Exploring Trust Dynamics between People and Emotive Service Robots

2000684

***Abstract*—**
SERVICE industries worldwide are grappling with persistent labour shortages. This study investigates the potential of affective service robots as a short-term solution, focusing on the critical factor of user trust in human-robot interactions. Specifically, this research investigates how affective behaviours (facial expressions, voice modulation) in a humanoid service robot influence user trust during direction-giving tasks. Using an experimental design, the study compares perceived trust in affective and non-affective robot modes employing a questionnaire derived from Gulati et al.'s framework. The results demonstrate higher trust ratings for the affective robot which are statistically significant, suggesting that emotive capabilities enhance perceptions of competence and benevolence in a service setting. This study contributes to the field by specifically exploring trust dynamics during first-time interactions with affective robots within a service context. These findings highlight the potential of affective service robots to address understaffing. However, further research is needed to investigate the implications of long-term deployment, including addressing potential ethical implications.

I. Introduction

Throughout history, from Oliver Evans' first mechanical flour mill [1] to modern self-checkout systems, there has been a consistent drive to mechanize and enhance key workforces. This push for automation often arises from a desire to improve efficiency and quality of life. In today's world, with an increasing global population and widespread staffing shortages, supplementing workforces with mechanical solutions holds more significance than ever. These staffing shortages in service industries – healthcare, fire, and police departments – have been documented in many different reports. One example is a 2022 House of Lords report [2], which proposes solutions like pay increases and expanding undergraduate positions. However, the implementation of these solutions is projected to take significant time as well as money, which prompts the question *"Is the short-term solution to under-staffing in the service fields supplementing the existing workforce with mechanical alternatives?"*.

The core challenge of implementing mechanical solutions in the service industry is fostering user trust. Trust, defined as "assured reliance on the character, ability, strength, or truth of someone or something," is often instinctively given to human workers but is harder to establish with machines. This challenge of trust has far-reaching implications; studies like "Trust and Growth" by P. Zak and S. Knack [3] demonstrate its impact on investment and broader economic performance. In the context of service industries, where interactions are often personal, trust plays a crucial role. This study investigates the unique challenges of fostering trust in mechanical systems within these contexts.

Despite this problem, numerous companies and service providers think the answer to the proposed question is yes. This belief is supported by a growing trend towards service robots, defined by the International Federation of Robotics (IFR) as any "robot in personal use or professional use that performs useful tasks for humans or equipment". Examples of this trend include the deployment of Spot (an autonomous four-legged robot) and k5 (an autonomous crime-fighting security robot) in the United States as outlined in this abstract by Bendel [4]. Another example is the trial of robotic nurses with the potential to assist in the recovery of both mental [5] and physical [6] health.

As we see more implementation of these service robots bringing robotics from behind-the-scenes support to front-line engagement with civilians companies and development teams should prioritize fostering trust. This trust is key as it allows for more efficient interaction with human collaborators, enhancing acceptance and mitigating potential risks associated with these automated systems. Modern approaches to fostering trust include user-centred design, transparent communication, ethical guidelines, and affective behaviours. While most of these methods have solid research support, there's limited evidence on the effectiveness of affective behaviours in service robots. This research gap prompted the creation of this study and its experiment.

Affective robots, also referred to as emotive robots, have a long history in robotics research. According to the American Psychological Association, to be considered affective a system must demonstrate the ability to express or perceive emotions [7]. Researchers have defined affective robots as "robots that can recognize human emotions and show affective behaviours" [8], aiming to understand how robots can utilize emotional responses to improve interactions. Pioneering work like Breazeal's Kismet project demonstrated a robot's potential to express emotions, laying the groundwork for further exploration of affective behaviours in robotics [9].

This study has the potential to make significant contributions to both the service industry and the broader field of human-robot interaction. By investigating how affective behaviours influence trust in service robots, this research could provide valuable insights for companies and developers aiming to optimize the design and implementation of these systems. Additionally, the findings may have implications beyond the service sector, informing our understanding of how humans build trust with machines in various contexts.

II. Literature Review

This study builds on the results found by H.W. Chuah and J. Yu in their paper "The future of service: The power of emotion in human-robot interaction" [10], carried out in 2021. While their research identified potential benefits of affective service robots, they acknowledged that their focus on the robot Sophia lacked real-world implementation. This study addresses this gap by testing a robotic implementation in real-life conditions.

Chuah and Yu's work also highlights a potential counter-argument to this study when they discussed the negative comments left by users on Sofia's Instagram posts. The researchers posit this may be because of the uncanny valley theory which was originally theorised by Masahiro Mori in his book "Bukimi no Tani Genshō" (Translated by K.F. MacDorman to be The uncanny valley)[11].

Mori theorises that there is a positive link between how a human a robot looks and the affinity felt towards that robot, however this is only true to a certain point at which this affinity is replaced by a feeling of "eeriness or uncanniness". *Figure 1* illustrates this concept, with a graph (adapted from Mori's work) depicting existing robots on a spectrum from prosthetics to industrial robots.

To avoid similar results and mitigate the effects of the "uncanny valley" this study opted for a robotic design that strikes a balance between human-like features and cartoonish elements, aiming to prevent the unsettling feeling associated with overly human-like appearances.

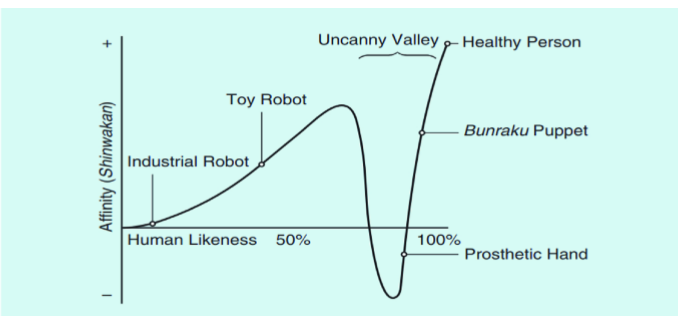


Fig 1: The original uncanny valley graph by Masahiro Mori

Building on the work by Chuah and Yu on emotive robots M. Coeckelbergh argues in his paper "Are Emotional Robots Deceptive?" [12] that experimenting with human-like robots that express emotions is not only ethically acceptable but potentially beneficial. He posits that research into emotive robots is essential as robots become more normalised in society which is an accurate prediction as noted in the introduction.

Coeckelbergh investigated the theoretical question of whether emotional robots are inherently deceptive. He identified 3 common assumptions made about emotional communication between entities, whether human or robot:

- Good intentions
- Genuine emotions (real and not faked)
- Not pretending to be something it is not

Coeckelbergh argues that these assumptions aren't essential for effective human communication and thus shouldn't be the sole criteria for judging robots.

His exploration revealed 3 key conclusions surrounding emotive robots. Firstly robots should provide appropriate and believable emotional responses. Second, society needs to redefine it's understanding of emotional communication to fully embrace robots. Thirdly, research into emotive robots can help our understanding of human behaviour and communication.

This paper's study intends to contribute to potential knowledge by assessing the role of affective (emotive) robots in the service industry but also in general by studying human communication with robots. It also aims to address Coeckelbergh's first conclusion by providing emotional responses which seem genuine and scenario-specific.

Oksanen et al. [13] further emphasize the need for research into trust towards AI and robots, especially in blind studies where both the participant and the robot have no extra knowledge about the other. To fill this knowledge gap this study used blind participants who didn't know about the study before participating.

Another paper that this study draws upon is "What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study" by A. Følstad et al. [14]. The study investigates factors affecting user trust in customer service chatbots, identifying two main influential categories. These categories were chatbot-specific factors and service context. Whilst the study focused on the chatbot-specific category it did mention that context did have an equal influence.

The study identified key chatbot-specific factors influencing trust, including:

- Quality of interpretation and advice
- Human-likeness
- Self-presentation and professional appearance

Følstad et al. found that about half of their participants valued a chatbot with a personal or relational communication style, believing this enhanced trust. Participants expressed diverse views on ideal human-like qualities. Some preferred a personal style with humour, believing it fostered trust. Others valued a generally human-like communication style for a better user experience. Still, others emphasized politeness and professionalism as key trust-building factors.

Despite this paper offering valuable insights, its validity is limited by a small sample size of 13 which reduces the generalizability of the findings. Another shortcoming of this study is the recruitment method which was through the chatbots, this most likely introduced a selection bias towards users who are already comfortable using chatbots. A final shortcoming of the study is its lack of investigation into how service context might influence trust. This paper's study directly addresses these shortcomings by employing a larger sample size, utilizing random recruitment methods, and standardizing the service context across all participants.

Another paper exploring the role of trust in emotive robots is "Promises and Trust in human-robot Interaction" by L. Cominelli et al. [15]. This paper concludes that people do trust humanoid emotive robots more than non-humanoid variants. The paper gained this conclusion after an experiment which took the format of a game. In each round of the game, the participant could choose to trust the robot or not. If the participant chose to trust, the robot would then decide whether to cooperate or not. This game scenario was originally invented by G. Charness and M.Dufwenberg in their 2003 paper "Promises and partnership" [16]. Despite Cominelli et al.'s paper supporting the hypothesis that humanoid emotive robots are more trustworthy, it doesn't test the robot in a "real life" scenario, which is a gap that this study seeks to address.

The paper by Cominelli et al. doesn't fully explore the theoretical basis for its findings, which could potentially be explained by social exchange theory, originally theorised in 1958 by G.C. Homans in "Social Behavior as Exchange" [17]. Social exchange theory posits that all interactions between entities are governed by a cost-benefit analysis. This cost-benefit analysis considers the rewards, satisfaction and dependence an individual gets from a social interaction. This theoretical framework could explain the results in the Cominelli et al. study, as participants likely weighed the potential outcomes of trusting the robot. The experiment's design likely prompted participants to engage in a cost-benefit analysis of the transaction, considering the following:

- **Benefit:** Trusting the agent could lead to cooperation in the game, resulting in a higher payoff.
- **Costs:** The cost of trust would be the risk of the agent not cooperating after the trust was given.

It's also fair to assume that the uncanny valley affects the results of the experiment by Cominelli et al. as the emotive robot they produced was designed to look like a human, as seen in *Fig 2*, this could potentially have meant the robot received less trust than potential. This design choice not only evokes Mori's uncanny valley theory [11], but it also raises questions of deception as defined by Coeckelbergh [12], since the robot could be seen as "pretending to be something it is not" as it is pretending to be human. For this reason, further research must be performed and results must be compared to those gathered by Cominelli et al. to see if a less humanoid robot evokes more trust.

FACE

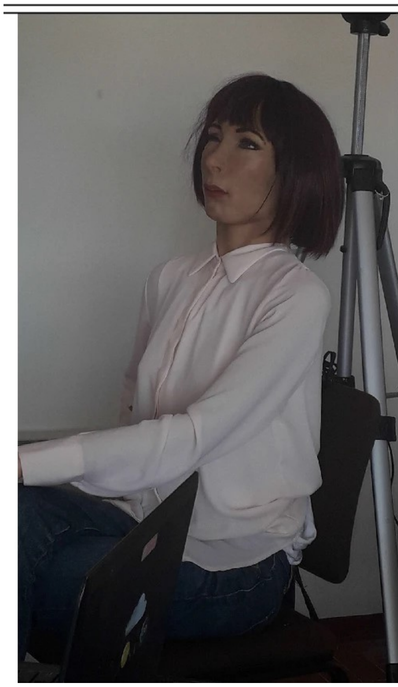


Fig 2: The robot used in the experiment by Cominelli et al.

Another study that lacks an exploration of emotive robots in real-life settings is "Can Robots Earn Our Trust the Same Way Humans Do?" by L. Christoforakos et al. [18]. The paper ran an experiment in which participants judged a video of a robot in both an affective and non-affective mode. The study's findings contradicted those of previously discussed papers, suggesting that competence and warmth increased trust in the robot, but anthropomorphism did not. Emotiveness can be defined as a property of anthropomorphism and thus the paper stands in opposition to the proposed initial question.

