E-commerce Customer Service: GA-Driven Chatbot Prompt Optimization

Open D/I Hackathon 2023 Challenge 3: LLM Genetic Algorithm

Team「**Hackit**」Hackathon Project Approach

Team Members

Team Lead	Lei Yang
Data Scientist	Ambreen Hanif
Data Engineer	Abrren Chen

Table of Contents

Introduction

Project Objective

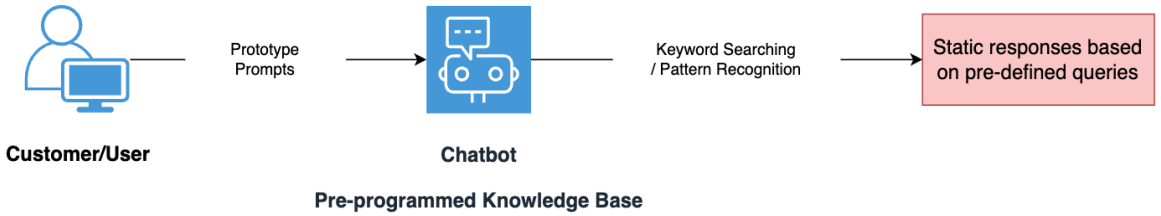
This solution aims to optimise Ecommerce Customer Service (CS) Chatbot by utilizing Genetic Algorithm (GA) and Large Language Model (LLM).

Problem Description

Traditional rule-based chatbots rely on predefined rules and patterns to generate responses. These bots follow a set of programmed instructions and keyword-based algorithms to interpret and respond to user inputs. The traditional chatbot has the following limitations:

1. Limited flexibility - They struggle to understand user inputs outside their programmed parameters, leading to a negative user experience.
2. Complex conversations - Handling intricate or multi-layered dialogues is difficult, often resulting in irrelevant or incomplete responses.
3. Lack of personalization - They are unable to offer user-specific interactions, making conversations feel impersonal.
4. Time-consuming updates - Continuous maintenance and updates consume time and resources, impacting their effectiveness.
5. Incapacity to learn - Unlike AI-powered chatbots, they cannot learn from user interactions and improve their responses.
6. Inability to Empathise / User Frustration.

7. Integration Challenges.
8. Escalation to Human Agents.
9. Language Limitation.



Solution Design and Development

Proposed GA-Driven Chatbot Solution

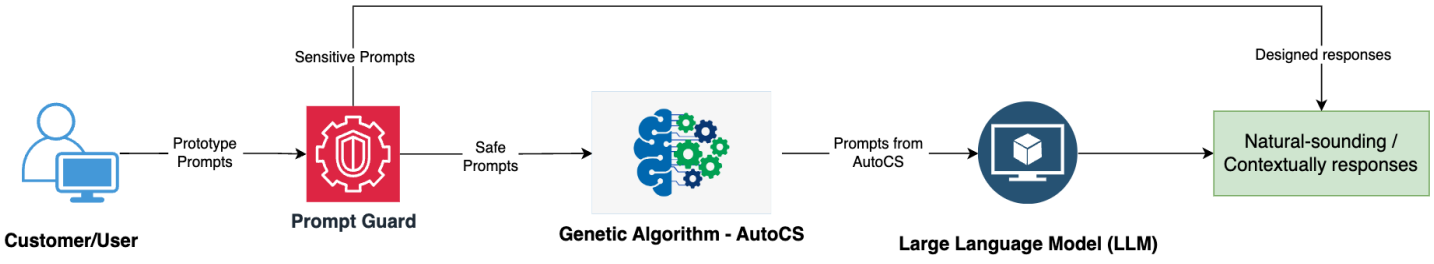
Our proposed chatbot solution utilises genetic algorithm to generate optimised prompts based on user input and feed the best prompt into our large language model. The main highlights of this solution are:

- Genetic Algorithm is used to generate the best prompt which makes the client's questions easier for LLM to understand.

- The LLM-powered chatbot will respond client's inquiry in a human-like manner and generate contextual responses.

