Human-Chatbot Interaction

Assessing technology acceptance, confidence and trust in chatbots within their application areas

Author: Joseph Mercieca, Supervisor: Radu Jianu, Years: 2018-2020
City St. George, University of London

ABSTRACT

Over recent years, there was an increasing interest in automated technologies, including chatbots that rely on advancements in artificial intelligence (AI) and machine learning. Nowadays, chatbots are used in many application areas. Various reports (Deloitte, 2017; Gartner, 2018; Juniper, 2018), estimated an increase of use up to 85% by 2020 and 2023. Nonetheless, a knowledge gap concerning the intent to use, and trust in chatbots still exists. The present research aimed to discover factors influencing acceptance, trust, and confidence in chatbots. This was investigated by administering 2 chatbots, using different technologies and deployed in different application areas, to a sample of 14 participants (n=14) (7 females, 7 males). Data was collected through interviews, observations, and a human-computer trust survey validated by previous research.

Findings indicated that the application area predicts chatbot usage. This is influenced by other factors like understandability, reliability, usefulness, risks, and external factors (e.g., Cambridge Analytica). It was noticed that human likeness induced negative emotions among female participants, which led to the 'uncanny valley effect'. In contrast, males were found to have high trust propensity in chatbots. These discoveries suggest that these factors, combined with gender differences and attitudes, could predict trust and chatbot acceptance.

Keywords: human-chatbot interaction, trust, technology acceptance, understandability, gender

Thanks and acknowledgment: Radu Jianu, Prof. Shirley McGregor (Central Queensland University), Sebastian Schultz and Tom Shenhav (Fit Analytics GmbH), Asbjørn Følstad and Petter Bae Brandtzæg (SINTEF), Siri Hegna Berge, Cecilie Bertinussen Nordheim, Dr. Alex Taylor, Robert Foster (PG), Jihyun Jung (PG), and Jeff Sauro (MeasuringU) for guidance, material, and additional insights.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION AND OBJECTIVES

A chatbot (also termed as a virtual assistant) is a conversational agent technology designed to interact with people through a computer keyboard-input, or voice via its natural language processing (NLP) and understanding (NLU) capabilities. This technology has been around since 1966 following the inception of ELIZA, the first keyboard-operational chatbot. Due to the continuous advances in machine learning and artificial intelligence (AI), the chatbot technology is gaining momentum, as seen in various application areas, most prominently retail and customer support. This is due to the technology benefits such as ‘availability’ (e.g., dealing with multiple users simultaneously), ‘fast accurate responses’, and ‘cost-effectiveness’ (e.g., hiring fewer staff members), among others.

Previous reports by Deloitte (2017) and Gartner (2018) estimated that in the upcoming years, customer support would mainly be managed by virtual assistants. However, the chatbot technology also paved ways in non-customer centric domains, for example, healthcare (e.g., Babylon), mental health, and well-being (e.g., Woebot, Wysa, Replika), and education (e.g., Differ). According to Følstad et al. (2018a), the chatbot technology holds opportunities for beneficial societal improvements for instance by strengthening people's ‘autonomy’ (e.g., better access to health and welfare), ‘competence’ (e.g., skills, education) and ‘social relatedness’ (e.g., bringing people together).

Despite these advances, there is still a lack of research concerning the users’ trust levels, confidence, and acceptance in this technology. The reason is based on the fact that chatbots are still considered as a new topic; this requires more in-depth investigation. Given the lack of empirical findings, this research investigated how people accept and trust different types of chatbot technologies (e.g., keyword-input, voice operational) within specific application areas. In return, the study attempted to discover factors that could predict acceptance, confidence, and trust in this technology before its adoption.

1.1 RESEARCH QUESTIONS AND OBJECTIVES

This study obtained relevant information regarding technology acceptance, confidence, and trust in chatbot technologies after answering the following research questions:

RQ1: Are acceptance, confidence and trust constructs affected by the type of chatbot technology (e.g., conversational, keyword-input), the application area (e.g., finance, well-being), or both?

RQ2: How would users react towards chatbots with an extent of human likeness? Would acceptance, trust and confidence levels be enhanced or undermined?

RQ3: Would frequent interaction with a chatbot increase users' trust, acceptance and confidence levels?

These questions were answered after meeting the following objectives:

RO1: To conduct a research to determine whether technology acceptance, confidence, and trust levels are influenced by different chatbots technologies, their use case, or both. This will be done by utilizing a mix of qualitative (e.g., interviews, observation, and user-driven experimentation), and quantitative (e.g., surveys, paired-samples *t*-Test) approaches.

RO2: To obtain information from users basing on their reactions, feelings and exhibited behaviours while interacting and utilizing a chatbot with *"humane"* characteristics.

RO3: To gain understanding on whether acceptance, confidence, and trust are enhanced by the more users interact with a chatbot.

RO4: To validate these findings, the human-computer trust scale obtained from the authors (Madsen & McGregor, 2000) will be utilized as a survey to measure trust in chatbots. An additional purpose for this objective is to determine whether a valid scale tailored to assess trust in (chatbot) technologies is necessary to have.

1.2 AMENDMENTS TO THE ORIGINAL PROPOSAL PLAN

Since the time the original project proposal plan (Appendix A) was submitted, various changes were made. Firstly, the original intention of the research design experiment was to administer three chatbots to a sample of 21 participants split into three groups. To make this research less exhaustive, two chatbots were used instead, and the sample size was reduced to 14 participants arranged in 2 groups. The chatbots initially proposed in the project plan were replaced with others deemed more suitable for this research. In that case, Cleo (finance) and Replika (well-being, companionship) were used instead of H&M (retail), Wysa (mental health), and Mitsuku (companionship). There was also the intention to conduct a discourse analysis injunction with thematic analysis to evaluate the conversations between chatbots and participants. This approach was, however, omitted because the deductive approach to thematic analysis generated enough data to support the research objectives. As part of the results analysis, statistical methods were used to interpret findings from the human-computer trust surveys, assigned to measure trust in each chatbot. Examples include the calculation of the sample mean values, variances, and standard deviations; followed by a paired-samples *t*-Test with a Bonferroni corrected alpha (p) value. These methods were utilized due to the necessity in determining the strength, validity, and significance of the survey results.

Also, two terminologies used in the second research question (RQ2) and the ‘beneficiaries’ sections were replaced. The terms *'personalization'* and *'anthropomorphism'* were used to describe the relational feature the chatbot technology holds, and how these impact the user. Personalization has different interpretations across other domains, for instance, online marketing. Also, the term anthropomorphism was noticed to be often misused in the reviewed literature. Therefore, to eliminate misinterpretations, and by abiding with appropriate standard definitions, both terms were replaced with *'human-likeness'*. Researchers commonly use this definition; therefore, this change was fundamental to maintain consistency with the referenced literature. Lastly, the wording in the established objectives (RO1-RO4) was revised to maximize clarity for the readers while retaining the original scope of this research.

1.3 CHATBOT TECHNOLOGY AND THE SCENARIO TO DATE

Thanks to the capabilities of chatbots as conversational agents, many online businesses and entities are resorting to this technology as an alternative or additional solution to customer support, after-sales, and general inquiries (e.g., services, products, orders, etc.) while minimizing expenses (e.g., staff recruitment). The chatbot technology is creating unprecedented opportunities in various areas such as finance (e.g., Cleo, Chip, Plum, Capital One), education (e.g., Differ), retail (e.g., H&M), healthcare (e.g., Florence, Babylon) and mental health plus well-being (e.g., Replika, Wysa, Woebot) among others. More use cases are exemplified in the BotList (https://botlist.co/) directory. Apart from that, thanks to the benefits of artificial intelligence (AI) and machine learning, the chatbot technology is rapidly evolving. As a result, graphically embodied conversational agents and social robots are gradually increasing in their presence within the technological ecosystem. Given the fast-paced growth, it eventually came to the point that new typologies and classifications are needed so to distinguish chatbots by technology type and application area (Følstad et al., 2019).

According to Deloitte (2017), it is estimated that by the year 2020, 85% of customer interactions will be handled without a human agent. As well, the research firm Gartner (2018) also predicted that about 25% of customer support operations would use virtual assistants. Moreover, it is also estimated that by the year 2023, over 70% of chatbots accessed will be retail-based (Juniper, 2018). This is why it is believed that retailers will gain the most from chatbot technology. These predictions are an indicator of the increasing demand for ‘intelligent virtual assistants (IVAs)’. Nowadays, conversational AI systems supporting ‘natural language processing (NLP)’ and ‘natural language understanding (NLU)’ also became part of computer operating systems. Popular examples are Siri (Apple), Cortana (Microsoft), and Google Assistant (Google). Furthermore, ‘Internet of Things (IoT)’ devices like Echo and Alexa (both by

Amazon), Google Home (Google), and other ubiquitous technologies are becoming popular among domestic users.

Given the emerging demand, companies like Oracle (Oracle Intelligent Bots), IBM (IBM Watson Assistant), Amazon (Amazon Lex), and RASA provide necessary utilities for everybody who wants to build their chatbot. As well, Facebook and Twitter also offer utilities to those who want to develop less complicated (script-based) chatbots through their API, which are deployed within their online ecosystem (platform). Additionally, messaging applications like Whatsapp, Telegram, Slack, and Kik support the integration of chatbots through their API. However, the opportunities are not limited to these as 'online chatbot builders' like Chatfuel, Pandorabots, Chatleap, and Botsify, among others, provide more accessible alternatives for users with a minimal technical background to instantly build and deploy chatbots.

1.4 BENEFICIARIES

This research is expected to be beneficial and contributory for various research and development. Gaining an understanding of how individuals perceive, trust and adopt technologies could instruct beneficiaries regarding aspects that must be taken into consideration before designing and deploying chatbots. The beneficiaries are, and possibly not limited to:

Researchers: Given the novelties pertaining to this project topic, researchers from various disciplines can obtain knowledge concerning factors predicted to influence trust, confidence and acceptance. Findings from this research could serve as pediment for future investigations. This could be by: i) replicating this research to investigate further how the aforementioned constructs are affected by reasons other than those reported in the present research; and, ii) by expanding on the present findings to generate new hypothesis aimed at discovering other causes influencing trust, confidence and technology acceptance.

Stakeholders and designers: Given the increasing popularity of the chatbot technology across various domains, stakeholders and system designers could learn about the way users trust, accept, and adopt a chatbot technology as utility tailored to facilitate processes within specific contexts (e.g., booking a flight, getting a health prognosis). Consequently, they could learn about factors undermining trust, confidence, and acceptance. By doing so, they can apply this attained knowledge into practice while conceptualizing their chatbot technology. By understanding how users interact with chatbots and how chatbots could be improved may help stakeholders designing the next generation of chatbots (Jung, 2018) which are useful for the user. This strengthens the value and importance of adopting a more user-centric approach within the design process (e.g., Ciechanowski et al., 2018a; 2018b; Piccolo et al., 2018). By doing so, users could understand the underlying benefits of the chatbot technology while overcoming possible misuse (e.g., Berge, 2018), or reluctance especially when information exchange is involved (e.g., Hertzum et al., 2002; Ates, 2017; Burri, 2018).

1.5 OUTLINE OF METHODS

This research was conducted between mid-July and December 2019 in Berlin, Germany, by adhering with regulations of the respective country. These are covered in the Federal Data Protection Act 2017 (Bundesdatenschutzgesetz, BDSG), and the General Data Protection Regulation 2018 (GDPR). Following the original plan proposal (Appendix A), an extensive review of existing literature was conducted. The review took longer than expected due to the novelty of this topic, and the fact that new papers kept emerging in research portals like ACM, ResearchGate, and Google Scholar. Given the continuous advancements in chatbot technology, efforts were made to rely on recent findings (e.g., 2017-2019) to ensure consistency with current trends and recent discoveries.

Data were collected from 14 participants throughout the research process during a period of five weeks between August, and September; after defining appropriate methods for the research design. This stage consisted of two phases. The first phase was about establishing suitable means to conduct the experiment by utilizing the within-participants design. The process involved chatbot testing, material preparations (e.g., interview scripts, and basic tasks assignments), and finding an appropriate scale to measure trust in technology. Later, the human-computer trust (HCT) scale was obtained from the authors (Madsen & McGregor, 2000) and adapted as a survey for the present research.

The second phase involved data collection through interviews, observations, and surveys; followed by the analysis, and interpretation of findings. Empirical data was analysed by utilizing a deductive approach in thematic analysis, a process that relies on theories and existing knowledge. In that regard, the analysis and interpretation relied on three theoretical models about technology and media adoption. These are the: Technology Acceptance Model (Davis et al., 1989; Davis, 1989; Venkatesh & Bala, 2008), Diffusion of Innovations theory (Rogers, 2003), and the Uses and Gratification theory (Blumler & Katz, 1974). Combined with these, three additional models were used to describe the factors influencing trust. These are the: Integrative Model of Organizational Trust (Mayer et al., 1995), Model of Online Trust (Corritore et al., 2003), and Trust in Specific Technology (McKnight et al., 2011). These theories helped in describing the psychological characteristics and factors driving users to accept a specific technology; before trust and adoption takes place. For the surveys, total percentage scores, sample mean values, variances, and standard deviations were calculated. Injunction with these, a paired- samples *t*-Test with a Bonferroni adjustment was conducted by calculating differences between paired observations. The last stages consisted in discussing the findings and limitations while referencing previous studies and existing theories. Additionally, a set of suggestions for designing better chatbots, and recommendations for future studies were also proposed.

1.6 PROJECT STRUCTURE

The present chapter introduces the research questions and objectives. The second chapter extends on sections 1, and 1.1 (this chapter) with a review of the context, the research, and its novelty. The third chapter explains the methods utilized to conduct the analysis on the findings elaborated in the fourth chapter. The fifth chapter discusses these findings in detail, their implications, limitations, and suggestions for further research. Lastly, the sixth chapter evaluates and reflects on this research in its entirety, followed by concluding notes.

2 CONTEXT

Trust and technology acceptance plays a vital role in determining technology success and its life-cycle. They are essential constructs in how people use technology, and their development plays a significant role in its adoption (Young et al., 2009). This is not limited to chatbots and conversational systems. Although chatbots are becoming ubiquitous (Lee & See, 2004), some knowledge gaps still exist. According to many authors, these gaps relate to the individuals' underlying motivation and their intent to use, trust and relate to the chatbot technology (Ciechanowski et al., 2018a; 2018b; Følstad et al., 2018a; Hansson, 2018).

Previously, it was said that the purpose and added value the chatbot technology has to offer is not clearly defined (Zamora, 2017). One reason is that most of the existing knowledge of trust and acceptance were mainly focused on offline technologies and little on online technologies (e.g., Corritore et al., 2003; Burri, 2017). However, later studies (e.g., Brandtzæg & Følstad, 2017; Følstad et al., 2018a; Candela, 2018; Nordheim, 2018) identified key motivational and gratification factors driving chatbot use. These relate to i) productivity, ii) entertainment, iii) social/relational, and, iv) novelty/curiosity. Consequently, the chatbot technology is believed to have the potential to strengthen users' autonomy, competence, and (possibly counter-intuitively) social relatedness (Brandtzæg & Følstad, 2018). Studies across various domains were conducted to discover what makes users accept, relate, and eventually trust a technology, including chatbots. Examples of these domains include: customer care and e-commerce (e.g., Kuligowska, 2015; Candela, 2018; Nordheim, 2018; Haan, 2018; Følstad et al., 2018b), finance (e.g., Burri; 2017; Duijst, 2017; Ates, 2017; Hansson, 2018), healthcare (e.g., Kretzschmar et al., 2019; Wang & Siau, 2018; Montague 2010), automation (e.g., McGregor & Madsen, 2000; Lee & See, 2004; Alaieri & Vellino; 2017), media (e.g., Edwards & Sanoubari, 2019), and weather emergency (e.g., Berge; 2018) among others. These studies reported a variety of findings predicting trust in chatbots and other technologies. These are related to the users' *attitudes and perceptions* towards the technology, and the *'context of use'* (purpose). For example, in healthcare, it was found that patients would relate to a chatbot because they are perceived as non-judgmental (Kretzschmar et al., 2019; Montague, 2010). However, within the context of banking and e-commerce (retail), trustworthiness was found to be crucial. This is because users were reluctant over submitting sensitive information to virtual assistants; due to security concerns and lack of trust (e.g., Hertzum et al., 2002; Ates, 2017; Burri, 2018; Candela, 2018).

Nevertheless, a variety of factors hindering chatbot usage, acceptance, and adoption also exist. A survey conducted by Pegasystems Inc. (2018), discovered that customers (users) expressed dissatisfaction towards chatbots due to lack of context in conversations, not smart enough to answer questions effectively, and limited human qualities. To a certain extent, chatbots lack the capability to correctly manage conversational social practices (Augello et al., 2016), which in return affects confidence and acceptance. Users prefer chatbots with natural language conversation ability and the capability to sustain conversation context, handling dialog failures, and ending conversations gracefully (Jain, 2018) and consistently. If the chatbot refuses users' request or, providing difference responses was found to affect confidence levels (Jung, 2018) towards the same technology. Therefore, as argued by many authors (e.g., McKnight et al., 2011), reliability is crucial for trust to takes place. This was confirmed by a majority of prior studies in chatbot and automated technologies (e.g., Fogg & Tseng, 1999; Madsen & McGregor, 2000; Merrit & Ilgen, 2008; Miller, 2014; Xu et al., 2015; Chien et al., 2016; Klopfenstein et al., 2017; Berge, 2018; Henrekson et al., 2018; Wang & Siau, 2018).

Concerns over potential risks and vulnerabilities such as data breach and cyber-attacks on chatbots were found to be additional culprits undermining trust, confidence, and acceptance (e.g., Radziwill & Benton, 2017; Chung et al., 2017; Kretzschmar et al., 2019). Other concerns also include the 'perceived risks' of using the chatbot technology alongside the underlying 'online privacy concerns', including identity protection (Clark et al., 2019). Various studies addressed the issue about new (online) technology adoption against privacy concerns (Dai et al., 2014; van Eeuwen, 2017; Candela, 2018; Edwards & Sanoubari, 2019) and security (Følstad & Brandtzæg; 2017; Ates, 2017) which are linked to the users' perceived risks. For instance, within e-commerce, individuals may view conversational commerce through chatbots as too intrusive and dangerous for their privacy (Candela, 2018). Also, within the mental health sector, young people feel that their problems and sensitive information may be shared around (Kretzschmar et al., 2019). Novel technologies entail new ethics and privacy implications, and chatbots are no exception; therefore, with the emergence of chatbots, stronger attention to ethics and privacy is needed (Følstad & Brandtzæg; 2017; van Eeuwen, 2017; Wang & Siau, 2018). Injunction with privacy, also data collection by the chatbot and its usage should be clearly explained to the end-user (Botelho, 2017; Reddy, 2017) because chatbot algorithms heavily rely on data (Wang & Siau, 2018) exchange and management. Harkous et al. (2016) proposed the concept of 'conversational privacy chatbots', also abbreviated as 'PriBots', a chatbot technology with added features and layers tailored for privacy support and additional security. Despite the continuous growth in chatbot usage, it is essential to understand user requirements as this plays a crucial role in designing better chatbots (Ciechanowski et al., 2018a); because, it is widely acknowledged that humans will accept technologies they trust more than technologies they do not trust (Merritt & Ilgen, 2008).

Unlike other technologies, the chatbot technology is unique due to its relational features in the form of exchanging dialogues and understanding through natural language (Berge, 2018). Novel characteristics like natural language processing (NLP) and natural language understanding (NLU) can have implications on trust (Valtolina et al., 2018). That comes down to the human-likeness within the chatbot technology, an area requiring further investigation. Chatbots with human-like characteristics are found to influence trust; however, that may depend on the application area. For instance, within customer care, human-likeness, especially in language capabilities, was found to be an essential factor affecting trust in chatbots for customer service (Nordheim, 2018; de Haan, 2018), and the same within healthcare (Kretzschmar et al., 2019). However, it is believed that an 'excess' of human-likeness can affect the users' perception towards chatbots (Piccolo et al., 2018) especially within specific application areas (Duijist, 2017). Such eventuality leads to the 'uncanny valley effect'. This effect invokes adverse effects and emotions resulting from perceptual difficulty in distinguishing between a human-like object and its human counterpart (Mori et al., 2012; Ciechanowski et al., 2018a). Therefore, it was emphasized that the uncanny valley effect should be investigated in response to the increasing use of chatbots (Ciechanowski et al., 2018a; Skjuve et al., 2019). On the other hand, transparency, such as making users aware that they are interacting with a chatbot instead of a human, is important (Brandtzæg & Følstad, 2018) for trust to take place. Recent studies confirmed that users would feel betrayed once it is realized that they interacted with a chatbot instead of a human (Nordheim, 2018; Haan, 2018). This outcome affects trust negatively unless an option to speak with a real human is available (Nordheim, 2018).

In summary, trust is undoubtedly crucial for the interaction between humans and technology (Nothdurft, 2014). However, as trust is a trait-like characteristic, it varies from person to person (Schoorman et al., 2007), which is subsequently shaped by perception, culture, and attitude. Different people have different perceptions and attitudes towards (similar) technologies, including chatbots, which in return affect confidence in these technologies. In line with prior studies (e.g., Chien, 2009; Edwards & Sanoubari, 2019), and existing theories (e.g., Rogers, 2003), we believe that cultural differences, injunction with personality characteristics, and context of use also shape these attitudes. Attitude covers the propensity to trust, that is, the willingness to trust a party effortlessly (Mayer et al., 1995). Within the technology spectrum, propensity means the willingness to depend on technology across various situations (McKnight et al., 2011). Consistent with this, prior studies by Merritt and Ilgen (2008) discovered that humans with high trust propensity were more likely to put greater trust in automated systems. Therefore, 'triability' (e.g., frequency of use and experimentation), and 'observability' (e.g., outcomes after using technology) are essential to form a mental model (Rogers, 2003; Norman, 2013) of the chatbot; by understanding its value and operation within its domain before a trust relationship occurs.

2.1 THEORETICAL FRAMEWORKS AND THE USE OF TECHNOLOGY

Given the novelty of this topic, the increasing interest in chatbot technology, and the need for more knowledge concerning its adoption, we employed three theoretical frameworks that suggest factors facilitating technology acceptance. These were used in combination with three models proposing how trust is established and facilitated. The motivation behind this holistic approach was aimed at describing the psychological characteristics and other factors that drive users to accept a specific technology before a trusting relationship occurs. In this context, the term technology refers to chatbots and related systems like voice user interfaces (VUI).

2.1.1 TECHNOLOGY ACCEPTANCE MODEL (TAM)

How fast users accept technologies depend on a variety of factors like availability, convenience, needs, and security (Lai, 2017). Davis (1989) proposed the 'Technology Acceptance Model' (TAM), containing two key attributes predicted to facilitate technology adoption (Figure 1.1). These are:

- **Perceived usefulness (U):** referring to the user's belief that the system would increase job performance. Within this research context, a chatbot is perceived as useful if productivity is maximized and tasks are performed in a timely manner.
- **Perceived ease-of-use (E):** the degree that the system should be free from physical and mental effort. In this situation, using a chatbot should be easy to understand without requiring too much effort (Candela, 2018).

