Design Document

A functional Personal Financial Portal (GitHub repo link for coders or exported workflow file for no-coders). A 1–2 page design document detailing agent roles, workflows, and pattern implementation rationale.

ˆ Reuse from Phase 1 report ?

Reflection Essay

Tool Selection Trade-offs

Decision is made for quick prototyping and be more involved in our learning experience, therefore we use LangChain and SQLite ... TBD

ˆ Reuse from Phase 1 report ?

Ethical Considerations

There are several ethical considerations that immediately come to mind:

Automation Bias

Users may develop excessive trust in AI-generated financial insights the agent. This automation bias risks diminishing independent financial decision-making, especially among users with limited financial literacy. Even with "not financial advice" disclaimers, users may uncritically accept AI interpretations that impact their economic wellbeing. The system should implement contextual disclaimers and encourage verification with qualified professionals to maintain user autonomy.

Data Privacy

Financial data reveals sensitive personal information about spending habits, income, and merchant relationships. Sticklet introduces multiple exposure points through:

- Receipt processing via Mistral's OCR
- Queries handled by OpenAI models
- Transaction storage in unsecured SQLite database

Each integration requires clear data transmission protocols and processing limitations. Users must be informed about how their financial data traverses multiple systems and what protections exist against unauthorized access or misuse. Depending on where the user reside, privacy laws must be taken into consideration.

Security Vulnerabilities

The RAG pattern implemented in Sticklet creates potential attack vectors, particularly through the `SQLQueryTool` which could be vulnerable to SQL injection without proper input sanitization. The multi-agent architecture may also obscure security gaps between agent handoffs. Validation must be performed. Moreover, there are inherent risks with Language Models based input and output, such as recent cases in Jailbreaking, harmful outputs, etc. Sufficient guardrails and controls should be put in place.

Accuracy and Responsibility

While the Agent aims to improve data quality through self-reflection and human-reflection, there still can be mistakes. Moreover, the nature of Language Models output may not be necessarily true (hallucinating). Care must be taken into account to make sure the user is aware of such risks.

Lesson Learnt

TBD