Simran Sinha
002475433

Prompt Engineering Impact on GenAI in Supply Chain Decision Making

Introduction

The Beer Game is a classic supply chain simulation that can be used to investigate demand variability, inventory, and decision-making in an uncertain environment. In this research, Generative AI is applied as a decision-making tool to decide on weekly order quantities for the retailer. Instead of modifying the original simulation code, various prompts are applied to influence the decision- making process of GenAI. Prompt engineering is considered a major factor in affecting GenAI's understanding of goals, limitations, and trade-offs. In this research, various prompt strategies are applied to analyze GenAI's decision-making processes in a multi-week environment.

Methodology

The experiment is specifically designed to test the retailer segment of the Beer Game supply chain.
The experiment will be conducted over 10 weeks of simulation (5 Iterations). Three prompt strategies will be tested. Each of these prompt strategies will have its own unique characteristics of constraint and guidance provided for the decision-making process.
For all prompt strategies:
- GenAI would be utilized to compute the order quantity for the week.
- The total costs would be computed for the above runs.

Prompt 1: Baseline (Full GenAI)
- The prompt is relatively flexible with less constraint.
- GenAI has the liberty to compute the order quantity.

Prompt 2: Cost-Aware Conservative
- Emphasis is given to minimizing costs.
- The approach is relatively conservative to prevent inventory build-up.

Prompt 3: Risk-Constrained Data-Driven
- The prompt is relatively structured.
- The decision-making process is relatively well defined.
- The decision variables are well constrained for the order quantity.

Evaluation Metrics
- Total cost calculated for all runs.
- Average total cost and Standard deviation of total cost.

Results

Prompt 1

Experiment 1	Prompt 1 (Baseline – Full GenAI)
Run 1	815
Run 2	590
Run 3	742
Run 4	796
Run 5	772
Average	743
Standard Deviation	89.7830719

Prompt 2

Experiment 2	Prompt 2 (Cost-Aware Conservative)
Run 1	1947
Run 2	1581

Run 3	2308
Run 4	2249
Run 5	1701
Average	1957.2
Standard Deviation	322.2874493

Prompt 3

Experiment 3	**Prompt 3 (Risk-Constrained Data-Driven)**
Run 1	6166
Run 2	3640
Run 3	4265
Run 4	4401
Run 5	6127
Average	4919.8
Standard Deviation	1156.087237

Analysis

- Prompt 1 (Baseline – Full GenAI) had the minimum average cost and variability. Hence, the decision-making capabilities of GenAI are good in this case. However, the results might be highly dependent on the demand pattern.
- Prompt 2 (Cost-Aware Conservative) had higher average costs with moderate variability. The conservative approach reduced the variability in inventory but resulted in inefficient decision- making in the case of an increase in demand.
- Prompt 3 (Risk-Constrained Data-Driven) had the highest average cost and maximum variability. The risk constraints resulted in inefficient decision-making capabilities by creating excess inventory.

Key Insights

The design of prompts can have a major impact on Gen AI decision results.

- Over constraining Gen AI can lead to increased costs due to too many safety buffers.
- Flexible prompts can be better than conservative ones if demand variability is moderate.
- Variability (or standard deviation) is as important as average cost in evaluating prompts.

Suggestions for Improvement

To experiment this model further, it might be helpful to improve its performance through the use of a hybrid approach, which will give the benefits of unconstrained GenAI with lightly applied constraints, which can be adaptive also. This means, rather than applying strict constraints, these safety buffers and ordering limits might be made dynamic depending on the volatility of demand and inventory levels. Another area for improvement with this model might be the inclusion of a feature for a short-term demand forecasting tool, rather than relying on historical averages, which might prevent inventory levels from getting out of hand. Further experimentation with these prompts, although with varying levels of volatile demand, might be a better measure for evaluating the viability of this system.

Conclusion

This experiment also demonstrates that the performance of GenAI in decision-making within the supply chain is significantly dependent on the structure of the prompt. Although the risk-aware prompt ensures the system's safety theoretically, this may also cause inefficiencies if the system's constraints are too tight. The balanced structure of the prompt is essential to ensure the system's cost efficiency and stability.