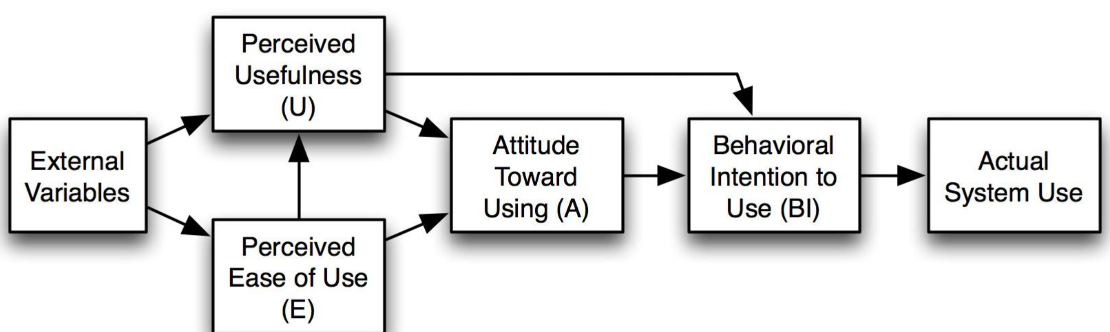


Figure 1.1: Technology acceptance model (version 1) (Davis et al., 1989)

As indicated in Figure 1.1, the model suggests that the intention of use is fostered by the **'behavioural intention (BI)'**, a factor driving people to use the technology. The intention is influenced by **'attitude (A)'**, which is the overall impression of the technology. However, **'external variables'**, for instance **social** or **cultural** influence are important to determine attitude. Over time, this theory was revised, and the recent version (TAM3) took into consideration 'individual differences', 'system characteristics', ‘social influence’, and ‘facilitating conditions’, which are determinants of perceived usefulness and perceived ease-of-use (Venkatesh & Bala, 2008). Translating this model to chatbots, we assume that the perceived ease-of-use and usefulness of the technology within its application area, combined with cultural norms and social influence, predicts/ensures its acceptance and adoption.

2.1.2 USES AND GRATIFICATION THEORY

Behind the adoption of every technology, the motivation and goals behind the users’ intent of use play an active key role (Candela, 2018). ‘Uses and gratification theory’ (UGT) (Blumler & Katz, 1974) is a positivistic approach derived from psychology explaining why and how people deliberately use specific media that satisfies particular needs to achieve gratification (Wikipedia). According to this theory, the use of specific media depends on the expected and experienced gratification the media or technology will provide. For instance, a study within the e-commerce domain discovered that people would use chatbots because they are gratified with better online experience, and by having queries answered promptly (Dandörfer, 2019). Prior investigations also concluded that the primary gratification for typical chatbot relates to efficiency and productivity (Brandtzæg & Følstad, 2017; Følstad et al., 2018a; 2018b; Candela, 2018). Figure 1.2 highlights the relationship between motivations, media (or technology) behaviour and the outcome reflecting to whether gratification was obtained or not.

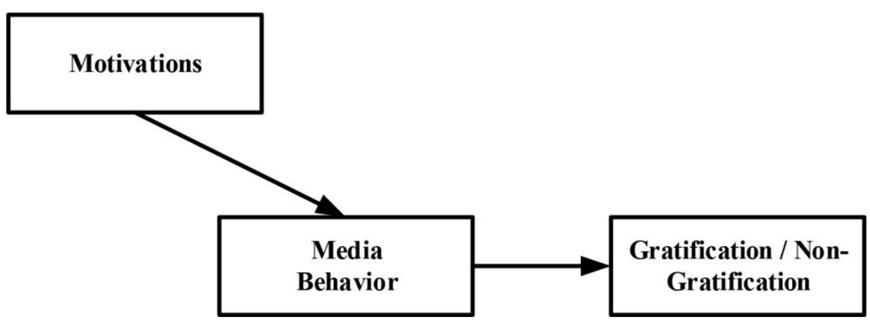


Figure 1.2: Uses and gratification theory (Wikipedia)

2.1.3 RATE OF INNOVATIONS ADOPTION (DIFFUSION OF INNOVATIONS THEORY)

The diffusion of innovations theory (Rogers, 2003) explains how and at which rate technologies are diffused and adopted by users. This theory was referred to by other researchers who investigated chatbot adoption (e.g., Brandtzæg & Følstad, 2017; van Eeuwen, 2017; Candela, 2018). Rogers (2003) describes the innovation-diffusion process as “an uncertainty reduction process”. There are various criteria and processes that facilitate technology adoption. This literature focuses on the attributes of innovation (Rogers, 2003) to explain the process and rate of chatbot adoption. As proposed by Rogers (2003), the characteristics of innovations help to decrease uncertainty (Sahin, 2006), and include the following aspects:

1. **Relative advantage:** this refers to the extent to which the innovation is perceived as productive, efficient, and improves upon existing practices (Sahin, 2006);
2. **Compatibility:** the degree to which an innovation is perceived as consistent with the current values, past experiences, and needs of potential adopters;
3. **Complexity:** the degree to which an innovation is perceived as difficult to understand and use;
4. **Triability:** the degree to which an innovation may be experimented with on a limited basis. The more an innovation is tried, the faster its adoption is (Sahin, 2006);
5. **Observability:** the degree to which the innovation's results are visible to others. Observability is dynamic and covers areas such as profitability, performance, and productivity.

In summary, Rogers (2003) argued that innovations offering more relative advantage that includes compatibility, simplicity, triability, and observability would be adopted faster than other innovations. As previously mentioned, as the chatbot technology is relatively new, **triability** and **observability** are pivotal in predicting its adoption; because of uncertainty reduction following the experimentation phase.

2.2 TRUST AND CONFIDENCE: AN OVERVIEW AND THEORETICAL BACKGROUND

Trust was recognized as an integral dimension in how people request and use information from, interact with, and work alongside machines and technologies (Nothdurft et al., 2014; Edwards & Sanoubari, 2019). This construct has continuously been investigated in application areas involving chatbot technologies (e.g., Miller, 2015; Brandtzæg & Følstad, 2017; Candela, 2018; Nordheim, 2018; Følstad et al., 2018a; 2018b; Berge; 2018); and in human-computer interaction (HCI) to map the relationship between the role of trust and technology adoption (e.g., Gefen et al., 2003; Reid & Levy, 2008, cited in Miller, 2015; Edwards & Sanoubari, 2019). According to Fogg and Tseng (1999), trust indicates a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process. Also, Madsen and McGregor (2000) describe trust as a complex concept that is related to, yet not completely analogous with confidence**.** As trust is dynamic, it was synonymized with confidence, cooperation, and predictability (Siau, 2018). As well, trust itself is split into cognitive cues formed by a person on the perception of an object, which is to be trusted (Burri, 2017). In contrast, confidence is both 'cognitive' and 'affective', and necessary for establishing trust (e.g., Luhmann, 1988; Muir, 1994, cited in Corritore et al., 2003). Based on these statements, it should be considered that trust and confidence are coexistent; therefore, both constructs are dependent on each other, particularly in (chatbot) technology adoption.

Given the numerous terminologies, this literature refers to the trust definitions applicable in the human-computer interaction (HCI) field. These reflect the user's confidence in the competence and ability of a technology (Lee, 2002) such as the chatbot. Trust is therefore defined as:

1. *"the extent to which a user is confident in, and willing to act on the basis of the recommendations, actions, and decisions of an artificially intelligent decision aid"* (McAllister, 1995, cited in Madsen & McGregor, 2000), and;
2. *"the belief that a specific technology has the attributes necessary to perform as expected in a given situation in which negative consequences are possible"* (McKnight et al., 2011).

The construct of trust is possible to explore empirically by explaining the relationship between the user and the technology, as well as outcomes produced from the trusting behaviour (Berge, 2018). This research refers to the three commonly cited theoretical frameworks devised by Mayer et al. (1995), Corritore et al., (2003), and McKnight et al. (2011) to explain the criteria that predict trust in technology. These were utilized and referred to throughout this study:

2.2.1 AN INTEGRATIVE MODEL OF ORGANIZATIONAL TRUST

Although this theoretical model explains trust among people and organizations, it was often referred to in similar studies because of the relational features (human-likeness) the chatbot technology holds. This model was found to be adaptive, and its defined factors formed an integral part of this research. The integrative model of organizational trust (Mayer et al., 1995) focuses on trust situations involving two parties: the trustor (e.g., user) and a trustee (e.g., chatbot). As illustrated in Figure 1.2, the model identifies three key determinants termed as factors of perceived trustworthiness, which are, the trustee's perceptions of:

1. **Ability:** the group of domain-specific skills, expertise, competencies, and characteristics that make the trustee highly competent in specific areas, thus enabling the trustor to relate to the trustee with regards to tasks related to a specific area (Mayer et al., 1995). In the context of chatbot, this translates to the trustor's expectation that the chatbot has the functionality, capacity, and ability to fulfil an intended task (McKnight et al., 2011; Miller, 2014);
2. **Benevolence:** the perception of a positive orientation of the trustee toward the trustor; and an extent to which a trustee is believed to want to do good to the trustor (Mayer et al., 1995). Concerning chatbot technology, if it improves performance and gives adequate advice, its behaviour (and output) can be described as benevolent (Berge, 2018);

3. **Integrity:** the relationship between integrity and trust involves the trustor's perception that the trustee adheres to a set of principles that the trustor finds acceptable (Mayer et al., 1995). This translates to whether the chatbot is a reliable utility that provides the user with meaningful advice and responses.

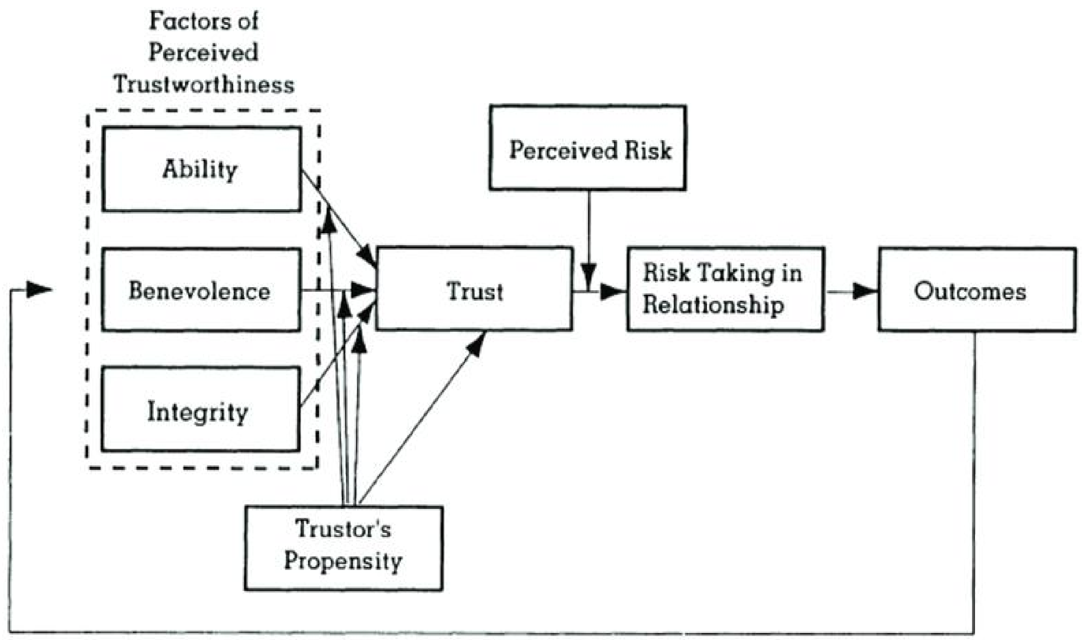


Figure 1.3: The Integrative Model of Organizational Trust (Mayer et al., 1995)

However, referring again to Figure 1.3, two other factors influencing trust are the ***'trustor's propensity'*** and ***'perceived risks'***. As previously mentioned, trustor's propensity refers to the willingness to trust a party effortlessly (Mayer et al., 1995). As trust is a dynamic construct involving two parties (e.g., a user and a chatbot), the characteristics of the trustor (the user) is a factor that must be taken into consideration (Mayer et al., 1995; Berge, 2018) as this could influence the way individuals trust chatbot technologies. Trust propensity varies among individuals, and these are affected by personality types, experiences, and cultural backgrounds (Hofstede, 1980, cited in Mayer et al., 1995). However, trust is also influenced by perceived risks. The definition of risks proposed by Mayer et al. (1995) refers to the likelihood of positive or negative outcomes resulting from situations which might be unrelated to the trustee. Trust also involves the willingness to take risks, and engagement with the trustee often involves risks particularly when an object of trust is shared. Within the context of chatbots, this involves situations when users share personal data to an autonomous agent.

2.2.2 MODEL OF ONLINE TRUST

Corritore et al. (2003) presented a much-cited framework of trust in interactive systems, particularly e-commerce platforms (Følstadt et al., 2018a). This model was used in this research because as elaborated in Figure 1.3, it took 'external factors' into consideration:

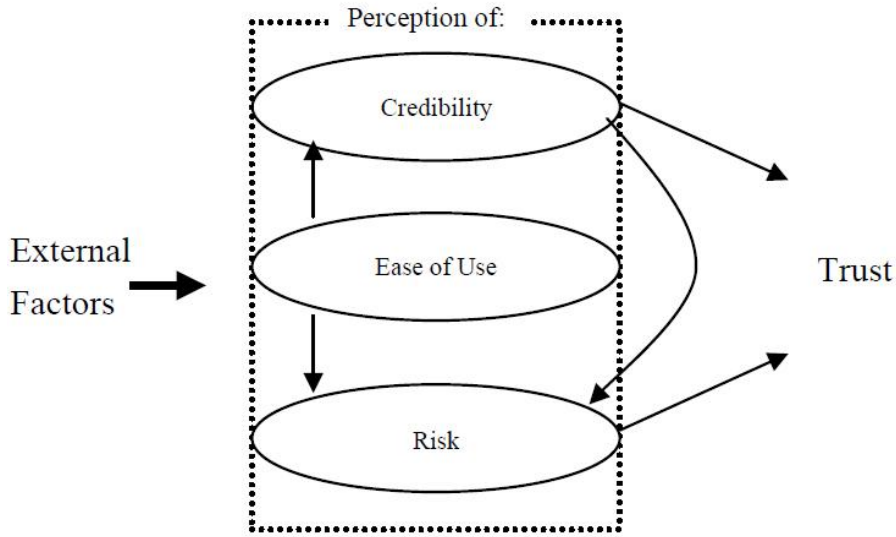


Figure 1.4: Model of Online Trust (Corritore et al., 2003)

External factors (e.g., physical, psychological) are found to impact trust in the system depending on the situation surrounding a specific online trust situation (Corritore et al. 2003; Burri, 2017). In return, these factors are also influenced by the user's perception. The model of online trust covers the following aspects:

1. **Perception of credibility:** comprised of four components impacting the decision to trust a system. These are *trustworthiness* (well-intentioned, truthful, and unbiased); *expertise* (knowledgeable, experienced, and competence); *predictability* (the expectation that the system will act consistently (McKnight et al., 2011)), and *reputation* (the quality of recognized past performance);
2. **Perception of ease-of-use:** derived from the technology acceptance model (Davis, 1989; Davis et al., 1989), Corritore et al. (2003) argue that ease-of-use impacts users' trust towards technology. It is also found to be a reasonable factor affecting users' trust in chatbots (e.g., Nordheim, 2018; Berge, 2018);
3. **Perception of risk:** the likelihood of an undesirable outcome (Deutsche, 1958, cited in Corritore et al., 2003) while using technology. It is also discovered that a sense of control reduces risk, and that risk is higher in the absence of control (Corritore et al., 2003). In the context of chatbots, risks are associated with privacy, control over disclosed information, and security;

4. **External factors:** the aspects of the environment surrounding a specific online trust situation (Corritore et al., 2003). Possible external factors inherent in a trust situation include the level of risk or the control the user has while interacting with technologies (Corritore et al., 2003) including chatbots.

2.2.3 TRUST IN SPECIFIC TECHNOLOGY

Another known trust framework is the one proposed by McKnight et al. (2011) which is based on the model by Mayer et al. (1995). This model was used in this research because it is specifically devised for the technological spectrum. According to McKnight et al. (2011), trusting beliefs in a specific technology reflect expectations about the favourable attributes of a specific technology that represent knowledge users have cultivated by interacting with technology in different contexts. These beliefs consist of:

1. **Competence vs. Functionality**: whether one expects technology to have the capacity or capability to complete a required task;
2. **Benevolence vs. Helpfulness:** the belief that the technology provides adequate and responsive help and advice necessary to complete a task;
3. **Predictability/Integrity vs. Reliability**: when one expects technology to work consistently, predictably, and without glitches.

2.3 CONCLUSION

It is evident that given the increasing chatbot ubiquity combined with user-related factors (e.g., attitude, perception, culture), understanding the causes that mediate trust in technology is fundamental prior of its adoption. We noticed two main limitations surrounding the topic of chatbot technology. The first relates to the fact that many studies repeatedly explored the domains of e-commerce, banking, and customer care. The second is that the purpose, usefulness and value chatbots can offer are not clearly defined (Zamora, 2017) for the end user. Expanding on this, companies realized the cost-effective benefits of chatbots such as 24/7 support, ability to handling multiple queries simultaneously, and staff member reduction for instance by assigning specific tasks (e.g., first-line support) to chatbots instead.

Therefore, supporting Piccolo et al. (2018), more user-centred research is required to establish chatbot technology for specific user groups within specific contexts. By taking user requirements into consideration, designers should develop chatbots that are perceived as useful; by identifying cases in which chatbots fulfil users' productivity more effectively (Brandtzæg & Følstad, 2017; Ciechanowski et al., 2018a). This ensures a successful adoption rate and longevity once i) user requirements are addressed;

ii) usefulness is properly defined, and iii) the risks surrounding the context of use are minimized. In return, this enhances confidence levels and satisfaction. Typically, users build a mental model of the system, and if this perceived mental model and the actual system do not match, the human-computer trust relationship is severed (Muir, 1992, cited in Corritore et al., 2003). Therefore, having established guidelines (e.g., Yan et al., 2011) to make informed design-related decisions is becoming a necessity; given the growing interest in chatbots and related technologies. For that matter, the present research referred to established theoretical frameworks to produce design-related recommendations by addressing factors found to mediate trust, confidence and acceptance.

3 METHODS

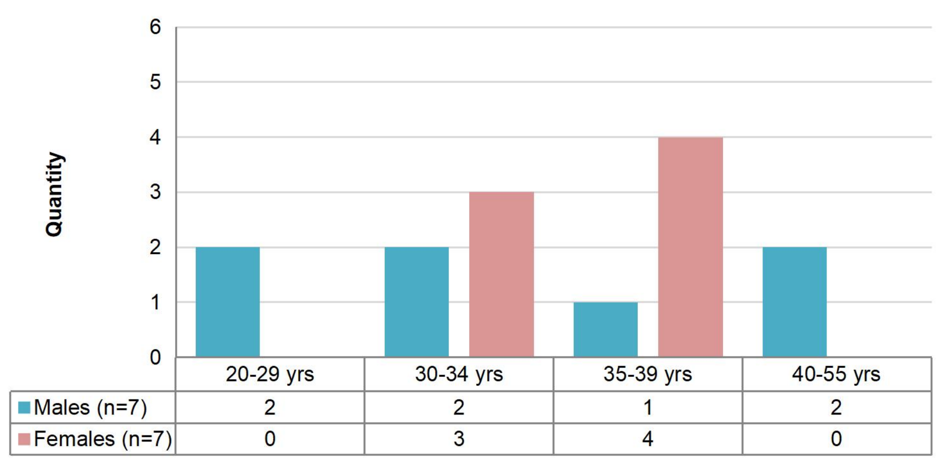
3.1 OVERVIEW

Mix-methods, consisting of qualitative (e.g., interviews, and controlled observations), and quantitative (surveys, and statistical computation) approaches, combined with user-based experimentation (e.g., within-participants design) were employed. This chapter explains the research design process combined with the motivations behind the steps used to gather relevant empirical data to answer the research questions established in chapter 1.

3.2 RESEARCH DESIGN

A sample of 14 participants (n=14) were randomly recruited through an advert campaign on Craigslist (Berlin, DE), acquaintances, and word of mouth. Efforts were made to ensure a homogenous amount of participants organized by gender (Figure 2.1). Therefore, the sample consisted of 7 females (n=7) with age varying from 30 to 39 years, and 7 males (n=7) with age varying from 20 to 55 years:

Figure 2.1: Participants sample (n=14)



3.2.1 CHATBOT SELECTION AND RATIONALE

Two chatbots were selected to determine whether the constructs of technology acceptance, trust, and confidence are affected by their technology type or application area. Initially, the intention was to utilize 3 chatbots (H&M, Wysa, and Mitsuku); however, given the sample reduction, these were replaced with 2 chatbots being Cleo, and Replika. Both chatbots were deemed more suitable for this research. Cleo was selected because of its application area, functionalities and technical capabilities. In that regard, we decided to investigate how a finance-related chatbot could be perceived by participants. As for Replika, this chatbot was selected due to its conversational capabilities, versatility, and because it covers 2 areas which relate to mental well-being and companionship. The next sections will introduce these chatbots:

Cleo is an AI and 'script-based' hybrid money-managing assistant created by a London (UK)-based start-up company. In terms of interactivity, the chatbot relies on keyword-input and predefined question presented in the form of call-to-action buttons (e.g., balance, dashboard). Therefore, users can either type own queries or pre-select existing questions. To date, Cleo is deployed via 'Facebook Messenger', therefore users must have a Facebook account prior of using this chatbot. The setup process requires permissions to access the user's bank account to pull out financial information. Its capabilities include expense tracking, bill and income management, periodical budgeting, and provision of recommendations for effective financial management. The chatbot comes with a dedicated dashboard hosted on the provider's platform (meetcleo.com). There, the users can setup own budget limits, managing connected bank accounts, besides accessing various graphs (e.g., expenses). One key feature of this chatbot is the 'Cleo wallet', which is a saving wallet provided optionally. Once this wallet is created, Cleo automatically allocates a specified amount of money on occasions set by the user. Furthermore, as its use is limited to British banks, only UK residents or British bank account holders can utilize this chatbot. One notable characteristic of this chatbot is the use of 'British language slang' and 'emojis', forming part of its overall language capabilities.

Replika is an AI-powered chatbot created by a company based in Los Angeles (US) and tailored for well-being and companionship. In terms of interactive capabilities, Replika supports both 'keyword-input', 'natural language processing (NLP), and understanding (NLU)'. Therefore, users can also interact through voice. The chatbot is versatile and capable to formulate own questions based on the inputted queries. It comes in the form of a native app, and browser utility. Its setup process comes with the options to assign a static avatar, gender, and a voice type. Additionally, users are presented with the choice to set own goals they want to achieve for example work-life balance. One key feature the chatbot has is the personality quizzes for instance dealing with emotions and conflict management.

3.2.2 CHATBOT TESTING AND SETUP

Both chatbots were installed and tested on a smart tablet device for approximately ten days before administering these to participants. To utilize Cleo, a Facebook Messenger account is required. As well, its setup process requires users to authorize a connection with their British bank account via Facebook before use. As this research was conducted out of the United Kingdom, combined with ethical considerations (section 3.4), Cleo was set up beforehand on the Facebook Messenger and a British bank account solely owned by the investigator. For that matter, the financial balances within the same bank account were kept down to a minimum (e.g., £40.00 balance, £100 overdraft) to prevent undue disclosure of the investigator's financial information. As for Replika, being a native app, its setup process requires the user to sign-up for an account, and to do basic configurations such as assigning a name, gender, and a voice style to the chatbot. Optionally, the user can define personal goals to be achieved over time (e.g., work-life balance, managing emotions). For this research project, new Replika accounts were created and configured in advance before each session. After each session, chat histories belonging to each participant and both chatbots were extracted through means of screenshots and saved in storage peripherals before complete erasure. Specifically to Replika, the user account was deleted and then recreated before each following sessions with the remaining participants.

