Evaluating User Trust and the Factors Affecting User Trust for the FarmBuddy Chatbot

Perez Eguiluz, Mario

Evaluating User Trust and the Factors Affecting User Trust for the FarmBuddy Chatbot

Mario Perez Eguiluz MP09112000@GMAIL.COM
Faculty of Humanities
Leiden University
Cleveringaplaats 1, 2311 BD, Leiden

Editor: Stephan Raaijmakers

Abstract

Background: Trust has been found to play a pivotal role in the way that interactions take place between users and chatbots, as well as significantly impacting technology adoption, continuance intention and user behavior. The evaluation of trust in relation to chatbots has been a subject of exploration in the existing literature. However, to my knowledge, there are currently no tools or instruments available, which are specifically designed for assessing user trust in chatbots used as information systems. Furthermore, gaining insight on the factors that affect user trust, and to what extent this occurs, can ultimately lead to improvements in the development or design of chatbots as information systems, such as FarmBuddy. Taking into account the novelty of the chatbot there are no previous studies assessing the perceived user trust and the factors affecting this trust towards FarmBuddy so it will also serve as a first evaluation towards its development.
Objective: Through this study we aim to develop a validated questionnaire that will not only assess the latent variable of user trust in chatbots as information system, but also provide an in-depth understanding corresponding to the factors affecting trust implemented in the survey. One of the objectives behind this study is for this validated questionnaire to be used by other researchers from the field and facilitating future research on this topic. In the literature review we delve into the importance of trust as a component of user experience as well as how this impacts the users' continuance intention towards chatbots. Finally, we will be using the validated questionnaire and semi-structured interviews in order to assess the factors influencing user trust in relation the FarmBuddy chatbot as well as evaluating user trust overall towards the aforementioned system. Via the use of a validated questionnaire and Partial Least Squares Structural Equation Modeling (PLS-SEM), this research seeks to enhance the validity and reliability of its findings, ultimately contributing to a more nuanced understanding of the dynamics of trust found in chatbot interactions.
Methods: The study distributed the questionnaire to FarmBuddy users and used the responses obtained to develop the following 11 latent constructs, via the use of a 7-point Likert scale; Familiarity, Propensity to Trust, Competence, Understandability, Appearance, Brand, Human-likeness, Likeability, Predictability, Transparency, and Risk. The process of validation is performed via PLS-PM leading to the final questionnaire displayed in *Appendix A*. The study uses PLS-SEM in order to analyze the data obtained regarding the factors affecting user trust. Furthermore, descriptive statistics are used to evaluate the overall user trust towards the FarmBuddy chatbot, and an ANOVA test in conducted to observe the significant differences on said trust based on the various demographic data obtained.
Results: The PLS-SEM model displayed the findings that Appearance (0.387), Risk

(0.335), Brand (0.308), Transparency (0.293), and Competence (0.265), resulted as being the most impactful factors affecting user trust for the FarmBuddy chatbot. Moreover, the descriptive statistics displayed a median and mean trust score of 4.800 and 4.514 respectively. Given that these values are above the 3.5 midpoint of the trust scale used, the participants or potential future users of the FarmBuddy chatbot generally deem the FarmBuddy chatbot as being trustworthy. The ANOVA showed no significant p-values for any demographic factor affecting user trust.
Discussion: The shift observed in the most influential factors affecting user trust in chatbots suggest a response to advancements in chatbot technology, since the increase in base functionality of the chatbots is accompanied with an increase in importance of what were previously seen as less relevant factors, according to the literature. Furthermore, perceived user trust towards the FarmBuddy chatbot is positive but shows room for improvement. The non significance of the demographic variables lead to a need for consistency in the chatbot across user groups. Moreover, user trust is observed to be an important component of user experience and continuance intention.

Keywords: trust, chatbot, FarmBook, factors affecting trust, questionnaire, information systems

Contents

1 Introduction

Through this study, we aim to address the existing gap concerning the evaluation of trust in chatbots used as a form of information or information retrieval system, as well as assessing the initial trust of potential future users of the FarmBuddy chatbot as a way of analysing the utility and effectiveness of the chatbot. FarmBuddy is a chatbot that is being developed and will be implemented in the EU-FarmBook project, in order to assist users in finding information related to the agriculture and forestry sectors. In order to do, we will be designing and validating a questionnaire to gather the necessary quantitative data to answer the aforementioned topics. Furthermore, we will be analyzing the way in which trust can be used as a measure of user experience and likelihood for continuance intention. To avoid confusion and provide context, this study defines continuance intention in regard to information systems as "an individual's intention to continue using an information system" (Bhattacherjee (2001, p.359), as cited in Yan et al., 2021), as well as the repeated use of that system.