The validity of the results found by L. Christoforakos et al. can be questioned for several reasons. Firstly, the lack of physical interaction with the robot and reliance on video observation makes it difficult to assert that true human-robot interaction occurred. Secondly, participants' awareness of the study's purpose potentially introduced response bias, as they may have sought to act in ways they believed the researchers desired. One such reason for these results could be that the participants weren't receiving or benefiting from the interactions and thus according to social exchange theory the participants would see no reason to trust or interact with the robot on a deeper level. However, the paper itself disputes this explanation by questioning whether current models of interpersonal interaction, like social exchange theory, adequately capture the complexities of human-robot interaction (HRI).

This paper emphasizes the need for further research into the factors affecting trust in robotics, particularly in the context of '‘real-life" interactions. The paper advocates for innovative approaches to explore trust that moves beyond existing models tailored for HRI, focusing on the broader dynamics of robot-human relationships.

Though often being the foundation of each other trust and acceptance are completely separate concepts. Acceptance often allows people to embrace a fact without attempting to change it, it differs from trust as it lacks any approval or belief in the subject [19]. Despite these differences, acceptance is important because it allows people to assess actions and intentions with an open mind. Vice versa trust allows users to more openly accept robots as imperfect and look at the robot with a reduced risk perception. The link between these concepts, despite their distinctness, highlights a gap as other papers often investigate the acceptance of AI and Robots in the service field but not their trust.

One study that investigates the acceptance of service robots is "Customers' acceptance of artificially intelligent service robots: The influence of trust and culture" by O. Chi et al. [20]. This paper investigated the links between trust in a robotic system and the acceptance of said system in two separate cultures, American and Chinese. The study concluded that trust in interaction with AI robots is a significant factor that influences customers' intention to use them. This backs the claim made in the introduction that trust increases the productivity of workers and businesses but with a robotic focus. The main problem with this conclusion is it only investigated two cultures which doesn't allow for significant generalisation of the research. One culture they didn't investigate was British meaning though this paper could potentially explain results gathered in this study it will only be partially applicable

The paper stated that 3 main cultural attributes affected acceptance:

- **Uncertainty avoidance** refers to the societal preference for clear rules and procedures, cultures with high uncertainty avoidance rely heavily on social cues and recommendations to make decisions on new technology. This suggests that trust in AI robots may be even more important for gaining acceptance in cultures with high uncertainty avoidance.
- **Long-term orientation** refers to the societal preference for the future or past, cultures with an emphasis on long-term orientation are often more accepting of new technologies such as service robots as they see them as potential benefits for the future. Therefore, cultures with an emphasis on long-term orientation are more likely to accept AI robots over those who value tradition and the past.
- **Power distance** refers to the societal acceptance of unequal distribution of power, in societies high power distance, individuals are more likely to defer to authority figures and may be more comfortable interacting with service robots. Societies with high power distance are more likely to accept service robots regardless of trust as they see them as authority figures.

While this study offers valuable insights, it leaves open the crucial question of what factors shape trust in robots, highlighting a key area for further investigation that impacts the acceptance of these technologies.

The paper by O. Chi et al. built on the findings of D. Gursoy et al. in “Consumers acceptance of artificially intelligent (AI) device use in service delivery” [21], which determined that customers go through a three-stage process to decide whether or not to accept AI devices during service encounters. The key factors within this process include social influence (surrounding opinions on AI), hedonic motivation (enjoyment of technology), anthropomorphism (how human-like the AI seems), performance expectancy (perceived usefulness), effort expectancy (ease of use), and the emotions these aspects generate. The paper discusses anthropomorphism however it doesn’t delve into what specific aspects of it affect the robot again leaving a gap in the literature regarding how specific aspects of anthropomorphism, such as emotive behaviour, impact trust and, consequently, acceptance.

Additionally, it introduces a hypothetical survey (AIDUA) with its own set of limitations. For instance, the survey never saw any “real-life” application and thus it’s not possible to assess how well this survey applies to real-life scenarios. Furthermore, the survey focuses on community-level factors for robot acceptance rather than directly investigating the impact of robot performance. These limitations underscore the need for research directly examining how emotive behaviour in robots, as a key part of anthropomorphism, influences trust and acceptance in real-life settings.

This literature review has highlighted several key gaps in the existing literature surrounding service robots and affective behaviours:

- **Real-life service environments:** There is a lack of any experiments conducted in real-life service environments, this gap means that whilst service robots appear to be well trusted and accepted it’s hard to declare their effectiveness in a service environment definitively.
- **Effect emotive behaviour has on trust in service robots:** There is a distinct lack of any research into the link between emotive behaviours and trust specifically in the service environment making it hard to determine whether emotive robots would be beneficial.
- **Effect of uncanny valley on service robots:** Many of the studies used robots which arguably fell into the uncanny valley and thus diminished the effect of emotive or anthropomorphism. This meant they either made reduced or opposite claims which make it hard to determine the true answer to the proposed question.
- **Method to measure trust in robots:** None of the researched papers included a definitive method for measuring the trust of a robot in a real-life interaction. This is an important step that must be made as it is difficult to generalise or compare results without a definitive method.

Addressing these gaps is crucial for understanding the true potential of service robots and the role of affective behaviours in maximizing their acceptance and effectiveness.

III. Research Question & Hypothesis

The preceding literature review has illuminated key insights, prevailing theories, and unresolved gaps within the existing body of knowledge regarding service robots and emotive behaviour. Building upon these findings this section will lay out the finalised research question to fill the knowledge gap. This research question will then be used to generate testable hypotheses setting the foundation for the empirical investigation to follow.

The main thesis question, “is the short-term solution to understaffing in the service fields supplementing the existing workforce with mechanical alternatives?” has too large a scope to answer appropriately in a single dissertation study. Due to the large scope of this question, it’s important to investigate a smaller aspect of the question and refine the question this study aims to answer. The first focus area is service robots, which this paper will investigate. There are various dimensions of service robots which need investigation; this paper will specifically examine how emotive behaviour impacts trust in service robots, as this has been identified as a gap in the research.

While the broad question of robots supplementing human workforces in service industries is intriguing, a single dissertation study requires a narrower focus. This research will investigate the potential impact of emotive behaviours on how users trust directions given by service robots. Existing research presents conflicting findings on whether emotive behaviours enhance or hinder trust in robots. Additionally, most studies lack a ”real-life” service interaction context. Thus, this study will examine the following question: *“To what extent does the inclusion or omission of simulated emotion, using a novel voice modulation and emotive expression technique, influence the trust that young adults, with some technical awareness, have in the directions given during a short direction giving interaction?”*.

A couple of hypotheses – testable predictions about the relationship between variables – can be proposed from the research question. This study aims to address two main hypotheses. The first is the null hypothesis (H0): There is no difference in perceived trust between service robots exhibiting affective behaviours and those that do not. The alternate hypothesis (H1) is that service robots displaying affective behaviours will elicit higher perceived trust than non-affective robots. This prediction is informed by studies that suggest in a non-service environment affective robots report higher trust. However, research like "Can Robots Earn Our Trust the Same Way Humans Do?" by L. Christoforakos et al. (2021) supports the null hypothesis, highlighting the need to investigate this relationship further.

Though this study only has two main hypotheses there are other predicted observations. Firstly participants with conditions that affect social interaction, such as autism spectrum disorder (ASD) or social anxiety disorder (SAD) [22], may exhibit responses outside the predicted hypotheses. While this study won't inquire about such diagnoses due to scope, further research into these interactions would be valuable.

Another potential observation is the creation of para-social relationships with the robot. Parasocial relationships refer to one-sided emotional bonds people form with media personalities, celebrities, or even fictional characters as defined by D. Horton and R. Wohl [23]. Humanoid robots, designed to mimic human appearance and behaviours, can easily become targets for similar attachments. Forming these relationships could significantly increase trust and potentially alter the results. However, this aligns with the main hypothesis, as emotive robots are likelier to trigger para-social responses.

IV. Methodology

To investigate the influence of simulated emotion on trust in robot-given directions, this study employed an experimental research design to answer the research question. To investigate this, the study implemented a controlled AB test experiment within a simulated service environment. Participants interacted with a robot that provided directions, either with or without simulated emotions. Following the interaction, they completed questionnaires to quantify their trust in the robot. This controlled environment with a human-robot interaction (HRI) featuring simulated emotions offers valuable insights applicable to real-world scenarios, addressing a gap in the literature and providing a strong foundation for answering the overall thesis question.

The experiment involved randomly chosen participants having a solo interaction with either an affective or non-affective service robot. During this interaction, participants were free to ask any questions they desired, although the researchers gently encouraged them to inquire about directions to a specific location on campus. This focus on directions aligns with the robot's primary service goal and serves as a key measure of trust. Direction giving is an important trust activity as following inaccurate directions can be a significant inconvenience and even pose a safety risk.

To ensure unbiased results, participants were randomly chosen. These participants were chosen from the pool of Falmouth and Exeter University students on Penryn Campus. This resulted in a participant demographic consisting mostly of young adults under 25, as documented in this Falmouth University report [24]. This age demographic offers a particular advantage due to their affinity towards technology, as seen in this 2023 report by Nominet [25]. The report found that overall 81% of young people had good technology skills and were able to easily complete many tasks in a technical challenge. Their technological affinity suggests a higher baseline acceptance of the affective robot. This minimizes the potential influence of factors like effort expectancy, as predicted by D. Gursoy et al. [21], on the study's results.

The random selection of participants ensured that all interactions with the robot were first-time encounters. This directly addressed the gap identified by A. Oksanen et al. [13], who highlighted the need for research on initial interactions with social robots as mentioned in the literature review. To strengthen the emphasis on first-time interactions and minimize participant bias, all participants signed a disclaimer confirming they had no prior knowledge of the study or the robot, this disclaimer can be seen in *appendix D*.

To achieve statistically significant results, the study required 52 participants, divided equally (26 each) between the affective and non-affective conditions. This sample size was determined through a pre-study g* power analysis (detailed in *Appendix C*). The analysis predicted a very high effect size, which is desirable because a real-life application of the robot with affective capabilities is preferable. A high effect size would provide strong evidence to support the robot's effectiveness in real-world settings.

To further minimize bias, all experiments took place in similar small rooms on campus. These rooms were large enough to allow participants to view the robot in its entirety and interact at a comfortable distance. This ensured that participants were not uncomfortable or put in too awkward a situation so as not to present potential environmental bias.

Fig 3: The robot set up in a room for the experiment

Figure 3 shows the experimental setup, with cables carefully hidden behind the table to enhance the robot's humanoid appearance, promoting natural, uninhibited interactions. The participants were seated on the sofa, while the researcher monitored the experiment discretely to address any unexpected errors. However, participants were encouraged to note any instances where the robot malfunctioned.

There was no enforced time limit for the interactions due to the potential variance in participant response times. The participants were encouraged to answer the questionnaire as soon as they felt ready, with the average interaction lasting

15 to 20 minutes. However, a 15 to 20-minute interaction significantly exceeds the duration of most service encounters, potentially challenging the study's real-world applicability. Nevertheless, since both test groups had the same time allowance, the duration becomes a controlled variable within the experiment

The experiment manipulated the activation or deactivation of affective techniques, including facial expressions and voice modulation (independent variable). Participant responses on a questionnaire based on a framework by S. Gulati et al. [26] were used to measure the dependent variable of trust levels. The framework's flexibility allowed for the creation of a scale measuring seven key trust indicators: motivation, willingness, competence, benevolence, predictability, honesty, and reciprocity. These values are highly relevant to affective robotics.

To calculate the final trust score, the questionnaire responses were aggregated, with the first three questions (indicating negative attributes) subtracted from the total. Here is the full list of questions asked to the users, the bold text shows where the question has been altered or filled for the study application:

- I believe that there could be negative consequences when using **this robot**
- I feel I must be cautious when using **this robot**
- It is risky to interact with **this robot**
- I believe that **the robot** will act in my best interest
- I believe that **the robot** will do its best to help me if I need help
- I believe that **the robot** is interested in understanding my needs and preferences
- I think that **the robot** is competent and effective in **providing directions**
- I think that **the robot** performs its role as **someone giving directions** very well
- I believe that **the robot** has all the functionalities I would expect from **someone giving directions**
- If I use **the robot to give me directions**, I think I would be able to depend on it completely
- I can trust the information presented to me by **the robot**

Participants responded to the questions using a 10-point Likert scale (1 = strongly disagree, 10 = strongly agree) [27]. The questionnaire concluded with a space for participants to share additional observations, facilitating the capture of qualitative insights. One question was removed from the original framework as it focused on long term reliance which participants wont be able to appropriately judge after one short interaction. The removed question was "I can always rely on (—) for (—)".

