

作品说明文档

一、目标用户画像与核心痛点

1.1 目标用户群体

本产品聚焦于两类典型“收支失衡”人群：


- 核心用户：18-30岁「隐形贫困人口」
 - 收入范围：月薪5,000-12,000元
 - 消费特征：线上支付占比 ≥85%，储蓄率 <5%
 - 行为模式：追求生活品质但缺乏财务规划，易陷入“月光”甚至“超前消费”
- 次级用户：31-40岁「高薪月光族」
 - 持卡数量：平均3.4张信用卡
 - 透支率：高达42%
 - 需求特征：对自动记账、预算预警、额度管理有强烈需求

1.2 用户核心痛点

痛点	数据支撑	用户影响
记一半就弃	传统记账工具手动录入步骤 >3 步时，次月留存率下降 58%	记账中断，数据不完整，失去参考价值
花了才后知后觉	76% 用户在账单日后才发现超支	预算失控，情绪焦虑，财务压力加剧
想省却找不到入口	购物节期间 64% 用户因信息过载错过优惠	“后悔支出”频发，错失省钱机会


1.3 典型用户故事

小李，24岁，互联网运营，月薪9k，曾是“发薪日富翁，月底吃土人”。自从用了「灵活账」，四位AI财务伙伴彻底改变了他和钱的关系。



清晨 · 小账（记账员）

通勤路上，小李顺手买了杯咖啡，语音说：“今天咖啡30块。”



小账立刻回应：“已记录至【餐饮】，本月餐饮预算已使用63%，今日还可花57元。”

——无需手动输入，记账零负担，消费即感知。

☀️ 午休 · 省省（省钱教练）

同事聊起周末短途旅行，小李心动却犹豫：“酒店好贵啊。”

他问：“我想订个周末的民宿，怎么省？”

👉 省省秒回：“推荐使用B平台‘新用户立减’+‘连住优惠’，比单订便宜220元；用C平台买景区联票还能返现8%。已帮你生成比价清单！”

——在消费决策前介入，提供可执行的省钱方案。

🌙 晚间 · 明查（财务洞察官）

睡前翻看App，小李好奇：“我这个月到底花了多少外卖钱？”

👉 明查展示：“本月外卖支出860元，占餐饮总支出的72%，比上月↑18%。你的‘健康饮食’目标可能需要调整。”

——数据说话，揭示隐藏模式，推动行为反思。

🌙 深夜 · 远谋（财务规划师）

小李感慨：“什么时候才能存够钱去日本玩啊？”

他随口一问：“我想一年后去日本旅行，大概要花1万，能实现吗？”

👉 远谋分析后回复：“根据你当前收支，每月多存833元即可达成。建议：1) 餐饮预算下调15%；2) 开启‘挑战储蓄’功能。我已为你生成6阶段计划，并协调省省提供专属省钱清单。”

——把愿望变成路径，让未来可预期、可抵达。

📅 17 月底总结

打开月度报告，小李惊喜发现：首次结余430元！

他笑着点击：“转入20%到货币基金”，并设置“下月预算+10%挑战”。

四位AI齐声祝贺：“小账已同步更新预算，省省准备了新挑战任务，明查将持续追踪，远谋祝你离目标更近一步！”

——从每一笔支出到每一次消费决策，再到长远财务目标，四位AI角色各司其职、协同联动，让小李第一次感受到：钱，原来可以被温柔而有力地掌控。

二、功能规划与优先级

2.1 核心功能（MVP阶段，本次开发）

在这4个agent中添加了联网搜索的mcp

功能模块	功能描述	优先级	对应Agent
✔ 智能记账与实时提醒	支持语音/文本输入，自动提取金额、类别、时间，实时反馈预算余额与消费建议	高	小账
✔ 财务数据查询与统计	查询历史记录、支出汇总、预算执行情况，支持按周期/类别筛选	高	明查
✔ 个性化省钱建议	基于消费数据，推荐替代方案、优惠渠道、省钱技巧，标注预估节省金额	中高	省省
✔ 财务目标规划与追踪	设定储蓄/消费目标，生成实现路径，定期追踪进度并动态调整	中高	远谋

2.2 未来拓展功能（二期规划）

功能模块	功能描述	优先级	技术依赖
➡️ SOON MCP支付自动同步	接入支付宝/微信等MCP协议，实现收支自动记录	中	支付平台API接入
➡️ SOON 每日账单自动生成	每日推送图文账单，含支出分布、预算使用率、消费趋势	中	定时任务 + 数据聚合
➡️ SOON 跨平台比价与优惠提醒	商品链接识别，自动比价、搜券、提示历史低价	中	爬虫/电商API
➡️ SOON 多端同步与语音播报	支持App、小程序、智能音箱多端联动，语音播报财务摘要	低	多端架构 + TTS集成
➡️ SOON 智能语音交互模块	为所有Agent提供语音对话能力以及不同音色，支持实时语音输入、转文本、调用模型处理、并返回语音响应，极大提升交互自然度与便捷性	中	语音识别(STT)/合成(TTS)服务、音频处理库、异步通信架构