First of all, we will be delving into the literature in order to provide some context towards the definition of trust, current positions on trust as a component of user experience in various contexts, the importance of trust in chatbots as information systems, the impact of trust on user engagement and continuance intention, and finally analysing the factors that affect user trust as a basis of the questionnaire design that will be performed. This review will be used to observe the way in which trust can be used as a measure of user experience and likelihood for continuance intention.

As mentioned, the questionnaire was designed on the thorough investigation of the literature through the use of a 7-point Likert scale, which was then presented to participants as a pilot test. The data obtained via this pilot test was statistically validated through the use of PLS-PM.

Furthermore, this validated questionnaire was later used in the final experiment to obtain the participants' perceived user trust in regards to the FarmBuddy chatbot as well as information regarding the factors affecting user trust. Through the use of PLS-SEM a model will be created to assess which of the factors obtained from the literature have the most influence on user trust towards the FarmBuddy chatbot. As well as this, descriptive statistics will be utilized to evaluate the perceived overall trust score of the participants in the chatbot, and an ANOVA test will be used to measure any statistical differences to this score based on demographics factors. Finally, these results will be discussed, considering the theoretical and practical implications that they entail.

Research question 1: What factors influence user trust towards the FarmBuddy chatbot, and to what extent do the users perceive the chatbot as being trustworthy?

Research question 2: How can trust be used as a measure of user experience and likelihood for continuance intention?

2 Literature Review

2.1 Introduction of theoretical background

The increasing prevalence of chatbots in various domains, such as healthcare and education (Alowais et al., 2023; Essel et al., 2022), highlights their important role in regard to facilitating interactions between users and information systems. Whether it be providing quick and easy access to information for college students, or assisting people with disabilities in their interaction with the system (Asfoura et al., 2023; Wijaya et al., 2020). Consequently, the importance of chatbot evaluation has seen a substantial increase in recent years, in order to ensure reliable results, as well as improving overall user experience, usability, and engagement. User trust as a factor for chatbot evaluation has also become more prevalent. However, to the best of my knowledge, at the time of writing this thesis this type of evaluation has not yet reached chatbots that have been designed as a form of information system. Through this literature review we will address the factors that have been found to influence trust, which will make up and form the basis of the FarmBuddy evaluation. Moreover, we will be delving into the literature in relation to the significance of trust in chatbots and information systems, trust as a component of user experience, and the relationship found between user trust and continuance intention towards the chatbot as a form of information system.

Trust in chatbots is a complex phenomenon which is influenced by various factors and plays a crucial role in regard to the optimization of users' interactions with chatbots. Possessing insights and understanding the concept of trust formation can be a powerful aid in the development and deployment of a more effective chatbot, which will build trust among its users. Even though trust in chatbots has been previously studied in order to optimize user experience and ensure the adoption of the chatbot, most of these studies have been related to chatbots in the field of customer service and not as information systems. In order to fill this gap, as well as relate the existing literature to the evaluation of the FarmBuddy chatbot, which was designed as a form of information or information retrieval system. The FarmBuddy chatbot is a knowledge-retrieval chatbot destined towards the agricultural and forestry domain that is currently under development. We will be showcasing various studies in order to gather insight on the multifaceted concept of trust in relation to chatbots, as well as chatbots as information systems. By understanding trust and the underlying mechanism of trust formation in human-computer interactions we can enhance the effectiveness and the acceptance of chatbots as a form of information system. The analysis of the impact of trust on continuance intention of chatbots as information systems will also lead to advancements of knowledge in the field, as well as improvements in chatbot design in order to optimize this process.

This review provides a rationale and background for selecting user trust as a form of evaluation or assessment for the FarmBuddy chatbot.