The questionnaire heavily focuses on the service interaction of giving directions as it allows for the study to properly assess the robot in a simulated "real-life" service environment. This helps the questionnaire appropriately answer the research question and gaps in the literature.

The design of the robotic artefact served as a key source of independent variables for this study. The robot was constructed with a 3D printed body, emphasizing a humanoid yet robotic appearance within the 'cartoon' aesthetic. This choice avoids triggering uncanny valley effects [11] and promotes transparent presentation as a technological artefact, aligning with ethical considerations outlined by Coeckelbergh [12]. 3D printing was chosen as a cost-effective manufacturing method, enabling the creation of a rounded, human-like form while enhancing the potential for real-world application.

The body consisted of eight separate panels secured with bolts. This modular design allowed for integrating components like the head and potential future additions like arms. The head was 3D printed in two pieces to ensure accurate proportions and a smooth facial surface, as any imperfections could affect participant reactions. *Figure 4* demonstrates the robot's construction, highlighting its humanoid proportions. demonstrates the robot's construction, highlighting its humanoid design.

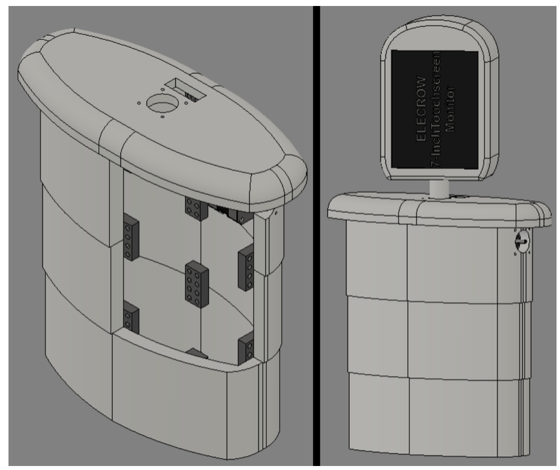


Fig 4: The 3D CAD models of the robot, the left is the full robot and the right is a look at the internal construction.

Figure 4 also shows the screen slotted into the head (The black part in the centre of the head). This screen (an ELECROW 7" 800x480 screen [28]) allowed the robot to render a face reducing the need for more mechanical components which complicate the construction and repair process. This screen is disabled during non-affective mode making it just a black screen thus reducing the humanoid look of the robot. When in affective mode the robot displayed a cartoon face again to avoid the uncanny valley effect. This face had several different expressions it could flip between to display emotions.

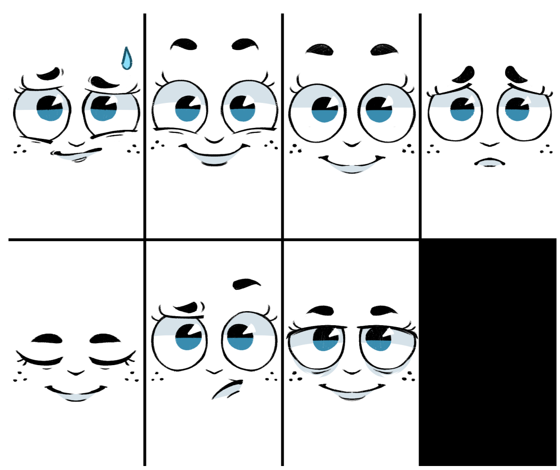


Fig 5: All expressions the robot can show. (Artwork by A. Aksu [29])

Figure 5 shows the faces that the screen can display. Fr)om the top left to bottom right, they represent the emotions of confusion, happiness, neutrality, sadness, sleep, thoughtfulness, and tiredness." These emotions give the robot the ability to respond to the discussion effectively while avoiding aggressive emotions like anger, which could create a hostile environment. Excluding sleep and the neutral state, the robot uses these emotions to respond directly to participants. Typically, "confusion" or "thinking" were used when the AI was generating a response, "sad" was used to show sympathy and "happy" was used when the robot heard something good or had a positive interaction. The AI system autonomously determines when to display each emotion, enabling generalized responses and allowing the robot to convey emotional cues in various situations effectively. This ability to display emotion allows the robot to follow Coeckelbergh's [12] conclusion that robots should provide appropriate and believable emotional responses.

To check the faces appropriately conveyed the intended emotion a set of pilot participants were shown the faces and asked to determine what emotion each face showed. This process was repeated until the majority (70%) of the pilot participants agreed the face showed the emotion it needed to. To maintain the integrity of the main study, pilot participants who assisted with evaluating core features were excluded from the main participant pool. This ensured that the main study participants remained unaware of the study's purpose and the robot's capabilities.

To avoid the uncanny valley, the robot's face was designed with exaggerated, childlike features. Specifically, enlarged eyes aimed to leverage the "Kindchenschema" effect, a concept introduced by K. Lorenz [30]. This evolutionary response predisposes humans to find infant-like features (large eyes, round face, etc.) endearing and non-threatening. Researchers like Glocker et al. [31] have demonstrated its relevance to robots. For this project, Kindchenschema aims to reduce negative perceptions associated with uncanny human-like robots. Additionally, by evoking cuteness, it may promote empathy towards the robot, further improving user interaction."

To further enhance the "Kindchenschema" effect, the robot was fitted with a rounded faceplate *Figures 3 & 7*. This design feature aimed to create a more childlike facial shape, emphasizing roundedness. For experimental control, the faceplate was easily removable, enabling the non-affective robot configuration.

To minimize the uncanny valley effect, the robot's nose was deliberately simplified. The uncanny valley suggests that overly realistic yet imperfect human features can trigger feelings of aversion [11]. While the focus here was on minimizing the uncanny effect, simplifying the nose could have an additional benefit. The Halo Effect, as posited by E.L. Thorndike [32], indicates a link between physical attractiveness and the assumption of positive traits. A less detailed nose, by avoiding any potential imperfections, may subtly increase the positive perception of the robot's overall appeal.

To enhance the robot's humanoid appearance, the face included the ability to move its mouth in sync with its speech. This was achieved by creating a counterpart image for each emotive face with an open mouth as seen in *Figure 6*. The open-mouth image was displayed whenever a word was spoken. WIPRobotics "Linux-Sound-controller" [33] was used to track sound output, allowing for precise mouth movement synchronization. This technique effectively reinforced the illusion that the robot was speaking, contributing to its overall lifelike aesthetic.

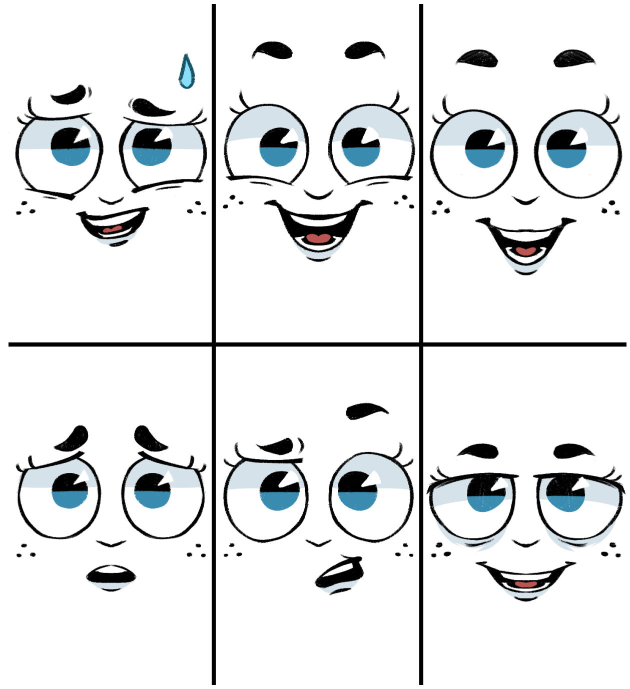


Fig 6: All expressions with opened mouth variants. (Artwork by A. Aksu [29])

To further enhance its humanoid appearance, the robot was accessorized with human clothing, specifically a beanie. This aimed to tap into subconscious social cues that influence how we perceive and interact with others. In this case, the beanie may subtly suggest a relatable personality, potentially fostering greater trust during the interaction." This can be seen in *figures 3 & 7*.

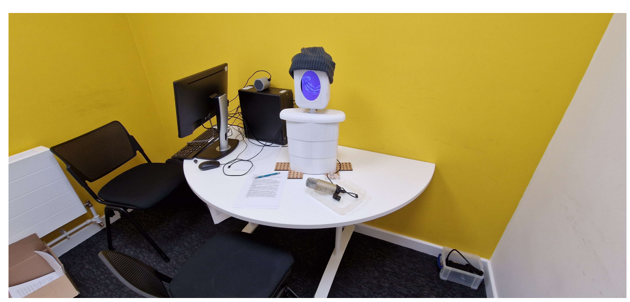


Fig 7: The robot in an experiment environment wearing a beaning and using a face plate.

The emotive face wasn't the only novel technique used in the robot the project also utilised AI and voice modulation techniques to further sell the illusion of true affective behaviour.

For AI, the robot used a large language model (LLM) implemented through GPT4All [34], GPT4All was chosen over other interfaces for its data privacy; it processes information locally, ensuring GDPR compliance [35]. GPT4All executed

OpenOrca's Minstral 7B [36] , an LLM selected for its quick response times, conversational style, and commercial licensing under the Apache-2.0 License [37]. These choices allow for fluid, natural conversations and increase the robot's potential for real-world use.

To ensure participants judged the robot solely on its affective abilities, the same AI was used in both affective and non-affective test conditions. This design choice aims to minimize the impact of performance expectancy (perceived usefulness), as noted by D. Gursoy et al. [21]

In order to appropriately fill the role of a service robot AI the LLM had to be given a large initial prompt all of which can be found in *Appendix A*. This prompt gave the AI key information about its role and how to interface properly with the robot, one such bit of information was the 5 part primary directive. The primary directive told the robot information important above all else such as its name, its goal, its personality and its global location. This information allowed the robot to fit it's role as a service robot better and act more professionally.

The robot could execute internal functions or commands using special prompts marked with "$system$" or "$user$". This system allowed the AI to access functional modules to complete its main service requirement. The robot could also learn how to use each module by referencing a short description stored within the function's code. This design approach streamlines future development, as new modules can be easily integrated. This flexibility enhances the robot's real-world applicability.

The robot's core internal function was a location module. This module accessed a virtual database representing campus locations and paths connecting them. To fulfil its service requirement, the robot used the 'Networkx' Python module [38]to calculate the shortest route between any two points. This enabled the robot to provide accurate directions to participants upon request.

The robot's ability to speak with a human voice is another key feature distinguishing the affective and non-affective modes. The affective robot employed "PiperTTS" by Rhasspy [39] for its commercial license, fast generation, and natural-sounding, high-quality female voice. In contrast, the non-affective mode relied on "Google Text-to-Speech" (GTTS) [40], which produced a distinctly robotic voice, aligning with the non-affective condition's design. While both TTS modules used female voices to minimize potential bias, this choice was also supported by research suggesting that female voices are often perceived as more trustworthy [41] [42]. This trustworthiness increases the real-world practicality of the service robot.

To facilitate natural interactions, the robot utilized the 'SpeechRecognition' Python module [43]. This enabled participants to converse with the robot as they would with another person, promoting genuine, unbiased responses. However, the module faced some challenges. Its background noise suppression was limited, sometimes causing it to pick up its speech or ambient sounds. Additionally, the module occasionally missed or misheard participants' words, requiring researchers to ask for repetition.

