***Comparative Analysis of NLP and OpenAI-based Chatbots: A User-Centric Evaluation***

**Vignesh Suvarna**

School of Computing Science

Sir Alwyn Williams Building

University of Glasgow

G12 8QQ

A dissertation presented in part fulfilment of the requirements of the

Degree of Master of Science at The University of Glasgow

August 2023

**Abstract**

The creation and implementation of chatbots have emerged as essential tools for seamless communication and information exchange in a time when human-computer interaction is quickly evolving. In this paper we discuss the development and compare the performance of two different chatbot models: an OpenAI-powered chatbot that uses sophisticated language models and an NLP-based chatbot that uses traditional NLP methods. The objective is to assess their respective performances and user preferences through a comprehensive user survey.

The study involves the design and development of both chatbots, each representing a different approach to conversational AI. The subsequent evaluation encompasses response accuracy, contextual understanding, and user satisfaction. A user survey gathers insights on clarity, coherence, usefulness, and overall experience.

This research examines the advantages and disadvantages of each strategy by contrasting the capabilities of state-of-the-art AI-driven chatbots with more conventional NLP-driven analogues. The research provides developers and organisations with subtle insights that may be used to make decisions that are appropriate for a certain use case. This study provides relevant insights into the changing chatbot technology environment as AI and NLP transform human-computer interaction. As AI and NLP continue to reshape human-computer interaction, this dissertation serves as a timely and insightful contribution to the ongoing discourse in the field.

**Education Use Consent**

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic format. Please note that you are under no obligation to sign this declaration, but doing so would help future students.

Name: Vignesh Suvarna Signature: Vignesh Suvarna

**Acknowledgements**

At the special moment of completing this dissertation paper, the one-year master’s life at the University of Glasgow is coming to an end. I experienced a wonderful time living and studying at the University of Glasgow.

I like to take this opportunity to express my sincerest gratitude to my supervisor, classmates and friends who have helped me.

Thanks to my supervisor, a patient and responsible professor. He has given a lot of guidance in the paper’s topic selection, conception and writing. The completion of this project is inseparable from my supervisor’s direction and discussion.

Thanks to my classmates and friends for helping me learn to use standard tools to write papers and reminding me of many precautions when I am busy. The on-time completion of this paper is inseparable from your patient reminders.

Finally, I would like to thank my family and girlfriend for their great help and care during my postgraduate life, making my life full of hope and vigour.

”The road ahead will be long. Our climb will be steep.” I will use this poem to spur myself to overcome obstacles and forge ahead on the road of life in the future.

Contents

[1 Introduction 5](#_Toc49458)

[2 Background 6](#_Toc49459)

[2.1 Numerical Weather Prediction 6](#_Toc49460)

[2.2 Successive Over-Relaxation Algorithm 6](#_Toc49461)

[2.3 Message Passing Interface 8](#_Toc49462)

[2.4 ArgoDSM 9](#_Toc49463)

[3 Design & Implementation 10](#_Toc49464)

[3.1 Data dependency analysis 10](#_Toc49465)

[3.2 Design of SOR Algorithm Based on MPI 11](#_Toc49466)

[3.3 Implementation of SOR Algorithm Based on MPI 12](#_Toc49467)

[4 Testing & Evaluation 14](#_Toc49468)

[4.1 Testing environment 14](#_Toc49469)

[4.2 Testing procedures 15](#_Toc49470)

[4.2.1 Summarize testing parameters 15](#_Toc49471)

[4.2.2 Deploy testing instances 16](#_Toc49472)

[4.3 Testing result evaluation 16](#_Toc49473)

[4.3.1 The effect of compiler optimization flag 16](#_Toc49474)

[4.3.2 The overall performance of each implementation 18](#_Toc49475)

[4.3.3 The SOR performance on fixed array volume 19](#_Toc49476)

[4.3.4 The effect of node count on MPI and ArgoDSM 20](#_Toc49477)