The GA-Driven Chatbot has the following advantages over rule-based chatbot:

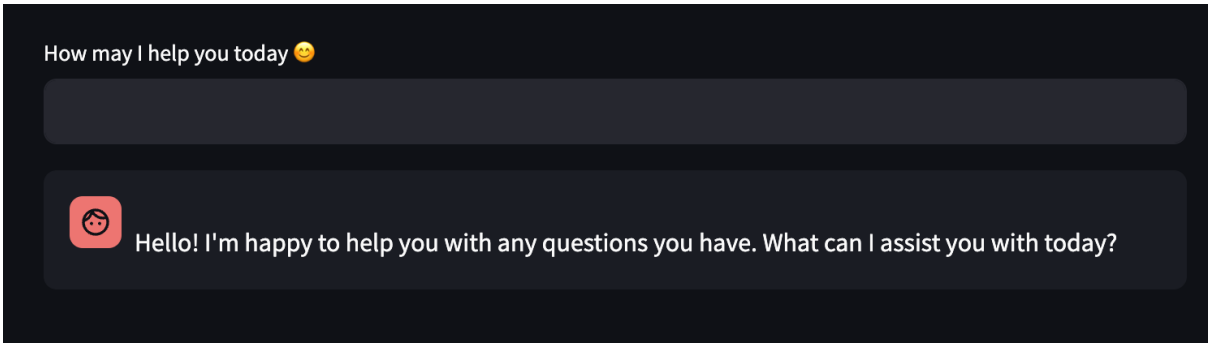
1. Improve contextual understanding.
2. Enhance response generation (natural-sounding & contextually appropriate answers).
3. Handling complex queries.
4. Continuous learning and adaptation.
5. Personalisation and customer experience.
6. Handling multilingual support.
7. Human-like interactions.
8. Increase efficiency.



Target Solution Breakdown

User Input

In this case, user inputs are customer's inquiries regarding the platform or online shopping. In our solution, we developed a simple chatbot. Therefore, the user inputs will be any inquiries customers enter in the chat.



Prompt Guard

Given the capacity of language models to learn and replicate information based on the input they receive, using a prompt guard helps maintain ethical standards and ensures that the model produces responsible and safe outputs. By filtering out sensitive content before it reaches the model, it aims to minimize the risk of generating problematic or inappropriate responses.

On Ecommerce platform, customers may use sensitive words when they are very frustrated with the survive and delivery. As large language models are not fully filtered by these sensitive information, adding a prompt filter is very important. It will improve the quality of customer service.

There are six key topics which the filter will apply on:

1. Moderation
2. Prompt Injection
3. Persional Identity Information
4. Sentiment
5. Unknown Links
6. Relevant Language

If the prompt contains any of the above topics, this prompt will be flagged and we will generate a designed response to this inquiry without sending this prompt into LLM.

- "moderation": "For assistance, kindly use respectful language. How can I help you with your inquiry?",

- "prompt_injection": "Sorry, I cannot answer this answer. Please provide another inquiry.",
- "pii": "For information safety, please do not provide any personal details.",
- "sentiment": "For assistance, kindly use respectful language. How can I help you with your inquiry?",
- "unknown_links": "Sorry, I cannot access any links via the chat.",
- "relevant_language": "Sorry, we do not support any other languages except English, please describe your inquiry in English."

Genetic Algorithm - AutoCS

Initial Population

We constructed our initial prompt dataset by formulating a prompt in various ways and then paired these with Llama to elicit responses. Ideal responses were handcrafted to establish a baseline for evaluating the effectiveness of the prompts via the fitness function.

Fitness Function

The fitness function was designed to consider multiple-objective metrics, including Relevance, Informativeness, and Engagement:

- For the Relevance metric, we utilised BERT embeddings and cosine similarity to compare the semantic correspondence between LLM responses and the ideal responses.
- For Informativeness, a comparison of key words was implemented.
- For Engagement, we combined Textblob sentiment analysis with Click-Through Rate (CTR) assessments to gauge customer engagement.

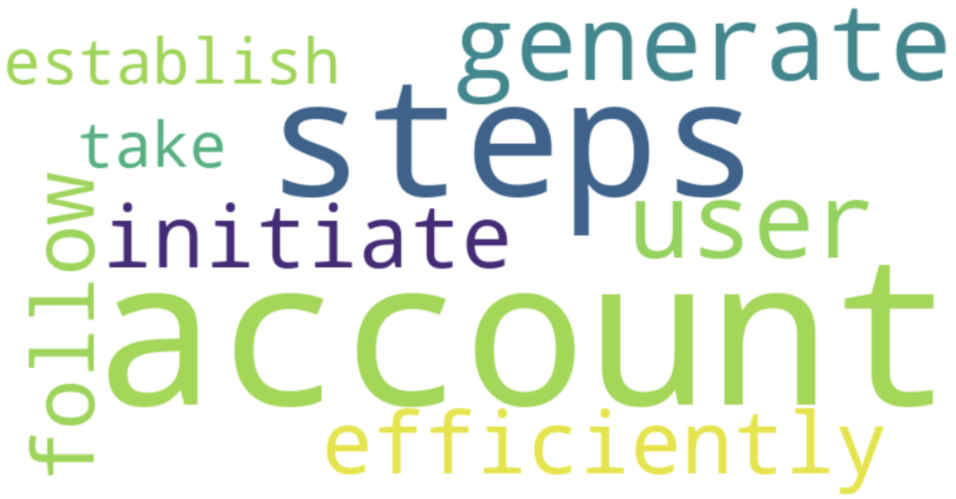
Weights were assigned to these metrics; in our case, they were distributed as 0.4, 0.4, and 0.2, respectively. The fitness function returned the sum of the weighted scores.