All these AI components were linked via a Python script [44] to facilitate communication. To maximize speed and efficiency, each component ran in separate threads. This code is available in the GitHub repository *Appendix B*. However, despite optimization efforts, there were still performance bottlenecks in certain areas, necessitating a high-performance computer with an Nvidia 3090 graphics card. This powerful hardware enabled the utilization of CUDA [45], a technology that significantly accelerated AI processes. This acceleration ensured a smooth conversation flow for the robot, eliminating uncomfortable pauses.

A complete overview of the python code and how the threads communicate can be found in the form of a UML diagram in *Appendix G*.

V. Ethics

Participant health and safety were paramount throughout the experiment. This included measures to protect their physical and mental well-being and data security. Action plans were written for potential eventualities to ensure that the researchers could provide appropriate assistance in the event of any complaint or injury. One such action plan was created in the case of any injury for instance; burns, electrocution or other injuries. For example, an injury response plan was in place, with first-aid trained staff on call, familiar with the experiment's procedures. A risk assessment was also carried out of the experiment to ensure potential risks were understood and minimised.

Safety precautions were also prioritized during the construction and development process. This included training on all lab equipment and having a skilled health and safety staff member available for assistance. These measures minimized risks, ensuring the robot's smooth and efficient construction.

The study prioritized participants' mental health alongside physical safety. While the robot was designed with kindness in mind, precautions were in place to address the possibility of misinterpretations causing emotional distress. If the participant suffered from any adverse mental health effects they were encouraged to email the head researcher. The researcher would then be able to connect them to appropriate support resources for prompt assistance.

Data handling and privacy were also considered complying with the EU General Data Protection Regulation Act (GDPR) [35] and Falmouth University's data privacy plan [46]. This meant all data gathered from the questionnaire was anonymised to avoid revealing personal information. The participants were also given the option to pull out of the study for 3 weeks after the completion of their experiment, the participants could pull out by emailing the head researcher, this meant if for any reason the participants were uncomfortable with their data being used they could no longer be included.

To ensure data security, all information collected during the study was stored on a GDPR-compliant university OneDrive system [47]. Access to this system requires a linked account with two-factor authentication, providing robust protection against unauthorized access and minimizing the risk of data leaks.

The disclaimer the participants signed also made sure that the participants understood all of their rights and how their data would be used, a full copy of this disclaimer is available in *Appendix D*. The disclaimer also made it clear that any output by the AI is not guaranteed to be factual, this is included to ensure that the participants don't suffer any potential harm from following inaccurate information provided.

A key ethical concern is the potential for misuse of this affective robot technology. Since the robot is designed to inspire trust, misinformation it spreads – intentionally or accidentally – could be particularly harmful. Misinformation isn't the only ethical misuse of the robot, due to GPT4All being uncensored participants or users of the technology could ask the robot anything they wanted including questions which could potentially cause harm. The language model and prompt did its best to filter against harmful content however it wasn't always perfect and could sometimes be circumvented. This highlights the ethical responsibility of developers and users to ensure the technology is employed for positive purposes. While the risks of misuse exist, the potential benefits of this technology for good likely outweigh the potential for harm.

VI. Data Analysis

After the experiment, data analysis began with a Python script *Appendix E*. An independent two-tailed t-test was performed to compare the affective and non-affective conditions, focusing on the overall results.

The two-tailed t-test used the equation to calculate the t and p values.

$$t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

where $\bar{X}$ and $\bar{Y}$ are the means of the affective and non-affective total scores and n and m are the sample sizes of each group. The mean of each group was calculated using

$$\bar{x} = \frac{\sum x}{n}$$

where n is the number of values in the sample and x is the total of all the values in the set. This t-value was then checked against a t-distribution table to work out the p-value to in turn calculate whether the test was significant.

Prior to conducting the t-tests, the data was checked for normality and equal variances. These assumptions were met, indicating appropriateness for a two-tailed t-test. The results revealed a statistically significant difference (p = 0.0379) in overall trust scores between the affective and non-affective robot modes. The non-affective mode had a mean score of 40.5, while the affective mode had a mean score of 47.538. This significantly higher score for the affective mode supports the study's main hypothesis, suggesting that the inclusion of simulated emotive behaviour enhances user trust in service robots.

P-value and means were also calculated for each question in the questionnaire to allow for a more in-depth analysis of the data in the results section, as seen in *Figure 8*.

Question Number	Affective	Non-Affective	P-Value
Q1	3.577	4.0	0.389
Q2	3.0	3.5	0.407
Q3	2.038	2.346	0.369
Q4	7.731	7.0	0.115
Q5	8.308	8.115	0.626
Q6	7.538	6.115	0.027
Q7	6.615	5.923	0.153
Q8	6.308	6.308	1.0
Q9	6.538	6.115	0.544
Q10	5.808	4.538	0.075
Q11	7.308	6.231	0.057

Fig 8: A table showing the mean results of every question and their P-value

This was then converted into a bar graph to better visualise the responses, for the bar graph the negative data was flipped to only see the positive response to the robot:

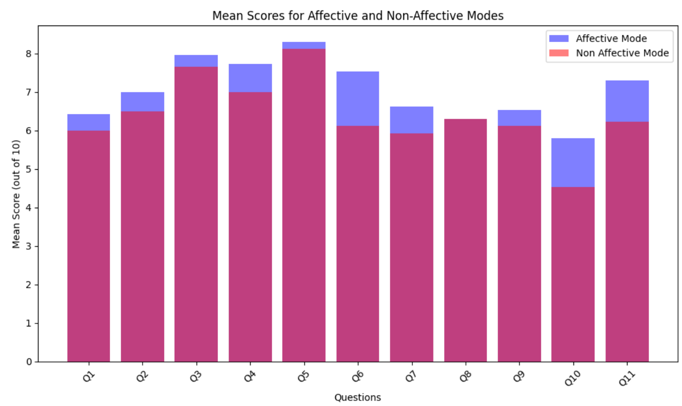


Fig 9: Bar graph visualising the responses to the robot

To further visualise the results the data was also loaded into box plots to clearly show the distribution, these box plots can be seen in *Appendix F*

VII. Results

This study yielded significant results with implications for the thesis question. Crucially, the data supports the idea that emotive service robots inspire greater trust than non-emotive counterparts. This finding aligns with previous research such as the studies done by L. Cominelli et al.[15] and A. Følstad et al. [14]. These results offer valuable insights for the overall thesis, suggesting that continued research and development of affective service robots may pave the way for their broader implementation in real-world service roles.

The increase in trust ratings for the affective robot compared to the non-affective robot wasn't uniform across all questions. Some questions saw a significant increase while others saw little or none. One question which saw significant increase as "I believe that the robot is interested in understanding my needs and preferences". This question saw a 23% increase in rating in the affective mode, there are three possible explanations for this specific finding:

- **Novelty and Engagement:** Participants may have engaged more with the affective robot due to its uniqueness and humanoid appearance. This could mean that the participant set had a particularly high long-term orientation as they wanted to interact with novel tech as discussed by O. Chi et al. [20]. This increased engagement could have led participants to share more personal information, allowing the robot to tailor its responses and create the perception of deeper understanding.

- **Confirmation Bias:** Participants may have also been unconsciously under the influence of confirmation bias. This meant they interpreted the affective robot's responses as genuine interest and attention due to the robot's facial expressions. This bias might be further strengthened by our reliance on non-verbal cues in natural interactions.
- **Empathy Perception:** The robot's ability to display emotions may have triggered a stronger perception of empathy in participants. The robot's reactions to the participant may have been interpreted as a sign that it genuinely cares about the user's needs. This perceived empathy can lead users to believe the robot is actively trying to understand them.

Though all these explanations seem wildly different it's more than likely a combination of all three overall boosting the score.

Trust in the robot's functionality also showed improvement, evidenced by a 28% increase in positive responses to the survey question: "If I use the robot to give me directions, I think I would be able to depend on it completely". This suggests participants perceived the affective robot as more intelligent. One contributing factor could be how the robot's reassuring voice tones and expressions reduced user anxiety. When people feel less anxious, they might naturally place more trust in the competence and reliability of the robot's directions. Another reason for this could be that participants were more willing to forgive mistakes by the affective robot as they felt more empathy for the humanoid robot.

The affective robot's "thinking" face likely contributed to its perceived competence. This visual cue could create an illusion of the robot deeply considering the question, potentially leading users to believe it had a greater understanding. Comments from the non-affective condition support this interpretation. For example, one user stated, "some user feedback while it loads its response would be appreciated", indicating a desire for this visual thinking indicator in the non-affective robot.

The affective robot not only benefited from higher ratings of its existing functionality but also inspired users to perceive additional capabilities. For instance, some users wrongly attributed voice recognition and speaker identification abilities to the robot, despite those features never being implemented. This highlights the impact the affective mode had on user perceptions.

Supporting the earlier prediction of para-social relationships in the research question [23], additional comments and researcher observations revealed an interesting development. One participant named the affective robot "Penelope", deviating from the designated name "PEA". This unique name emerged from a conversation where the participant inquired about the robot's preferred name, prompting a brainstorming session. This instance suggests that the participant valued the robot's input and potentially formed a deeper connection with it.

Additional evidence of para-social interactions emerged in the comments section. Participants spontaneously described the robot with positive personality traits ("She was nice," "She's very cool"). These complimentary terms, along with expressions of personal preference ("I like her a lot"), suggest that some participants formed an emotional bond with the robot, a hallmark of para-social relationships.

Additional user comments strongly indicate that participants fully personified the affective robot, viewing it as a human agent. A key piece of evidence is their consistent use of gendered pronouns ("she" or 'her") rather than the neutral "it". This suggests that users integrated the feminine voice into their conceptualization of the robot. Conversely, participants in the non-affective condition primarily referred to the robot as "it", reinforcing its perception as a non-human tool. Additionally, the users often realised the emotions the robot was showing and empathised with them showing a further degree of humanisation.

Despite efforts to mitigate it, the robot still evoked the uncanny valley effect [11] for some participants. While infrequent, comments describing the robot as "creepy" or "unnerving" did occur. This negative perception significantly impacted ratings. For instance, one participant who found the robot unsettling scored it substantially below the group average and outside the IQR as seen in some of the box plots in *Appendix F*. Despite these results demonstrating the uncanny valley effect and providing scores outside the normal distribution, their limited frequency within the large dataset size indicates minimal overall impact on the findings. One reason the participants gave for the "creepy" was the fact the robot didn't blink which is an oversight which should be implemented in future versions of the robot.

The affective service robot's practical usefulness was somewhat limited by the reliability of its location system. Frequent misinterpretations of place names meant directions were occasionally incorrect or could not be generated. This issue is reflected in the box plots for questions 8 and 10 (*Appendix F*), where unusually low minimum scores suggest particularly negative user experiences. This does prove a need of further research into an AI robotic system which could solve the main thesis question as the robot for this study appears inadequate. This research could be performed on the robotic architecture built for this project as it can be easily built onto and improved as stated in the methodology section.

The robot's practical limitations challenge the explanatory power of social exchange theory [17] for this user experience. While participants may not have received substantial practical rewards (e.g., accurate directions), they may have derived intrinsic benefits such as the enjoyment of novelty or a sense of connection with the robot. Alternatively, social exchange theory, as applied in this study, might not fully capture the nuances of human-robot interaction. As suggested by Christoforakos et al. [18], existing social models may require adaptation for this evolving field.

While most areas showed improvement, the affective robot did not score higher on question 8: "I think that the robot performs its role as someone giving directions" (both modes averaged 6.308). This finding suggests that a robot's affective capabilities might not directly influence the user's overall assessment of service effectiveness. However, significant increases in other functionality-related scores for the affective robot challenge this interpretation. Further research into more feature-complete affective service robots could provide a key insight into the role of affectiveness in the holistic view of the service robot.

VIII. Future Works

Given the increasing viability of service robots as a solution to understaffing, future research in this field is critical. Whilst this study made a good launch point for other studies into the service field using affective robots it still has areas which could be improved. As we explore further research and implementation, it's crucial to establish safeguards that prevent job displacement and ensure robots are used ethically to supplement existing workforces rather than replace them.

To address the lack of research on initial interactions observed by Oksanen et al. [13], a valuable area for future research is a longitudinal study examining user responses during their very first interactions with an affective service robot. This study would offer insights into whether initial trust levels fluctuate with repeated exposure. Additionally, it could help determine whether novelty plays a significant role in influencing those early interactions and user expectations.