[4.3.5 The effect of dsmNX and dsmNY on ArgoDSM 22](#_Toc49478)

[5 Conclusion & Future Work 23](#_Toc49479)

**Chapter 1**

**Introduction**

In an era defined by rapid technological evolution, the convergence of artificial intelligence (AI) and natural language processing (NLP) has led to the proliferation of chatbots, revolutionizing the way humans interact with machines. This paper launches thorough research into the capabilities of two distinctive paradigms within the chatbot landscape: the cutting-edge OpenAI model and a Natural Language Processing (NLP) model powered by TensorFlow, a flexible machine learning framework. Chatbots have developed in the modern environment from simple scripted responses to dynamic beings capable of contextual comprehension and natural interaction. These virtual companions now have a wide range of uses, including customer service in e-commerce, smartphone virtual assistants, and even mental health care. This growth underscores their crucial function in speeding information transmission and improving user experiences in a variety of industries.

Chatbots have orchestrated a transformative shift in the internet landscape, ushering in a new era of interactive digital experiences. By transcending traditional static web interfaces, these intelligent agents have made it possible for dynamic, real-time discussions to mimic human interactions. They have a significant positive effect on people, improving customer service, boosting information accessibility, and personalising user experiences. Chatbots have increased user engagement by promoting quick responses and cutting down on wait times, providing a seamless link between users and digital services. As virtual companions, they have improved operational effectiveness while also changing how people interact with technology, making the internet more user-friendly, approachable, and responsive.

In the fields of artificial intelligence (AI) and machine learning, OpenAI is a well-known research institution. They have significantly contributed to creating sophisticated language models like GPT. These models are designed to understand and generate human-like text by training on vast amounts of diverse textual data.

Natural Language Processing (NLP) makes it possible for computers to meaningfully and accurately read, interpret, and produce human language. NLP techniques span a wide range of methodologies, including rule-based systems and machine learning-based approaches. Language translation, sentiment analysis, text parsing and analysis to extract information, among many other things, are some of these techniques. TensorFlow, a cornerstone of modern machine learning, stands as a vital enabler, facilitating the creation of sophisticated neural architectures that underpin the NLP model's ability to discern language nuances and formulate coherent responses.

This research aspires to decipher the capabilities and constraints of two divergent paths of chatbot development: the forefront of OpenAI's innovation and the functional prowess of NLP techniques bolstered by TensorFlow. By meticulously evaluating their performances and soliciting user perspectives, this study not only illuminates their inherent potential but also navigates the complex interplay between innovation and established methodologies.

The subsequent chapters of the dissertation are as follows:

* Chapter 2
* Chapter 3

**Chapter 2**

**Background**

2.1 Evolution of Chatbots

The emergence of chatbots is an enthralling story that shows the ongoing interaction between technical innovation and human communication demands. Its origins can be traced back to early AI studies intended to replicate human conversation. ELIZA, created by Joseph Weizenbaum in the mid-1960s, is one of the pioneering examples. ELIZA, functioning as a Rogerian psychotherapist, engaged users in text-based dialogues that mimicked empathetic responses, even though it operated on simple pattern-matching rules. In the late 1990s, ALICE (Artificial Linguistic Internet Computer Entity) emerged, which incorporated a broader rule-based framework and broadened the repertory of conversational themes. Jabberwacky, which was developed at the same time, took a more user-driven approach, learning from user inputs to generate its responses. Despite their shortcomings in understanding context and producing meaningful dialogues, these early chatbots set the framework for future endeavours.