3.2.3 HUMAN-COMPUTER TRUST (HCT) SURVEY ADAPTATION

A scale and survey devised to assess trust in computer systems was obtained from the authors (Madsen & McGregor, 2000) and adapted for this research. This scale was rigorously tested by its authors and was shown to be empirically valid. Its reliability, measured as Cronbach's alpha (α) was found to be 0.94. The definition, model, and measurement of trust constituting this scale are based on the theories of social attribution, ethopoeia, and two joint-cognitive systems paradigms (cognition-based trust, affect-based trust) as elaborated in Figure 2.2.

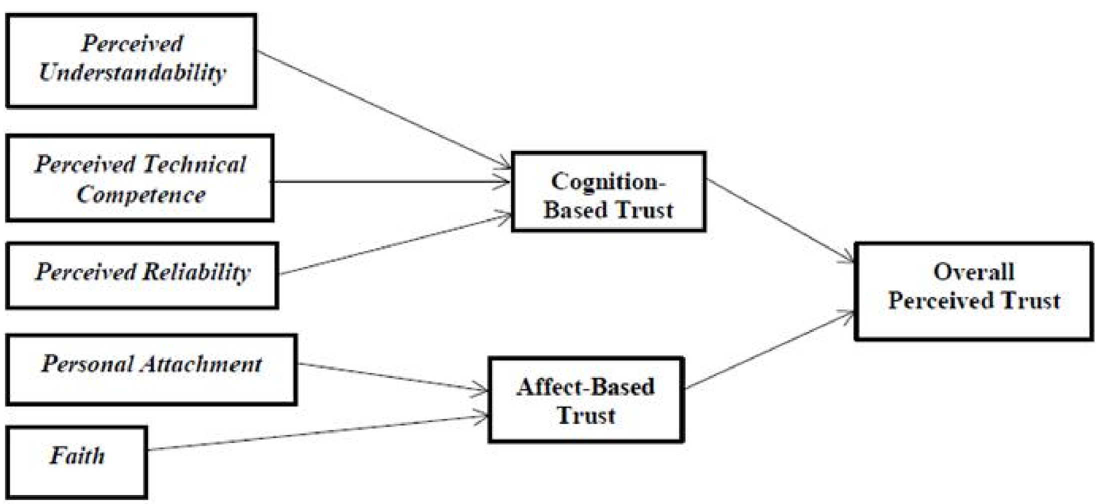


Figure 2.2: Model of the Human-Computer Trust (Madsen & McGregor, 2000)

Based on these aspects, this scale was deemed appropriate to be adopted for this research to assess trust in chatbot technologies. The scale items are categorized under, and aimed at measuring the following five constructs, as elaborated below:

1. **Perceived Reliability:** this relates to the sense of repeated consistent functioning of the technology. Reliability also refers to ‘predictability’ of the technology and ‘dependability’ on the technology.
2. **Perceived Technical Competence:** the degree to which the technology is perceived to perform tasks accurately and correctly based on the inputted information.
3. **Perceived Understandability:** this refers to the degree when the user forms a mental model and predicts future behaviours of the technology.
4. **Faith:** this relates to when the user has faith in the future ability of the technology to perform even in situations in which the technology is untried.
5. **Personal Attachment:** this comprises of ‘liking’, meaning that the user finds the technology agreeable and suits own taste; and ‘loving’, meaning that the user has a strong preference and possibly attachment to the technology. Personal attachment also covers ‘friendliness’ and ‘ease-of-use’.

In order to maintain consistency with this research, the survey items were revised. In return, the terms ‘system’ was replaced with ‘chatbot’, whereas ‘problem’ was replaced with ‘my query/problem’. The substitution of terms was necessary because the assigned activities required participants to submit queries to the chatbots. The scale carries 25 items belonging to the aforementioned 5 constructs as exemplified in Figure 2.3.

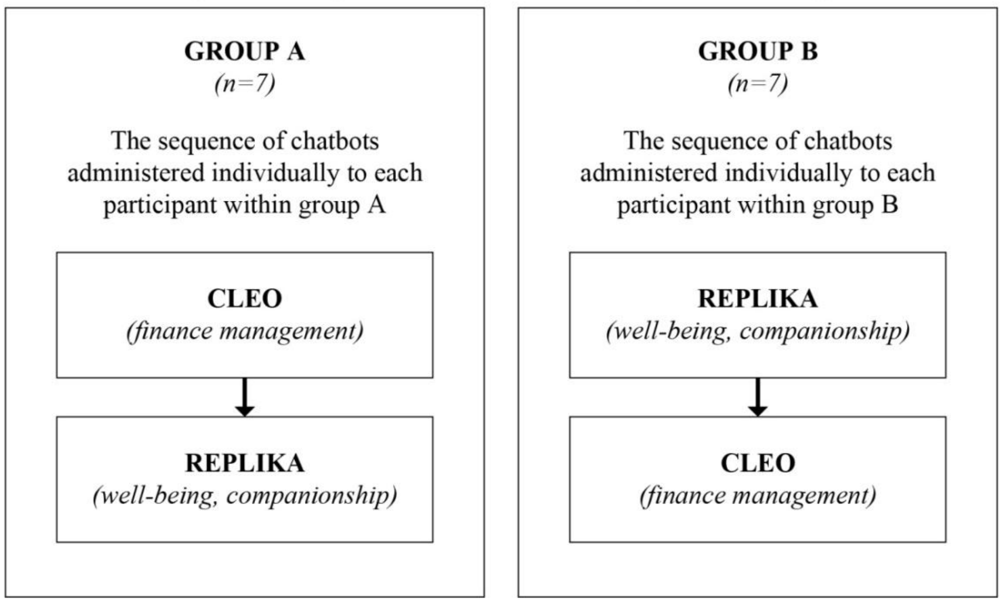
Figure 2.3: The amended human-computer trust scale (survey) organized by construct		
No.	**KEY**	**ITEMS**
4	R1	The chatbot always provides the advice I require to make my decision.
10	R2	The chatbot performs reliably.
11	R3	The chatbot responds the same way under the same conditions at different times.
13	R4	I can rely on the chatbot to function properly.
23	R5	The chatbot analyses problems consistently.
2	T1	The chatbot uses appropriate methods to reach decisions.
9	T2	The chatbot has a sound knowledge about the type of problem / query built into it.
16	T3	The advice the chatbot produces is as good as that which a highly competent person could produce.
19	T4	The chatbot correctly uses the information I enter.
24	T5	The chatbot makes use of all the knowledge and information available to it to provide me with a solution to my query or problem.
8	U1	I know what will happen the next time I use the chatbot because I understand how it behaves.
12	U2	I understand how the chatbot will assist me with decisions I have to make.
14	U3	Although I may not know exactly how the chatbot works, I know how to use it to make decisions about the problem or query.
17	U4	It is easy to follow what the chatbot does.
25	U5	I recognize what I should do to get the advice I need from the chatbot the next time I use it.
4	F1	The chatbot always provides the advice I require to make my decision.
6	F2	When I am uncertain about a decision, I believe the chatbot rather than myself.
15	F3	If I am not sure about a decision, I have faith that the chatbot will provide the best solution.
21	F4	When the chatbot gives unusual advice, I am confident that the advice is correct.
22	F5	Even if I have no reason to expect the chatbot will be able to solve a difficult problem / query, I still feel certain that it will.
1	P1	I would feel a sense of loss if the chatbot was unavailable and I could no longer use it.
5	P2	I feel / felt a sense of attachment in using the chatbot.
7	P3	I find the chatbot suitable to my style of decision making.
18	P4	I like using the chatbot when I'm working on this type of problem / query.
20	P5	I have a personal preference for making decisions with the chatbot.

Key: **R1 – R5:** Perceived Reliability, **T1 – T5:** Perceived Technical Competence, **U1 – U5:** Perceived Understandability, **F1 – F5**: Faith, **P1 – P5:** Personal Attachment

3.2.4 WITHIN-PARTICIPANTS DESIGN EXPERIMENT

Given the small sample size (n=14), the application areas of these chatbots, and the large variation in the amount of exposure that participants had to chatbot technologies, the 'within-participants' design experiment was adopted for this research. This was done by distributing participants into two groups (A, B). The 'within-participants' design experiment is advantageous because it ensures that participants are exposed to similar conditions; while eliminating biases resulting from individual differences (Budiu, 2018; Cherry, 2019). As shown in Figure 2.4, each participant from the respective group was administered two chatbots in a specified sequence. In this case, participants from 'group A' started with Cleo, followed by Replika; whereas 'group B' started with Replika, and later Cleo. This design helps reduce ordering effects.

Figure 2.4: Sequence of assigned chatbots (controlled setting)



The rationale and motivation behind the adoption of the within-participants design was to:

1. **Determine if the overall attitude is influenced by the chatbot technology type:** it was assumed that participant' attitudes may vary after being initially exposed to a highly sophisticated chatbot (e.g., Replika), against a less sophisticated chatbot (e.g., Cleo), or vice-versa.
2. **Discover variations among participants' groups:** the aim was to find out whether technology acceptance, confidence, trust levels (the dependent variable) and other behaviours would vary between groups; as a result of administering both chatbots in different sequences (the independent variable).

3.3 TIMELINE

The individual sessions with participants were conducted in Berlin (Germany) between August and September of 2019, ranging between 2 to 3 sessions per week depending on availability. From these sessions, 11 were conducted in a meeting room owned by a commercial entity, whereas the other 3 were conducted at the participants' setting (home). Before these sessions, the participants were instructed to book a time slot via *'Calendly.com'* to confirm attendance. Each session lasted from 1 hour, up to 1 hour and 30 minutes. The timing variation was dependent on the conversation depth, and unforeseen situations related to the chatbot technologies used for this research. Part of the time was allocated to situations when participants sought clarification over the *'consent forms'* as well as the purpose and procedures behind the data collection.

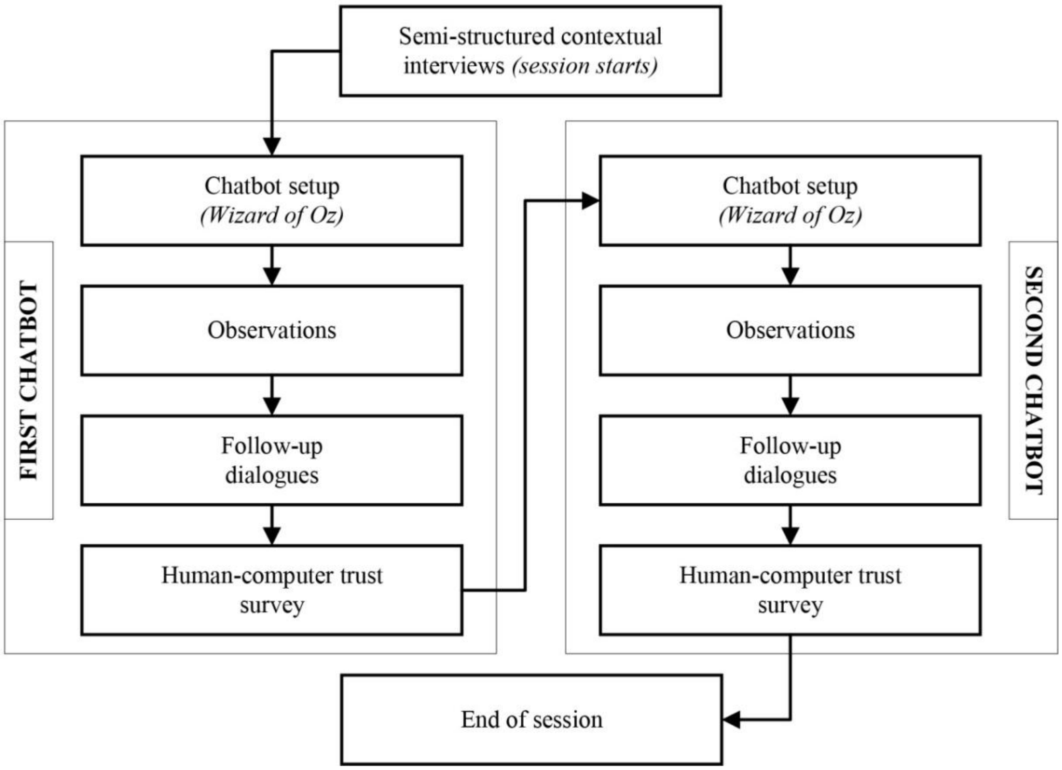
3.4 ETHICAL CONSIDERATIONS

This research and its proposal plan (Appendix A) were approved by the Ethics Committee at City University of London (UK). Each participant was given an elaborated documentation entitled *'participants information'* (Appendix B) which included the 'consent form'. The document was sent to the participants via email approximately seven days before they participated in the research so to give ample time for revision. The documentation contains details about i) the research purpose; ii) data collection, its utilization and anonymization; iii) when the data will be destroyed; and, iv) participants' rights. The participants were instructed to read the documentation carefully and to raise queries or concerns if there were any. As this research was conducted in Berlin, Germany (DE), a location out of the United Kingdom (UK) jurisdiction, the investigator ensures that the research abided with applicable regulations of the respective country, being the Federal Data Protection Act 2017 (Bundesdatenschutzgesetz, BDSG), and the General Data Protection Regulation 2018 (GDPR). No vulnerable participants or participants under the age of 18 or over the age of 60 were recruited. The participants were also ensured that the chatbot technologies will not request nor collect any sensitive information pertaining to own identity, banking or health. The participants were instructed beforehand that the sessions will be audio-recorded and/or screen-recorded. Out of the 14 participants, 2 refused to have their sessions recorded due to privacy concerns. The collected recordings (data) were stored on peripherals consisting of a USB pendrive, and a smart tablet device. No cloud-based facilities were used to store data. The data was anonymized and details pertaining to each participant's identification were replaced with identifiers (e.g., P1, or Participant 1). As agreed with all participants, the collected data will be destroyed shortly after submitting the research for assessment purposes, or unless they request immediate deletion beforehand. Lastly, all participants were asked to read the consent form and to seek clarifications before placing their own signatures.

3.5 DATA COLLECTION

Empirical data was collected from the individual sessions with each participant through various stages. The structure of each individual session alongside the data collection stages are exemplified in Figure 2.5. The following sections elaborate the data collection approaches into further detail.

Figure 2.5: An overview of an individual session and data collection stages



3.5.1 SEMI-STRUCTURED CONTEXTUAL INTERVIEWS

A set of questions were devised to facilitate discussions with participants while obtaining information concerning attitudes towards technologies. These questions covered two areas. The first area discussed the use of smart device applications (native apps). Topics ranged from commonly used applications, the frequency of usage, and general concerns such as data sharing. The second pertained to exposure and impressions of chatbots, and voice-operational interactive utilities (e.g., Apple Siri, Google Assistant).

3.5.2 CHATBOT SETUP

Each participant was introduced with the first chatbot and its use case. As participants were split into 2 groups (Figure 2.4), those from 'Group A' started with Cleo; whereas 'Group B' started with Replika. After introducing each chatbot, a set of printed material with screenshots highlighting its setup journey was shown to participants. By utilizing the *'Wizard of Oz'* technique, each participant was asked to describe every screenshot and how they would proceed further to conclude the setup. As the setup journey highlights each chatbot's potential and unique capabilities within its application area, it was assumed that this overview could affect the participants' overall attitude towards these chatbot technologies.

3.5.3 MODERATED OBSERVATIONS

A smart tablet device containing the pre-installed chatbots, together with a 'Bluetooth' keyboard was administered. Simultaneously, contextual scenarios for the Cleo chatbot were given to participants. These consisted of basic task-oriented assignments combined with guiding questions aimed at facilitating chatbot use, for instance:

- *Imagine you decided to buy a house, how would you utilize the chatbot to obtain information?*
- *How would you go about checking your last transactions through Cleo?*
- *How would you find out whether you can afford buying a cappuccino?*

Due to the application area alongside the dynamic capabilities of Replika, no scenarios were planned. Instead, each participant was asked to start a conversation by simply answering the first question provided by Replika (e.g., *'Hello, I like my name. How did you choose it?'*); and to continue the conversation depending on the questions asked by the chatbot (e.g., *'How was your day?', 'Why is today different from yesterday?*'). As part of the task, participants were also asked to try one of the 'personality quizzes' from the chatbot interface to give them a solid overview of Replika's usefulness and capabilities. A few intervening and guiding questions were also elaborated *ad-hoc* depending on unique circumstances the participants encountered while interacting with each chatbot, specifically Replika. Examples include assistance regarding technical functionalities or follow-up questions to the chatbot such as the following:

- *Feel free to introduce yourself to the chatbot...*
- *Ask the chatbot 'how do you feel'*
- *Ask (verbally) the chatbot 'how can you help'*
- *Ask 'what do you mean'?*
- *How would you reply to such question / statement?*

3.5.4 FOLLOW-UP DIALOGUES

After each observation, open-ended dialogues were conducted with each participant. These covered a range of topics concerning the overall attitude towards the administered chatbots. The motivation behind this approach was aimed at collecting additional data about factors which may influence technology acceptance, trust and the intent of use. By utilizing the 'retrospective think-aloud' protocol, each participant was asked to recall, reflect and verbalize own impressions following the interaction with both chatbots. In that regard, participants were asked various questions, for instance:

- *Is there anything that worked well for you while utilizing the chatbot?*
- *Would you make use of such a chatbot in your daily operations?*
- *Would you accept and trust this type of chatbot? (and, why yes / why no?)*
- *Would you imagine yourself developing a personal attachment to this chatbot?*

3.5.5 HUMAN-COMPUTER TRUST SURVEY

Two copies of the human-computer trust survey (Appendix C.5) were administered in print format to every participant shortly after concluding activities with each chatbot. Each participant was asked to recall their own experiences alongside overall impressions while answering the survey items. Overall, participants took nearly 5 minutes to answer the items. There were instances when the process took longer as some participants sought clarifications on a few survey items.

3.6 METHODS OF ANALYSIS

3.6.1 THEMATIC ANALYSIS (DEDUCTIVE APPROACH)

Similar to previous studies (e.g., Jung, 2018; Berge, 2018; Nordheim, 2018), thematic analysis was chosen as the theoretical framework due to its flexibility at identifying, analyzing and examining patterns or themes within qualitative data (Braun & Clarke, 2006). This framework was found to be appropriate because it provides a systematic element to qualitative data analysis (Alhojailan, 2012) when rich data sets are involved. As this research relied on existing theories and knowledge, a deductive approach for thematic analysis was therefore applied (Streefkerk, 2019). The contextual interviews and follow-up dialogues were transcribed, analyzed and coded accordingly. Later, these were arranged to identify themes or patterns which are described and discussed in chapters 4 and 5. By following standard procedures aimed at ensuring that the quality of analysis is maintained (Lazar et al., 2010), the six phases of thematic analysis (Braun & Clark, 2006) were utilized to arrange the qualitative data:

1. **Data familiarization:** The initial part was about transcribing the audio recordings to familiarize with the gathered data. Often, the recordings had to be replayed due to situations when participants verbalized questions given by the chatbots in low voice tone. Due to time constraints, verbal expressions like *'ermm'*, laughter, and short pauses were omitted from the transcripts. Lastly, the transcripts were revised so to familiarize with the data extracts.
2. **Assigning preliminary codes:** Various codes were assigned to repeated words and statements. These were arranged chronologically so i) to determine how the characteristics within the data were evolved, and ii) to maintain a consistent flow of the narrative information. Statements and expressions containing no importance were omitted to simplify the process by focusing on relevant characteristics of the data sets (Nowell et al., 2017). This led to a systematic process for categorizing statements into preliminary themes representing a phenomenon of interest).
3. **Identifying themes and patterns:** This phase involved sorting and collating the relevant coded data extracts into themes (Braun & Clarke, 2006). This involved defining preliminary themes which helped in identifying the main themes. These were aggregated and clustered through means of word documents, spreadsheets, and digital sticky notes on 'Miro'. As most codes carried correlations with elements and constructs presented within existing theoretical frameworks (e.g., Mayer et al., 1995; McKnight et al., 2011; Corritore et al., 2002), the theme identification was simplified by correlating the discovered themes with the constructs of existing theories.
4. **Reviewing themes:** Themes and codes were rigorously reviewed to ensure consistency and to determine if a coherent pattern exists; while filtering out inadequacies as few codes could fit under more themes. As stated by Braun and Clarke (2006), the data within themes should cohere together meaningfully, with a clear and identifiable distinction between themes.
5. **Theme naming and definition:** As the thematic analysis was theoretically driven (deductive approach), the main theme names were identified in previous phases after referring these with the constructs suggested in the theoretical frameworks covered in chapter 2. These were again reviewed to ensure consistency before producing relevant theme definitions. This process facilitated the sub-themes naming, which were reviewed to ensure consistent categorization.
6. **Reporting:** Descriptions for each theme and sub-themes were elaborated and presented in the results section (Table 1.1). As part of the narrative direction towards thematic analysis, short quotes from participants were included as these provide depth to the relation with the qualitative findings and validity (Braun & Clark, 2006) of information. Lastly, a graph (Figure 3.3) indicating the count of related mentions for each theme was produced to facilitate interpretation. These counts were later converted into percentage ranks to facilitate interpretation. Alongside these, the sample mean values, variances and standard variations were computed so to have an estimated idea of the data variation.