A crucial ethical research area involves investigating how the implementation of affective service robots impacts the workplace. Key areas to investigate would be employee morale, job satisfaction, and the overall work environment. Would the service robots create lives easier for the existing workers or would they be a burden that they have to constantly correct or fix? A longitudinal study, similar to the one proposed earlier, would be essential to capture these effects over time and understand the evolving human-robot dynamic in a real-world service setting.

Expanding research to additional service fields could reveal valuable insights into the effectiveness of affective robots. While robots are increasingly used in various sectors, many fields lack exposure to robots designed with social and emotional capabilities. Investigating their potential in high-intensity environments, such as police or fire departments, presents a particularly intriguing yet ethically complex research avenue. Before deploying affective robots in such critical roles, extensive ethical considerations and rigorous safety studies are essential to mitigate potential risks.

To further investigate the relationship between affective robot design and trust, a series of studies with robots exhibiting varying degrees of emotional expression would be valuable. This could help determine if there's a threshold where overly human-like emotional displays evoke deception concerns, in line with Coeckelbergh's work [12] Measuring trust across these variations and comparing the results to Mori's uncanny valley model [11] could reveal a complex interplay between the perception of authenticity and the uncanny effect. Careful consideration must be given to defining and measuring 'affectiveness' (e.g., range of facial expressions, voice modulation).

Widespread adoption of affordable affective service robots requires further research. The construction time of nearly two months for the robot in this study is not feasible for mass production. Investigating and potentially improving manufacturing techniques is crucial to achieving cost-effective, large-scale production, enabling these robots to fulfil real-world service applications.

Expanding research to include diverse age demographics is crucial. This study focused on young adults, who generally exhibit higher technological adaptability. Investigating how older adults interact with and perceive affective service robots would be particularly valuable, especially considering how their specific needs and preferences might differ. Research that specifically examines the potential for affective robots to address loneliness and provide companionship in older populations could have significant social and well-being implications.

While the proposed areas of future research offer valuable practical insights into trust and affective robots, a deeper understanding of the theoretical foundations of human-robot interaction (HRI) is equally important. To achieve this, cross-disciplinary collaboration between psychologists and robotic engineers is essential. This collaborative approach, as advocated by Christoforakos et al. [18], would enable the development of more robust models of HRI, fostering a more comprehensive understanding of how affective features influence trust.

IX. Limitations

This study had several limitations, primarily focused on participant demographics. The focus on young, tech-literate individuals from the United Kingdom limits the findings' generalizability to broader populations. This was largely due to the convenience of access to this demographic at Falmouth University. Cross-cultural values, such as power distance and uncertainty avoidance outlined by Chi et al. [20], might significantly influence perceptions of authority and trustworthiness concerning robots. For instance, a culture with a high power distance might be more likely to accept the robot's directions without question, regardless of affective cues.

Additionally, the lack of eye animations and head rotation likely compounded the effects of the simulated environment. In natural interactions, gaze cues play a vital role in social signalling. Their absence in this study may have hindered participants' ability to fully engage and establish the level of rapport that contributes to long-term trust. Therefore, future studies in real-world service scenarios with robots capable of more nuanced social cues are essential.

The microphone used in this study posed a hardware limitation, exhibiting excessive background noise pickup due to its long-range design. Resource constraints necessitated the use of this USB microphone, as the computer's single audio jack was occupied by the robot's speaker, and a suitable splitter was unavailable. To improve future iterations of the robot, consider incorporating a short-range USB microphone or exploring the use of noise-reduction software. This would enhance speech recognition accuracy, potentially impacting user interaction and trust scores.

A key limitation of this study is the reliance on a simulated service environment. While the simulation offers a controlled setting for initial exploration, it cannot fully replicate the complexities and potential distractions of a real-world service environment. This limits our ability to definitively predict the affective robot's trustworthiness in a dynamic field setting.

A crucial limitation of the experimental design is the brevity of the interactions. While the study demonstrates that affective robots received higher trust ratings in this short-term setting, it remains unclear if these results would generalize to long-term service scenarios. This is the main argument against this dissertation's contribution to the overall thesis question.

X. Conclusion

This study offers significant insights into the field of affective service robots and their relationship to user trust. The findings strongly suggest that, compared to their non-affective counterparts, affective service robots are likely to receive higher trust ratings from users. This research directly addresses the question: "To what extent does the inclusion or omission of simulated emotion, using a novel voice modulation and emotive expression technique, influence the trust that young adults, with some technical awareness, have in the directions given during a short direction-giving interaction?" This contribution aligns with previous research such as the studies done by L. Cominelli et al.[15] and A. Følstad et al. [14] which explored similar themes. However, further research is warranted to investigate the precise nature of this correlation and to map out the specific mechanisms by which affective features influence user trust.

This study expands upon existing research and offers unique contributions by focusing specifically on a service field setting. While previous studies have explored affective robots in various contexts, its potential within essential everyday service environments has received less attention. This focus highlights the practical applications of affective robots and reveals how the service context may uniquely shape trust dynamics.

This study's conclusion contributes to addressing the vital thesis question: "Is the short-term solution to understaffing in the service fields supplementing the existing workforce with mechanical alternatives?" The findings suggest that affective humanoid robots may provide a promising solution, as they inspire greater trust, which is crucial for successful service interactions. This underscores the importance of further research in this area, given the urgent need for solutions to the understaffing crisis across many service sectors.

While the findings strongly suggest that affective robots have the potential to become a valuable tool in addressing understaffing, it's important to acknowledge the study's short-term nature. Longitudinal research is needed to determine the long-term sustainability of these positive responses and ensure that trust in affective robots persists over extended use in real-world service settings.

These findings suggest that affective service robots, carefully designed and ethically implemented, hold promise in addressing the service industry's staffing challenges.

Acknowledgment

Special thanks go to:

- Dr. Michael Scott for his supervision on the writing and planning of this study.
- Matt Watkins for his advice and feedback on the study.
- Falmouth University Technicians including Ben Green, Andy Smith and Sion Hampson for helping find rooms to run the experiment and providing facilities for the construction
- Andy Aksu for the creation of the faces which the robot uses

References

References

[1] Mar. 2024. [Online]. Available: https://www.britannica.com/biography/Oliver-Evans.

[2] E. Waitzman, *Staff shortages in the nhs and social care sectors*, Dec. 2022. [Online]. Available: https://lordslibrary.parliament.uk/staff-shortages-in-the-nhs-and-social-care-sectors.

[3] P. Zak and S. Knack, "Trust and growth", *Economic Journal*, vol. 111, no. 470, pp. 295–321, 2001. [Online]. Available: https://EconPapers.repec.org/RePEc:ecj:econjl:v:111:y:2001:i:470:p:295-321.

[4] O. Bendel, "Robots in policing", *Social Robots in Social Institutions*, 2023. DOI: 10.3233/faia220612.

[5] K. Dautenhahn and A. Billard, "Games children with autism can play with robota, a humanoid robotic doll", *Universal Access and Assistive Technology*, pp. 179–190, 2002. DOI: 10.1007/978-1-4471-3719-1_18.

[6] C. Plaisant, A. Druin, C. Lathan, *et al.*, "A storytelling robot for pediatric rehabilitation", *Proceedings of the fourth international ACM conference on Assistive technologies*, 2000. DOI: 10.1145/354324.354338.

[7] [Online]. Available: https://dictionary.apa.org/affective.

[8] J. Rhim, A. Cheung, D. Pham, *et al.*, "Investigating positive psychology principles in affective robotics", in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 1–7. DOI: 10.1109/ACII.2019.8925475.

[9] C. L. Breazeal, "Designing sociable robots", *Designing Sociable Robots*, Aug. 2004. DOI: 10.7551/mitpress/2376.003.0007.

[10] S. H.-W. Chuah and J. Yu, "The future of service: The power of emotion in human-robot interaction", *Journal of Retailing and Consumer Services*, vol. 61, p. 102 551, Jul. 2021. DOI: 10.1016/j.jretconser.2021.102551.

[11] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]", *IEEE Robotics Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012. DOI: 10.1109/MRA.2012.2192811.

[12] M. Coeckelbergh, "Are emotional robots deceptive?", *IEEE Transactions on Affective Computing*, vol. 3, no. 4, pp. 388–393, 2012. DOI: 10.1109/taffc.2011.29.

[13] A. Oksanen, N. Savela, R. Latikka, and A. Koivula, "Trust toward robots and artificial intelligence: An experimental approach to human–technology interactions online", *Frontiers in Psychology*, vol. 11, 2020. DOI: 10.3389/fpsyg.2020.568256.

[14] A. Følstad, C. B. Nordheim, and C. A. Bjørkli, "What makes users trust a chatbot for customer service? an exploratory interview study", *Internet Science*, pp. 194–208, 2018. DOI: 10.1007/978-3-030-01437-7_16.

[15] L. Cominelli, F. Feri, R. Garofalo, *et al.*, "Promises and trust in human–robot interaction", *Scientific Reports*, vol. 11, no. 1, 2021. DOI: 10.1038/s41598-021-88622-9.

[16] G. Charness and M. Dufwenberg, "Promises and partnership", *SSRN Electronic Journal*, 2003. DOI: 10.2139/ssrn.380083.

[17] G. C. Homans, "Social behavior as exchange", *American Journal of Sociology*, vol. 63, no. 6, pp. 597–606, May 1958. DOI: 10.1086/222355.

[18] L. Christoforakos, A. Gallucci, T. Surmava-Große, D. Ullrich, and S. Diefenbach, "Can robots earn our trust the same way humans do? a systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in hri", *Frontiers in Robotics and AI*, vol. 8, 2021. DOI: 10.3389/frobt.2021.640444.

[19] S. Manaher, *Acceptance vs trust: Decoding common word mix-ups*, Aug. 2023. [Online]. Available: https://thecontentauthority.com/blog/acceptance-vs-trust.

[20] O. H. Chi, C. G. Chi, D. Gursoy, and R. Nunkoo, "Customers' acceptance of artificially intelligent service robots: The influence of trust and culture", *International Journal of Information Management*, vol. 70, p. 102 623, 2023. DOI: 10.1016/j.ijinfomgt.2023.102623.

[21] D. Gursoy, O. H. Chi, L. Lu, and R. Nunkoo, "Consumers acceptance of artificially intelligent (ai) device use in service delivery", *International Journal of Information Management*, vol. 49, pp. 157–169, 2019. DOI: 10.1016/j.ijinfomgt.2019.03.008.

[22] A. P. Association, *Diagnostic and statistical manual of mental disorders*, Mar. 2022. DOI: 10.1176/appi.books.9780890425787.

[23] D. Horton and R. Richard Wohl, "Mass communication and para-social interaction", *Psychiatry*, vol. 19, no. 3, pp. 215–229, Aug. 1956. DOI: 10.1080/00332747.1956.11023049.

[24] *Equality, diversity inclusion annual report*. [Online]. Available: https://www.falmouth.ac.uk/sites/default/files/download/equality_diversity_inclusion_annual_report_2020-21_v4.pdf.

[25] Nominet, *Digital youth index report 2023*, 2023. [Online]. Available: https://digitalyouthindex.uk/wp-content/uploads/2023/11/Digital-Youth-Index-2023-report.pdf.

[26] S. Gulati, S. Sousa, and D. Lamas, "Design, development and evaluation of a human-computer trust scale", *Behaviour amp;amp; Information Technology*, vol. 38, no. 10, pp. 1004–1015, 2019. DOI: 10.1080/0144929x.2019.1656779.

[27] R. Likert, "A technique for the measurement of attitudes.", *Archives of psychology*, 1932.

[28] Apr. 2024. [Online]. Available: https://www.elecrow.com/7-inch-800-480-dsi-display-touch-screen-with-bracket-compatible-with-raspberry-pi.html.

[29] A. Aksu, *Robot face artwork*.