In the late 1990s, ALICE (Artificial Linguistic Internet Computer Entity) emerged, which incorporated a broader rule-based framework and broadened the repertory of conversational themes. Jabberwacky, which was developed at the same time, took a more user-driven approach, learning from user inputs to generate its responses. Despite their shortcomings in understanding context and producing meaningful dialogues, these early chatbots set the framework for future endeavours. The late 2000s marked the rise of virtual assistants like Apple's Siri, which introduced a novel dimension by enabling voice-based interactions. These systems incorporated speech recognition, natural language understanding, and access to vast data repositories to answer queries and perform tasks. With the advent of cloud computing, these voice-driven chatbots became increasingly sophisticated and capable.

The turning point arrived with the surge of machine learning and NLP advancements in the 2010s. Chatbots like IBM's Watson and Microsoft's Xiaoice combined machine learning techniques with language understanding to provide more contextually relevant responses. However, it was the introduction of deep learning techniques and neural networks that truly revolutionized chatbot capabilities. The release of OpenAI's GPT series, beginning with GPT-1, signalled a revolution in chatbot development. Massive datasets and transformer topologies were used in these models to generate text that was strikingly human-like in its coherence and contextuality. This trend was followed by GPT-2 and GPT-3, with GPT-3 having an unprecedented amount of parameters and the ability to engage in a variety of conversational contexts.

2.2 Chatbots' impact on the digital landscape

Chatbots have significantly changed the online landscape by incorporating interactivity into digital interfaces. Their ability to hold real-time discussions with users has transformed customer service, e-commerce, and information dissemination. Chatbots improve user experiences, promote engagement, and optimise operational efficiency by providing personalised recommendations and reacting quickly to concerns. E-commerce has been invigorated by personalized recommendations, simplifying the shopping experience. Additionally, chatbots' ability to swiftly disseminate information and respond to queries has revolutionized information accessibility, making the digital realm more user-centric and responsive than ever before.

2.3 Types of Chatbots

The landscape of chatbots is diverse, encompassing various types that cater to distinct user needs and interaction scenarios. By embracing machine learning, these chatbot variants transcend traditional static responses, paving the way for more engaging, efficient, and human-like conversations. These types can be broadly categorized as follows:

2.3.1 Rule-Based Chatbots

Rule-based chatbots operate on predefined rules and scripts. They follow a decision-tree structure to match user inputs with predetermined responses. While simple to implement, they are limited in their ability to handle complex conversations and adapt to dynamic user queries. However, machine learning adds a layer of adaptability, enabling them to refine answers based on user interactions and feedback. Machine learning comes into play by observing user inputs and learning from their behaviour over time. The chatbot can then adjust its responses, assigning weights to different rules based on user preferences.

2.3.2 AI-Powered Chatbots

AI-powered chatbots utilize machine learning and natural language processing to understand user inputs and generate responses. These chatbots can grasp context, making interactions more dynamic and human-like. OpenAI's GPT-3 exemplifies this category, pushing the boundaries of AI-driven conversational agents. Unlike rule-based chatbots, AI-powered variants do not rely solely on predefined scripts. Instead, they leverage NLP models and deep learning algorithms to understand the nuances of user inputs. These chatbots process text data through layers of neural networks, allowing them to grasp context, sentiment, and intent. By analysing vast amounts of training data, AI-powered chatbots can recognize patterns in language usage, enabling them to generate human-like responses that resonate with users.

A diagram of types of chat bots

Description automatically generated

**Figure 1**: Types of Chatbots

2.4 Related Works

Several chatbots have developed as examples of seamless user involvement and assistance in the sector of e-commerce customer service. These examples highlight the importance of AI-powered conversational bots in improving client experiences in the e-commerce landscape.