2.2 FarmBuddy chatbot as a type of information system

Throughout this section, I will be providing the reasoning behind classifying the FarmBuddy chatbot as a type of information system. According to Encyclopædia Britannica (Zwass, 2024) an information system can be defined as "an integrated set of components for collecting, storing, and processing data and for providing information, knowledge, and digital products". Additionally, Zwass (2024) describes the set of components that make up information systems as; "computer hardware and software, telecommunications, databases and data warehouses, human resources, and procedures". Similarly, Ceri et al. (2013) acknowledged that information retrieval systems have the purpose of obtaining information that might be considered relevant or useful to the user via the representation, storage, organization, and access to information items.

Taking these definitions into account, we can make an informed decision as to whether the FarmBuddy chatbot can be considered a type of information system and information retrieval system. The FarmBuddy chatbot operates as an information system by collecting, storing, processing, and providing information through an integrated set of technological components, such as such as a Large Language Model, and a knowledge base. The chatbot also relies on human resources, such as the developers who are responsible for maintain and updating the chatbot, its knowledge base, and its algorithms.

However, even though FarmBuddy appears to adhere to these definitions, it is crucial to evaluate its functionality and effectiveness compared to other information systems. For instance, a potential limitation that can emerge is the dependency on the quality of its knowledge base. Fortunately, in the case of the FarmBuddy chatbot, the EU-FarmBook platform serves as the basis of its knowledge base, which aims to provide a certain level of quality. The EU-FarmBook platform, as defined in its website of the same name, "is a Horizon Europe project that is working at regional, national, and European (EU) levels to build an Online Platform. Gathering and sharing agriculture and forestry knowledge" (n.d.). In relation to the definition provided by Ceri et al. (2013), the chatbot functions as an information retrieval system given that it manages the representation, storage, organization, and access to agricultural and forestry information items or knowledge objects. Moreover, it aims to retrieve useful or relevant information for users, thereby meeting the core objective of information retrieval systems.

Taking into account everything stated above, we can observe that FarmBuddy qualifies as both an information system and an information retrieval system, given that it meets the criteria outlined by the aforementioned definitions provided by Zwass (2024) and Ceri et al. (2013). While the aim of the chatbot is to retrieve useful and relevant information for users, this should be empirically tested in order to ensure the relevance and usefulness of the retrieved information. User feedback should be used to assess whether the FarmBuddy chatbot meets its objectives effectively. This highlights the necessity for designing and validating a questionnaire that is specific to chatbots that are designed in this manner, in order to provide an accurate assessment of the FarmBuddy chatbot. Furthermore, it addresses the importance of considering non-technical aspects such as user trust, and overall user experience, as a way of assessing the relevance and usefulness of the retrieved information.

2.3 Definition of trust

In order to achieve the aforementioned, it is paramount that we first provide a definition of "trust" in the context of chatbots that best fits our criteria. Firstly, a general and widely accepted definition of trust, which was first proposed and conceptualized by Mayer et al. (1995, p. 712), states that trust "is the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustier, irrespective of the ability to monitor or control that other party". Authors often utilize this definition of trust presented by Mayer et al. (1995) to refer to trust in various domains such as information systems (Söllner et al., 2016), or a modified version of the aforementioned definition, "the willingness of the trustor to "behaviorally depend on a piece of software (e.g., a statistical system) to do a task..." (McKnight, 2005, as cited in Li et al., 2008, p. 42). This version emphasizes the chatbots' ability to perform a specific function in an effective manner, which is crucial in the context of chatbots. In relation to chatbots, trust is frequently described or framed as the confidence a user holds in the reliability and credibility of recommendations and responses made by the AI agent (Shin, 2021, as cited in Baek & Kim, 2023). We can clearly observe the emphasis given by this definition to the users' reliance on the accuracy and usefulness of the chatbot's capabilities.

However, while these definitions provide a solid foundation to consider for our specific needs, they each have certain limitations to be considered. The definition provided by Mayer et al. (1995) is relatively broad, and as a consequence does not address some of the relevant nuances of trust in the context of human-computer interactions. On the other hand, even though McKnight (2005) does provide a definition that is more specific to software, it may be considered to still be too general to accurately be applied to this context.

For this study of the FarmBuddy chatbot, we have chosen what can be consider as a combination of the above, which is the definition of trust in relation to chatbots provided by Choudhury and Shamszare (2023), as "the user's willingness to take chances based on the recommendations made by this technology". Through this definition, we believe that an integration is achieved of both the elements provided by Mayer et al. (1995), as well as the emphasis on behavioural dependence made in the definition conceptualized by McKnight (2005). This ultimately provides a definition that captures the willingness of the user to rely on the chatbot's recommendations for agricultural information, which is the essence of the FarmBuddy chatbot.