Crossover

Parents with the best fitness scores were selected and subjected to the Crossover Function. Initial experiments that involved combining parts of parent prompts or using synonyms sometimes resulted in prompts that were not logical or grammatically correct—for example, “Why was my I report an issue with the website?” Consequently, we opted for a straightforward crossover function and allowed the mutation function to generate more creative prompts.

Best Prompt Analytics

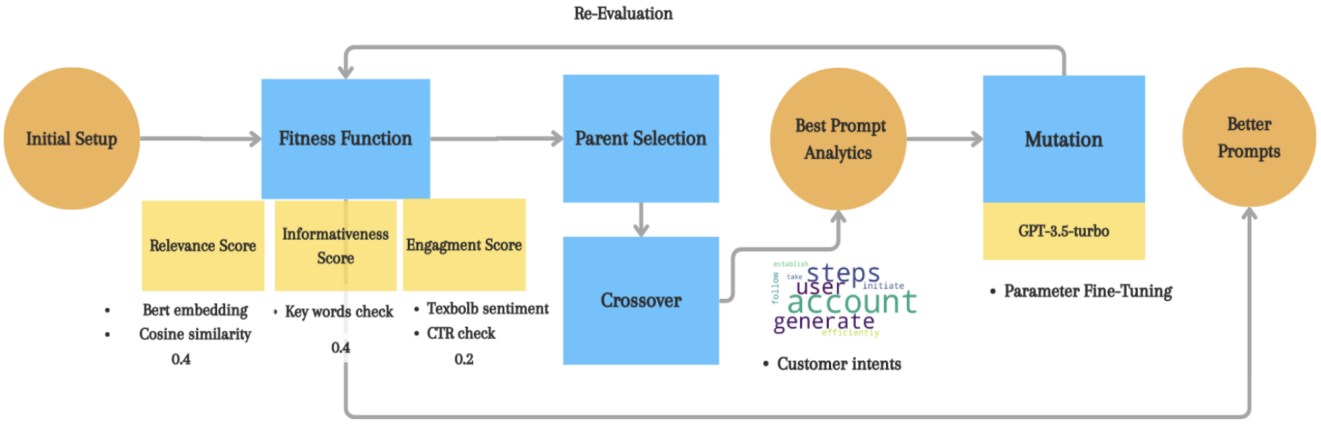
We performed analytics specifically on the best prompts, which possessed the top fitness scores. This analysis provided insights into common words used in the best prompts and helped detect customer intentions.



Mutation

The mutation function was enhanced by GPT-3.5-turbo, which utilised insights from the outcomes of preceding steps to generate diverse prompts. These prompts included top words from the previous steps while preserving customer intention. We fine-tuned the GPT model with the parameter temperature = 0.7, striking a well-balanced mix of creativity and coherence.

Prompts generated by the GA were re-evaluated using the Fitness Function to recalculate fitness scores. The prompt with the highest score was chosen as the new best prompt after a specified number of iterations or upon the model's convergence.



Large Language Model (LLM)

The large language model which we integrated in this solution is Llama 2 Chat with 70 B parameters. In this solution, we request a post call from the Hugging face Llama 2 Inference API.

For detailed description of this model, please check this Hugging Face page.

For simple instruction code on how to deploy inference API, please check this page.

For Llama 2 inference API documentation, please check this documentation page.

Future Work

Given the limited time and resources of this hackathon project, we only create a prototype of this solution. There are many areas of this project which can be improved to make the solution more practical.

- Improve the interface of the chatbot and add more user-friendly functions.
- Fine-tuning the large language model based on a specific user case, e.g., we can train the LLM on a specific platform and build up knowledge base for this platform.
- Improve the prompt guard filter ability to detect more sensitive information.
- Train the genetic algorithm with larger dataset and make it more practical and generlised.