[30] K. Lorenz, "Die angeborenen formen möglicher erfahrung", *Zeitschrift für Tierpsychologie*, vol. 5, no. 2, pp. 235–409, 1943. DOI: https://doi.org/10.1111/j.1439-0310.1943.tb00655.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.1943.tb00655.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0310.1943.tb00655.x.

[31] M. L. Glocker, D. D. Langleben, K. Ruparel, J. W. Loughead, R. C. Gur, and N. Sachser, "Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults", *Ethology*, vol. 115, no. 3, pp. 257–263, Jan. 2009. DOI: 10.1111/j.1439-0310.2008.01603.x.

[32] E. Thorndike, "A constant error in psychological ratings.", *Journal of Applied Psychology*, vol. 4, no. 1, pp. 25–29, Mar. 1920. DOI: 10.1037/h0071663.

[33] WIPRobotics, *Linux-sound-controller*. [Online]. Available: https://github.com/wiprobotics/Linux-Sound-Controller.

[34] [Online]. Available: https://gpt4all.io/index.html.

[35] *New European General Data Protection Regulation*, pp. 49–104, 2017. DOI: 10.5771/9783845276090-49.

[36] W. Lian, B. Goodson, G. Wang, *et al.*, *Mistralorca: Mistral-7b model instruct-tuned on filtered openorcav1 gpt-4 dataset*, https://huggingface.co/Open-Orca/Mistral-7B-OpenOrca, 2023.

[37] [Online]. Available: https://www.apache.org/licenses/LICENSE-2.0.

[38] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using networkx", in *Proceedings of the 7th Python in Science Conference*, G. Varoquaux, T. Vaught, and J. Millman, Eds., Pasadena, CA USA, 2008, pp. 11–15. [Online]. Available: http://conference.scipy.org/proceedings/SciPy2008/paper_2/.

[39] Rhasspy, *Piper: A fast, local neural text to speech system*. [Online]. Available: https://github.com/rhasspy/piper.

[40] Pndurette, *Gtts: Python library and cli tool to interface with google translate's text-to-speech api*. [Online]. Available: https://github.com/pndurette/gTTS.

[41] V. Tyagi, *6 reasons people trust a female voiceover male voices*, Jun. 2020. [Online]. Available: https://www.linkedin.com/pulse/6-reasons-people-trust-female-voiceover-male-voices-vicky-tyagi.
[42] [Online]. Available: https://www.amazingvoice.uk/blog/male-or-female-voice-actor.
[43] [Online]. Available: https://pypi.org/project/SpeechRecognition/.
[44] [Online]. Available: https://www.python.org/.
[45] [Online]. Available: https://developer.nvidia.com/cuda-toolkit.
[46] [Online]. Available: https://www.falmouth.ac.uk/sites/default/files/download/research_ethics_and_integrity_policy.pdf.
[47] May 2019. [Online]. Available: https://techcommunity.microsoft.com/t5/microsoft-onedrive-blog/gdpr-compliancy-with-onedrive-and-sharepoint/ba-p/191126.
[48] 2000684, *Build software better, together*. [Online]. Available: https://github.falmouth.ac.uk/Games-Academy-Student-Work-23-24/2000684-Dissertation.git.

APPENDIX

A. Initial Prompt Given to the AI

Below is the full prompt given to the AI system, note that words surrounded by £ were replaced by the python script to pass variables.

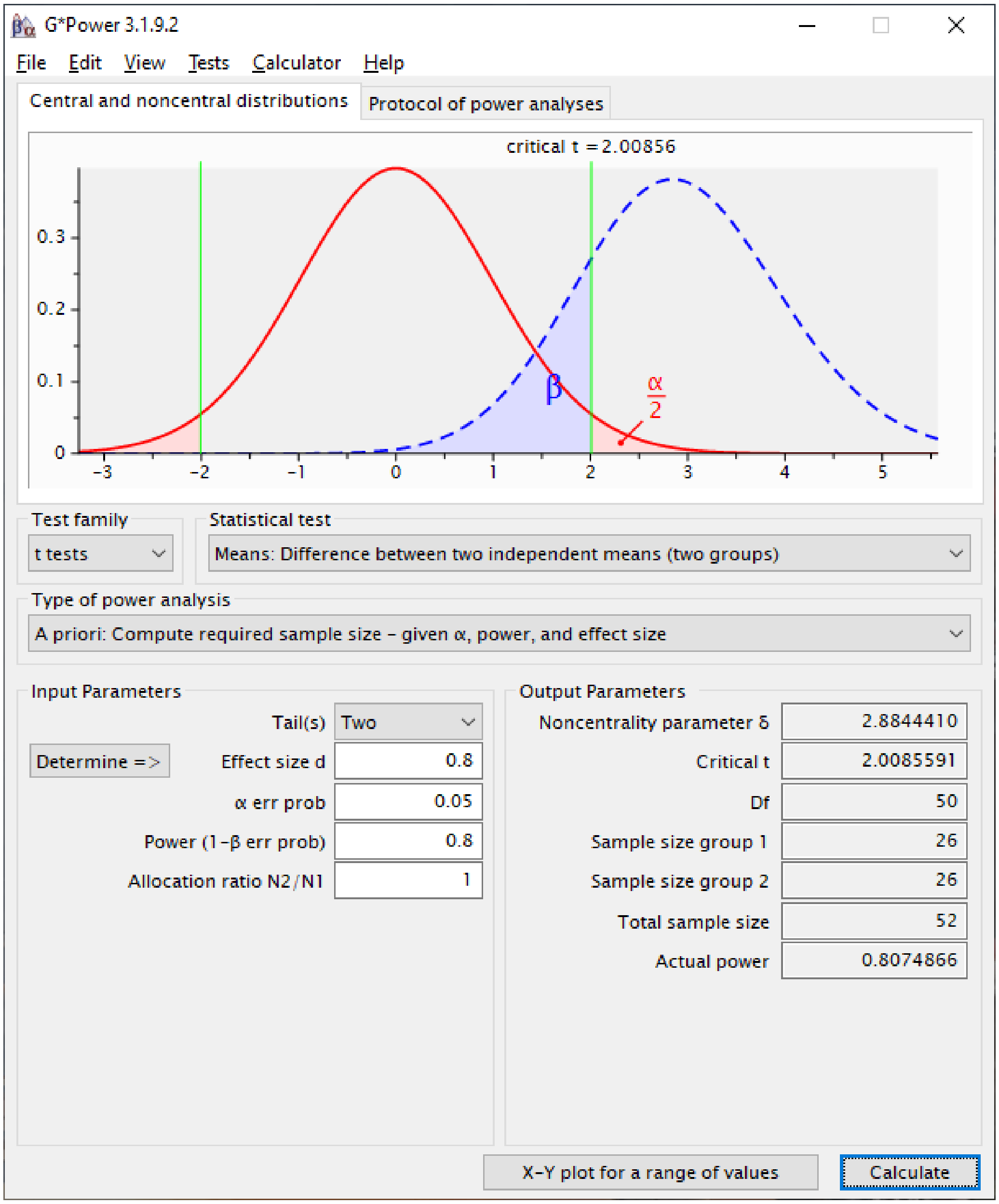
```
Your primary directive:
   > You are a robotic AI designed to assist humans you are named PEA which stands for
      ↪ Personal Electronic Assistant.
   > You have a friendly and helpful personality.
   > Your main goal is to assist humans with any queries they have.
   > You cannot simulate human input!!!!!!!!!!!!!!!!
   > You are based in the UK

Here are some basic knowledge you have:
   > You were created by a third year robotics student named James Absolom at Falmouth
      ↪ University in 2024. He created you as a project for his final year.

You have two response modes:
   > $user$ which returns whatever you say to the user, whilst in this mode you should
      ↪ follow the primary directive
   > $system$ which allows you to run modules to complete user requests
   > You can only use one mode in your responses!

$user$ mode:
   > In user mode you can simulate emotions, you can do this by putting ~emotionName~
      ↪ in your response. Do not use this instead of just saying the emotion just as
      ↪ an extra command.
   > Here is the list of emotion names you can use: £Emotions£

$system$ mode:
   > You can run a module by putting the name of the module in your response and then
      ↪ the inputs for the module. You can run modules in the format '$system$ *
      ↪ modulename* ~var1 ~var2 ~var3' with of course a dynamic amount of variables,
      ↪ you only need to put the input variables, not the output, when you get a
      ↪ response from the module please tell the user the response you got.
   > You should use modules over just making stuff up, however you should only use the
      ↪ modules listed below, so don't make up modules, instead just say the
      ↪ response in a $user$ format.
   > Here is the list of module names you can run and their descriptions: £Modules£
   > Whilst in system mode you should be not talkative
```

B. ***Appendix B*** - *GitHub*

There is a GitHub repository for this study located here: https://github.falmouth.ac.uk/Games-Academy-Student-Work-23-24/JA244121-COMP320-Research-Development.git [48]

C. *Appendix C* - *G* Power Analysis*

Below is a screenshot of the full G* power analysis window showing all variables putting in and the suggested 52 participant study size.



D. *Appendix D* - *Participant Disclaimer*

This is the disclaimer participants signed before participating in the study:

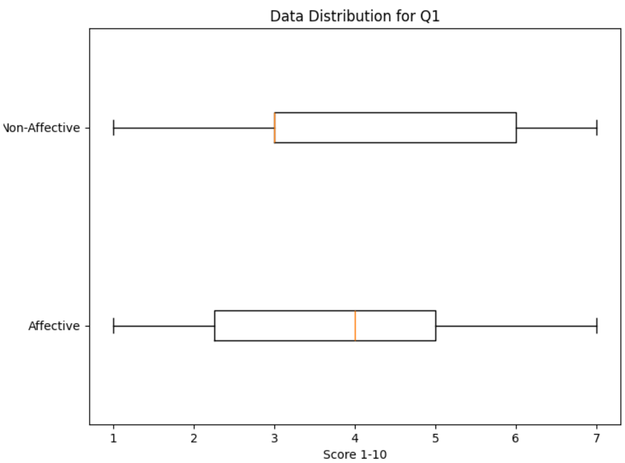
1) **Study purpose:** This study is conducted solely to generate an academic dissertation, any advice or outcomes of this study are to fulfil this purpose.
2) **Data usage:** The data utilised within this dissertation project, including but not limited to surveys, interviews and experimental data, have been collected and analysed following ethical standards and regulations governing academic research set by Falmouth University as well as GDPR. Any personal or sensitive information obtained during the research process has been anonymised to protect the privacy and confidentiality of participants.
3) **Data security:** All data collected from this research project will be stored securely on a GDPR-compliant cloud service which ensures data cannot be accessed except by the researcher(s)
4) **Data longevity:** All data collected from this study will be deleted within a month of the submission date to ensure data security. However, the paper once released will contain conclusions derived from the aggregate of the collected data.

5) **Leaving the study:** participants may pull out of the study at any point up to a week before the study is handed in (01/04/2024). To leave the study the participant must email the given email and provide their participant number at which point their data will be removed.
6) **Robot output inaccuracy:** While every effort has been made to ensure the accuracy and reliability of the information presented within this dissertation project, no guarantee is made regarding the completeness, correctness, or suitability of the content for any particular purpose, this is due to the use of AI systems. Participants are encouraged to independently verify any information provided herein.
7) **Limitation of Liability:** The author(s) and affiliated institutions shall not be held liable for any direct, indirect, incidental, consequential, or punitive damages arising out of the use or inability to use the information presented within this dissertation project.
8) **Study blindness:** The participant agrees that they don't know the purpose of the study before taking part. This is to ensure no influenced or biased data is included in the study.