Via the adoption of the definition given by Choudhury and Shamszare (2023), we aim to accurately capture the essence of user trust most specific to the FarmBuddy chatbot. This choice is justified given the combination of the theoretical robustness of the definitions presented before, as well as some practical considerations related to the functionality of the chatbot. Ultimately, it provides a comprehensive and contextually relevant framework, which will guide the assessment of user trust in the FarmBuddy chatbot. Nevertheless, it is vital to acknowledge potential limitations, such as the potential need for empirical validation of this definition in diverse user contexts.

2.4 Trust as a component of user experience

The development and design of chatbots should be performed with the objective in mind of strengthening the user experience (Følstad & Brandtzaeg, 2020). Trust plays a vital role in shaping user experiences with chatbots, influencing interactions, perceptions, and adoption rates. Through this subsection we will explore the important role that trust plays within the context of user experience in regard to general chatbot applications, as well as chatbots which are specifically designed as information systems. User experience is considered to encompass all of the users' perceptions, preferences, emotions, and responses that arise before, during, or after the engagement with an interactive system (ISO 2019, as cited in Haugeland et al., 2022). This includes pragmatic qualities such as usefulness, effectiveness, and efficiency, and hedonic qualities such as pleasure, stimulation, and identification (Hassenzahl, 2018, as cited in Haugeland et al., 2022). Achieving a well rounded experience in the context of chatbots has been found to be dependent on balancing both of these types of qualities as a means to meet the user's need and preferences effectively (Følstad & Brandtzaeg, 2020). Nevertheless, for task oriented chatbots such as FarmBuddy, usefulness is considered to be the main determiner behind a positive user experience. However, negative hedonic attributes, particularly manifested in responses which are deemed to be rude or out-of-place, can significantly influence the user experience of chatbots, ultimately resulting in the development of strong negative emotions by the user (Følstad & Brandtzaeg, 2020).

As we will further observe in this paper, users' trust in a chatbot greatly influences their willingness to engage with it, their continuance intention towards it, and affects the users' overall satisfaction in the chatbot. Several studies have highlighted the important impact that trust has as a key component or factor in the overall user experience of the chatbot (El Bakkouri et al., 2022; Kvale et al., 2020; White et al., 2022). One factor influencing trust in chatbots is their level of humanlikeness or anthropomorphic features. In spite of the potential negative impact on trust stemming from the the uncanny valley effect, an increase in human like traits has been shown to strengthen the hedonic qualities in chatbots and subsequently the user experience (Smestad and Volden, 2018 as cited in Haugeland et al., 2022). Making the interaction more engaging and enjoyable for users via the integration of these hedonic aspects found in user experience can also significantly strengthen trust. Additionally, the importance of humanlikeness in both user experience and trust has been researched and validated in domains such as customer service (Nordheim et al., 2019). Taking this into account, we can observe that the impact of humanlikeness on users' trust and overall experience should be further investigated given the complexity that arises whilst balancing human likeness to avoid the uncanny valley effect. This refers to the sense of unease or discomfort that can arise when an individual interacts with a human like virtual entity (Brenton et al., 2005).

In the realm of customer service, Kvale et al. (2020) found that, aside from humanlikeness, the increase in trust that takes place through accurate, efficient, and helpful responses leads to a better overall user experience. However, it is noted in this study that the chatbot's ability to provide these responses is also influential for general chatbots' user experience, so it is not exclusive to customer service. This highlights the fact that even though accuracy and efficiency metrics are of great importance, they may not fully capture the broader range of factors that contribute towards adequate user experience.

Whilst assessing the factors that impact user experience of COVID-19 chatbots, in the study published by White et al. (2022) various similarities where found to the factors affecting trust. For example, the level of engagement and trust in a chatbot was significantly determined by its clear and appropriate branding and association with a reputable organization, such as public health authorities. Moreover, an increase in transparency or disclosure by the chatbot, ultimately increased engagement, perceived warmth, and the likelihood of self-disclosure by the user (White et al., 2022). These factors impacting user trust and engagement of COVID-19 chatbots are crucial in shaping user experience.