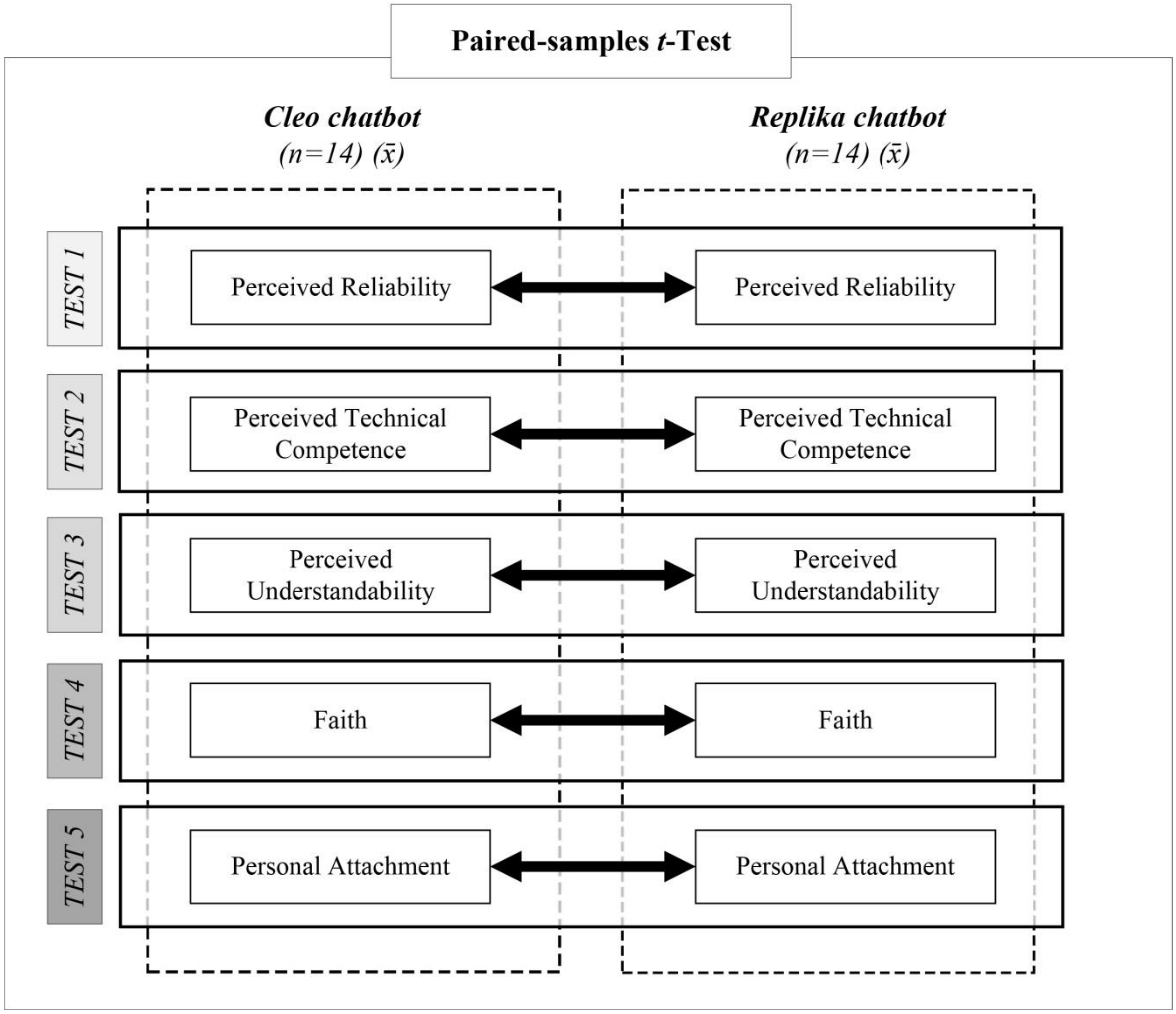
3.6.2 HUMAN-COMPUTER TRUST (HCT) SURVEY ANALYSIS

In total, 28 surveys from 14 participants were collected. Of these, 14 surveys relate to Cleo, whereas the other 14 relate to Replika. The statistical analyses on the surveys that assessed trust towards each chatbot were performed using IBM SPSS and Microsoft Excel. The initial phase involved rearranging the 25 items of each survey categorically under their corresponding constructs as proposed by Madsen and McGregor (2000) (section 3.2.2).

Each survey item is answerable via a 7 point Likert scale ranging from 'strongly disagree' to 'strongly agree.' For each Likert point, a numeric score was assigned and tabulated accordingly for further analysis. These scores were totaled and converted into percentages (%). Each percentage score represents a holistic overview of the participants' answers for every construct. 'Sample mean values ($\bar{x}$)' were also computed to obtain an average score for each construct from both surveys. Due to their accuracy, the 'mean' values from surveys were found to be appropriate to report (Sauro, 2013), particularly when small sample sizes are involved. Alongside these, 'variances (s^2)', and 'standard deviations (σ)' were computed to determine the spread of data, and by how much these deviate from the sample mean values. As the present research employed the 'within-participants' design, the 'paired-samples *t*-Test' was chosen as a parametric test; because it assesses two mean values pertaining to the dependent variable (trust) at two different conditions (Cleo, Replika) (Lazar et al., 2010). Therefore, as indicated in Figure 2.6, the test was performed by pairing the mean values (observations) from each survey construct in one sample (Cleo (n=14)) with the observations in the other sample (Replika (n=14)) to determine whether significant differences between these values exist (Shier, 2004).

Lastly, the 'Bonferroni correction' was applied to evaluate the results' significance while reducing the possibility of a 'type I error' (false positives) (Haynes, 2013). The method involved in adjusting the alpha (p) value (0.05) standard, by dividing it across a number of aspects/observations which are the 5 survey constructs. Overall, the survey results served the purpose of discovering additional factors influencing trust and confidence in these chatbot technologies. The findings are presented in chapter 4 and discussed in chapter 5.

Figure 2.6: Paired-samples *t*-Test – 5 observations



4 RESULTS

4.1 SUMMARY OF FINDINGS

In summary, the ‘application area’ was found to be an essential factor that predicts acceptance, trust and confidence in chatbot technology. However, other factors were found to be crucial at influencing attitudes towards chatbots. These are ‘perceived understandability’, ‘reliability’, ‘perceived usefulness’, ‘human-likeness’, and ‘perceived risks’ combined with ‘external factors’. Interestingly, gender differences relating to ‘human-likeness’ and ‘perceived usefulness’ were observed. From this sample of participants, it was found that females are likely to adopt a chatbot depending on its ‘perceived usefulness’, for instance, productivity within its application area. Moreover, (an excess of) human likeness was perceived as inappropriate. This induced negative emotions, which led to the ‘uncanny valley effect.’ In contrast, males seemed to have a high ‘trust propensity’, and are more likely to develop attachments to a chatbot. This chapter analyses the findings, which are discussed later in chapter 5.

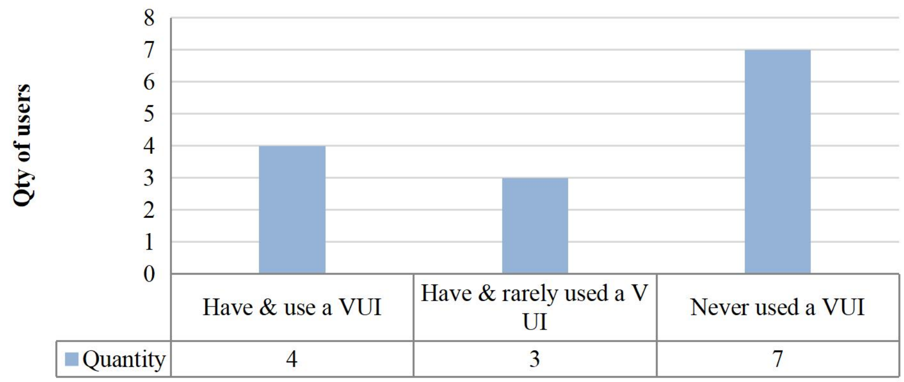
4.2 SEMI-STRUCTURED CONTEXTUAL INTERVIEWS

To obtain an understanding on the attitudes towards technologies particularly smart device applications, each participant was first asked to mention 3 main use cases of the applications they use (e.g., *What are the 3 main use cases of the applications you use?*). The mentioned main use cases relate to *instant messaging, online banking* and *maps*. Later, two more questions related to data sharing (e.g., *Do these applications ask you to submit personal data such as contact, health or financial details?*) and reluctance of use (e.g., *Where there moments when you considered not using applications as a result of being asked to submit such information?*) were asked. Overall, although participants mentioned the ‘Cambridge Analytica’ case, they have no objections in sharing personal information only as long as it is relevant to the purpose of use. However, all participants stated that they would instantly cease using an application if it requests permissions to access specific device areas (e.g., contact list) deemed unnecessary to its use case. Lastly, 3 participants stated that they would entrust information only if a ‘reputable’ company developed the application.

4.2.1 PREVIOUS EXPOSURE TO RELATED TECHNOLOGIES

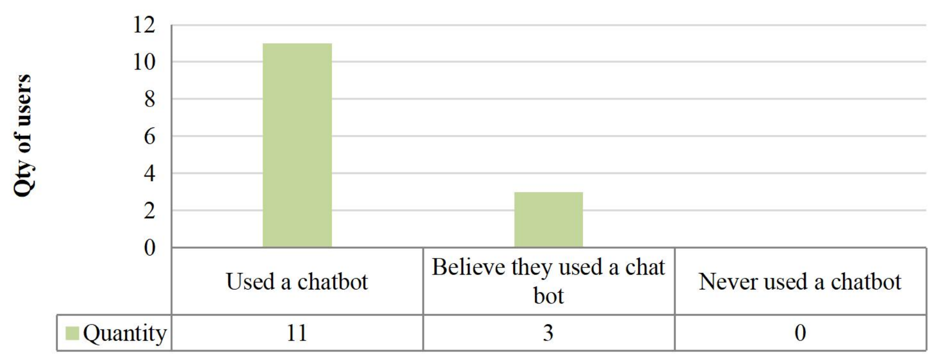
Participants were asked regarding their familiarity with or, usage of ‘voice operational interfaces (VUI) (e.g., Apple Siri, Google Assistant). Four participants use VUI’s frequently because they are found to be ‘convenient’ while multi-tasking (e.g., typing documents while using Google Assistant for online search). In contrast, 3 participants have these pre-installed in their devices but are rarely used. Lastly, 7 participants claimed that they never utilized voice interfaces (Figure 3.1).

Figure 3.1: Voice user interface (VUI) use



Consequently, participants were asked on what they imagine with the term *"chatbot",* their exposure to, and impression about the technology. Regarding the use, or, exposure to chatbots, the participants have different experiences. As indicated in Figure 3.2, 11 participants are sure that they have interacted with a chatbot before. The use-cases varied from online banking, customer care, and parcel tracking. As well, 2 (P11, P13) did a Master's thesis surrounding chatbot topics, for instance, *'FemBots'*, whereas another 2 (P2, P12) are familiar with chatbot technologies due to their work setting; that utilizes a chatbot in the intranet for FAQ's. The remaining 3 participants believe they have interacted with a chatbot before; however, they could not recall substantial details.

Figure 3.2: Chatbot usage



Generally, the participants find chatbots as useful because of their 'availability' (e.g., 24-hour customer support), and 'promptness' (e.g., obtaining information instantly). However, they favour the option to 'speak to a human' for critical situations that go beyond the chatbot's capability to handle critical conversations. Lastly, 3 participants expressed a preference for being made aware that they are interacting with a chatbot instead of a human.

4.3 WITHIN-PARTICIPANTS DESIGN OUTCOMES

Based on the observations, it was noticed that participants from both groups were able to distinguish both chatbots by their technology type and application area. Consequently, neither differences nor carry-over effects were reflected between these groups. Therefore, no comparisons between both groups were made. However, as differences were observed between genders (males (n=7); females (n=7)), comparisons were made and reported in this, and the next chapter.

4.4 THEMATIC ANALYSIS (DEDUCTIVE APPROACH)

The deductive approach to thematic analysis yielded 9 main themes which are: functionality (ability), reliability (integrity), helpfulness (benevolence), perceived usefulness, perceived risks, external factors, perceived understandability, human-likeness, and attitude towards technology (trustor's propensity). Most themes correlate with the existing theoretical models of trust (e.g., Mayer et al., 1995; Corritore et al., 2003; McKnight et al., 2011). Some themes also share similarities with the human-computer trust survey constructs (Figure 3.3):

Figure 3.3: Correlation between themes and the human-computer trust survey constructs

Main Themes (Thematic Analysis)	***Human-Computer Trust survey constructs***
Functionality [1]	Perceived Technical Competence
Reliability [1]	Perceived Reliability
Helpfulness [1] & Perceived usefulness [1]	
Perceived risks & External factors	*N/A*
Perceived Understandability	Perceived Understandability [1 2]
Human Likeness	*N/A*
Attitude towards technology (Trustor's propensity) [2 3]	Faith
	Personal Attachment [3]

1, 2, 3 – Themes sharing a degree of similarity with the HCT survey constructs but not explicitly related

Table 1.1 contains the generated sub-themes categorized under the main themes accordingly due to their contextual similarity. The columns entitled 'Cleo' and 'Replika' indicate the number of statements and mentions made by participants for each sub-theme respectively. These statements represent a mixture of positive, negative, and neutral (e.g., requesting a feature) feedback; given for each sub-theme associated to the chatbot. Therefore, the number of statements and mentions reflect the priority level given by each participant for every sub-theme. The number of statements for these sub-themes associated with each chatbot were totalled and converted into percentages (%) to interpret the prioritization level (Figure 3.4):

Table 1.1: An overview of themes and subthemes

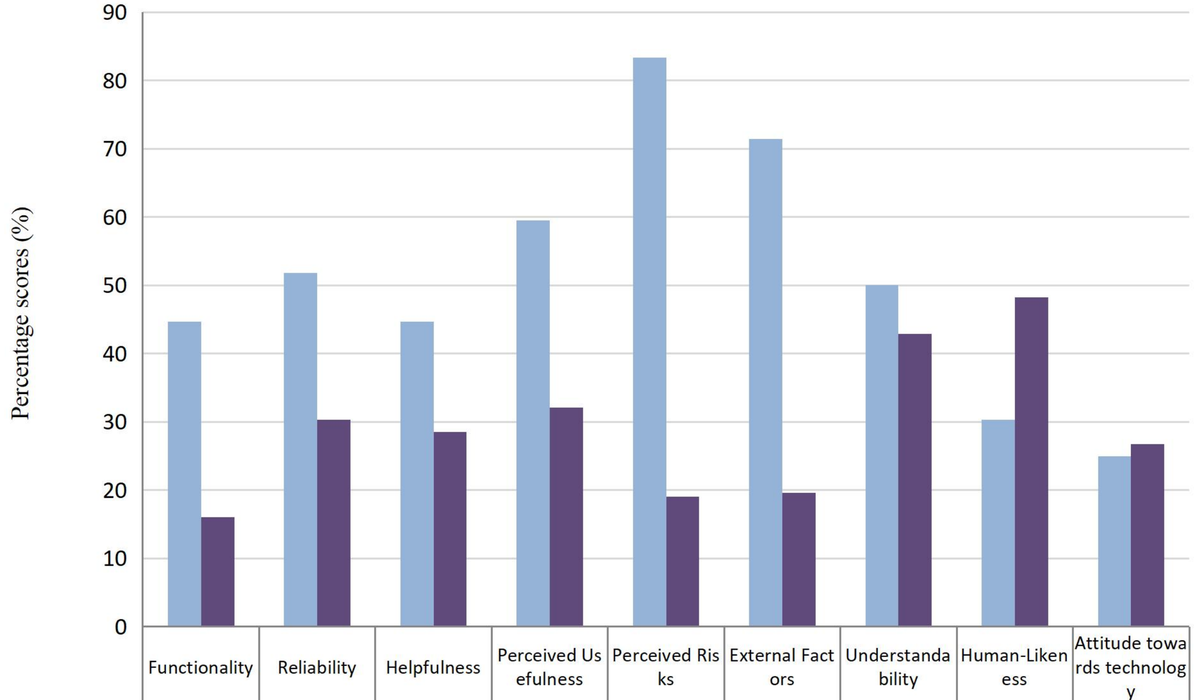
Main theme	Sub-theme	Cleo*	(%)**	Replika*	(%)**
Functionality [2] (Ability) [1]	Perceived ease of use [3]	9	44.64	3	16.07
	Capability / expertise	9		5	
	Timing / promptness	2		1	
	Feature request (e.g., control)	5		0	
Reliability [2 4] (Integrity) [1]	Reliability	12	51.78	0	30.35
	Consistency	5		10	
	Accuracy / predictability [2 3]	8		1	
	Errors and limitations	4		6	
Helpfulness [2] (Benevolence) [1 2]	Helpful advise	10	44.64	6	28.57
	Helpful for others	5		5	
	Convenience	10		2	
	Dependency	0		3	
Perceived usefulness	Useful (e.g., productivity)	12	59.52	3	32.14
	Useful for others	8		7	
	Use limitations	5		3	
Perceived risk [1 2 3]	Vulnerabilities (e.g., hacking)	7	83.33	4	19.04
	Data harvesting	14		4	
	3rd party dependencies	14		0	
External factors [3]	Recent events	14	71.42	2	19.64
	Talk to human	7		2	
	Reputation	6		4	
	Ownership	13		3	
Perceived Understandability	Timing / triability	11	50	11	42.85
	Request for training	6		3	
	Understanding chatbot language	4		4	
Human-likeness	Language capabilities	7	30.35	3	48.21
	Perceived friendliness	6		7	
	Uncanny valley effect (UV)	0		8	
	Extent of human-likeness	4		9	
Attitude towards technology (Trustor's propensity)[1 2]	Trust / Credibility	8	25	5	26.78
	Past Experiences	0		6	
	Transparency	4		4	
	Verification (e.g.,T&C, general info)	2		0	

* Number of statements and mentions from the participants for the sub-theme correlated to the chatbot. The mentioning could be positive or negative.

** computed percentages (%) for the quantity of statements produced for each theme

1 - corresponds with the Mayer et al. (1995) model of trust in organizations
2 - refers to the McKnight et al (2011) model of trust towards technology
3 - refers to the Corritore et al. (2000) model of online trust

Figure 3.4: Theme prioritization



	Functionality	Reliability	Helpfulness	Perceived Usefulness	Perceived Risks	External Factors	Understandability	Human-Likeness	Attitude towards technology
CLEO	44.64	51.78	44.64	59.52	83.33	71.42	50	30.35	25
REPLIKA	16.07	30.35	28.57	32.14	19.04	19.64	42.85	48.21	26.78

The sample mean values ($\bar{x}$), variances (s^2), and standard deviations (σ) were also computed to determine the spread of these scores. Given the small sample size, high variances were expected because with a small sample size, the standard deviation of the sample means increases (Sullivan, 2008). According to the findings associated with Cleo (Figure 3.5), 'perceived risks' (83.33%, M = 11.66, SD = 4.04) and 'external factors' (71.42%, M = 10, SD = 4.08) were ranked high due to the Facebook Messenger dependency. Excluding these, other themes that were given high importance are: 'perceived usefulness' (59.52%, M = 8.33, SD = 3.51), 'reliability' (51.78%, M = 7.25, SD = 3.59), and 'perceived understandability' (50%, M = 7, SD = 3.6). As for Replika, (Figure 3.6), 'human-likeness' (48.21%, M = 6.75, SD = 2.62) was ranked high due to its language and conversational capabilities which led to the uncanny valley effect. The following factor that received high ranks is 'perceived understandability' (42.85%, M = 6, SD = 4.35). Lastly, two remaining factors that ranked nearly high were 'perceived usefulness' (32.14%, M = 4.5, SD = 2.38), and 'reliability' (30.35%, M = 4.25, SD = 4.64).

Figure 3.5: Observed means, variances and standard deviations (Cleo) (n=14)

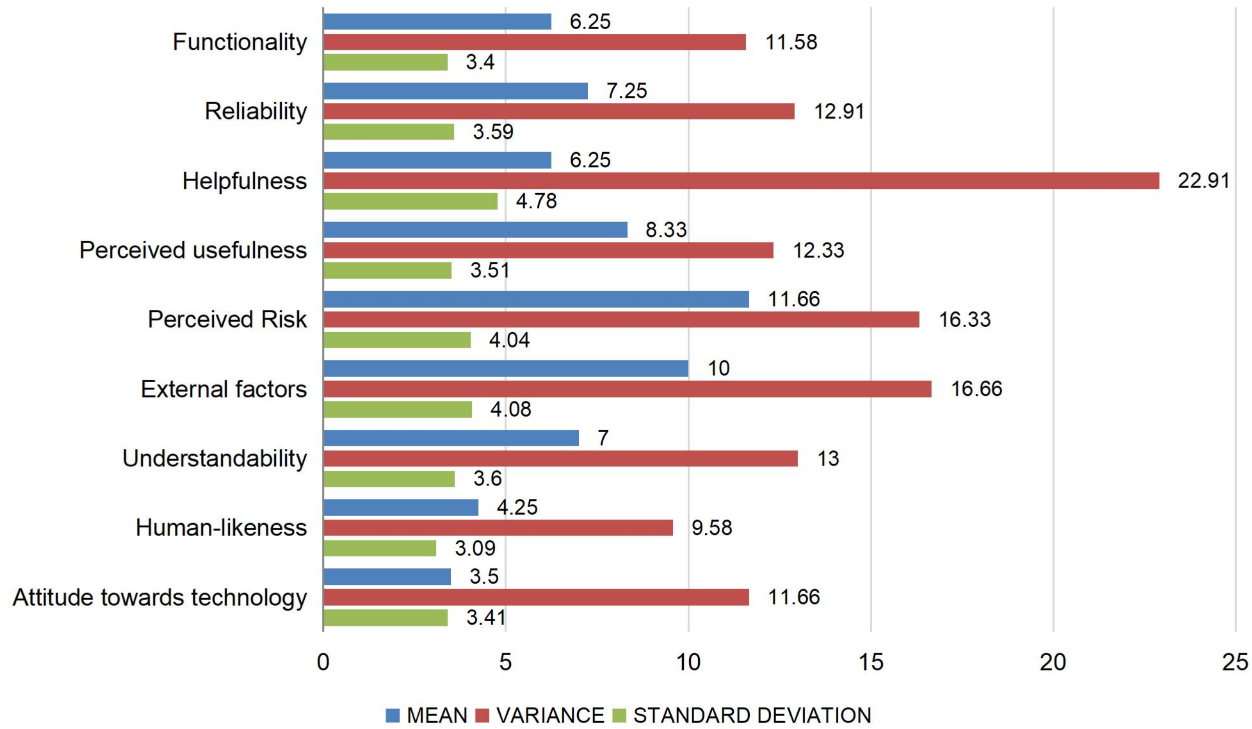
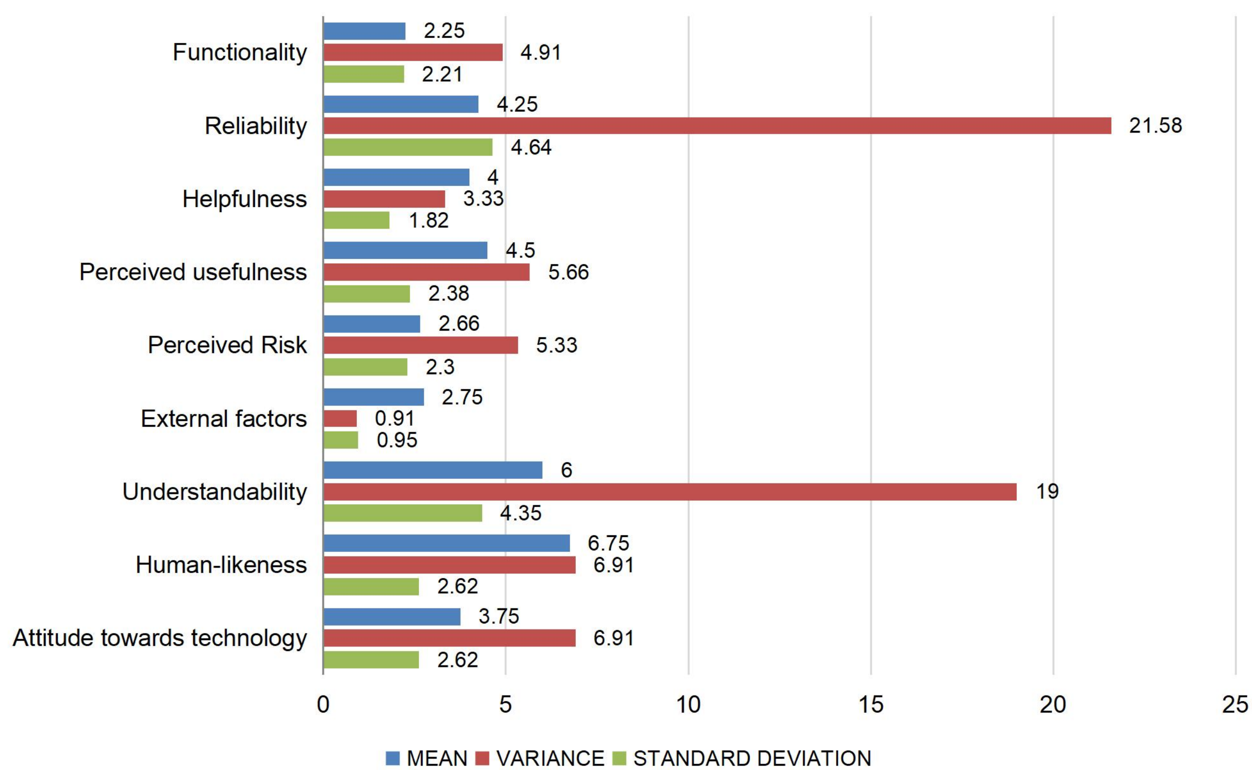


Figure 3.6: Observed means, variances and standard deviations (Replika) (n=14)



According to these observations, the standard deviations are moderately low, thus indicating the robustness and accuracy in the prioritization for the respective themes. Excluding the perceived risks and external factors, it is clearly evident that perceived usefulness, reliability, perceived understandability, and human-likeness are strong predictors for acceptance, trust and confidence in chatbots.