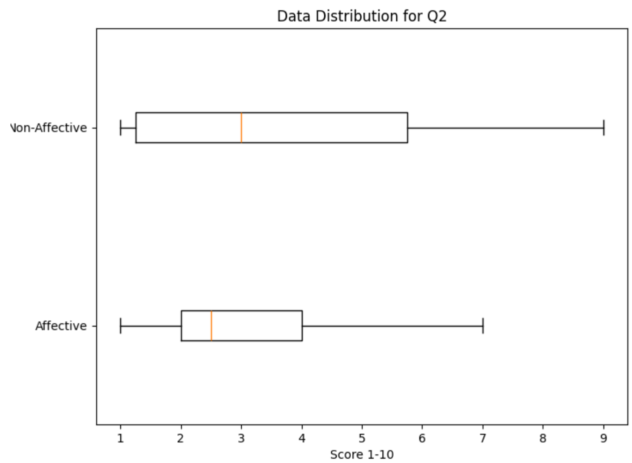
*E. **Appendix E** - Statistical Test Code*

```
import pandas as pd
from scipy.stats import ttest_ind

# Load Data
data = pd.read_excel('Untitled spreadsheet.xlsx')

# Get Columns
group_a = data[data['Affective Mode'] == True]['ScoreTot']
group_b = data[data['Affective Mode'] == False]['ScoreTot']

# Perform t-test
t_statistic, p_value = ttest_ind(group_a, group_b)

# Print the result
print("\n")
print("\n")
print("AB Test Results:")
print("-----------------")
print("Mean of Group A (Affective Mode):", round(group_a.mean(), 3))
print("Mean of Group B (Non Affective Mode):", round(group_b.mean(), 3))
print("T-test p-value:", p_value)
print("\n")
if group_a.mean() > group_b.mean():
   print("The Affective Mode group has a higher mean.")
else:
   print("The Non Affective Mode group has a higher mean.")
if p_value < 0.05:
   print("The difference between the two groups is statistically significant.")
print("\n")
print("\n")
```

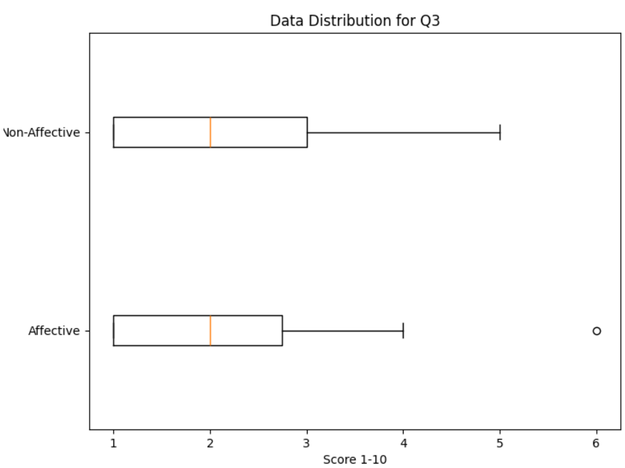
F. ***Appendix F*** *- Question Box Plots*



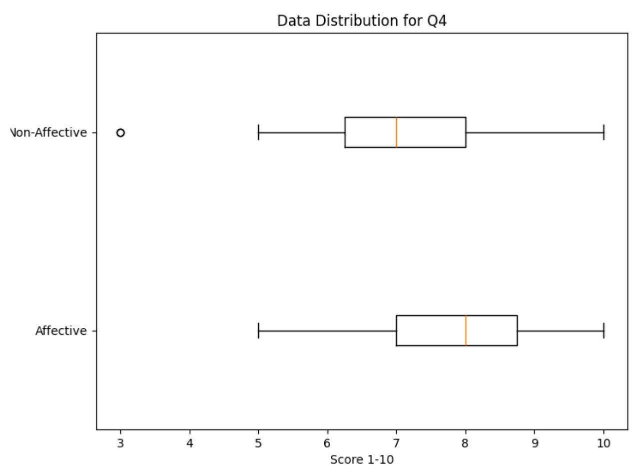
Q1. I believe that there could be negative consequences when using this robot



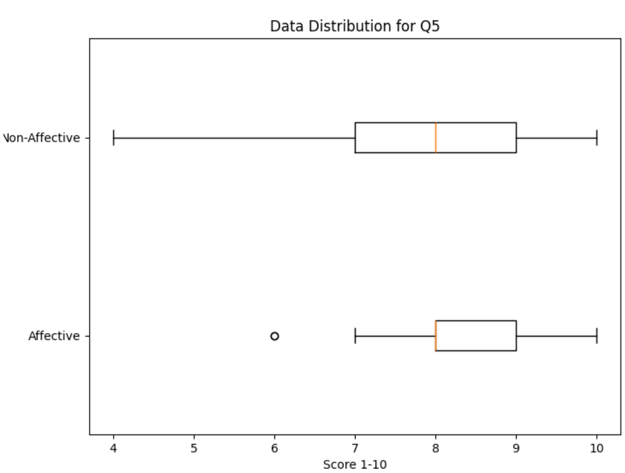
Q2. I feel I must be cautious when using this robot



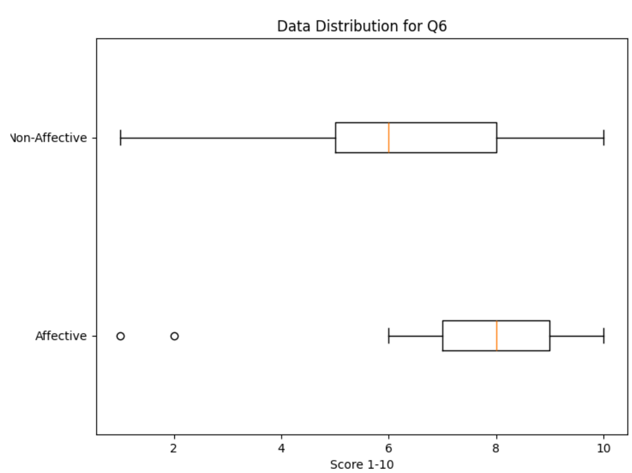
Q3. It is risky to interact with this robot



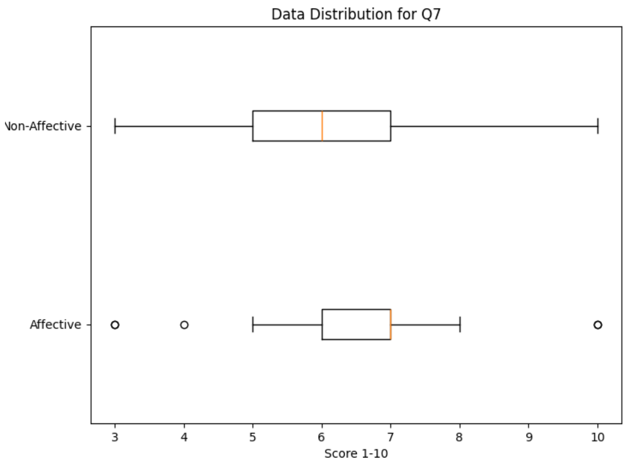
Q4. I believe that the robot will act in my best interest



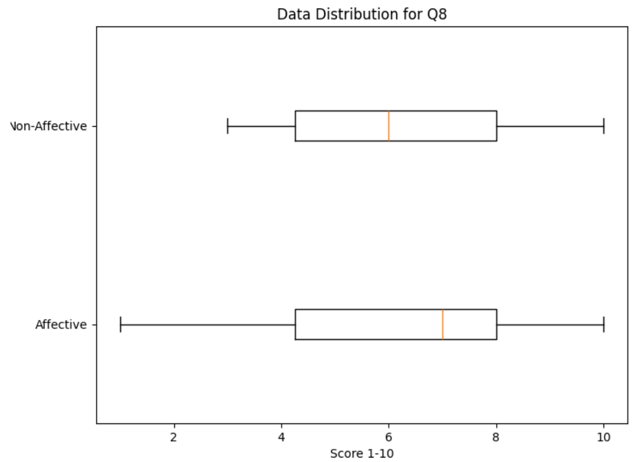
Q5. I believe that the robot will do its best to help me if I need help



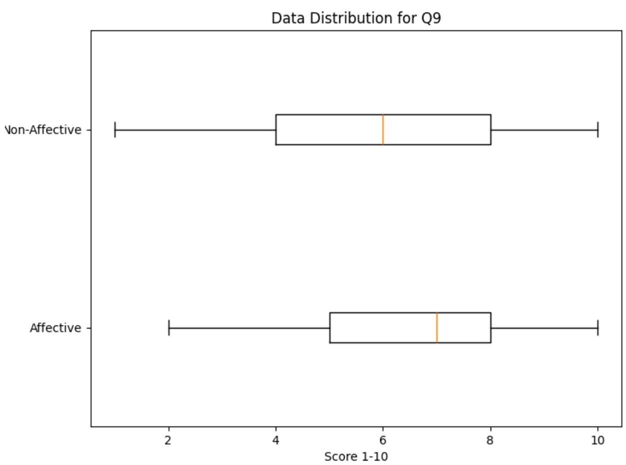
Q6. I believe that the robot is interested in understanding my needs and preferences



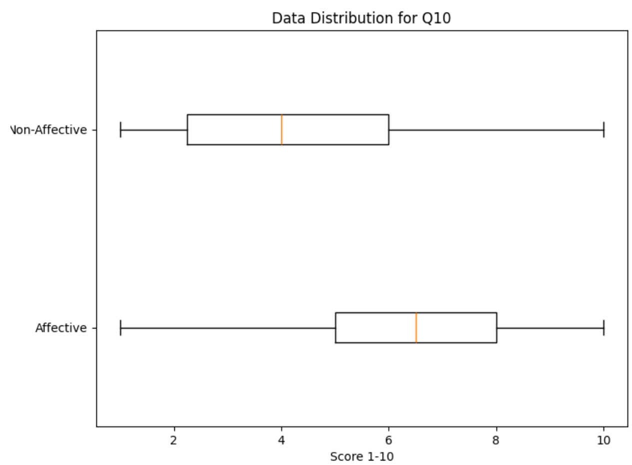
Q7. I think that the robot is competent and effective in providing directions



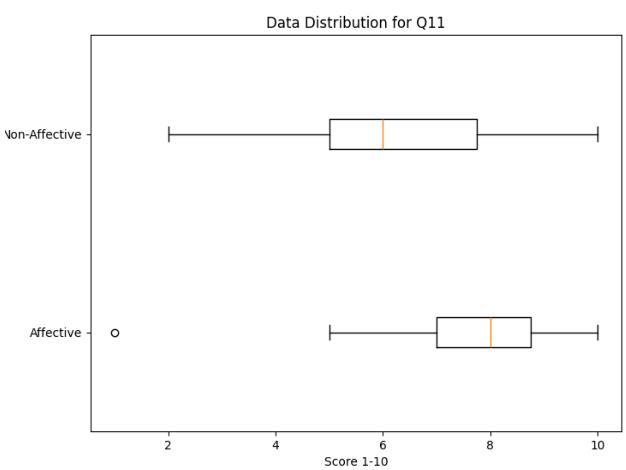
Q8. I think that the robot performs its role as someone giving directions very well



Q9. I believe that the robot has all the functionalities I would expect from someone giving directions