The use of chatbots as information systems enhances the importance of trust as a component of user experience, due to the reliance from the users on the chatbots as a source of accurate and reliable information. According to Følstad and Brandtzaeg (2020), aiding the user in achieving their objective is critical to providing good user experience. In the context of information systems like the FarmBuddy chatbot, this pertains to acquiring the relevant information requested by the user. Taking this into consideration, it is important for user experience in chatbots as information systems to account for trust-building factors such as competence, understandability or ease of use, etc, during the developmental process of the chatbot. It must also be noted that, in accordance to the use of AI in the FarmBuddy chatbot, the role that artificial intelligence has on user experience has also been statistically proven in the study performed by Daqar and Smoudy (2019). Daqar and Smoudy (2019) found a positive significant relationship between AI and Customer Experience, with AI accounting for 26.4% of the explanation of the variance in customer experience. In other words, this highlights the substantial impact of AI on customer or user experience since it contributes to over quarter of the variation of the score obtained for customer experience in this study.

Considering the aforementioned points, trust, within the context of chatbots as information systems, has been demonstrated to be a significant component of user experience greatly influencing user satisfaction, engagement, and continuance intention.

2.5 Importance of trust in chatbots as information systems

Trust has been found to play a pivotal role in the way that interactions take place between users and chatbots, as well as significantly impacting technology adoption and user behaviour. The importance of evaluating trust in chatbots as a whole, stems from the fact that trusting technology and the intentions that the user has towards employing said technology are positively correlated (Jeng, 2019; Van der Heijden et al., 2003, as cited in Salah et al., 2023), emphasizing the importance of trust in regard to user behaviour. Moreover, Przegalinska et al. (2019) denotes trust as "the focal point of successful human chatbot interaction", underscoring the influence of trust on facilitating an interaction that can be considered to be meaningful. Furthermore, trust is considered by many scholars to be of the utmost importance for the adoption of a new type of technology in the online market (Corritore, Kracher and Wiedenbeck, 2003, as cited in Nordheim et al., 2019).

However, whilst these studies underscore an important consideration, which is the significance of trust, there are possible limitations that may arise. For instance, the generalization of these findings across different types of chatbots can be problematic, given that trust may manifest in different manners depending on various aspects, such as the chatbot's purpose and the users familiarity with these systems.

Concerning users' willingness to share what might be regarded as personal or sensitive information. A lack of trust in a chatbot can also hinder information sharing, due to security concerns, which may result in a sub optimal use of the chatbot and a failure to extract its full capabilities (Chung, Joung, and Ko, 2017b, as cited in Przegalinska et al., 2019). The use of natural language in the chatbot's interaction with the user has become increasingly more prevalent with novel advances in the field of natural language processing and deep learning. Further increasing the importance of the users' perceived trust in the chatbots' ability (Holtgraves, Ross, Weywadt, and Han, 2007, as cited in Nordheim et al., 2019). This relationship between the technological advancements and user trust may require further exploration to fully understand the influence of these advancements on trust across various domains, in this particular case the agricultural and forestry domain.

Aside from the importance of trust in regard to maintaining a successful interaction, the psychological well being of the user whilst engaging with a chatbot, more specifically ChatGPT, also seems to be affected by the trust that the individual has towards said chatbot. This is thought to occur due to an increased sense of control and reduced anxiety or stress whilst using the chatbot (Abd-Alrazaq et al., 2019; Inkster et al., 2018; Schiff et al., 2020, as cited in Salah et al., 2023).

In the context of chatbots as a type of information system, trust plays an important role in regard to technology acceptance throughout the field of information systems(e.g., Gefen et al, 2003b; van der Heijden et al, 2003; Pavlou and Gefen, 2004; Wang and Benbasat, 2005; Connolly and Bannister, 2007; Datta and Chatterjee, 2008, as cited in Söllner et al., 2016), which is thought to occur due to "the value of trust as a mechanism to reduce social and technical complexity" (Luhmann, 1979; Gefen, 2000; Lee and See, 2004, as cited in

Söllner et al., 2016). This becomes even more prevalent through the increased complexity of the automation of these systems (Lee and See, 2004, as cited in Söllner et al., 2016).

Furthermore, an increase in trust seems to aid in overcoming perceptions of risk and uncertainty that appear in the use and acceptance of new technology(Gefen et al., 2003; Pavlou and Gefen, 2004, as cited in Li et al., 2008). Additionally, trust has been observed to be positively related with the intention of acting upon a recommendation made by another entity to address a specific user need. This can be considered to be further proof of trust's role in shaping not only user behavior, but also shaping the users' ultimate action (Matook et al., 2015).