Shopify's Kit, a trailblazing chatbot, stands as a testament to the transformative influence of AI in e-commerce operations. Kit goes above and beyond standard customer encounters, helping organisations streamline many aspects of their operations. Kit demonstrates how AI can handle complex jobs effectively, improving the operational effectiveness of e-commerce businesses. Kit handles duties including developing marketing campaigns, administering social media promotions, and tracking orders. With Kit's assistance, merchants can allocate more time to core business activities, as the chatbot tackles routine tasks. This innovative approach to customer support transforms the role of chatbots from mere query responders to dynamic allies in driving business success. As AI technology advances, Kit stands as a model for how chatbots can evolve into indispensable tools that not only enhance customer interactions but also contribute to the overall growth of e-commerce ventures.

eBay's ShopBot stands as a testament to the potential of AI-driven chatbots in facilitating guided and personalized shopping experiences. ShopBot engages users in dynamic engagements that go beyond simple browsing within the comfort of Facebook Messenger. ShopBot acts as a virtual shopper by providing individualised product recommendations, facilitating pricing comparisons, and supporting decision-making. Its seamless integration into well-known messaging services redefines convenience by giving users access to insightful buying information at their fingertips. ShopBot embodies the paradigm change in user engagement, where AI uses conversational interactions to improve customer pleasure, support informed decisions, and streamline the difficulties associated with online purchasing. By conversing with users in a natural and interactive manner, ShopBot creates an immersive shopping experience. Through natural interactions, users may ask questions about products, get personalised suggestions, and learn about pricing patterns. The shopping experience is made more interesting and educational by this interface, which democratises access to eBay's extensive inventory. The AI-powered chatbot ShopBot is a prime example of how AI-powered chatbots can change shopping from a transactional chore into a stimulating adventure where consumers receive tailored help and make informed decisions in real-time.

Lidl's Margot shines as an illuminating example of how AI-driven chatbots can transcend traditional customer interactions to offer specialized assistance within distinct product domains. Margot's emphasis on wine suggestions exemplifies how precisely chatbots may respond to certain user demands, boosting the personalised buying experience. Margot's involvement starts with pleasant questions about consumers' wine tastes, inviting them to open up in a conversational manner about their taste preferences and desired qualities. This friendly manner not only makes conversations fun, but also guarantees that the recommendations fit with each person's interests. Margot uses natural language processing techniques to read and understand the subtleties of users' responses as they converse, gathering information that informs its wine recommendations. Lidl's Margot is emblematic of the evolving role of chatbots in e-commerce, transcending mere transactional exchanges to become interactive shopping companions. By providing personalized product suggestions, Margot demonstrates how AI-driven chatbots can transform the shopping experience into a dialogue, fostering a deeper connection between users and brands. Margot establishes a fantastic example for AI-powered e-commerce chatbots by streamlining wine selection and boosting customer engagement and happiness.

H&M's Chatbot, an AI-powered virtual stylist, epitomises the fusion of AI and fashion, heralding a new era of personalised shopping experiences. It plays the part of an expert fashion counsellor by having fun dialogues with users about their personal style preferences, outfit requirements, and fashion goals. It exchanges have a sympathetic tone that gives people the impression that they are speaking with a real stylist. Users can choose their desired looks and situations, and even add reference photographs. This chatbot uses sophisticated natural language processing to evaluate these inputs and derive insightful information that serves as the basis for its suggestions. ability to transform user choices into carefully chosen outfit ideas from H&M's vast selection is what gives it its true brilliance. It displays ensembles with components that correspond to users' inputs and appeal to their sense of fashion. It's AI-driven capabilities allow it to comprehend fashion nuances and provide recommendations for everything from casual attire to formal outfits. H&M's AI chatbot redefines how consumers interact with companies online by transforming the purchasing experience from transactional to immersive. It fosters a personal relationship between customers and the H&M brand by providing tailored fashion recommendations. This virtual shopping assistant, exemplifies the potential of artificial intelligence to bring humanity to online shopping by providing individualised advice that takes into account consumers' uniqueness and preferences.