三、LLM 模型选型说明

3.1 模型对比维度

从“功能适配性、场景匹配度、成本可控性”三个核心维度，对比主流 LLM 模型：

对比维度	阿里千问3	豆包（字节跳动）	百度文心一言 4.0
核心能力	文本理解与分析能力强，支持基础工具调用	文本 + 语音双模态处理（语音转文字 / 文字转语音），工具调用响应速度快	文本生成与逻辑推理能力强，支持多轮复杂对话
场景适配性	适合纯文本财务分析，但缺乏语音交互能力，无法满足“语音记账”核心需求	完美匹配“语音记账 + 文本分析”场景，语音识别准确率≥96%，响应延迟<1.5 秒	适合复杂财务规划，但语音功能需额外集成，开发成本高
成本可控性	按调用次数计费，语音功能需额外接入第三方 API，总成本较高	一站式提供文本 + 语音能力，无额外集成成本，中小规模调用（日均 1000 次内）成本比千问低 15%	基础调用成本高，且语音集成需额外付费，总成本最高
开发难度	需单独对接语音 API，开发周期约 2 周	无需额外集成，API 接口统一，开发周期约 3 天	需单独对接语音 API，且接口文档复杂度高，开发周期约 3 周

3.2 最终选型：豆包模型

选择豆包模型的核心原因的在于“场景深度适配”与“开发效率优先”：

- 1. **满足核心需求：**本产品“语音记账”是 MVP 阶段的核心功能（解决“记一半就弃”痛点），豆包自带的语音转文字 / 文字转语音能力，无需额外集成第三方工具，可直接实现“用户语音输入→Agent 识别记账→语音反馈结果”的闭环；
- 2. **提升用户体验：**豆包的语音识别支持方言（如四川话、广东话）与模糊表述（如“昨天晚上吃火锅花了两百多”），识别准确率比千问 + 第三方语音 API 的组合高 8%-10%，更符合年轻用户的交互习惯；
- 3. **降低开发与成本门槛：**统一的 API 接口减少了 50% 的开发工作量，且语音功能无额外收费，相比“千问 + 第三方语音 API”的组合，每月可节省 30% 的调用成本，适合项目初期快速验证。

四、AI 角色技能拓展规划

小账（理财记录员）：作为精准记录者，将强化收支信息的自动提取能力，支持从用户语音、文本乃至发票或支付截图等图片中识别金额、类别、日期和支付方式；同时具备数据校验功能，可自动检测

重复录入或异常金额（如“10000元早餐”），并主动提示确认；未来还将支持微信、支付宝及银行卡账单的多账户自动同步，真正实现“无感记账”。

明查（财务洞察官）：作为专业分析师，将深入挖掘用户消费行为，识别高频模式（如“周末奶茶成瘾”“每月5号网购高峰”）；能对比实际支出与预算差异，精准定位超支原因（如“因3次外卖溢价订单导致餐饮超支300元”）；并自动生成包含趋势图、占比饼图等可视化元素的月度财务报告，支持交互式查看明细，帮助用户直观理解财务状况。

省省（省钱行动教练）：作为实用建议师，将提供高度个性化的省钱方案，例如基于用户常购商品推荐“会员日低价提醒”；针对高消费项目提供性价比替代建议（如用便利店挂耳咖啡替代连锁品牌饮品，每月节省180元）；并在购物节期间实时查询收藏商品的历史低价、优惠券与平台补贴，计算最优购买时机，主动提示“618期间下单可省800元”。

远谋（财务目标规划师）：作为长期规划师，将具备目标可行性分析能力，能评估用户设定的大额储蓄目标是否现实，并量化所需调整（如“需每月多结余2167元”）；可将长期目标拆解为可执行的周/月任务，并在进度偏离时动态调整计划（如某月超支后，自动分配补足额度至后续月份）；同时协调其他Agent资源，形成闭环支持，确保财务目标不仅被设定，更能被达成。

五、系统架构设计与技术规划

5.1 整体架构概述

本系统采用模块化分层架构，核心围绕四个AI Agent（小账、明查、省省、远谋）构建，通过统一的交互层接收用户输入，基于选型的豆包大模型进行智能处理，最终提供个性化的财务助手服务。

代码块

1	用户交互层（Gradio界面）	→	智能处理层（四大Agent）	→	工具调用层	→	数据持久层（SQLite）
2	↑		↓		↓		↓
3	多模态输入		意图识别		专业能力调用		数据库操作

5.2 语音交互模块详细设计（未来扩展）

5.2.1 模块概述

模块名称：语音交互接口模块 | **负责人**：张兆颀 | **当前状态**：原型测试完成，待集成

核心功能：为所有Agent提供语音对话能力，实现语音输入、转文本、调用豆包模型、返回语音响应的完整链路

5.2.2 技术实现方案

核心处理流水线

1 用户语音输入 → 音频预处理 → STT转换 → 豆包模型处理 → TTS合成 → 语音输出

已完成的原型验证

代码块

```
1  # 测试程序核心架构
2  class VoiceInputModule:
3      - 语音采集（麦克风输入）
4      - 音频预处理（降噪、格式转换）
5      - 语音转文本（STT服务）
6      - 豆包模型对话接口
7      - 文本转语音（TTS输出）
```