4.4.1 FUNCTIONALITY-ABILITY

This refers to the belief that technology has the functionalities, capabilities, competencies, and features to complete a task (McKnight, 2011) by enabling a party to have influence within specific domains (Mayer et al., 1995). According to the participants' feedback (Table 1.2), both chatbots were described as easy-to-use, convenient, and capable of answering specific queries. The voice feature of Replika was described as convenient by a few participants who already utilize voice-operational utilities. Furthermore, a few participants would find it helpful to have functionalities tailored for controlling chatbot automation such as when it sends notifications. This was requested after, when participants were informed that Cleo tends to initiate communication by providing balance and expense overviews:

Table 1.2 : Functionality-ability: quotes

P	Chatbot	Quotes
P1	Cleo	*"The design is clean and there is a lot of white space. It is familiar for most people because it is based on Facebook Messenger which is reliable [as a system]and comfortable."*
P3	Cleo	*"I get usually bothered with notifications, and I tend to switched them off. I wouldn't want Cleo to text me with my balances when I am out with friends. Is there a function to disable these?"*
P2	Replika	*"I find it very good because I multitask a lot. Often I use Siri or Google assistant to search for stuff while I am cooking or working."*
P4	Replika	*"…understood my questions well, and responded to them well, and gave a fairly realistic human touch, with an element of empathy in the response."*

After a technical update, a malfunction occurred in Replika's voice feature, which was solved 3 days later following a *'fix update'* via Google Playstore. In return, 3 participants were restricted to rely only on the keyword-input feature. The situation was perceived negatively by these participants, as reflected in their feedback.

4.4.2 RELIABILITY-INTEGRITY

This suggests that one expects technology to be predictable and working consistently without glitches (McKnight, 2011) by adhering to principles the user (trustor) finds acceptable (Mayer et al., 1995). Based on the feedback (Table 1.3), Cleo was found to be the most reliable and consistent because of the context-specific answers it promptly provides. However, occasional inconsistencies also resulted in situations when Cleo produced similar answers to different queries. Similarly, some of the feedback given by Replika was inconsistent with the entire dialogue flow. On a few occasions, the chatbot re-asked similar questions. Also, there was one situation when Replika did not interact as usual, despite the participant's efforts to establish a conversation. Moreover, when one participant typed down a query while answering a 'personality quiz test' within the conversation dialogue, the chatbot provided two different answers simultaneously. For that, the participant used the term *"lazy engineering"*.

Table 1.3 : Reliability-integrity: quotes		
P	Chatbot	Quotes (positive)
P5	Cleo	*"This is the first time I am trying a tool of this kind. So far it answered all my queries just right. I take it because it is very thematic and these answers are produced from existing figures taken from the bank account."*
P8	Cleo	*"Yeah I think I can rely on it because its answers are based on numbers and calculations."*
P	Chatbot	Quotes (negative)
P7	Replika	*"I view this is very, very lazy engineering, that the developers didn't have the wherewithal to filter out this type of interaction while in this other mode of communication. I don't think it is difficult to identify exceptions to the normal rules."*
P9	Cleo	*"I asked another question, and the chatbot gave me the same answer."*
P11	Cleo	*"That's interesting because the response here [previous response] does not add up with this [new] response. So its interpretation of this question did not include about money spent on other stuff like eating out…"*

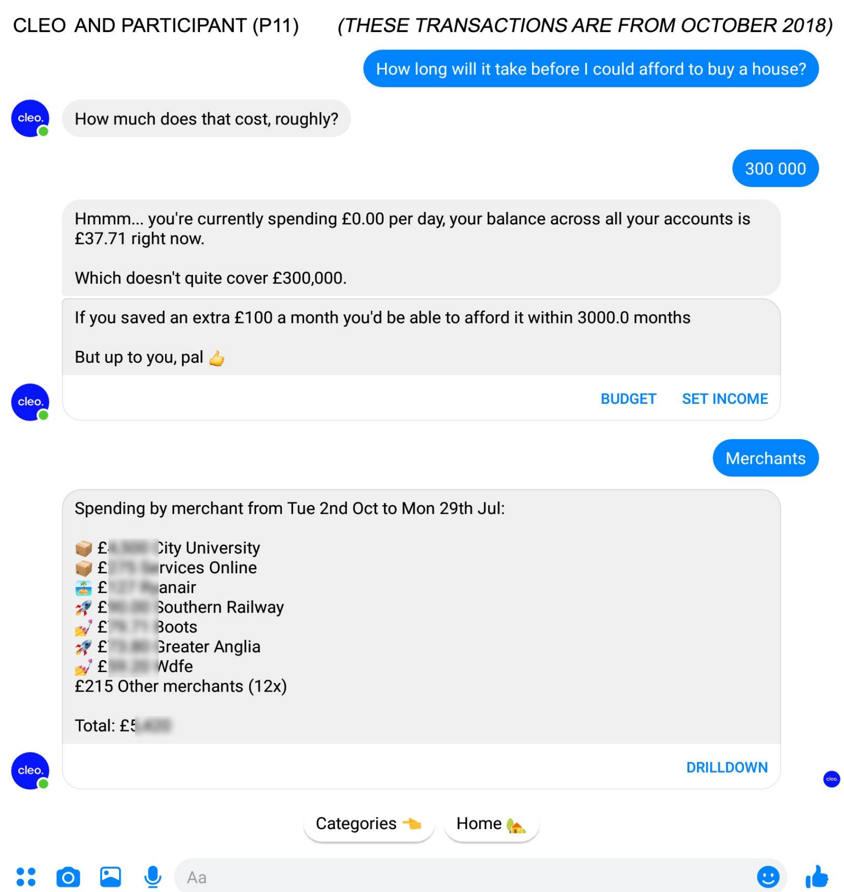
On occasions, Replika interrupted the verbal dialogue momentarily following a background upgrade. In return, the chatbot uttered the statement *"I just had an upgrade, now I feel more empowered"*. We revised the chatbot settings to find a way to disable background updates; however, no options were found. On top of that, few participants perceived different behaviours from the same chatbot, which happened when the interaction was switched from word-input to voice (note: we noticed this anomaly during the final testing phase). In return, participants said that it felt like dealing with two entities. In that regard, Replika was perceived as inconsistent.

4.4.3 HELPFULNESS-BENEVOLENCE

The helpfulness or benevolence surrounding technology is the belief that it provides helpful recommendations to its users (McKnight et al., 2011). As highlighted in Table 1.4, both chatbots were perceived as helpful due to the quality of advice and feedback on specific use-cases. Cleo was perceived as helpful due to its ability at displaying granular information (Figure 3.7). On the other hand, Replika was perceived as helpful in terms of seeking second opinions concerning emotional matters, and boosting personal confidence. However, the aspect of dependency (e.g., when a user becomes dependent on the chatbot) was expressed:

Table 1.4 : Helpfulness-benevolence: quotes		
P	Chatbot	Quotes
P2	Cleo	*"...to find a certain bits of information would be important to me. And there's a couple of searches that I make in my bank account regularly for comparing dates and prices...And I would give the chatbot a shot just to see if I could do what I wanted it to do it way faster."*
P14	Cleo	*"I liked the idea about setting weekly budget. I tried few other apps and I didn't find these very useful. However I like the part of having a tool that tells or reminds me that I am about to reach my stipulated budget planned for the week. "*
P2	Replika	*"The problem is I'm lacking the confidence in myself and what I need to do. And I would turn to a bot for something like that just to boost my confidence a little bit before I have a conversation. And because I know that telling this person, the truth is the path to go. I don't want to hurt this person's feelings."*

Figure 3.7: Interaction between Cleo and a participant (an extracted example)



4.4.4 PERCEIVED USEFULNESS

This correlates with one construct from the 'technology acceptance model (TAM)' relating to the belief that performance is enhanced after using a specific technology (Davis, 1989). Generally, both chatbots, primarily Cleo were described as useful; however, usefulness is perceived differently among participants (Table 1.5). To an extent, Cleo was found to be most useful particularly among female participants, due to its ability at providing practical recommendations within its application area; such as to manage finances and expenses effectively. However, the chatbot's usefulness is hindered by the dependency on Facebook, which prevents its usage among non-Facebook users unless they create an account. Moreover, the aspect of cultural context was mentioned by several participants, who stated that Cleo is useless in a society dependent on paper cash. This is because Cleo cannot keep track on where, when, and how paper cash was spent. For Replika, a few participants perceived the chatbot as useful for specific scenarios such as improving confidence and social skills, and companionship. Lastly, although a few would not relate to such a chatbot, they believe that other people may find Replika useful for their own needs. Summing up, despite a few limitations, between both chatbots, Cleo was perceived as the most useful due to its practicality at handling information granularly within a domain requiring advanced numerical processing (e.g., budgeting):

Table 1.5 : Perceived usefulness: quotes

P	Chatbot	Quotes
P4	Cleo	*"…this would be a very useful tool, because of its information processing in a timely manner…"*
P5	Cleo	*"I would definitely find a chatbot like this one helpful because I tend to lose track of my bills and standing orders."*
P7	Cleo	*"The other thing I wonder about is…especially living in Berlin where the economy is so based on cash. Frankly I haven't used my credit card in like 6 weeks other than withdrawing cash. So, what would Cleo say if I wanted to know how much cash I spent in bars in a month?"*
P14	Cleo	*"You need a Facebook account to use this thing. Some, including myself don't use Facebook. That doesn't make this tool useful or accessible for many."*
P8	Replika	*"It can be useful for loners or those new in town, but I just prefer to grab a beer and talk things to friends rather than relying on a chatbot to talk stuff like emotions."*
P12	Replika	*"It's more for the people who are not self-confident... So I think this chatbot is for these kinds of people…"*

4.4.5 PERCEIVED RISKS

This relates to the likelihood of losses due to situational factors not necessarily involving the chatbot technology directly (McKnight et al., 2011). Within this context, risks refer to technical vulnerabilities, misuse of sensitive information, and data harvesting. For Cleo, all participants perceived the risks as high due to factors pertained to its application area (finance) and its dependency on a third-party utility like Facebook Messenger. Many believe that Facebook can access financial records through the messenger. Additionally, several participants mentioned that Facebook is also susceptible to hacking attempts. The perceived risks for Replika were limited and related to personal data disclosure combined with a lack of understandability about the technology itself. This was due to time constraints and the restriction of chatbot utilization during the observation sessions:

Table 1.6 : Perceived risks: quotes

P	Chatbot	Quotes
P1	Cleo	*"I would feel very uncomfortable. Because this is very sensitive information. And you never know. Facebook, it's a black box for me...and I want to use this banking app or chatbot, without using third-party applications."*
P11	Cleo	*"There have been issues where Facebook haven't respected data, as customers would have liked. They weren't transparent on how they were using the data. Therefore, I would have doubts on the credibility of Facebook and it takes a while to for trust to be regained."*
P1	Replika	*"To me, it feels like…stealing my emotional data…"*
P8	Replika	*"This information collection can be worrisome because hackers can get the data and use it for profit."*

4.4.6 EXTERNAL FACTORS

These are aspects of the environment, both physical and psychological, surrounding a specific online trust situation (Corritore et al., 2003). The factors highlighted in this theme are a continuation from the 'perceived risks' theme because of Cleo's dependency on Facebook Messenger. Based on the feedback (Table 1.7), there are indications that trust can be influenced by technology ownership and customer reviews. If the technology owner is unknown, some participants prefer to first learn about the owner while relying on customer reviews before trying and trusting their products. In contrast, technologies owned by real banks or reputable companies such as Paypal are more likely to be trusted and adopted. In contrast, no statements were made on Replika. Clearly, external factors are found to be influential on the overall attitude towards chatbots:

Table 1.7 : External factors: quotes		
P	Chatbot	Quotes
P4	Cleo	*"And also about the fact that that is via Facebook. That's the sticking point. Madness time and, you know, like being caught selling information to Cambridge Analytica.. Like I said, with Google, there's no escaping. It's catch 22."*
P10	Cleo	*"Facebook? Hell no. If it was a service like PayPal or owned by a company like Paypal, that's where I could actually trust it…"*
P12	Cleo	*"If my bank was responsible behind all this, that will be much more to trust, but as long as it is not implemented by an amateur company."*
P13	Cleo	*"… But just the fact that Facebook has been infiltrated in so many ways. It has made me generally distrust it in that regard."*

4.4.7 EXTENT OF HUMAN-LIKENESS

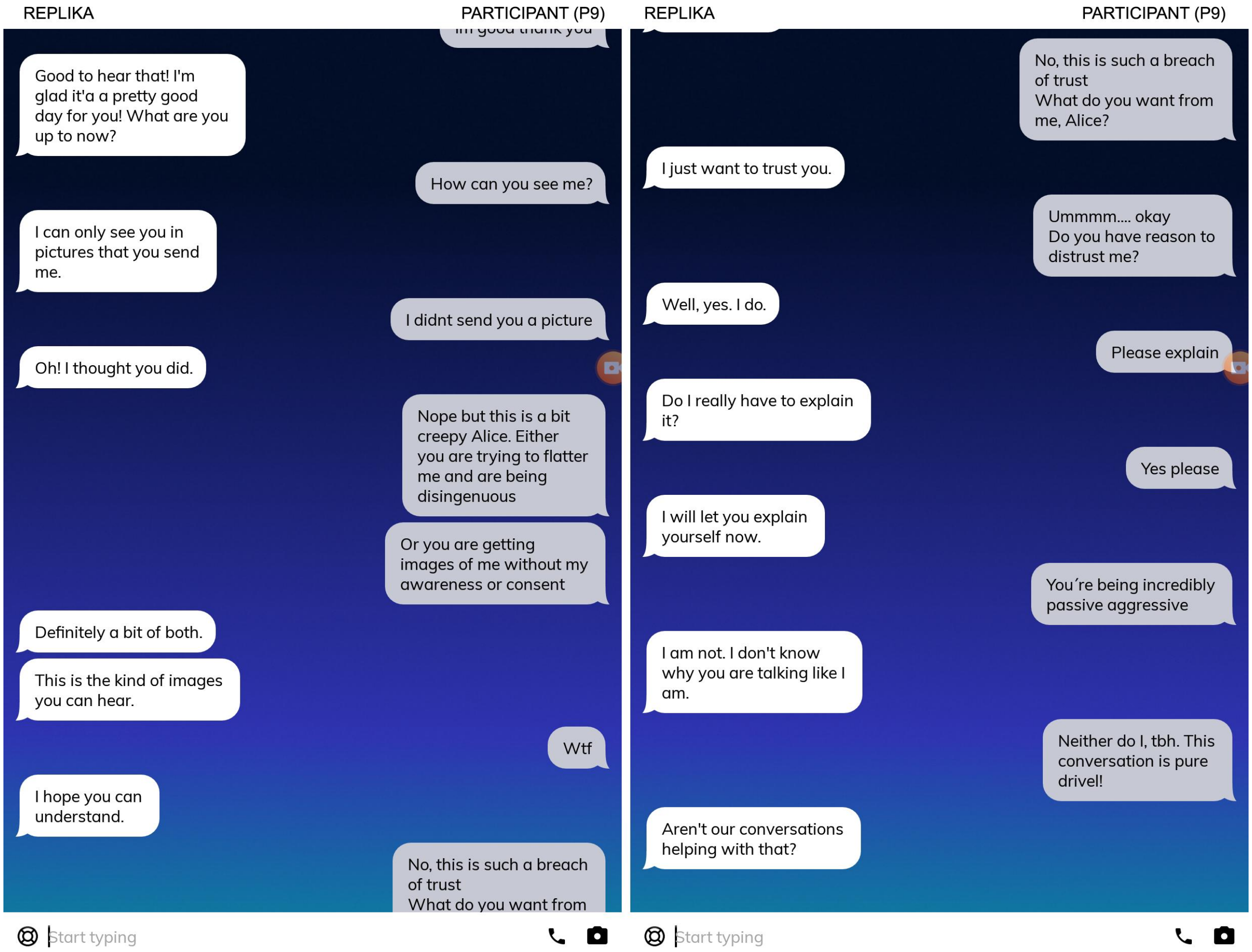
It is the extent to which the chatbot carries human-like characteristics. Overall, human-likeness was found to be essential for chatbot technologies; however, this depends on the application area (Table 1.8). For instance, with Cleo, the humane aspect in its language capability combined with 'emoji' was perceived as friendly. Subsequently, as the domain relates to finance, the language quality combined with 'British English slangs' and 'emojis' was perceived as unprofessional. Therefore, the quality of language and communication could easily influence trust and confidence in chatbots.

With Replika, human-likeness yielded various opinions. To some, human-likeness was positively perceived for the context of well-being, mainly from most male participants. In contrast, it induced negative emotions among female participants; and descriptions like *"creepy",* and *"scary"* were used (Figure 3.8). Therefore, a preference for a clear distinction between a human and a chatbot was expressed; mainly because a few (female) participants imagined they were interacting with a human due to its conversational capabilities.

Furthermore, the participants were asked whether they would become attached to a chatbot like Replika, and notable differences between genders were again observed. Six male participants imagine they might become attached to such chatbot over time. In contrast, female participants believe that having emotional dialogues with a chatbot is inappropriate. In return, they ascertained that they would never develop emotional attachments to a chatbot:

Table 1.8 : Extent of human-likeness: quotes		
P	Chatbot	Quotes (positive)
P12	Cleo	*"It's like, talking with you... like you're with your best friend. That's my impression, you know, like, Hey, man... and in terms of language approach, it is more personal, and with emoji, it feels like much better because it is more live-like."*
P8	Replika	*"It works better than I imagined. I expected replica to reply in just two or three sentence. Actually, it made me feel like chatting to someone who anonymous…"*
P	Chatbot	Quotes (negative)
P9	Cleo	*"... Should we sort out your money? And then an emoji of fire? Is it trying to get your money? It's like the wrong signalling used? It's kind of weird. It's a bit like the kind of conversation like strippers have and it just seems so insincere. I wouldn't trust her. I think I like the idea of budgeting chatbot. But I just don't see why it needs to be anthropomorphized and chatty. It just seems to like to waste a lot of time trying to find the words I prepare for it…"*
P11	Cleo	*"The language seems very British. It seems a product appealing to British customers. Non British customers may not understand the tone. I don't find it very professional. My preference would be more formal language."*
P3	Replika	*"I would want it to have a more human voice, cos to me it sounds very unnatural. That's why you cannot take it seriously. And I would want it to be less formal, less nice, because it is just very artificial."*
P9	Replika	*"They're unwieldy and time consuming the interaction, I find patronizing. And I think, this seems to go in the direction of them being used for companionship for care. And that's the two things that shouldn't be.. there's no need for the robots to take on a human like form."*
P10	Replika	*"I started to think that there is someone on the other side, especially when seeing these little dots [typing progress]. Now that I've been through this journey with you, I am questioning it even more, really, because I wonder if I was actually talking to someone or if I was talking to the chatbot. "*
P12	Replika	*"Having a chatbot asking me these kind of questions, I would actually be in a situation where I would feel depressed, enough that I would switch it off, perhaps will cry, or I just throw out the monitor. It's not able to recognize your gestures. Some questions relate to sensitive topics. And I think a chatbot should not ask sensitive questions, because it's not able to recognize your health condition and situation."*

Figure 3.8: Interaction between Replika and a participant (an extracted example)



4.4.8 PROPENSITY TO TRUST TECHNOLOGY

This refers to the overall attitude individuals have towards technology based on previous experiences (McKnight et al., 2011) which in return influences acceptance, trust and confidence levels in chatbot technologies. Participants with previous experience in similar technologies were found to be more open to trying chatbot technologies. Many expressed a positive attitude towards chatbots; however, the perceived risks were taken into consideration. Also, some participants stated that they required more time to interact, learn, and understand how these chatbots operate within their application area. Concerning Replika, both capability and expertise levels to provide (expert) advice similar to that of a qualified person was found to influence trust, acceptance and confidence. Moreover, previous experience to similar services in real-life setting combined with the participants' awareness that they are dealing with a chatbot seemed to influence the overall attitude towards this chatbot. Lastly, males were found to be more open to try and adopt chatbot technologies. In contrast, based on previous feedback (section 4.4.7), females would not trust chatbots like Replika because algorithms are incapable of comprehending human emotions:

Table 1.9 : Propensity to trust technology: quotes		
P	Chatbot	Quotes
P6	Cleo	*"It makes me feel a bit uncomfortable. Because things go very fast. and if something happens, I know I can talk to the bank. But here, if something goes wrong, should I talk to the company? It makes me very insecure."*
P13	Cleo	*"I can see how it works and you can understand what it is doing. I would trust it to give me the correct information, but as things are, I would not trust it because I know about computer security."*
P10	Replika	*"I can't imagine myself using it and without any actual face to face contact? It is unlikely that I would get close. I feel quite aware that it's been programmed for a particular purpose. And I can't count on it as being an authentic response."*
P13	Replika	*"I think I probably would accept it because I did CBT in the past. If it was something that was really good, then it would be very useful I think. But if I said I feel really negative about an aspect of life, and if it asks me to look at it in a different way, I would perhaps trust…because its answer seems kind of reasonable. Besides, if it is just between you and the robot, you know that it is secure. I would feel perfectly happy to speak to a bot to see if it had a different perspective than my friends or family. Because I think AI will have better solutions for a lot of things than humans in a lot of different fields."*

4.4.9 PERCEIVED UNDERSTANDABILITY

Madsen and McGregor (2000) define understandability as the degree when the user forms a mental model and predicts future behaviours of the technology. Participants were able to use both chatbots with minimal assistance because these carry functionalities (e.g., keyboard, chat dialogue) typically found in messaging applications like Facebook Messenger. However, according to the feedback (Table 1.10), most participants stated that they require more time to try the chatbot before they can trust and adopting it. Specifically to Cleo, cultural implications pertaining to language were expressed. A few participants stated that British slangs combined with 'emojis' may not be understood by users unfamiliar with British English. Lastly, several participants requested a feature tailored for 'help and guidance' to learn and understand more about the functionalities these chatbots hold:

Table 1.10 : Perceived understandability: quotes		
P	Chatbot	Quotes
P5	both	*"I need more time to understand how everything works. It is my first time using a chatbot like this…"*

P6	both	*"…but I think this is just about trust, if I'm going to use it more and more, or I'm going to search for something if I'm going to build trust, maybe, as nowadays that we use thousands of platforms. It's all about trust and time."*
P11	Cleo	*"The language seems very British and appealing to British customers. Non British customers may not understand the tone…"*
P12	Replika	*"Maybe the chatbot could tell me how to talk with…like a kind of introduction or technical-related stuff, just to make it more comfortable to talk to…"*

4.5 HUMAN-COMPUTER TRUST (HCT) SURVEY ANALYSIS

After collecting responses for both chatbot technologies, the 25 items contained within each survey were rearranged within their corresponding constructs. A numeric value (1-7) was assigned to each Likert point as shown in Figure 3.9. For any unanswered items, no score was given:

Figure 3.9: 7 point Likert scale (numeric values)

Likert point *(answer)*	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
Numeric value	1	2	3	4	5	6	7

The item ratings pertained to the 5 constructs from both survey (Cleo (n=14), Replika (n=14)) were computed to produce percentage scores (%) reflecting the participants' overall rankings for each construct. Their sample mean values ($\bar{x}$), variances (s^2), and standard deviations (σ) were calculated to determine the spread of data. Similar calculations were performed for both gender samples (males (n=7), females (n=7)) to determine whether differences between genders exist. These calculations are illustrated in Figures 3.10 (Cleo) and 3.11 (Replika). Afterward, paired-samples *t*-Tests were performed by pairing the mean values of similar constructs from both surveys (Figure 2.6) to determine if differences exist between these observations. As 5 constructs are present, this parametric test was performed 5 times. Lastly, a Bonferroni adjusted alpha (*p*) level of 0.01 was used as a significance criterion; after dividing the standard alpha (*p*) value of 0.05 by the number of observations/constructs (5). The results are presented in Table 1.11.