ASOS' Enki emerges as a striking embodiment of the synergy between AI and e-commerce, ushering in a new era of personalized fashion discovery. Enki, a chatbot that can be accessed through Facebook Messenger, acts as a virtual stylist, revolutionising how users discover and interact with fashion trends. The core strength of Enki is its capacity to have relatable and authentic conversations with users. It gains useful insights about consumers' fashion tastes, social contexts, and outfit needs by including them in dynamic interactions. With this contextual knowledge in hand, Enki continues to develop personalised fashion recommendations based on specific preferences. These recommendations cover a wide range of fashion accessories, footwear, apparel, and makeup, resulting in a thorough and engrossing fashion experience. Enki serves as a proactive stylist with its proactive involvement, providing trend insights, outfit suggestions, and even allowing users to visualise whole ensembles. It fills the gap between users' desires for fashion and ASOS's vast selection by focusing choices on those that complement each user's individual style the best. The efforts of Enki go beyond merely suggesting products. It forges an emotional bond between customers and the ASOS brand by encouraging fashion exploration and encouraging users to try out various looks. It exemplifies the function of chatbots in fostering consumer loyalty and involvement, enhancing ASOS' standing as a fashion authority that values individual tastes. Enki exemplifies the transformational effect of AI-powered chatbots in the dynamic e-commerce industry. It increases productivity while also adding a sense of companionship and direction to the shopping experience. Enki, an AI-powered conversational interface from ASOS, is an excellent example of how the blending of conversational interfaces, e-commerce platforms, and AI can transform the process of shopping into a fun, exciting voyage of fashion discovery.

Chatbots have become crucial tools for boosting customer service and engagement in the world of e-commerce. The above mentioned chatbots illustrate the potential of conversational agents by exemplifying the capabilities of AI-driven assistance, guiding users through shopping journeys, offering personalised recommendations, and creating an engaging and effective shopping experience.

**Chapter 3**

**Design and Implementation**

This chapter outlines the intricacies associated with creating and deploying both the NLP-driven and OpenAI-integrated chatbots. The exhaustive overview covers the design specifications, technology decisions, and the integration of OpenAI's capabilities. The chapter also elaborates on the user survey methods used to obtain insightful viewpoints on the relative effectiveness of both chatbots.

3.1 System Architecture

The architecture of the chatbot comparison study encompassed two distinct models: an OpenAI-integrated chatbot and an NLP-driven chatbot. Each model consisted of frontend and backend components, facilitating user interaction, processing of inputs, and generation of contextually relevant responses.

A diagram of a chatbot

Description automatically generated

**Figure 2**: Working of a Chatbot

3.1.1 OpenAI Integrated Chatbot Architecture

The architecture of the OpenAI-integrated chatbot engages users in an interactive dialogue facilitated by a seamless client-server model. This framework capitalizes on modern web technologies, harmonizing HTML, CSS, and JavaScript to create a user-friendly interface that empowers users to engage in dynamic conversations. The front end serves as the gateway for user inputs, composed of a well-crafted blend of HTML and CSS that fosters an intuitive and visually appealing environment. Within this interface, users are enabled to input messages and interact with the chatbot effortlessly. User inputs transition seamlessly to the backend through streamlined HTTP requests, orchestrated by a combination of JavaScript and other components. The backend facilitates the creation of contextually pertinent responses by acting as a bridge between the front end and OpenAI's API. Upon receiving user inputs, the backend communicates with OpenAI's API using meticulously crafted HTTP requests. These requests encapsulate the user's message and relevant model specifications, harnessing the prowess of the GPT-3 model for response formulation. The responses created by the OpenAI model are sent back to the front end and stitch a conversation together. These responses elegantly integrate within the chat interface, allowing users to witness the exchange evolve in real time. This architecture ensures the OpenAI-integrated chatbot leverages the complete capabilities of the GPT-3 model while offering a user experience that is both intuitive and engaging. The harmonious synergy of frontend, backend, and the OpenAI API delivers an interactive and insightful exchange, exemplifying the capacity of AI-driven chatbots to elevate user interactions in a seamless and sophisticated manner.