关键技术特性：

- **多模态输入支持：**语音、文本、录音文件三种输入方式统一处理
- **实时流式处理：**支持边说话边转译的低延迟交互
- **上下文记忆：**基于豆包模型的对话历史维护

5.2.3 与现有系统集成方案

接口设计

代码块

```
1  class VoiceIntegration:
2      def __init__(self, main_app):
3          self.main_app = main_app  # 主程序实例
4          self.voice_processor = VoiceProcessor()
5
6      async def process_voice_input(self, audio_data):
7          """处理语音输入并返回主程序可理解的指令"""
8          text = await self.voice_to_text(audio_data)
9          # 调用现有Agent处理逻辑
10         return await self.main_app.process_message(text)
```

集成策略

1. **输入层扩展：**在现有文本输入基础上增加语音输入通道
2. **处理层适配：**语音转文本结果无缝对接现有Agent处理流水线
3. **输出层增强：**在文本回复基础上增加语音播报功能

5.2.4 应用场景赋能

通过语音模块，极大提升四大Agent的交互体验：

Agent	语音交互场景示例	用户体验提升
小账（记小账（理财记录员）账员）	用户在忙碌（如驾驶、家务）时，直接说出：“记一下，刚才买咖啡花了28元。”系统通过语音识别并调用“小账”Agent完成记录，并语音回复：“好的，已为您记录一笔28元的餐饮支出。”	实现“动口不动手”的极速记账，记账效率提升300%以上，尤其适合多任务场景。
明查（财务洞察官）	"本月交通费多少？" → 语音播报统计结果 用户想快速了解财务状况，无需打开App筛选查询，直接语音询问：“我这个月在吃饭上花了多少钱？和上个月比是多了还是少了？”系统通过语音播报统计结果和对比分析。	将复杂的多步数据查询操作简化为一句问话，信息获取效率大幅提升，交互更加直观。
省省（省钱行动教练）	"推荐便宜的加油方式" → 语音给出省钱建议 用户在购物时，可实时咨询：“我想买这款手机，有什么省钱的门路吗？”系统结合实时网络搜索，通过语音给出比价结果、优惠券信息或替代方案建议。用户可随时语音查询长期目标进展：“检查一下我年底存5万块钱的目标，完成得怎么样了？”系统语音汇报当前进度、预测完成情况，并给出鼓励或调整建议	实时建议更自然在消费决策的关键时刻提供实时、主动的省钱指导，从“被动记录”变为“主动规划”，有效帮助用户省钱。
远谋（财务目标规划师）	"年底存款目标进度" → 语音汇报实现情况	目标追踪更直观使抽象的财务目标变得可感知、可追踪，通过语音播报增强目标达成的仪式感和动力，提升用户坚持完成目标的意愿。

5.3 开发里程碑规划

5.3.1 第一阶段：核心功能实现（当前）

- ✔ 四大Agent基础功能开发
- ✔ 文本交互界面实现
- ✔ 豆包模型集成调试
- ✔ 基础数据管理功能

5.3.2 第二阶段：语音模块集成（下一步）

周期：预计7-10天

- 接口对接（2-3天）：定义模块边界和API契约，完成基础通信
- 功能集成（3-4天）：语音输入接入消息循环，实现端到端测试
- 体验优化（2-3天）：性能优化和错误处理完善

5.3.3 第三阶段：体验优化（未来）

- 语音识别准确率优化（针对数字、金融术语）
- 多设备兼容性测试
- 个性化语音交互体验打磨

5.4 技术风险评估与应对

5.4.1 语音模块专项风险

- 音频质量问题：准备降级方案（自动转为文本输入）
- 网络延迟影响：实现本地缓存和智能重试机制
- 模型响应一致性：制定严格的回复格式规范

5.4.2 系统集成风险

- API兼容性：制定版本管理策略和回滚机制
- 数据格式差异：设计通用的数据转换适配层
- 性能影响：模块化部署，支持功能开关控制

5.5 测试验证策略

5.5.1 语音模块专项测试

- 单元测试：各语音处理组件独立功能验证
- 集成测试：与主程序接口的连通性测试
- 场景测试：四大Agent语音交互场景全覆盖
- 性能测试：并发压力和响应延迟测试

5.5.2 用户体验验证

- 真实用户场景测试（驾驶、家务等场景）
- 不同音频质量下的识别率测试
- 长时间语音交互的稳定性验证

5.6 总结与展望

语音交互模块的引入将使人机交互方式从"图形文本"升级到"自然对话"，为产品带来以下核心价值：

1. **交互革命**：从手动输入到自然语音，大幅降低使用门槛
2. **场景扩展**：支持驾驶、家务等不便打字场景，扩大使用范围

3. **技术储备：**为未来向智能音箱、车载设备等平台扩展奠定基础

4. **体验升级：**提供更人性化、更便捷的个人财务管家体验

通过模块化、分阶段的实施策略，确保语音功能平稳集成到现有系统中，同时为产品的长期发展构建坚实的技术架构基础。