoSCoW**（Must/Should/Could/Won’t）优先级：

**Must（首发就要有）**

* **多角色会话切换**（6 角色预置，图标+简介+一键进入场景）
* **快捷短语面板**（每角色≥5 条，一键插入输入框）
* **结构化引导提示**（如福尔摩斯“已知/未知/待验证”表单；成步堂“证据清单”表单）
* **即时纠错/建议模式**（语言学习：发音/表达替换；健身：动作要点/常错提醒）
* **语音输入/播报**（ASR+TTS）
* **基本记忆**（轻量会话记忆：学习目标、训练进度、偏好语气）
* **导出/复制**（将对话关键输出导出为 Markdown/文本）

**Should（次优先）**

* **模板库**（法律沟通话术、风险测评问卷、复盘模板、饮食日程）
* **进度与打卡**（语言打卡、健身周计划、情绪微仪式统计）
* **多模态**（图片动作纠错、合同截图要点标注）
* **RAG 知识库**（导入用户自己的合同/训练记录/投资记录进行私有化回答）

**Could（可选/增强）**

* **团队分享**（将“推理方案/法律清单/训练计划”生成可分享链接）

**Won’t（暂不做）**

* **医疗诊断/法律结论性意见**（严格作为信息与教育，不替代专业服务）
* **实盘交易/法律文书自动生成并提交**（涉及合规与风险）

**本次计划开发**

* Must 全部 + Should 中的 **语音输入/播报**、**模板库**（精简版）。
* 简化的**本地记忆**（仅保留“目标/偏好/阶段”）。
* 不做用户登录与云同步，先本地缓存（便于快速验证价值）。

**最终开发和实现**

* 使用静态html作为基础，可切换角色，并实现语音输入，在后端转换成文本；
* 本地计算机显存不足导致在kaggle实现在线部署，使用jsonl文件作为数据集传入，针对不同的角色数据集对LLM进行微调，生成lora文件，调用时采用加载ChatGLM3-6B和lora文件结合处理文本；
* 根据选择的角色，采用不同类型的语音朗读基于文本处理函数生成的回复。

**3.你计划采纳哪家公司的哪个 LLM 模型能力?你对比了哪些，你为什么选择用该 LLM 模型?**

**ChatGLM3-6B（智谱）**

**优点：**

* 中文对话自然度高，日常问答与多轮上下文衔接稳定。
* 推理条理性在同量级里不差；指令跟随友好，易拿到“步骤化”答案。
* 生态成熟：本地部署与微调资料多，社区经验丰富，兼容 CPU/GPU 多平台。
* 6B 体量对设备更友好，LoRA/QLoRA 案例多，做小样本对话风格微调比较省心。

**短板/注意：**

* 英文与代码能力不如同量级的部分西方模型；长文本结构化能力一般。
* 安全对齐相对保守，个别场景可能“过度拒答”，需做提示词优化。

**Baichuan2-7B-Chat（百川）**

**优点：**

* 中文阅读理解与知识问答表现稳；训练数据广，学科覆盖较全。
* 指令对齐到位，问答“又快又直给”，适配 ToB/ToC 常见场景。

**短板/注意：**

* 英文与代码生成较中庸；复杂推理链条与工具编排能力一般。
* 对话风格变化不算强，做角色扮演时需要更精细的提示词或轻微微调。

**Qwen-7B-Chat（阿里）**

**优点：**

* **中英双语**都比较稳定，角色扮演与风格迁移能力强；多轮记忆相对可靠。
* 工具调用（函数/检索）社区范例多，易做“表单→结构化输出”。

**短板/注意：**

* 生码与长篇英文写作略逊同量级英系模型；极长上下文需要仔细评估具体权重版本。
* 某些安全阈值偏保守，少数边界问题会触发拒答。

**Mistral-7B-Instruct（Mistral）**

**优点：**

* 英文**推理稳、结构化强**，高质量指令跟随；摘要/改写/归纳表现亮眼。
* 推理“链式”可读性好，适合做“步骤化解决方案”。

**短板/注意：**

* 中文口语化对话略显生硬；中文知识问答不如中系 7B。
* 中文角色扮演需要更多提示词或小规模微调。

**LLaMA2-7B-Chat（Meta）**

**优点：**

* **社区生态最大**：教程、微调脚本、适配器模型最丰富；做基线/教学/二开极合适。
* 英文对话与常识问答稳定，作为“参照系”好用。
* 资料与示例极多，入门成本最低；许可证为社区许可，商用需遵条款。

**短板/注意：**

* 代际已老，综合能力被 Mistral/Qwen 等后发 7B 追平或超越。
* 中文与代码生成较一般；要高质量输出需二次微调+提示工程。

**Yi-6B-Chat（01.AI）**

**优点：**

* **中文任务强**，问答简洁清晰；在 6B 体量中性价比高。
* 兼容性好，量级更小，低配硬件运行难度更低。

**短板/注意：**

* 英文与复杂推理能力不如 7B 新秀；内容创作的多样性一般。
* 对超长上下文与工具编排的支撑有限。

**Deepseek-Chat 7B（Deepseek）**

**优点：**

* **代码理解/生成**与“工具型任务”支撑较好；对话稳定。
* 指令对齐不错，执行类任务（写脚本、改代码、解释报错）体验佳。

**短板/注意**

* 纯中文文案的自然度与文学性不如中系模型；情感化对话略硬。
* 超过中等长度的中文长文组织力一般。

最后选择：**ChatGLM3-6B**

1. **你期望 AI角色除了语音聊天外还应该有哪些技能?**

**A. 结构化产出与表单驱动**

* 一键生成/维护：
  + 语言：地道表达替换清单、7/14/28 天打卡计划。
  + 推理：已知/未知/待验证表、证据强弱分级表。
  + 理财：三档资产配置卡、止盈止损纪律卡、风险清单。
  + 法律：争议焦点表、证据清单、路径对比表（诉讼/仲裁/和解）。
  + 健身：RPE 周计划、替代动作库、最低维持方案卡。
  + 情绪：情绪命名卡、5 分钟呼吸与 Grounding 指南、界限话术库。
* 导出为 Markdown / PDF（前期先 Markdown + 复制）。

**B. 工具调用（Function Calling）与轻量 RAG**

* 工具：计时器（番茄钟）、单位换算、BMI/宏量计算、日期与提醒生成。
* 轻量 RAG：将用户自己的合同/训练记录/学习词书上传后，做私有检索与引用（首期可只做本地缓存 + 引用片段标注）。
* 引用可视化：在答案末尾列出“来自你资料的段落”。

**C. 多模态（逐步引入）**

* 图片动作纠错（健身）、合同截图条款定位（法律）、手写笔记抓要点（学习）。
* 先做占位：上传图片→文本建议（不必首发就做关键点可视化）。

**D. 安全与合规护栏**

* 法律与医疗相关：**不提供实务性结论/诊断**，只做教育与信息整理；必要时提示“寻求专业人士”。
* 投资：标注“非投资建议”，输出风险披露与决策自查清单。

**E. 记忆与个性化**

* 角色内记忆：学习目标/口音偏好、体脂目标与伤病史、情绪触发因素与有效安抚法。
* 语气可调：正式/鼓励/严谨三档。
* 连续性：打卡天数、复盘节奏、小胜利记录。

**F. 性能与成本策略**

* 多模型路由：轻任务→小模型（如 GPT-5 mini），深推理/长文→主模型/Claude。[OpenAI+1](https://openai.com/index/introducing-gpt-5/?utm_source=chatgpt.com" \t "_blank)
* 输出控制：先结构后展开；允许“摘要/详解/一步步”三档。