**Chatbot Conversations: A Comparative Study on User Interaction and Preference: Project Proposal (working title)**

**Introduction:**

In the digitized era we live in, our relationship with artificial intelligence has become increasingly intricate, presenting many unique challenges to navigate. Joseph Weizenbaum’s 1967 program, ELIZA, while simplistic by today’s standards, was known to have elicited some feelings of attachment in those that used it, so one can only imagine the plethora of emotions that a modern or future program could evoke.

With the speed at which artificial intelligence and natural language processing is developing, it is becoming more and more vital to establish a solid foundation, backed by research, upon which we base our developments. Within the scope of this project specifically: the creation of chatbots with Natural Language Processing.

Chatbots have the potential to revolutionize many aspects of our lives, whether through virtual assistants at work, mental health assistance, or even serving as substitutes for human companionship. Hence it is necessary that we implement them in the right way – thoughtfully and effectively.

This project aims to provide an understanding of what people want from their interactions with chatbots as well as establish any correlations between types of users and their preferences. The objectives that need to be completed in order to achieve this include:

* Developing (prompt engineering) two distinct chatbots for users to interact with
* Composing a study to collect data from a user after they have interacted with a chatbot
* Creating a simple machine learning program that, given some data about a user, could predict what type of chatbot the user would prefer

What some researchers refer to as HMC (human-machine communication), is an emerging yet underdeveloped field (especially when framed around AI), so any contribution to it is a step in the right direction.

We live in exciting times when it comes to Artificial Intelligence. While the concept of an “artificial intelligence” has been around for a long time, only in the past few years have we really seen an acceleration in its progress and accessibility, with the likes of NLP models such as ChatGPT 4.0 and stable diffusion image generation to name a few examples. I have a deep desire to help contribute to this new technological boom, with the hopes that I can, in any way, better the lives of people in their day-to-day.

Overall, this project has the ability to act as the perfect ending to my academic journey, incorporating all I’ve learnt over the years, whether that’s SQL databases in first-year, web-based applications in second-year, machine learning algorithms in third-year, and even phycology/research methods that I learnt at A-level.

**Literature review:**

**Central Texts:**

*GAAIS:*

* [Full article: The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust (tandfonline.com)](https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2085400)

*Importance of researching human-AI interaction:*

* [Frontiers | Editorial: On the “Human” in Human-Artificial Intelligence Interaction (frontiersin.org)](https://www.frontiersin.org/articles/10.3389/fpsyg.2021.808995/full)

*Trust in AI:*

* [The effects of personality and locus of control on trust in humans versus artificial intelligence (cell.com)](https://www.cell.com/heliyon/pdf/S2405-8440(20)31416-X.pdf)
* [Frontiers | Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online (frontiersin.org)](https://www.frontiersin.org/articles/10.3389/fpsyg.2020.568256/full)

*Other:*

* [Guidelines for Human-AI Interaction - Microsoft Research](https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/)

**General Texts:**

*Psychology of trust:*

* [Gender, Stereotypes, and Trust in Communication | SpringerLink](https://link.springer.com/article/10.1007/s12110-020-09376-3)

*General psychology:*

* [Effects of taboo conversation topics on - ProQuest](https://www.proquest.com/docview/2457326127?forcedol=true&pq-origsite=summon)
* [The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power - PubMed (nih.gov)](https://pubmed.ncbi.nlm.nih.gov/27055049/)
* [Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS - ScienceDirect](https://www.sciencedirect.com/science/article/abs/pii/S0092656616301325)

*General AI Research:*

* [ELIZA—a computer program for the study of natural language communication between man and machine (acm.org)](https://dl.acm.org/doi/pdf/10.1145/365153.365168)
* [Medical students' attitude towards artificial intelligence: a multicentre survey | SpringerLink](https://link.springer.com/article/10.1007/s00330-018-5601-1)

**Methodology:**

**Creating the two chatbots (prompt engineering):**

When considering how to implement the chatbots, the first step is choosing what model to use. A new model made specifically for this research does have its advantages, however the cost and complexity of training a new model capable of doing what we need (having detailed and varied conversations) would be too high.

Out of the existing models, ChatGpt-3.5 seems to be the best suited for this task. It is capable of handling a diverse range of conversation style and topics, has a large context window, and the use of system messages allows for further direction. Additionally, I already have some experience integrating it into a web-based application.

When it comes to the bots needed for this project, both bots will be limited in what they can talk about, making it much easier to compare results (reduces variance – some users might be more engaged when trying to push the bot to its limits). Research suggests that ‘taboo’ subjects can leave a negative impression after the conversation, but equally the topic should be interesting as to not bore the user so *‘Natural Geography’* has been settled on as it strikes a good balance.