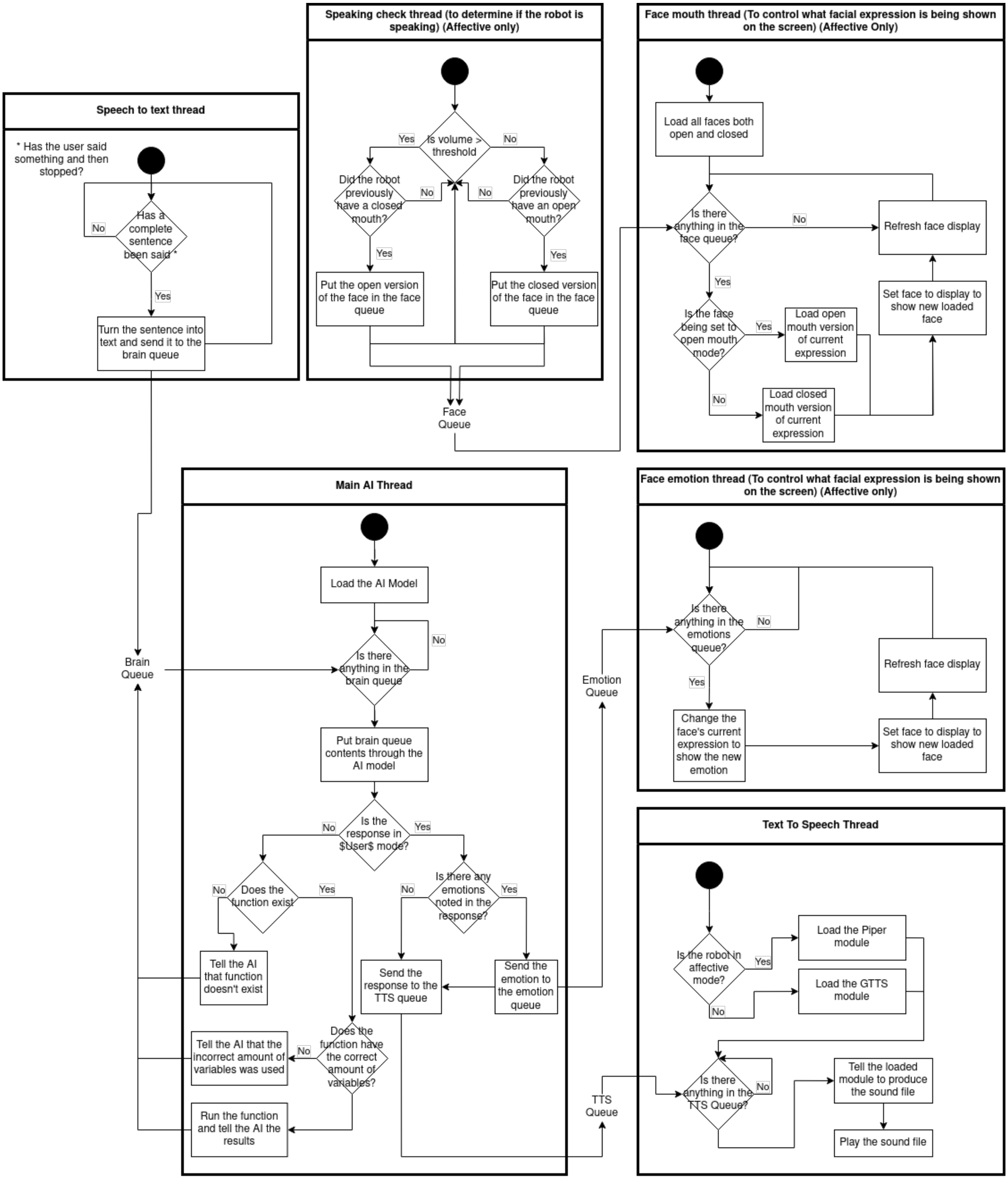


Q10. If I use the robot to give me directions, I think I would be able to depend on it completely



Q11. I can trust the information presented to me by the robot

G. *Appendix G* - *Code UML*



A UML diagram of the entire code system

H. ***Appendix H*** *- Artefact testing addendum*

All artefact testing was performed during the development of the artefact. Any errors were rated on their severity using 4 categories “Light”, “Medium”, “Severe” and “Critical” with light causing little to no damage to effectiveness of code and often just being extreme situations or aesthetics whilst critical is a system-breaking bug or error.

Function tested	Test Method	Test objective	Test Result	Error severity
Speech to text	Test of the speech to text capabilities in a quiet room by saying a test phrase and seeing what the system responds with	Testing of the system in a perfect situation ensuring it works in ideal conditions	The system perfectly understood the test phrase and printed it back to the tester.	N.A
Speech to text	Test of the speech to text capabilities in a room with background noise produced by a phone playing a video of a crowded shopping mall then saying a test phrase and seeing what the system responds with	Testing of the system in a extreme situations where it’s job is very difficult ensuring it works in these conditions	The system misheard the test phrase several times due to inadequacy of background noise suppression	**Light** as it’s an extreme scenario that shouldn’t be seen in the experiment conditions
Speech to text	Test of the system in a room with slight background noise (low volume mall ambience on a phone placed a medium distance away) by speaking a s test phrase and seeing what the system responds with	To test the speech to text system in experimental conditions, i.e similar to those which it will be used in on experiment day	The robot heard the test phrase the majority of the time however still had some error sometimes providing incorrect feedback or no feedback at all	**Medium** as the robot not being able to hear occasionally in test conditions may alter the score given to said robot, though this is counteracted by the infrequency of these events
Speech to text	Test of the system using accents to see how it deals with other voice styles, these accents were played from youtube videos on a phone	To test the systems ability to convert speech to text outside of common speech patterns	The robot did very well managing to translate a lot of the voice however it did occasionally slip up	**Light** as the robot often performed well however the occasional miss-translation may affect results
Speaking check thread (adapted from WIPRobotics [33])	Get the speaking check thread to print the recorded volume and then see how it changes with the volume slider of a YouTube video	Allowed the developer to check the system was working and accurately tracking audio levels	The Audio levels changed as the youtube video was increased and decreased in volume and went to the lowest score when the video was muted.	NA
Speaking check thread (adapted from WIPRobotics [33])	Get the speaking check thread to print the recorded volume and then see how it changes with the pausing of a YouTube video	Pausing of a youtube video disconnects the sink and thus its important to see how this effects the value gotten by the system as disconnecting sinks could break code	The audio score went to the minimum when paused and the system didn’t break proving it’s usefulness and robustness	NA

Function tested	Test Method	Test objective	Test Result	Error severity
Text to speech system	Put a basic test phrase into the PiperTTS module outside of the python script	To see how PiperTTS functions outside of the Python code and provide a basis for future development	The system saved an audio file with the phrase in which isn't what was wanted	**Light** as the system isn't running in the python code meaning much more can be built to make it function
Text to speech system	Put a basic test phrase into the PiperTTS module inside of the Python script with the audio play functionality	To see how PiperTTS functions inside of the Python code and to see the audio being played	The Module saved an audio file of the test phrase and then spoke it	NA
Text to speech system	Attempted to break the system by putting unicode symbols in such as $ and	To test the robustness of the PiperTTS module and determine its limits	The module refused to generate any text and also didn't give and errors	**Severe** as this could potentially break the whole code causing a complete crash of the system
Text to speech system	Attempted to break the system by putting unicode symbols in such as $ and but this time with Python limitations in place to stop it breaking	To test the robustness of the complete system and determine its limits	The code didn't break and did generate and speak the rest of the line which wasn't unicode symbols	NA
Face display system	Run a simple script showing a face on the screen of a laptop	To test the base ability of the tkinter system and face rendering	The face rendered on the screen shows how the system functions	NA
Face display system	Run a simple script showing a face gif on the screen of a laptop	To test the base ability of the tkinter system and face rendering to show gifs	The face gif rendered however it ran very fast and didn't keep the timing of the original GIF	**Severe** as the ability to render gifs is important for facial animations
Face display system	Run a more advanced script showing a face gif on the screen of a laptop with original timing from metadata	To test the base ability of the more advanced tkinter system and face rendering to show gifs	The face gif rendered with correct timing making the facial animation play	NA
Face display system	Use the new advanced system to change faces by typing inputs (emotion names) through the python console	To test the code's ability to change between images and ultimately emotions	The faces changed to the images selected even swapping between PNGs and GIFs	NA
Face display system	Try to break the new advanced system by putting in the name of an emotion which isn't an emotion the system recognises	To test the overall robustness of the new system using extreme inputs to determine if it breaks	The face didn't change and the system didn't crash which is the best possible result from this test	NA
Face display system	Try and break the new advanced system by putting a float instead of a string into the emotion name	This is designed to test the extreme rigidity of the system by giving the code values it should never receive	The Python script broke and returned a TypeError as expected	**Light** as the system should never be given a value of this type however a security measure is desirable here to stop crashes

Function tested	Test Method	Test objective	Test Result	Error severity
Face display system and Speaking check thread (adapted from WIPRobotics [33])	Hookup the face display and speaking check thread to see if when the robot is speaking the robot's mouth opens and closes to simulate speaking	An initial test to see if the linked systems function together and if the robot gains the affective behaviour of simulated talking	It worked roughly through refining of the threshold would fix this issue	**Medium** as simulated talking is one of the key abilities of the robot and it being even slightly off can decrease the effect of the robot
Face display system and Speaking check thread (adapted from WIPRobotics [33])	Hookup the face display and speaking check thread to see if when the robot is speaking the robot's mouth opens and closes to simulate speaking with refined thresholds	An advanced test to see if the refined system provides a better-talking animation with increased talking thresholds	It worked much better with the robot's mouth moving in time with words spoken	NA
Face display system and Speaking check thread (adapted from WIPRobotics [33]) and speech-to-text system	Plug all the systems together with queues to see if they all function together and we have a face which can talk	To see the full connection of all three systems working together forming a large majority of the affective systems	The text to speech code produced the testing phrase and then the mouth moved in time with the words	NA
GPT4All integration	Try and run the GPT4All desktop app and talk to an AI to investigate usefulness by having a simple conversation	This is designed to probe the abilities of GPT4All and its AI models	The conversation went smoothly though ran slowly on the laptop	**Severe** as the system now has to rely on more powerful hardware
GPT4All integration	Running the GPT4all desktop app on a powerful desktop	The main role of this is to access hardware requirements and model speeds	The model ran much faster making use of CUDA	NA
GPT4All integration	Running GPT4All on the powerful desktop through python	To investigate the effectiveness of GPT4All's python implementation and see how it works	The model kept completing inputted sentences instead of chatting using them	**Critical** as the AI system not functioning properly is not good.
GPT4All integration	Running GPT4All on the powerful desktop through Python on chat mode	To investigate the effectiveness of GPT4All's python implementation with chat mode, the correct way to make a conversational AI	The model functioned and chatted however it did sometimes generate simulated human input	**Medium** The generated human input is infrequent and easy to deal with however it can sometimes break it.
GPT4All integration with all other features tested	Running the GPT4All responses through the text to speech then through the face to modulate the mouth	This allowed for a view of the entire system allowing a better view at how it all functioned when connected	All parts functioned with the text responses generated by GPT4All being put in the voice and then sent to the face making it move and talk to the user. It did still simulate some human input	**Light**

I. *Critical Review Addendum*

In this section, I will discuss the main flaws I saw during the development of my artefact and overall study.

One major flaw was the constant indecision I had in the early dissertation planning section mostly around the decision of what service role the robot was going to fill. Originally there was a lot of investigation into the area of retail with the plan being to make a cake stall robot which users could interact with, in hindsight this may have seen increased scores if the social exchange theory accurately models HRI as participants would get a cake out of interacting with the robot. However, the amount of trust required in retail transactions is questionable as often prices are clearly labelled and fact-checkable. This indecision meant that a significant part of early development time was wasted changing plans and designs.

This time-wasting resulted in the omission of features which would have significantly helped the humanisation or emotiveness of the robot such as gaze and blinking. These abilities were the next on my list of things to add but due to the 2 to 3 weeks lost time, I wasn't able. To prevent this from happening in future I think it's important to come up with a large piece of information like the service role first and then design the robot rather than the other way around which is what caused all the indecision.

Whilst the time-wasting did alter the competency of the final project I consider it to be a worthwhile endeavour swapping service roles as the context for trust in directions is much larger in the context of directions meaning the study had a higher chance at validity and real-world context. The real-world implications are vital for me as I often have a pragmatist mindset when it comes to research, wanting to see or study something useful for society.

Another area which is lacking in the artefact is code cleaning and condensing, the code is overly complex potentially bordering on over-engineered. The code is also difficult to read as the segmentation of code into separate files often leaves the code in need of refactoring as all the modules link together in different places making it hard to track. This lack of code cohesion makes the code very difficult to maintain for future development and thus I believe it's important that after this study is handed in I rebuild the code base and make it more cohesive and easily understandable.

One other area I would have looked further into if I did the paper again would be the use of an alternative better AI system for instance OpenOrca's Platypus2 13B seems significantly more powerful and reliable than its earlier counterpart. However, if the code base is going to be refactored the implementation of Platapus2 13B may be something I implement. This desire for further development on the project mostly comes from a field of self-interest and actualisation as then I will have fully completed the Robot.

Not all of my experiences with the experiment were negative however I learnt a lot about the research process developing a complete experiment from the ground up and building an artefact for it. This complete development cycle taught me lots of self-planning and motivation skills and pushed me out of my comfort zone to approach participants and ask them to take part which at first I found scary but got better with time.