FTEC 5660 Agentic AI for Business and FinTech HW1 Report

Name: Bowen Liu Student ID:1155254007

1. Introduction

This report presents a solution for Individual Homework 01 in the course **Agentic AI for Business and FinTech (FTEC5660)**. The system is designed to process supermarket receipt images and respond to user queries regarding financial information. The two supported query types are:

- **Query 1:** Total amount spent across receipts.
- **Query 2:** Original amount before discount.

  The system is also capable of rejecting irrelevant or out-of-domain queries.

2. Methodology

2.1 System Architecture

The solution uses Google's Gemini-2.5-flash model via Lang Chain with three components: (1) Query Classification Module to determine query type, (2) Image Processing Module to extract financial data, and (3) Calculation Module to aggregate values across receipts.

2.2 Query Classification

The system classifies queries into three categories using prompt-based LLM classification:

Table 1: Query Classification Categories

Category	Description	Example Queries
query1	Total amount spent	" How much money did I spend in total for these bills"
query2	Original amount before discount	" How much would I have had to pay without the discount?"
irrelevant	Unrelated questions	"What's the weather?"

2.3 Amount Extraction

For each query type, specialized prompts guide the LLM to extract appropriate values: (1) Load image as base64 data URL, (2) Construct query-specific prompt, (3) Invoke LLM, (4) Convert to float and aggregate.

3. Implementation

The system is implemented in Python using Google Colab, with the following key components and workflows:

3.1 Query Type Detection

The query classification module utilizes a prompt-based approach with Gemini-2.5-flash to categorize user queries into three distinct types.

Process Flow:

1. User query is received as text input
2. The prompt template is populated with the query
3. Gemini-2.5-flash processes the text prompt
4. The model returns a single classification label
5. System routes to the appropriate extraction module based on classification

3.2 Total Amount Extraction (Query 1)

For queries requesting the total amount spent, the system employs a specialized vision prompt that instructs the LLM to locate and extract the final payment amount from each receipt image.

Extraction Logic:

1) The model is instructed to find the "Total" or "Amount Due" field on each receipt
2) Emphasis is placed on extracting the final payable amount after all calculations
3) The prompt specifically excludes pre-discount amounts, taxes, or subtotals

Processing Steps:

1. Each receipt image is converted to base64 encoding
2. The specialized prompt is sent to Gemini-2.5-flash along with the image data
3. The LLM extracts the total amount value from each receipt
4. All extracted values are converted to floats and summed across all receipts
5. The aggregated total is formatted and presented to the user

3.3 Original Amount Extraction (Query 2)

For queries regarding the original amount before discounts, the system uses a different prompt strategy to locate pre-discount totals.

Extraction Logic:

- The model is directed to find the "Subtotal", "Original Amount", or "Total Before Discount" fields
- The focus is on amounts before any promotional discounts, coupons, or loyalty reductions
- If multiple amounts are present, the system selects the largest value that appears before discount-related fields

Processing Steps:

1. Image preprocessing and base64 conversion (same as Query 1)
2. Application of the original amount extraction prompt
3. LLM analysis of receipt structure to identify pre-discount totals
4. Extraction and conversion of numerical values
5. Aggregation across all provided receipts
6. Presentation of the total original amount that would have been paid without discounts

4. Results

4.1 Query 1: Total Amount Spent

Query 1 Result: Total = $1,974.30 | Expected = $1,974.3| ≤2

4.2 Query 2: Original Amount Before Discount

Query 2 Result: Total = $2,348.20 | Expected = $2,348.2 | ≤2

5. Conclusion

The system successfully demonstrates multimodal LLM application for intelligent document processing. Key achievements include: (1) Accurate query classification into three categories, (2) Precise financial extraction from receipt images using specialized prompts, (3) Correct aggregation across multiple receipts, and (4) Passing all test cases with exact matches. The solution leverages Gemini-2.5-flash's vision capabilities with structured prompting. The modular design enables future extension for additional query types.