**Designing the questionnaire:**

The questionnaire/survey that users will answer in order to provide data for the research will be split into two parts. The first part revolves around gathering general user info and will take place before speaking to any chatbots. It covers general demographic information (age, gender, etc…), the BFI-2-XS (an abridged, 15 question version of the Big Five Inventory 2), and the positive and negative GAAIS

The second part of the questionnaire occurs twice after the user has spoken to each of the chatbots. It will use a 1-5 scale (Disagree strongly, disagree a little, Neutral no opinion, agree a little, Agree strongly) similar to that of the BFI-2-XS and the GAAIS for simplicity and uniformity. It will cover the users’ thoughts and feelings about the bot/conversation, and will look something like this:

* I felt confident in the accuracy and reliability of the information provided by the chatbot
* The chatbot was a useful tool in learning about Natural Geography
* The conversation with the chatbot kept me engaged and interested
* I am satisfied with the quality of the conversation I had with the chatbot
* I, at any point, felt as though what the chatbot was telling me was untruthful or unexpected

How long the user talked to each of the chatbots, as well as how many messaged they sent will also be recorded to measure engagement, though the user will not be explicitly asked about this.

**Creating (and using) the application:**

Upon entering the application (which will be a web application), the user will be met with a disclaimer that details:

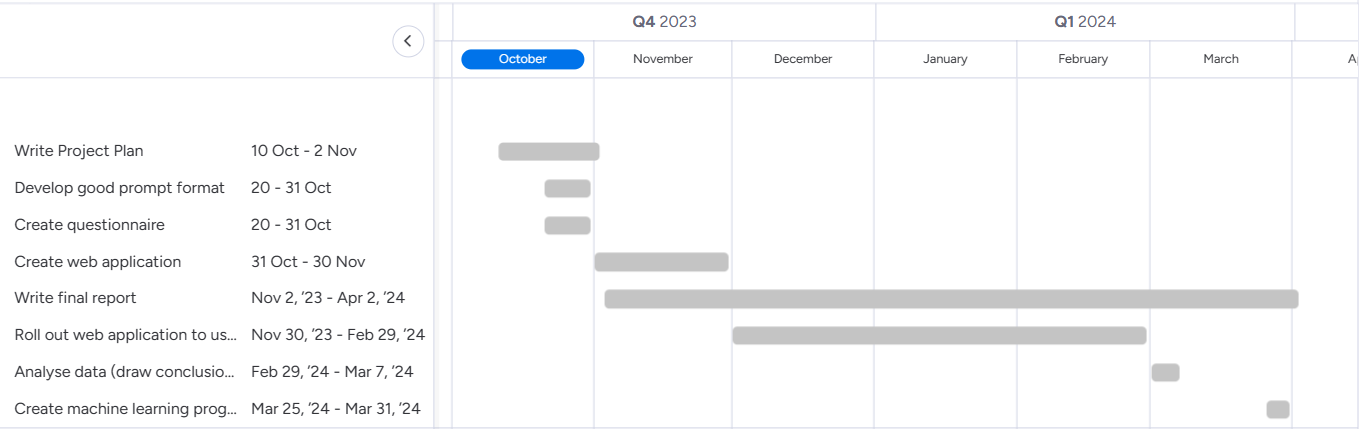
* What the project aims to do and how the website will aid it (specifically what their data will be used for)
* How their data will be kept secure and used only for the purpose of analysis within this project
* What exactly the user is going to partake in (as well as roughly how long it will take), explaining how they will answer a few general questions, speak to the chat bots, and then answer a few questions about the conversations
* A warning about how the chatbots may “hallucinate” and produce responses that are not-aligned with its intended purpose. As well as a reminder for the user to ‘take note’ of any of these occurrences to report later

If the user accepts the outlined terms, they will then be presented with the first part of the questionnaire (general questions) and then start talking to one of the two chatbots (which one they speak to first is randomized to try and reduce any order bias). The conversation will have a capped time limit, as to not deplete the tokens attached to the ChatGPT key, that will be somewhere in the 5-10 minutes range (will be determined by early testing). After the time limit is exceeded (or the use clicks an ‘end conversation’ button) the user will then be asked questions regarding the conversation they just had. This process is repeated: speaking to the second bot and answering the same questions.

**Data Collection and Analysis:**

Data collected from participants through the web application will be stored in an SQL database to organize the responses. From here we can compute the means and standard deviations of the demographic data, add correlate any of that initial data against the responses after talking to one of the bots (differentiating between and comparing results of the two different chatbots), displaying most of the data in graph form.

**Plan:**

**Project Plan (and roadmap):**

**Resource Plan:**

*Resources that are needed to complete the project include:*

* A valid **ChatGPT API key** (a temporary one can be generated for free, and this process can be repeated a few times for the duration of the project)
* A service capable of hosting a web-application that can handle node.js (for GPT API calls) as well as SQL databases (for storing the collected data), an example of this is **Amazons AWS EC2**
* A service like **Jupyterlab** for training a learning model (The University has granted me access to theirs through the LH Neural Computation course)
* A source of **users** to complete in the questionnaire (a lot of them can be sourced via networking: friends, family, university societies, as well as possibly using a paid service)

**Contingency Plan:**

*Aspects of the project that may need to be changed:*

* If it becomes an issue surveying users and saving the data via an SQL database on the application, then a simple Google form could be used instead
* If there are not enough users contributing data for analysis then a few things could be done: the amount of time the survey is available for could be extended, and a push for more users (from more sources) could be made. Other options include limiting the scope of the research so that a smaller sample is feasible (Eg: the research could just look into how introverts interact with AI)
* If time is becoming an issue, the predictive machine learning programming can be omitted from the project as it is not a central part of the research