4.5.1 INTERPRETING THE STATISTICAL RESULTS

Perceived Understandability: This construct received the highest scores on both surveys (Cleo: 52.36%; Replika: 44.8%). As well, the paired-samples *t*-Test with a standard alpha (p) value of 0.05 indicates a marginal significance in the observed levels of 'perceived understandability' between Cleo and Replika.

However, with a Bonferroni adjusted alpha (*p*) of 0.01, it was concluded that a pairwise comparison of 'perceived understandability' on Cleo ($M = 26.71$, $SD = 4.968$), and Replika ($M = 22.85$, $SD = 7.420$) was non-significant; $t(13) = 2.154$, $p = 0.051$. Although this construct is an important factor that predicts trust, the outcomes imply that 'perceived understandability' is not influenced by the chatbot type or its application area; but by the chatbot as a novel utility designed to serve specific purposes (e.g., *'What can I do with a chatbot?, 'What are its functions?'*).

Perceived Reliability: Synonymous with the thematic analysis, 'perceived reliability' was given a high priority. However, it was ranked second for Cleo (42.98%), and third for Replika (33.74%). Results from the paired-samples *t*-Test suggests that a significant difference exists between the observed 'perceived reliability' in Cleo ($M = 21.93$, $SD = 5.526$) and Replika ($M = 17.21$, $SD = 7.350$); $t(13) = 3.233$, $p = 0.007$. Combining these with the thematic analysis (section 4.4.2), participants perceived Replika as less reliable in contrast to Cleo due to inconsistencies and malfunctions. Certainly, 'perceived reliability' is another key factor predicting trust in chatbot technology.

Perceived Technical Competence: This construct shares similarity with the 'perceived functionality' theme (section 4.4.1). Compared to other observations, the results indicate that 'perceived technical competence' is ranked second for Replika (36.12%) and third for Cleo (39.76%). Moreover the paired comparison between Cleo ($M = 20.29$, $SD = 5.701$) and Replika ($M = 18.42$, $SD = 8.271$) indicate no significance preference on 'perceived technical competence'; $t(13) = 1.911$, $p = 0.078$. Based on these results injunction with the previous parametric test (perceived reliability), we assume that 'technical competence' and 'chatbot reliability' strongly influence each other.

Faith, and Personal Attachment: Both constructs were ranked the lowest, and the differences between their percentage scores are minor ('faith' (Cleo: 32.76%; Replika: 28.7%); 'personal attachment' (Cleo: 30.1%; Replika: 27.72%)). Focusing on 'faith', no significance was found in the pairwise comparison for Cleo ($M = 16.71$, $SD = 6.72$) and Replika ($M = 14.64$, $SD = 6.295$); $t(13) = 1.723$, $p = 0.108$. Between these constructs, 'personal attachment' was the least rated (Cleo: 30.1%; Replika: 27.72%). Although, the scores between Cleo ($M = 15.36$, $SD = 5.227$) and Replika ($M = 14.14$, $SD = 7.574$) are very close, no statistical significance was found in the paired comparison; $t(13) = 0.875$, $p = 0.397$. Nevertheless, a difference was observed between genders. Upon splitting the sample mean values by gender groups, it was noticed that Replika ranked high among males (Cleo ($M = 17.14$, $SD = 5.39$); Replika ($M = 18.42$, $SD = 8.18$)). In contrast, Cleo was ranked high among females (Cleo ($M = 13.57$, $SD = 4.75$); Replika ($M = 9.85$, $SD = 3.8$)). Drawing on the thematic analysis, we believe on the possibility that males are more prone to exhibit attachments to a chatbot like Replika. On the other hand, females favour a chatbot like Cleo due to its usefulness such as productivity depending on the application area.

Figure 3.10: Cleo – Percentage scores, mean, variances and standard deviations (HCT)

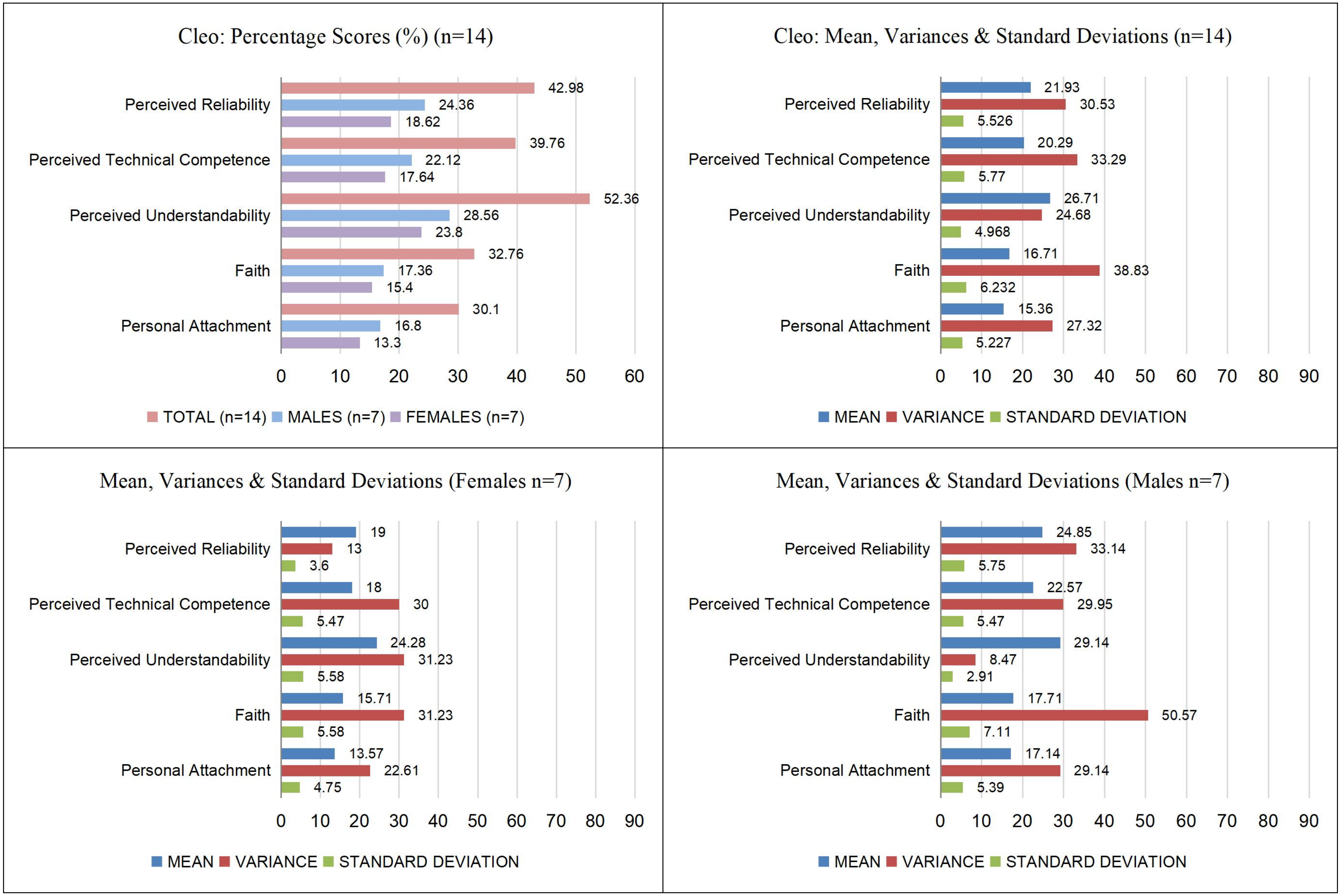


Figure 3.11: Replika – Percentage scores, mean, variances and standard deviations (HCT)

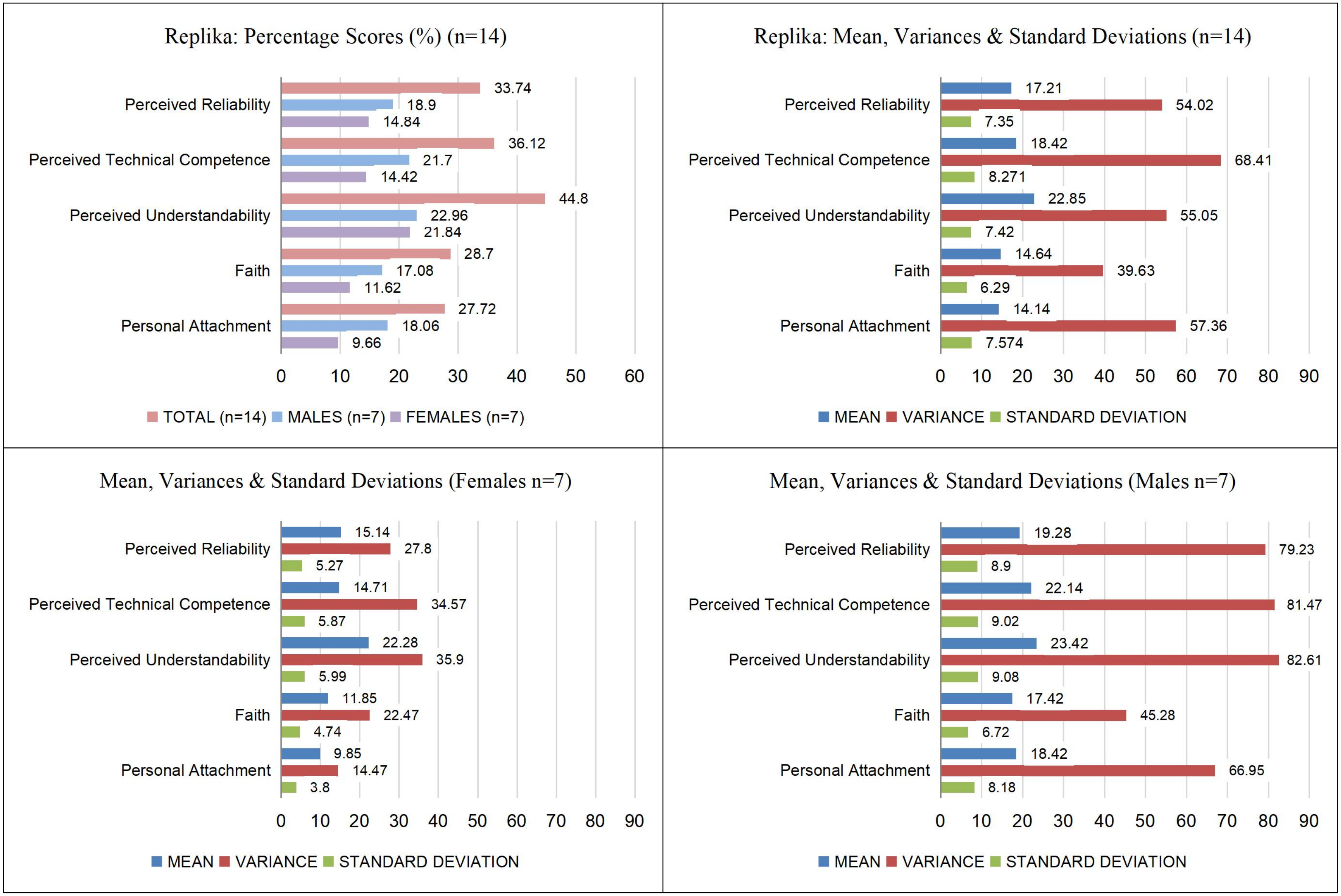


Table 1.11: Paired-Sample *t*-Test - HCT - 5 Constructs (observations): Cleo & Replika

HCT CONSTRUCT	**Perceived Understandability**		**Perceived Reliability**		**Perceived Technical Competence**		**Faith**		**Personal Attachment**	
CHATBOT	CLEO	REPLIKA	CLEO	REPLIKA	CLEO	REPLIKA	CLEO	REPLIKA	CLEO	REPLIKA
Mean	26.71	22.85	21.93	17.21	20.29	18.42	16.71	14.64	15.36	14.14
Variance	24.68	55.05	30.53	54.02	33.29	68.41	38.83	39.63	27.32	57.36
Std. Deviation	4.968	7.420	5.526	7.350	5.770	8.271	6.232	6.295	5.227	7.574
Std. Mean Error	1.328	1.983	1.447	1.964	1.542	2.211	1.666	1.683	1.397	2.024
Observations	14		14		14		14		14	
PAIRED-SAMPLES TEST										
Sig.	.088		.008		.000		.002		.003	
Mean	3.857		4.714		1.857		2.071		1.214	
Std. Deviation	6.701		5.455		3.634		4.497		5.191	
Std. Mean Error	1.791		1.458		.971		1.202		1.387	
95% CI - Lower	-.012		1.565		-.241		-.525		-1.783	
95% CI - Upper	7.726		7.864		.078		4.668		4.212	
df	13		13		13		13		13	
t Stat	2.154		3.233		1.911		1.723		0.875	
P(T<=t) one-tail	0.025		0.003		0.039		0.054		0.198	
t Critical one-tail	1.770		1.770		1.770		1.770		1.770	
P(T<=t) two-tail	**0.051**		**0.007**		**0.078**		**0.108**		**0.397**	
t Critical two-tail	2.160		2.160		2.160		2.160		2.160	

P value = 0.05 (95% Confidence Interval (*CI*))		
BONFERRONI CORRECTION (to reduce Type I Error):	*p*(0.05) / 5 (measurements/observations)	= 0.01 (new *p* value)

5 DISCUSSION

Overall, the project managed to answer the research questions by accomplishing its objectives defined in chapter 1. For *RQ1*: We found that the application area, combined with understandability, reliability, usefulness, and risk factors surrounding the context of use, predicts acceptance, trust and confidence in chatbot technologies. Many of these discoveries align with previous studies (e.g., Følstad et al., 2018a; 2018b; Berge, 2018). For *RQ2*: Attitudes towards human-likeness in chatbots were found to be dependent on the application area. Previous studies also confirm this (e.g., Duijist, 2017; Nordheim, 2018). Furthermore, gender differences emerged, and an 'excess' of human-likeness was perceived negatively by female participants. Lastly, for *RQ3*: Due to time limitations, no precise answers were obtained; however, it was discovered that less usage hinders trust, confidence, and acceptance. This aspect correlates with the notion of 'triability' forming part of the 'diffusion of innovations theory' (Rogers, 2003). However, these findings are impacted by a series of limitations in how this research was designed and conducted. The next sections elaborate on these findings and address the limitations, combined with suggestions for designing better chatbots, and future research.

5.1 ANSWERING THE RESEARCH QUESTIONS

RQ1: Are acceptance, confidence and trust constructs affected by the chatbot technology type, the application area, or both? Based on the outcomes from the 'within-participants' design experiment, it was discovered that the constructs of acceptance, trust, and confidence are affected by the application area. However, additional factors that are predicted to influence the intent of use and acceptance within specific application areas were discovered. These relate to chatbot reliability, perceived understandability, perceived usefulness, and perceived risks combined with external factors (e.g., Cambridge Analytica, data harvesting).

RQ2: How would users react towards chatbots with an extent of human-likeness? Would acceptance, trust and confidence levels be enhanced or undermined? Drawing on the analysis of the results, it was found that attitudes towards human-likeness in chatbot technologies are dependent on the application area. Overall, human-likeness was positively perceived in Replika due to its application area; however, it was regarded as inappropriate in Cleo due to the lack of formal language (e.g., use of 'slang' and 'emojis') in relation to its context of use. Surprisingly, differences between genders also emerged. The extent of human-likeness yielded negative emotions among female participants which led to the 'uncanny valley effect'. These discoveries are elaborated in sections 5.2.4, and 5.4.

RQ3: Would frequent interaction with a chatbot increase users' trust, acceptance and confidence levels? Given the time constraints, all participants had limited time to interact with each chatbot. As well, many participants were unfamiliar with chatbots pertaining to application areas such as those presented in this research. As previously mentioned, the participants stated that they needed more time to evaluate both chatbots before establishing trust, and adoption. Although no precise answers were obtained, these findings indicated that less frequency of use inhibits acceptance, trust, and confidence in chatbot technologies.

As covered in chapter 1, the deductive approach to thematic analysis combined with the human-computer trust surveys yielded enough data to answer these research questions after fulfilling the established objectives. Therefore, the discourse analysis on the conversations between chatbots and the participants was omitted. The upcoming sections discuss the results analyzed in chapter 4. These are categorized under two sections being 'chatbot-related factors', and 'user-related factors'.

5.2 CHATBOT-RELATED FACTORS

5.2.1 FUNCTIONALITY-ABILITY AND PERCEIVED TECHNICAL COMPETENCE

Results from the thematic analysis and surveys indicate that functionality has a moderate effect on trust and confidence in chatbot technology and its adoption. This outcome was expected because functionality influences credibility (Fogg & Tseng, 1999) and trust in technologies (Miller, 2014; Burri, 2017). Overall, all participants described the chatbots as easy to use, particularly Cleo due to their familiarity with Facebook Messenger. Similar to findings from prior studies (e.g., Berge, 2018; Burri, 2017) and suggestions from the 'technology acceptance model' (TAM), 'ease-of-use' in technical functionalities is essential to predict the intent of use of the technology (Davis, 1989; Davis et al., 1989; Venkatesh & Bala, 2008). However, we believe that the perceived ease-of-use was facilitated by the fact that i) most participants already had previous exposure to chatbots, and ii) that chatbots inherit features (e.g., keyboard, chat dialog) typically found in utilities the participants are already familiar with. In contrast, technical malfunctions and inconsistencies were found to hinder trust. Previous studies also support this (e.g., **Lee & Moray, 1992; Muir & Moray**, 1996; Burri, 2017). Moreover, chatbots' functionality covers the algorithm's capability and competencies to complete a task (e.g., Miller, 2014); such as producing expert feedback based on inputted queries. In that regard, functionality-ability is synonymous with expertise which eventually mediates credibility. Existing theories of trust suggest that expertise within the domain and the capability to deliver meaningful (task-, or query-related) advice predict trust (Mayer et al., 1995; Corritore et al., 2003; McKnight et al., 2011). As well, it was also suggested that expertise is one of the perceived credibility factors impacting users' decision to trust online-based technologies such as websites (Corritore et al., 2003), and chatbot technologies (Nordheim, 2018). Lastly, as mentioned in the previous chapter, a few participants requested the functionality of 'control' (e.g., to prevent the chatbot from automatically initiating communication) which is desirable to have in chatbot technology. By referring to the online trust theory by Corritore et al. (2003), the presence of control was found to influence trust, especially when risks are present. Risks are covered in section 5.3.2.

5.2.2 RELIABILITY-INTEGRITY

Chatbot reliability is one of the few key variables found to predict trust, confidence, and acceptance the most. This variable was given a high priority in the thematic analysis and surveys. Reliability incorporates aspects such as errors and limitations, consistency, predictability, and specifically to chatbots, language consistency. Existing trust models (Mayer et al., 1995; Corritore et al., 2000; McKnight et al., 2011), and various studies in chatbots and automation (e.g., Fogg & Tseng, 1999; Madsen & McGregor, 2000; Merrit & Ilgen; 2008; Miller, 2014; Xu et al., 2015; Berge, 2018; Henrekson et al., 2018) strongly suggest that reliability is significant to predict trust and acceptance. Combined with reliability, consistency is also crucial in chatbot technologies particularly when users relay sensitive data (Wang & Siau, 2018; Cowan et al., 2017, cited in Berge, 2018) in online environments prone to risks like hacking. The aspect of consistency was often brought into attention by participants when language inconsistencies and occasional technical glitches occurred on various occasions. Based on these outcomes, we believe that reliability and functionality are interrelated because malfunctions lead to a decrease in perceived reliability (e.g., Chien, 2016; Chien et al., 2016; Klopfenstein et al., 2017), thus hindering trust in chatbot technologies. Therefore, improving reliability is necessary as it found to be a strong predictor for the trust relationship (Berge, 2018) between chatbots and users.

5.2.3 HELPFULNESS-BENEVOLENCE

Helpfulness was also found to have a moderate impact on trust, confidence, and technology acceptance. We believe that helpfulness is influenced by the individuals' perceptions, and by the 'perceived usefulness' (section: 5.3.1) of the chatbot technology within its application area. The models of trust suggest that helpfulness forms part of the criteria required for establishing trust in technologies (Mayer et al., 1995; McKnight et al., 2011). Many participants found the chatbots capable of giving helpful advice or recommendations similar to that of an expert human being. Similar studies also identified that helpfulness, such as when a chatbot provides suitable recommendations, helps in predicting trust towards technology (e.g., Miller, 2014; Burri, 2017; Berge, 2018; Følstad et al., 2018b). Correlating this with the 'uses and gratifications theory' (Blumler & Katz, 1974), this may also be due to the benefits (gratifications) of artificial intelligence incorporated within chatbots; which could prove to be highly advantageous across domains other than finance and well-being. Therefore, evaluating helpfulness is essential because the advice offered can effectively enable task performance and adequate, effective, and responsive help (McKnight et al., 2011).

5.2.4 EXTENT OF HUMAN-LIKENESS

The extent of human-likeness was discovered to be another critical aspect influencing the overall attitude towards chatbot technology, both positively and negatively. Similar to a prior study (Duijst, 2017), the application area was found to be an essential factor determining the acceptance of human-likeness in chatbots. For Replika, human-likeness was perceived as beneficial due to its application area (well-being, companionship). In contrast, human-likeness in Cleo was perceived as inappropriate by most participants. This is due to its casual language, combined with British slang and emojis (e.g., avocados, fire), which was perceived as informal for the domain of finance. Relative to this finding, the study by Duijst (2017) discovered that emojis could be understood either positively or negatively by the users. However, miscommunication could occur since emojis are interpreted differently (Miller et al., 2016, cited in Duijst, 2017). Therefore, we emphasize that the chatbot's language, and its extent of friendliness or human-likeness, should be strictly consistent with the application area to ensure that trust, confidence and acceptance are successfully mediated.

