Report on Automated Financial Receipt Auditing via Large Language Models

Ruiqian Luo,1155244663

Summary

In the assignment, For the two queries, the llm model should be able to perform 2 queries:

1.Total Spend Calculation: Extract the actual payment amount from each receipt and sum them up.

2.Original Price Reconstruction: Based on the logic of "actual payment amount + discount amount," reconstruct the original total price before discounts.

My llm demonstrates exceptional robustness when processing explicit data, such as the total expenditure in Query 1, achieving 100% accuracy. However, when processing implicit inter-row data (e.g., original price calculations in Query 2), the system still encounters accuracy limitations. Despite employing sequential auditing and reflection mechanisms, subtle in-line discounts are occasionally overlooked by the visual model, resulting in certain deviations in the final output.

Discount information on receipts often appears in various formats (such as negative numbers between lines, bottom totals, or percentage reductions). When llm perform discounted price calculation, there will be a precision loss during multi-step floating-point calculations, leading to results that deviate from the ground truth. I've tried numerous prompt writing approaches, including routing and reflection, but the results from extracting the discount value and performing calculations consistently show errors.

For queries outside our scope of interest, the prompt performs quite well. The main challenge in this assignment lies in the large language model consistently failing to extract discount prices accurately across different receipt types, resulting in persistently high error rates for query 2. I can't come up with a good way to solve this problem.

Code

The coding part mainly follows the instructions of the Notebook 1, The main changes are in the block of image input to the llm. Details are in the ipynb files.