3.1.2 NLP-Driven Chatbot Architecture

The NLP-driven chatbot architecture leveraged Python's NLTK library and TensorFlow for natural language processing and machine learning capabilities. Similar to the OpenAI-integrated chatbot, the frontend interface enabled users to interact with the chatbot through text inputs. The front end communicated user messages to the backend server via HTTP requests. The backend, developed in Python and integrated with Flask, performed NLP tasks using NLTK and TensorFlow. It involved tokenizing user messages, analysing context, and generating responses. The TensorFlow model, trained on a curated dataset of conversations, played a pivotal role in comprehending user inputs and producing coherent replies. Responses generated by the TensorFlow model were relayed back to the front end, where they were displayed to the user. This architecture showcased the chatbot's ability to understand user inputs, process them through NLP techniques, and generate contextually relevant responses.

A screenshot of a computer

Description automatically generated

**Figure 3**: Front-end UI of Chatbot

A screenshot of a computer

Description automatically generated

**Figure 4**: Chatbot generating a response to the user’s query

3.2 Data Collection and Pre-processing

Quality and relevancy of the training data form the basis of any successful chatbot. In the pursuit of building a robust NLP-driven chatbot, meticulous data collection and pre-processing were pivotal stages. The chatbot's ability to understand user intents precisely is shaped by the training data, which also has an impact on the chatbot's conversational skills. The data collection process commenced with a purpose-driven approach, focusing on extracting insights from the most frequently asked questions (FAQs) encountered on e-commerce platforms. By taking this strategy, we made sure that the chatbot would be prepared to respond to frequent customer inquiries, improving user experience. Sources for frequently asked questions included official websites, forums for customer service, and discussion boards for online shopping.

After gathering the FAQ data, the process of pre-processing was initiated to refine the dataset for training. The text was tokenized into individual words using the Python Natural Language Tool Kit (NLTK) package, which helped break down phrases into their basic linguistic components. Then, to remove variances and improve the model's capacity to generalise across synonyms and related keywords, lemmatization was used to transform words into their basic forms. This process helped to keep the vocabulary's core meaning while minimising its dimensionality.

A "bag of words" representation for each training instance was carefully created to assure the training dataset's effectiveness. This representation captured the occurrence of words in each sentence, effectively transforming textual data into numerical features that could be ingested by machine learning algorithms. Each occurrence was also linked to a corresponding intent label, which made supervised learning easier.

The full dataset was then divided into NumPy arrays to simplify the model training process and randomised to remove any inherent order bias. The training of the sequential model, which involved the neural network being iteratively altered to acquire the intricacies of language patterns, associations, and purpose classifications, was greatly aided by this transition.

In conclusion, the procedure for gathering data and pre-processing represented a thorough effort to convert real-world e-commerce FAQs into a specific training corpus. The training data was prepared for the sequential model's learning process by using techniques like tokenization, lemmatization, and the creation of a "bag of words" representation, allowing the chatbot to understand and respond to user queries in a way that is both informative and contextually relevant.

A diagram of a data processing process

Description automatically generated

**Figure 5**: Data Preprocessing Diagram

3.3 Model Training and Integration

The heart of the NLP-driven chatbot's prowess lies in the meticulous training and integration of its underlying model. The implementation of this section harmoniously combines Python's NLTK library and TensorFlow to orchestrate an intelligent dialogue with users, culminating in a cohesive and responsive conversational experience. Extraction of data from the provided intents file marks the beginning of the training phase. Tokenizing user patterns skilfully by the Python script results in a larger vocabulary. Once this is done the data preprocessing takes place as explained in section 3.2, Throughout the training, the model learns to associate patterns of words with specific intents.

For this particular model architecture, the use of stochastic gradient descent with Nesterov accelerated gradient yields the best convergence. The accuracy metrics and categorical cross-entropy loss function are used to assess the model's performance. Upon convergence, the model encapsulates a wealth of linguistic insights, allowing it to comprehend user inputs and generate relevant responses. The combined NLTK and TensorFlow deep learning capabilities are demonstrated via the integrated NLP model. A mix of cutting-edge technology with linguistic understanding, this dynamic synergy enables the NLP-driven chatbot to translate user inputs into meaningful interactions.