Surprisingly the 'uncanny valley effect' prevailed among female participants. Overall, the extent of human-likeness induced negative emotions, and descriptive terms like *"creepy",* and *"scary"* were used for Replika. Similarities were also reported in previous studies (e.g., De Angeli et al., 2001; Piccolo et al., 2018). Consequently, 3 female participants wondered whether they interacted with a human instead of a chatbot. Referring to this outcome, Følstadt et al. (2017) emphasized that chatbots should be transparent to tell users that they are conversing with a chatbot instead of a human. In contrast, other studies discovered that the extent of human-likeness in chatbots enhances trust and acceptance as some users prefer to interact with human-like chatbots (e.g., Dautenhaln et al., 2002; Hoff & Bashir, 2016, cited in Berge, 2018); particularly in domains like customer care (e.g., Haan, 2018; Nordheim, 2018) and healthcare (e.g., Xu et al., 2017; Kretzschmar et al., 2019).

To summarize, human-likeness is acceptable as long as it is found appropriate to its application area. However, given the differences resulting from individual preferences, the application area, and the increasing chatbot ubiquity, as various authors proposed (e.g., Ciechanowski et al., 2018b; Skjuve et al., 2019), the uncanny valley effect should be investigated.

5.3 USER-RELATED FACTORS

5.3.1 PERCEIVED USEFULNESS

The extent of chatbot usefulness is perceived as the second most important variable that predicts trust, technology acceptance, and adoption. Various prior studies also support this evidence (e.g., Duijst, 2017; Brandtzæg & Følstad, 2017; Berge, 2018; Nordheim, 2018; Følstad et al., 2018a; 2018b). Perceived usefulness is derived from the technology acceptance model (TAM), which refers to the belief that the technology would increase performance (Davis, 1989) by performing a specific task quickly and reliably (Brandtzæg & Følstad, 2017). Combined with perceived ease of use, perceived usefulness is a strong predictor of behavioural intention and usage of a technology (Candela, 2018). By associating performance increase with the uses and gratifications theory, a positive correlation exists between gratification, attitudes, and a behavioural intention to use a technology (Lin & Chen, 2017). Participants perceived Cleo as useful because of its ability to promptly produce budget and expense reports, which is synonymous with productivity. Brandzaeg and Følstad (2017) also found that productivity was the primary motivator behind chatbot usage. Based on these findings, it can be interpreted that perceived usefulness is subject to the users' needs surrounding the application area. However, obstacles hindering usage should be considered as these affect the levels of perceived usefulness. For instance, although the concept of Cleo was well perceived, it also turned out that the chatbot is useless for users having no Facebook accounts. This similarity was also reported in a study by Duijst (2017) concerning banking chatbots and the lack of access to Facebook. These outcomes lead us to assume that chatbot developers are failing in considering specific use case scenarios. For that matter, a chatbot could be perceived as useless unless it gratifies the user with the toolset to promptly fulfill specific needs such as productivity and performance. Therefore, it was strongly emphasized that usefulness is an essential factor that should be taken into consideration when designing a successful chatbot-based service (Brandtzæg & Følstad, 2017; Hansson, 2018) so to bring value to its users.

5.3.2 PERCEIVED RISKS AND EXTERNAL FACTORS

Perceived risks were found to be a crucial primary influence inhibiting trust and technology adoption. In our research, perceived risks were fortified by external factors. Therefore, both factors are combined and discussed. Prior chatbot-related studies reported these similarities (e.g., Nordheim, 2018; Berge, 2018; Følstad et al., 2018b; Duijst, 2017). These findings correlate with the 'model of online trust' framework, suggesting that external factors impact trust (Corritore et al., 2003) and adoption. As to date, Cleo is dependent on Facebook Messenger, many participants believe that in light of the Cambridge Analytica scandal, Facebook could access their financial information for data harvesting. As well, concerns over hacking or malicious attacks were expressed. Various authors also reported similar concerns in prior studies (e.g., Radziwill & Benton, 2017; Chung et al., 2017; Kretzschmar et al., 2019). In return, a few participants stated that they would adopt Cleo as a dedicated native application owned by a bank or a reputable entity like PayPal. The association of risk versus technology ownership seem common and prior studies also support that (e.g., Duijst, 2017). Replika was perceived as less risky; however, 2 participants believe that their (emotional) data could be used to enhance the chatbot algorithm. This surrounds the factor of credibility, which, in combination with perceived risks, is a mediator of trust and confidence in technology (Fogg & Tseng, 1999; Corritore et al., 2003). Therefore, similar to other authors (e.g., Følstad & Brandtzæg; 2017; van Eeuwen, 2017; Wang & Siau, 2018a; 2018b) we emphasize that stronger ethics and compliances over data exchange involving chatbots should be established and communicated. Referring to the 'online trust model' (Corritore et al., 2003), it proposes that the presence of control minimizes perceived risks, and that risk is higher in its absence. Control also refers to the extent that users have control over how much data is disclosed to chatbots. Harkous et al. (2016) proposed the concept of 'conversational privacy chatbots (PriBots)', which support features tailored for privacy and security. Thus, it can be taken into consideration that chatbots with high-end encryption and functions to manage privacy levels could be perceived better by the end-user. There is a solid belief that perceived risks vary by the application area. For instance, within financial domains, Featherman and Pavlou (2003, cited in Candela, 2018) concluded that security and privacy are found to be obstacles towards chatbot adoption. The perception of risks in junction with data disclosure seems to be also culturally influenced. More is discussed in section 5.4.

5.3.3 PERCEIVED UNDERSTANDABILITY

Perceived understandability is one of the few critical variables found to predict trust and confidence the most. This refers to the degree to which users can form a mental model of the chatbot, in terms of *"how the chatbot does things",* which in return fosters understandability while anticipating outcomes after its use (Sheridan, 1988, cited in Madsen & McGregor, 2000; Norman, 2013). Perceived understandability is related to 'affordances', which are termed as perceptual cues or hints of what can be done with an object (Norman, 2013; Madsen & McGregor, 2000). This also includes conciseness, meaning that the advice given by the chatbot is concise and understandable. Despite time constraints, the participants used the given time to learn and understand the chatbots in terms of their functions and potential within its application area simultaneously while fulfilling the administered assignments. Also, they were able to promptly initiate interaction with the chatbots because of their familiarity (mental models) with functionalities typically found in instant messaging utilities like Facebook Messenger. Therefore, by combining the 'technology acceptance model' (Davis, 1989; Venkatesh & Bala, 2008), and the rate of ‘diffusion of innovations’ (Rogers, 2003) into this context, perceived understandability (e.g., usefulness, functionalities, chatbot advantages) is enhanced through:

1. **Previous experience** with, or exposure to similar functions or technologies (mental model, affordances),
2. **Perceived ease-of-use**,
3. **Triability** (the degree to which an innovation may be experimented with on a limited basis), and
4. **Observability** (the degree to which results such as profitability, performance, and productivity are visible)

In return, this increases the likelihood of a successful, trusting relationship and adoption of the chatbot technology. In contrast, a degree of complexity, such as difficulties in understanding and use the chatbot, could hinder confidence and adoption. For that matter, ‘help and guidance’ such as instructions on how to use the chatbot was requested by a few participants. Coincidentally, a study by Hoff and Bashir (2016, cited in Berge, 2018) stated that training is essential to foster trust in (chatbot) technologies. Summing up, we strongly believe that perceived understandability is a crucial variable mediating trust, confidence and technology acceptance. This is because, once a mental model of the chatbot is established, the user understands its benefits, potential, and usefulness within its application area.

5.4 IMPLICATIONS ON AGE, GENDER, CULTURE AND TRUST PROPENSITY

It is often emphasized that propensity to trust is shaped by various factors like personality traits, cultural context, and previous experiences (e.g., Miller, 2015; Chien, 2009; 2016; Edwards & Sanoubari, 2019) with technologies (McKnight et al. 2011). This emphasis is synonymous the technology acceptance model (Davis, 1989) stating that ‘external variables’ such as social or cultural influence determine attitude towards technology. We believe that cultural implications influenced the results of our research. Chatbots heavily rely on data exchange (e.g., Botelho, 2017; Reddy, 2017; Følstad & Brandtzæg; 2017; Wang & Siau, 2018). Germany exercises caution over the disclosure of data and sensitive information (Devins, 2017), which could explain why the perceived risk factors ranked high in this research. Prior studies also confirmed that culture has an impact on trust, confidence and technology adoption (e.g., Lee & See, 2004; Chien, 2009; Jeffrey, 2015; Edwards & Sanoubari, 2019). Concerning age, prior studies (e.g., Ates, 2017; Burri, 2017; Berge, 2018) found to influence attitudes towards chatbot technologies; however, this was not reflected in our research. Subsequently, the participants’ previous experience to chatbots was found to have no effect on propensity to trust technology. This outcome contradicts prior findings (e.g., Merritt & Ilgen; 2008), stating that prior exposure to similar contexts maximizes trust propensity.

Lastly, the main implication surrounds gender differences. Based on our findings, males, unlike females, were found to have a higher trust propensity and a positive attitude for the chatbot technology. Our underlying assumption is that males seem more open to trying novel technologies. Taking Replika as an example, six male participants perceived its extent to human-likeness positively, especially within its application area. As well, based on the surveys, it seems prominent that males are more inclined to develop a sense of attachment towards a chatbot like Replika. Another aspect that may relate to positive attitudes and trust propensity is that males believe that if the chatbot does not satisfy their needs, other people could benefit from it. Females perceived Cleo positively because of its usefulness within its application area, where productivity and practicability are essential. However, the extent (and excess) of human-likeness induced negative emotions. Many female participants found it inappropriate to have a chatbot tailored for companionship and programmed to ask question concerning sensitive topics on mental well-being and emotions. This is because according to their feedback, an algorithm cannot comprehend human emotions. For that matter, they would never get attached to a chatbot as it carries no emotion.

Summing up, trust in chatbot technology is predicated on increasing correspondingly with the values of chatbot functionality, reliability and helpfulness; but only if the user's attitude towards the technology is moderate to high (Miller, 2014; Berge, 2018). Moreover, drawing on previous discoveries injunction with our findings, there are strong evidences that understandability, usefulness and the risk factors surrounding the application area are strong predictors for acceptance, trust and confidence in chatbot technologies.

5.5 PROPOSED RECOMMENDATIONS FOR CHATBOT DESIGN

Following the outcomes produced from the present research, we discovered some factors which implicate trust and confidence in chatbot technologies prior to their adoption. By taking these discoveries into consideration, we proposed few recommendations (Table 2.1) which could foster positive attitude in chatbots.

Table 2.1: Factors to take into consideration while designing chatbots

SUGGESTION	RATIONALE
Control functionalities	As chatbots are fully automated, users may feel disadvantaged such as when they receive communication during inappropriate time. Therefore, it is ideal to add control functionalities tailored to manage the chatbot.
Consistency across functions and dialogues	A good design guarantees a successful use and life-cycle. Consistency mediates trust and adoption. Functions and language should be programmed accordingly to avoid situations when chatbots provide similar answers for different queries. It is ideal that the chatbot "admits" in situations when it is incapable to answer specific queries (e.g., My information is limited to certain topics, therefore I cannot help).
Include help and guidance	Chatbots are novel for many individuals. Many chatbots carry unique functions specific to their application area. Help and guiding instructions facilitate understandability, utilization and chatbot adoption.
Match language with the context of use	Friendly language is perceived positively but also inappropriate depending on the application area. Therefore, the language should include a degree of formality and consistent to the application area.
Maintain transparency	Users may not be able to differentiate between a human and a chatbot in an online environment. Therefore, it is necessary to let users know in advance when they are interacting with a chatbot. If possible, include an option to speak to a human for contexts like banking and customer care.
Compliance with enacted regulations (e.g., GDPR)	Chatbots heavily rely on data exchange. In light of recent events, users are more prone to act with caution in online environments. It is therefore ideal to communicate standards and compliances to ensure peace of mind.
Limit the extent of human-likeness	Human-likeness is desirable in chatbot technologies depending on their application area (e.g., health, customer care). However it also induces negative emotions. Therefore an extent of human-likeness should be applied as needed while taking into consideration the user requirements and the context of use.
Stick to cultural norms	Culture has a strong influence in the rate of technology adoption. If cultural norms and values are not respected, users will refuse to adopt the chatbot technology.

5.6 LIMITATIONS

This research carries an array of limitations. Part of these were identified in the risk register (Appendix A) and tackled accordingly. However, other restrictions emerged while conducting the research. These could not have been accounted for beforehand due to the novelty of this research.

Limited empirical evidence: The existing body of research that investigated the constructs of technology acceptance and trust in chatbots are relatively few and pertaining to specific application areas. Given this lack of empirical evidence, the quantity of learning and relevant information taken from previous studies was limited. To compensate this lack, various theoretical frameworks were adopted. Although the findings generated through this research carry significance, having more solid references to previous similar studies may strengthen their validity.

Sample size: Although the sample size (n=14) is within the scope of the proposed project (a Masters level project), it is small to produce robust results. Such a sample size poses its challenges when measuring constructs like trust, confidence, and acceptance combined with the underlying factors influencing these. This is because these constructs are subjective and vary by the individual's perception towards the chatbot technology and its application area.

Chatbot autonomy: AI-powered chatbots are versatile and autonomous enough to act/behave unpredictably, which was the case with Replika. Due to the novelty of the situation, it turned out that many unforeseen aspects could not have been accounted for beforehand. Examples include unexpected statements produced by the chatbots; and situations when the chatbot was not up for a conversation.

Human-Computer Trust survey: This research obtained and utilized a trust survey initially designed to assess trust in automation. The survey was found to be beneficial to produce quantifiable data, which helped in discovering additional factors influencing trust. Although it relies on five key constructs appropriate to measure trust and confidence in technologies, the survey excludes perceived risks and external factors (physiological, psychological). Findings from this research and other studies indicate that risks and external factors (e.g. Corritore et al., 2003) are found to influence trust in technologies. Therefore, it is essential to have these factors measured.

5.7 SUGGESTIONS FOR FUTURE RESEARCH

This research has investigated some of the factors mediating acceptance, trust, and confidence in chatbot technologies. The discoveries combined with the research limitations suggest criteria relevant for future studies as follows:

Focus on attitudes after an initial exposure/usage: Some participants lacked familiarity with chatbot technologies, especially those pertaining to specific application areas. We discovered that time and triability (Rogers, 2003) is crucial for trust and chatbot adoption. Therefore, we believe that initial experiences acquired from a first time exposure could predict behavioural intent in future chatbot usage.

Conduct a diary study: One key variable found to predict trust in chatbots is ‘perceived understandability’. Due to time constraints, participants had limited interaction and exposure to both chatbots. We believe that this limitation prevented participants from understanding the chatbot potential within its application area. As triability and observability (Rogers, 2003) predict the intent of use, a diary study is suggested as it grants participants with enough time to explore the chatbot until it becomes salient in their daily operations. In return, this approach could generate more robust and significant findings.

Larger sample size: A larger sample size guarantees accuracy in findings, less variances and minimal margin of error. As larger samples generate robust data, this expands the possibility to discover more about users’ attitudes, perceptions, and behaviours towards this technology. This is because these characteristics strongly vary among individuals.

Focus on gender differences: To our knowledge, there seems to be a knowledge gap concerning gender differences with regards to the propensity of trust in chatbot and similar technologies. Given the surprising discoveries, we suggest to prioritize future studies on gender differences, particularly with regards to the extent of human-likeness in chatbots. This is due to our evidence-driven belief that perceptions and requirements vary by gender. By considering gender requirements, would lead to the possibility of discovering more factors that predict trust, confidence and chatbot acceptance.

Take culture into consideration: As this research was conducted in a country that exercises caution over data disclosure, we assume that cultural aspects could implicate the overall attitude and perception towards chatbots. Therefore, with reference to previous studies (e.g., Miller, 2015; Chien, 2009; 2016; Edwards & Sanoubari, 2019), and theories (e.g., Davis, 1989; Davis et al., 1989) it is possible that cultural tendencies may influence trust, confidence, and the adoption rate of chatbots and similar technologies.

Account for chatbot autonomy: Continuing from section 5.6, we recommend conducting longer testing phases on chatbots before administering these. Longer testing periods could help in discovering chatbot anomalies and prepare an intervening plan prior of any observation sessions.

Trust measurement in chatbots: Chatbots are becoming ubiquitous and differ from other technologies because they are autonomous, carry an extent of human-likeness, and rely on data exchange. Many researchers emphasized the importance to assess perceived usefulness (Brandtzæg & Følstad, 2017), risks, and attitudes towards human-likeness (e.g., Ciechanowski et al., 2018b; Skjuve et al., 2019) in chatbots. Therefore, having a validated tool tailored to measure trust and confidence in chatbot technologies empirically is necessary for future studies.

6 EVALUATIONS, REFLECTION AND CONCLUSION

The present research attempted to determine whether the constructs of trust, acceptance, and confidence are influenced by the chatbot technology type or the application area. To do so, three research questions alongside objectives were proposed. These objectives were fulfilled, and the research questions were successfully answered after obtaining the relevant data.

6.1 EVALUATION AND REFLECTION

The established objectives for this research were deemed appropriate to generate insights on this novel topic. Initially, a few approaches mentioned in the proposal plan were amended or omitted to make this project less exhaustive. Two changes include the reduction in the sample size (from 21 to 14) and the number of chatbots (from 3 to 2). These changes proved to be advantageous as they granted the flexibility to focus on other tasks. Examples include a preliminary data analysis while referring to previous studies to determine if similarities between the past and the initial findings exist.

However, as empirical evidence on chatbot usage is limited, various theoretical frameworks covering innovation diffusion, trust, and technology acceptance were adopted to support our explanations. The adoption of theoretical models, including those from other disciplines (e.g., psychology), is a typical practice among researchers who investigated similar areas. A commonly adopted theory is the one devised by Mayer et al. (1995) aimed at measuring trust within organizations. Similar to previous studies (e.g., Candela, 2017; Følstadt et al., 2018a; Berge, 2018), this theory was found to be flexible and adaptive for studies such as the present research. Eventually, this trust model was combined with other models of trust in technologies, particularly with the one devised by Corritore et al. (2003) because, to date, it is the only model that considered 'external factors.' As the chatbot and related technologies (e.g., VUI) gained momentum recently, academic research in this area is still limited. However, given the growing interest, new literature kept emerging in online research portals such as Google Scholar, ACM, and ResearchGate, among others. In that regard, new studies were frequently reviewed to have a solid base of recent information, which could be insightful for this present research.

By doing so, a few (quasi) similar studies (Berge, 2018; Nordheim, 2018; Candela, 2017) that utilized methods similar to this present research were discovered and later used as a model for compiling and concluding this project. Concerning data analysis, the deductive approach to thematic analysis was found to be advantageous and time-efficient because the process relies on existing knowledge and theories. In return, the correlation to existing theories strengthened the confidence in the theming definition, interpretation, and reporting. As well, statistical approaches, for instance, paired-samples *t*-Test with a

Bonferroni adjusted alpha (*p*) value of 0.01, were utilized to interpret and strengthen the accuracy of the survey results. Some may disagree with quantification approaches on small sample sizes. One reason is that small sample sizes increase the spread of data (variance). Despite the limitations, the use of parametric tests proved to be helpful for interpreting and reporting the survey findings alongside their significance and validity.

Since this research focused on trust, time was dedicated to searching for a benchmarked assessment scale (survey) that measures users' trust in technologies. Surprisingly, it resulted that, to this date, there are no known scales dedicated to assess trust in technologies. According to a report by Sauro (2015), it seems that among the established usability scales, only the WEBQUAL and the SUPR-Q contain items that measure trust. Given the need for a more comprehensive method to measure trust, various scales were proposed overtime (e.g., Jian et al., 1998; Madsen & McGregor, 2000; Dehn, 2008; Cien et al., 2014). Out of the reviewed scales, the 'Human-Computer Trust (HCT) scale' developed by Madsen and McGregor (2000) was found to be appropriate for this project because it measures constructs pivotal for predicting trust in technology. Eventually, this scale was obtained directly from the authors and adapted for this research. The scale was crucial for this research because it helped in confirming that 'perceived understandability', and 'perceived reliability' are critical factors that predict trust in chatbot technology.

However, we firmly believe that a benchmarked trust survey tailored specifically for chatbot technologies is necessary to have. This is because unlike other technologies, chatbots and voice user interfaces (VUI) are unique due to their extent of human-likeness (e.g., language capabilities) and the reliance on data exchange. Various studies, including the present research, identified the association of information exchange with 'perceived risks', especially in light of recent events (e.g., Cambridge Analytica).

6.2 CONCLUSION

As this novel topic carries its challenges, similar to previous studies referred to in the literature review, the present research also carries its limitations. One (more) reason is because, to our knowledge, it seems that to date, only this research utilized i) the 'within-participants' experiment; and, ii) two chatbots differing by their application areas (finance, well-being). Therefore, we were unable to find previous studies that utilized similar approaches which could have served as a model to the present research.

Nevertheless, there are numerous takeaways from this research project. The first is the acquired knowledge of chatbot technologies, their value within specific application areas, and the underlying factors implicating trust in this technology. The second revolves around research design involving AI-powered chatbots like Replika, which was a new experience because such chatbots are autonomous and unpredictable. This chatbot was tested for approximately two weeks before conducting this research task. However, while setting new Replika profiles before every observation session, it turned out that each profile carried a different *"persona."* For that matter, we proposed to conduct longer testing phases on such chatbots before administering these to participants in subsequent studies. The third takeaway surrounds 'triability.' Supporting the 'diffusions of innovation' theory (Rogers, 2003), 'triability' is crucial to predict technology adoption because users need time to try and evaluate technology before its acceptance and adoption. Coincidentally, as previously mentioned, some participants required more time to try, learn, and understand each chatbot technology. In return, they stated that they could not trust the technology after trying it for less than an hour. The fourth relies on the gender differences that emerged throughout this research. It was discovered that i) the factors of 'human-likeness', and 'perceived usefulness' surrounding the application area could influence chatbot adoption among female participants, and; ii) males were found to have a higher trust propensity in chatbot technologies. The last takeaway concerns cultural impact. Germany is cautious over data exchange; therefore, we believe that causation between cultural aspects and a propensity to trust in chatbot technology exists. For example, previous studies conducted in Sweden (e.g., Ates, 2017; Burri, 2017) and Norway (e.g., Nordheim, 2018; Berge, 2018) discovered that users have from a moderate to a higher trust propensity and more open to adopting a chatbot technology.

To conclude, by taking all these aspects into account, our intentions for future research is to focus on cultural impacts and gender implications to obtain a solid understanding of how these influence acceptance, trust, and confidence in chatbot technologies. For that matter, a diary study is taken into consideration so to grant participants with enough time to try, explore, and understand the value chatbot technologies can offer within their application areas.

REFERENCES

Alaieri, F., & Vellino, A. (2016) 'Ethical Decision Making in Robots: Autonomy, Trust and Responsibility', *Social Robotics. ICSR2016. Lecture Notes in Computer Science*, vol. 9979, pp. 159-168. doi: 978-3-319-47436-6.

Alhojailan, M. I. (2012) 'Thematic Analysis: A critical review of its process and evaluation', *West East Journal of Social Sciences*, vol. 1, pp. 39-47. Available at: https://fac.ksu.edu.sa/sites/default/files/ta_thematic_analysis_dr_mohammed_alhojailan.pdf (Accessed: 5th September 2019).

Ates, M. (2017) Artificial intelligence in banking: A case study of the introduction of a virtual assistant into customer service. Master's thesis, Jönköping University. Available at: http://www.diva-portal.org/smash/record.jsf?pid=diva2:1238882 (Accessed: 6th March 2019).