3.4 Frontend Development

The frontend development of both chatbots played a pivotal role in ensuring a seamless and engaging user experience. The user interface (UI) was meticulously designed to facilitate effortless interactions and convey a sense of familiarity to users. HTML supplied the structural foundation for designing the interface, making it possible to separate parts for user inputs, chat history, and displayed responses. The UI's visual appeal and layout were improved with CSS, resulting in a consistent and visually acceptable design across a variety of devices. JavaScript, the backbone of interactivity, enabled dynamic updates of the chat window as users engaged in conversations.

To maximise user comfort and engagement, the UI of both chatbots was created to seem similar to popular messaging apps. An emphasis was placed on the conversational presentation of responses, the availability of chat history, and the natural arrangement of input fields. The chatbots' real-time interaction with users was made possible through integration with Flask backend systems. JavaScript code easily communicated data to the relevant backend when users submitted queries, where the processing and response creation took place. The frontend received the responses from the backend and changed the chat window accordingly.

The frontend development and integration process adhered to principles of usability, responsiveness, and accessibility. It attempted to achieve a balance between aesthetics and functionality, enabling users to easily engage in meaningful conversations with the chatbots. The achievement of this integration was crucial in developing an engaging and participatory user experience, laying the groundwork for thorough testing and performance evaluation in later stages of the study.

**Chapter 4**

**Testing and Evaluation**

This chapter presents a detailed analysis of the test and evaluation phase conducted to assess the performance and usability of the OpenAI-integrated and NLP-driven chatbot architectures. The primary focus was on participants' responses to the System Usability Scale (SUS) questionnaire and additional questions, shedding light on their perceptions, preferences, and experiences with the chatbots across different scenarios. This chapter casts a spotlight on the strategic foundation of our study, a crossover experiment methodology that aims to extract rich and nuanced insights from participants' interactions with chatbots.

4.1 Methodology

The design of a crossover experiment served as the basis for the methodology used in this step. With the use of this design strategy, we were able to weave a story that explored several user experience characteristics, maximising the depth and breadth of our investigation. Each path was designed to highlight a different aspect of the chatbot designs as it led two different groups of users down diverse directions. The journey of the participants was orchestrated through two distinct sequences, carefully choreographed to expose them to different scenarios while interacting with the chatbots. The first group was presented with a scenario in which they first interacted with the OpenAI chatbot to exchange their product and then with the NLP chatbot to track their order. The second group took a different route that was complementary to that of the first group, starting with the OpenAI chatbot to track their order and moving on to the NLP chatbot to exchange their merchandise. These are two of the situations that most clients dealing with online shopping encounter most frequently.

This methodology choice sought to capitalize on the inherent variability in participants' experiences by exposing them to contrasting scenarios. By doing this, we hoped to record a wide range of user interactions, taking into account various factors such as conversation flow, answer accuracy, usability, and more. The intentional variety replicates the diverse range of real-world events, enhancing the sincerity of users' interactions with the chatbots.

The brilliance of the crossover design resides in its capacity to maintain viewpoint equilibrium. The strategy minimises potential biases resulting from individual preferences or circumstances by guaranteeing that each participant interacts with both chatbot models. This strategy strengthens the reliability of our assessment by enabling us to evaluate the relative advantages and disadvantages of each building from a variety of perspectives. The chosen methodology serves as the compass that guides our exploration of user experiences. It empowers us to paint a panoramic portrait of participants' journeys through the chatbot scenarios, ensuring that no brushstroke is left unexamined.

A diagram of a group

Description automatically generated

**Figure 6**: Crossover Experiment Survey

4.1 System Usability Scale (SUS)