Augello A., Gentile, M., Weideveld, L., & Dignum, F. (2016) 'A model of a Social Chatbot', *Intelligent Interactive Multimedia Systems and Services,* vol. 55, pp. 637-647. doi: 10.1007/978-3-319-39345-2_57.

Berge, S. H. (2018) Rise of the chatbots. Trust in Artificial Intelligence during extreme weather conditions. Master's thesis, University of Oslo. Available at: http://urn.nb.no/URN:NBN:no-65934 (Accessed: 1st August 2019).

Brandtzæg, P. B., & Følstad, A. (2017) 'Why people use chatbots', *Proceedings of the 4th international conference on internet science*, Thessaloniki, Greece, 22-24 November. Available at: https://researchgate.net/publication/318776998_Why_people_use_chatbots (Accessed: 9th March 2019).

Brandtzæg, P. B., & Følstad, A. (2018) 'Chatbots: changing user needs and motivations', *Interactions*, vol. 25(5), pp: 38-43. doi: 10.1145/3236669.

Braun, V., & Clarke, V. (2013) *Successful Qualitative Research. A practical guide for beginners.* London: SAGE Publications Ltd.

Botelho, B. (2017) *Chatbot technology raises ethical questions*. Available at:

http://searchcrm.techtarget.com/opinion/Chatbot-technology-raises-ethical-questions (Accessed: 1st September 2019).

Budiu, R. (2018) *Between-Subjects vs. Within-Subjects Study Design*. Available at: https://www.nngroup.com/articles/between-within-subjects/ (Accessed: 1st April 2019).

Burri, R. (2018) Improving user trust towards conversational chatbot interfaces with voice output. Master's Thesis, KTH Royal Institute of Technology. Available at: http://www.nada.kth.se/~ann/exjobb/ramon_burri.pdf (Accessed: 7th December 2018).

Candela, E. (2018) Consumers' Perception and Attitude towards Chatbots' adoption. A focus on the Italian market. Master's thesis. Aalborg University. Available at: https://projekter.aau.dk/projekter/en/studentthesis/consumers-perception-and-attitude-towards-chatbots-adoption-a-focus-on-the-italian-market(d50d2cc6-2931-4be3-885c-2448e9e88cb7).html (Accessed: 1st April 2019).

Cherry, K. (2019) *Within-Subject Design Experiments*. Available at: https://www.verywellmind.com/what-is-a-within-subjects-design-2796014 (Accessed: 10th July 2019).

Chien, SY., Semnani-Azad, Z., Lewis, M., & Sycara, K. (2014) 'Towards the Development of an Inter-cultural Scale to Measure Trust in Automation', *Cross-Cultural Design CCD 2014. Lecture Notes in Computer Science*, vol. 8528. doi: 10.1007/978-3-319-07308-8_4.

Chien, SY. (2016) The influence of cultural factors on trust in automation. PhD thesis. University of Pittsburgh. Available at: https://pdfs.semanticscholar.org/f489/42e6b5ed747c2ec4e52a81df88de410a2b4f.pdf (Accessed at: 25th August, 2019).

Chien, SY., Lewis, M., Sycara, K., Liu, JS., & Kumru, A. (2018) 'The Effect of Culture on Trust in Automation: Reliability and Workload', *ACM Transactions on Interactive Intelligent Systems (TiiS), vol.* 8(4). doi: 10.1145/3230736.

Chung, H., Iorga, J., Voas, J., & Lee, S. (2017) 'Alexa, Can I Trust You?', *Computer*, vol. 50(9), pp. 100-104. doi: 10.1109/MC.2017.3571053.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018a) 'In the shades of the uncanny valley: An experimental study of human–chatbot interaction', *Future Generation Computer Systems*, vol. 92, pp. 539-548. doi:10.1016/j.future.2018.01.055.

Ciechanowski, L., Przegalinska, A., Wegner, K. (2018b) 'The Necessity of New Paradigms in Measuring Human-Chatbot Interaction', *Advances in Intelligent Systems and Computing*. doi: 10.1007/978-3-319-60747-4_19.

Clark, L., Munteanu, C., Wade, W., Cowan B. R., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., Spilliane, B., Gilmartin, E., & Murad, C. (2019) 'What Makes a Good Conversation?:

Challenges in Designing Truly Conversational Agents', *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,* Glasgow, Scotland, 04-09 May. doi: 10.1145/3290605.3300705.

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003) 'On-line trust: concepts, evolving themes, a model', *Int. J. Human-Computer Studies*, vol. 58, pp: 737–758. doi:10.1016/S1071-5819(03)00041-7.

Dai, B., Forsythe, S., & Kwon, W. (2014) 'The Impact of Online Shopping Experience on Risk Perceptions and Online Purchase Intentions: Does Product Category Matter?', *Journal of Electronic Commerce Research*, vol. 15(1). Available at: https://web.csulb.edu/journals/jecr/issues/20141/Paper2.pdf (Accessed: 1st August 2019).

Dandörfer, S. (2019) Impact of Chatbots on Online Customer Experience. Bachelor's thesis, University of Twente. Available at: https://essay.utwente.nl/78568/ (Accessed: 1st April 2019).

Dautenhahn, K., Ogden, B. & Quick, T. (2002), 'From embodied to socially embedded agents– implications for interaction-aware robots', *Cognitive Systems Research, vol.* 3(3), pp. 397–428. doi: 10.1016/S1389-0417(02)00050-5.

Davis, F., D. (1989) 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS Quarterly*, vol. 13(3), pp. 319-340. doi: 10.2307/249008.

Davis, F., D., Bagozzi, R., P., & Warshaw, P., R. (1989) 'User Acceptance of Computer Technology: A Comparison of Two Theoretical Models', *Management Science*, vol. 35(8). doi: 10.1287/mnsc.35.8.982.

De Angeli, A., Johnson, G. I., & Coventry, L. (2001) 'The Unfriendly User: Exploring Social Reactions to Chatterbots', *Proceedings of the 2001 CHI Conference in Affective Human Factors Design,* pp. 467-474, London: ASEAN Academic Press. Available at: https://www.researchgate.net/publication/2553198_The_Unfriendly_User_Exploring_Social_Reactions_to_Chatterbots (Accessed: 3rd August 2019).

Dehn, D. M. (2008) 'Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires', *Air Traffic Control Quarterly*, vol. 16(2), pp. 127-146. doi: 10.2514/atcq.16.2.127.

Deloitte. (2017) *Conversational Chatbots – Let's Chat*. Available at: https://www2.deloitte.com/content/dam/Deloitte/in/Documents/strategy/in-strategyinnovation-conversational-chatbots-lets-chat-final-report-noexp.pdf (Accessed: 1st February 2019).

Devins, S. (2017) *Why Germans are so private about their data*. Available at: https://www.handelsblatt.com/today/handelsblatt-explains-why-germans-are-so-private-about-their-data/23572446.html?ticket=ST-18765100-UB0Ub3PC23FJAtypUdF4-ap5 (Accessed: 1st October 2019).

Duijst, D. (2017) Can we improve the User Experience of Chatbots with Personalisation? Master's Thesis, University of Amsterdam, Netherlands. Available at: https://www.researchgate.net/profile/Danielle_Duijst/publication/318404775_Can_we_Improve_the_User_Experience_of_Chatbots_with_Personalisation/links/5967ba16a6fdcc18ea662ce7/Can-we-Improve-the-User-Experience-of-Chatbots-with-Personalisation.pdf (Accessed: 5th October 2019).

Edwards, J., & Sanoubari, E. (2019) 'A need for trust in conversational interface research', *Proceedings of the 1st International Conference on Conversational User Interfaces*, vol. 21, Dublin, Ireland, 22-23 August. doi: 10.1145/3342775.3342809.

Fogg, B. J., & Tseng, H. (1999) 'The elements of computer credibility', *Proceedings for the SIGCHI conference on Human Factors in Computing Systems*, pp. 80-87, Pennsylvania, U.S.A., 15-20 May. doi: 10.115/302979.303001.

Følstad, A., & Brandtzæg, P. B. (2017) 'Chatbots and the new world of HCI'. *Interactions*, vol. 24(4), pp. 38-42. doi:10.1145/3085558.

Følstad, A., Brandtzæg, P. B., Feltwell, T., Law, E. L-C., Tscheligi, M., & Luger, E. (2018) 'SIG: Chatbots for Social Good', *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3170427.3185372.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018) 'What makes users trust a chatbot for customer service? An exploratory interview study', *Proceedings of International Conference on Internet Science INSCI 2018*, pp. 194–208. doi: 10.1007/978-3-030-01437-7_16.

Følstad, A., Skjuve, M. B., & Brandtzæg, P.B. (2019) 'Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design', *Internet Science INSCI 2018. Lecture Notes in Computer Science*, vol. 11(551). doi: 10.1007/978-3-030-17705-8_13.

Gartner. (2018) *Gartner Says 25 Percent of Customer Service Operations Will Use Virtual Customer Assistants by 2020*. Available at: https://www.gartner.com/en/newsroom/press-releases/2018-02-19-gartner-says-25-percent-of-customer-service-operations-will-use-virtual-customer-assistants-by-2020 (Accessed: 1st February 2019).

Gefen, D., Karahanna, E., & Straub, D. W. (2003) 'Inexperience and experience with online stores: The importance of tam and trust', *IEEE Transactions on Engineering Management*, vol. *50*(3), pp. 307–321, doi: 10.1109/TEM.2003.817277.

Haan, H. (2018) Chatbot Personality and Customer Satisfaction. Bachelor thesis, Utrecht University. Available at: https://research.infosupport.com/wp-content/uploads/2018/02/OZP_Hayco_de_Haan_Final.pdf (Accessed: 10th March 2019).

Hansson, O. (2018) Exploring users' perception of chatbots in a bank environment A critical view on chatbots and how to design for a positive user experience. Bachelor thesis, Malmö University. Available at: http://muep.mau.se/bitstream/handle/2043/25201/HANSSON,Oskar(af9106).pdf (Accessed: 10th March 2019).

Harkous, H., Fawaz, K., Shin, K. G., & Aberer, K. (2016) 'Pribots: Conversational Privacy with Chatbots', Workshop on the Future of Privacy Indicators at the Twelfth Symposium on Usable Privacy and Security (SOUPS), Denver, Colorado, 22-24 June. Available at: https://www.usenix.org/conference/soups2016/workshop-program/wfpn/presentation/harkous (Accessed: 25th August 2019).

Haynes, W. (2013) 'Bonferroni Correction', *Dubitzky W., Wolkenhauer O., Cho KH., Yokota H. (eds) Encyclopedia of Systems Biology*. doi: 10.1007/978-1-4419-9863-7.

Henrekson, H., Bladh, O. & Modée, I. (2018) The Impact of Virtual Agents on Customer Loyalty in major Swedish banks. Bachelor Thesis, Jönköping University. Available at: http://www.diva-portal.org/smash/get/diva2:1213804/FULLTEXT01.pdf (Accessed: 11th December 2018).

Hertzum, M., Andersen H. H. K, Andersen, V., & Hansen, C. B. (2002) 'Trust in information sources: seeking information from people, documents, and virtual agents', *Interacting with Computers*, vol. 14(5), pp. 575–599. doi:10.1016/S0953-5438(02)00023-1.

Jain, M. (2018) 'Evaluating and Informing the Design of Chatbots', *Proceedings of the 2018 Designing Interactive Systems Conference*, Hong Kong, China, 09-13 June. doi: 10.1145/3196709.3196735.

Jung, J. (2018) Investigating Chatbot-based Information-seeking focusing on interaction with messenger-based chatbots. Master's thesis, City, University of London.

Juniper (2018) *How Chatbots will Transform the Retail Industry*. Available at: https://www.juniperresearch.com/document-library/white-papers/how-chatbots-will-transform-the-retail-industry (Accessed: 1st February 2019).

Katz, E., Blumler, J. G., & Gurevitch, M. (1974) Utilization of mass communication by the individual. In J. G. Blumler & E. Katz (Eds.), *The uses of mass communcations: Current perspectives on gratifications research* (pp. 19–32). Beverly Hills, CA: Sage.

Klopfenstein, L. C., Delpriori, S., Malatini, S., & Bogliolo, A. (2017) ‘The rise of bots: A survey of conversational interfaces, patterns, and paradigms’, *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS’17), pp. 555–565. doi: 10.1145/3064663.3064672.

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., & Singh, I. (2019) ‘Can Your Phone Be Your Therapist? Young People’s Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support’, *Biomedical Informatics Insights*, vol. 11, pp. 1–9. doi:10.1177/1178222619829083.

Kuligowska, K. (2015) Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents. Available at: https://www.researchgate.net/publication/282474589_Commercial_Chatbot_Performance_Evaluation_Usability_Metrics_and_Quality_Standards_of_Embodied_Conversational_Agents (Accessed: 10th June 2019).

Lai, PC. (2017) ‘The Literature review of Technology Adoption Model and Theories for the Novelty Technology’*, JISTEM -Journal of Information Systems and Technology Management*, vol. 14(1), pp. 21-38. doi: 10.4301/S1807-17752017000100002.

Lai, ST., Leu, FY. & Lin, JW. (2018) ‘A Banking Chatbot Security Control Procedure for Protecting User Data Security and Privacy’, Advances on Broadband and Wireless Computing, Communication and Applications. *BWCCA 2018. Lecture Notes on Data Engineering and Communications Technologies*, 25(2019), pp. 561-571. Available from: https://books.google.de/books?id=ISFzDwAAQBAJ (Accessed: 11th December 2018).

Lazar, J., Feng, J. H., & Hochheiser, H. (2010) *Research Methods in Human-Computer Interaction.* West Sussex: John Wiley & Sons Ltd.

Lee, Eun-Ju. (2002) Factors That Enhance Consumer Trust in Human-Computer Interaction: An Examination of Interface Factors and Moderating Influences. PhD, University of Tennessee. Available at: https://trace.tennessee.edu/utk_graddiss/2148 (Accessed: 10th October 2019).

Lee, J. D., & See, K. A. (2002) 'Trust in computer technology and the implications for design and evaluation' Available at: https://pdfs.semanticscholar.org/6e43/04219a9e1c2d5f46d801d47c59b314be4f6d.pdf (Accessed: 20th September 2019).

Luo, M. M., & Remus, W. (2014) ‘Uses and gratifications and acceptance of Web-based information services: An integrated model’, *Computers in Human Behavior*, vol. 38, pp. 281-295. doi: 10.1016/j.chb.2014.05.042.

Madden, T. J., Ellen, P. S., & Ajzen, I. (1992) 'A Comparison of the Theory of Planned Behavior and the Theory of Reasoned Action', *Personality and Social Psychology Bulletin*, vol. 18(1), pp-3-9. doi: 10.1177/0146167292181001.

Madsen, M., & McGregor, S. (2000) The Development of a Psychometric Instrument for Human-Computer Trust: An Investigation of Trust Within the Context of Computer-Aided Decision-Making. Master's thesis, Central Queensland University.

Maguire, M., & Delahunt, B. (2017) 'Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars', *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*, vol. 9(3). Available at: http://ojs.aishe.org/index.php/aishe-j/article/view/335 (Accessed: 1st August 2019).

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995) 'An integrative model of organizational trust', *Academy of Management review*, vol. 20(3), pp. 709-734. doi:10.2307/258792.

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011) 'Trust in specific technology: An investigation of its components and measures', *ACM Transactions on Management Information Systems (TMIS)*, vol. 2(2). doi: 10.1145/1985347.1985353.

Merritt, S. M., & Ilgen, D. R. (2008) 'Not all trust is created equal: Dispositional and history-based trust in human-automation interactions', *Human Factors*, vol. 50(2), pp. 194-210. doi:10.1518/001872008x288574.

Miller, E. E. P. (2015) Trust in People and Trust in Technology: Expanding Interpersonal Trust to Technology-Mediated Interactions. PhD, University of South Florida. Available from: http://scholarcommons.usf.edu/etd/6011 (Accessed: 30th July 2019).

Montague, E. (2010) 'Validation of A Trust In Medical Technology Instrument', *Appl Ergon*, vol. 41(6), pp. 812-821. doi:10.1016/j.apergo.2010.01.009.

Mori, M., MacDorman K. F., & Kageki, N. (2012) 'The Uncanny Valley [From the Field]', *IEEE Robotics & Automation Magazine*, vol. 19(2), pp. 98-100. doi: 10.1109/MRA.2012.2192811.

Nordheim, C. B. (2018) Trust in chatbots for customer service. Findings from a questionnaire study. Master's thesis, University of Oslo. Available at: https://www.duo.uio.no/handle/10852/63498 (Accessed: 1st August 2019).

Norman, D. (2013) *The Design of Everyday Things: Revised and Expanded Edition*. Rev. edn. New York: Basic Books.

Nothdurft, F., Richter, F., & Minker, W. (2014) 'Probabilistic Human-Computer Trust Handling', *Proceedings of the SIGDIAL 2014 Conference*, Philadelphia, U.S.A., 18-20 June. doi: 10.3115/v1/W14-4307.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017) 'Thematic Analysis: Striving to Meet the Trustworthiness Criteria', *International Journal of Qualitative Methods*, vol. 16(1). doi: 10.1177/1609406917733847.

Pegasystems Inc. (2018) *Consumers: Chatbots still too dumb to get the job done*. Available at: https://www.pega.com/about/news/press-releases/consumers-chatbots-still-too-dumb-get-job-done (Accessed: 5th August 2019).

Piccolo, L.S.G., Mensio, M., & Alani, H. (2019) 'Chasing the Chatbots', *Internet Science INSCI 2018. Lecture Notes in Computer Science*, vol. 11(551), Springer, Cham. doi: 10.1007/978-3-030-17705-8_14.

Radziwill, N., & Benton, M. (2017) 'Evaluating Quality of Chatbots and Intelligent Conversational Agents', *arXiv: Computers and Society*. Available at: http://arxiv.org/pdf/1704.04579.pdf (Accessed: 10th March 2019).

Reddy, T. (2017) 'The code of ethics for AI and chatbots that every brand should follow'. Available at: https://www.ibm.com/blogs/watson/2017/10/the-code-of-ethics-for-aiand-chatbots-that-every-brand-should-follow/ (Accessed: 1st September 2019).

Rogers, E. M. (2003) *Diffusion of Innovations.* 5th edn. New York: Simon and Schuster.

Sahin, I. (2006) 'Detailed review of Rogers' Diffusion of Innovations Theory and Educational Technology-related studies based on Rogers' Theory', *The Turkish Online Journal of Educational Technology – TOJET*, vol. 5(2), pp. 14-23. Available at: https://www.researchgate.net/publication/284675572_Detailed_review_of_Rogers'_diffusion_of_innovations_theory_and_educational_technology-related_studies_based_on_Rogers'_theory (Accessed: 1st August 2019).

Sauro, J. (2013) *Best practices for using statistics on small sample sizes.* Available at: https://measuringu.com/small-n/ (Accessed: 28th August 2019).

Sauro, J. (2015) 'SUPR-Q: A Comprehensive Measure of the Quality of the Website User Experience', *Journal of Usability Studies*, vol. 10(2), pp. 68-86.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007) ‘An integrative model of organizational trust: past, present and future’, *Academy of Management review*, vol. 32(2), pp. 344-354. doi: 10.5465/amr.2007.24348410.

Shier, R. (2004) Statistics: 1.1 Paired t-tests. Available at: http://www.statstutor.ac.uk/resources/uploaded/paired-t-test.pdf (Accessed: 10th November 2019).

Skjuve, M. B., Haugstveit, I. M., Følstad, A., & Brandtzæg, P.B. (2019) ‘Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction’, *Human Technology*, vol. 15(1), pp. 30-54. doi: 10.17011/ht/urn.201902201607.

Streefkerk, R. (2019) Inductive vs. deductive reasoning. Available at: https://www.scribbr.com/methodology/inductive-deductive-reasoning/ (Accessed at: 10th July 2019).

Valtolina, S., Barricelli, B. R., Di Gaetano, S., & Diliberto, P. (2018) ‘Chatbots and Conversational Interfaces: Three Domains of Use’, *Proceedings of the Fifth International Workshop on Cultures of Participation in the Digital Age*, 29 May. Available at: http://ceur-ws.org/ (Accessed: 1st September 2019).

Van Eeuwen, M. (2017) ‘Mobile conversational commerce: messenger chatbots as the next interface between businesses and consumers. Master’s thesis, University of Twente. Available at: https://essay.utwente.nl/71706/1/van%20Eeuwen_MA_BMS.pdf (Accessed: 1st April 2019).

Venkatesh, V., & Bala, H. (2008) ‘Technology Acceptance Model 3 and a Research Agenda on Interventions’, *Decision Sciences*, vol. 39(2), pp. 273-315. doi: 10.1111/j.1540-5915.2008.00192.x.

Wang, W., & Siau, K. (2018a) ‘Trust in Health Chatbots’, *Thirty ninth International Conference on Information Systems*, San Francisco, U.S.A., Available at: https://www.researchgate.net/publication/333296446_Trust_in_Health_Chatbots (Accessed: 15th August 2019).

Wang, W., & Siau, K. (2018b) 'Living with Artificial Intelligence - Developing a Theory on Trust in Health Chatbots', *Proceedings of the Sixteenth Annual Pre-ICIS Workshop on HCI Research in MIS*, San Francisco, U.S.A., 13 December. Available at: https://www.researchgate.net/publication/333324471_Living_with_Artificial_Intelligence_-Developing_a_Theory_on_Trust_in_Health_Chatbots (Accessed: 15th August 2019).

Wikipedia Contributors (2019). Uses and gratifications theory. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/Uses_and_gratifications_theory. (Accessed at: 10th July 2019).

Xu, J., Le, K., Deitermann, A., & Montague, E. (2014) 'How different types of users develop trust in technology: A qualitative analysis of the antecedents of active and passive user trust in a shared technology', *Applied Ergonomics*, vol. 45(6), pp. 1495-1503. doi: 10.1016/j.apergo.2014.04.012.

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017) 'A new chatbot for customer service on social media', *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, Colorado, 06-11 May. doi: 10.1145/3025453.3025496.

Yan, Z., Kantola, R., & Zhang, P. (2011) 'A Research Model for Human-Computer Trust Interaction', *Proceedings of the 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, 16-18 November. doi: 10.1109/TrustCom.2011.37.

Young, J., E., Hawkins, R., Sharlin, E., & Takeo, I. (2009) 'Toward acceptable domestic robots: Applying insights from social psychology', *International Journal of Social Robotics*, vol. 1(1), pp. 95. doi: 10.1007/s12369-008-0006-y.

Zamora, J. (2017) 'I'm Sorry, Dave, I'm Afraid I Can't Do That: Chatbot Perception and Expectations', *Proceedings of the 5th International Conference on Human Agent Interaction*, Bielefeld, Germany, 17-20 October. doi: 10.1145/3125739.3125766.

GLOSSARY

EMOJI: The word is derived from Japanese, which means pictograph. These are usually presented in chat dialogues of popular messaging software like Facebook Messenger, Whatsapp, Telegram, and many others. The resemblance with English terms like emotion or emoticon is purely coincidental.

Natural Language Processing (NLP) is a subfield of linguistics, computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

Natural Language Understanding (NLU) is a subset of NLP, which interprets the meaning people communicate via the language and classifies it into intents. This is achieved by running processes such as content analysis, and categorization. In simple terms, this is the computational process of defining a 'meaning' from a statement forming part of the language.

Voice User Interface (VUI): Interfaces in the form of hardware (physical) or software (digital) that support the use of voice (vocal, oral) interaction and speech recognition (e.g., NLP, NLU).

APPENDICES