In conclusion, user trust has been shown to be a crucial component to the effectiveness and adoption of chatbots as information systems. It influences overall user experience, technology acceptance, and user behavior. In addition, user trust significantly reduces the perceived risks and uncertainties that arise whilst using chatbots leading to an increase in engagement and reliance on the chatbots recommendations. This shows the importance of enhancing trust during the development of the chatbot when dealing with the aforementioned concepts. However, future research may benefit from filling the existing gaps, particularly concerning the generalizability of the findings accross different types of chatbots, as well as the rationale behind the influence of technological advancements on trust accross various domains.

2.6 Impact of trust on user engagement and continuance intention

Fostering long-term interactions between users and chatbots is an important consideration to have whilst designing a chatbot, since this can enhance the implementation and user adoption of the chatbot and increase the chatbot's utility and value. (Przegalinska et al., 2019) even considered retention rate to be an important metric for assessing the success that a chatbot has overall. In this section we will be delving into the literature in order to obtain a more comprehensive understanding of the role that trust plays in the dynamics of user retention, continuance intention, and reusage of chatbots as information systems.

In regard to automation in general, trust is believed to influence the reliance that the user has towards it and the potential rejection of what they consider to be untrustworthy automation (Lee and See, 2004, as cited in Körber, 2019). If a service is highly trusted, this results in a sustained relationship between customers and services. On the contrary, in low trust situations, the user places more emphasis on calculating the benefits and costs of the service for the continuation of said relationship, which may ultimately lead to the termination of the relationship if it is deemed unworthy (Kim & Chang, 2020). The users' reliance and readiness to collaborate with a conversational agent is also dictated by the users' trust in said agent, which has been found to be true in various papers(Hayashi and Wakabayashi, 2017; Kundinger et al., 2019; Sundar and Kim, 2019, as cited in Rheu et al., 2021). In the agriculture sector, the positive influence that can be found between trust and product reuse intention is of great significance towards the development of the FarmBuddy chatbot, and a significant inspiration behind the evaluation of this phenomenon in this study (Lin et al., 2018).

However, these studies often overlook the nuanced differences that arise in trust dynamics across various chatbot applications and domains. Fortunately, the literature suggests that the importance of trust as a factor of user retention for chatbots spans various fields. In the banking sector trust was found to have the strongest effect on continuance intention towards banks' chatbots (Nguyen et al., 2021). The influence of trust on customer retention was also found to be prevalent in the online fashion industry (Funke et al., 2023) and the health sector (Liu et al., 2024). These findings presented in the literature seem to suggest that the importance of trust for user retention is not only multidisciplinary but universal to different types of systems. For instance, users of recommender systems appear to exhibit a greater inclination towards reusing the system when their trust level is higher (Acharya et al., 2022). Pertinent to the case of the FarmBuddy chatbot, which has a use akin to information systems, information system users' continuance intention has also been found to be positively related to trust (Vatanasombut et al., 2008).

Due to the advancements that have taken place in Generative Artificial Intelligence in the last few years and the importance of these changes in regard to chatbots as information systems, we must also acknowledge how trust affects user retention in this context. The study performed by Choudhury and Shamszare (2023) provides empirical evidence of the direct effects found between trust, and intent to use and actual use. Furthermore, this study found a significant indirect effect of trust on actual use in relation to ChatGPT, which was mediated by intent to use. Trust in Generative Artificial Intelligence has been proven to foster user commitment or loyalty towards AI chatbots, and positively influencing usage intention and engagement (Mostafa and Kasamani 2022 Baek & Kim, 2023). Additionally, trust not only has a significant effect on future use of the chatbot by a user, but also on the indirect effect between the service quality of the AI chatbot and the loyalty that the user has towards the brand that developed the aforementioned chatbot (Shahzad et al., 2024).

One of the main objective of this study is to evaluate the initial trust that users have towards the FarmBuddy chatbot. The decision to follow the initial trust measurement method used in this study instead of a more longitudinal approach is due to the initial trust being what users take into consideration in order ”to determine the extent to which future interactions will take place” (McKnight et al, 2002b; Koufaris and Hampton-Sosa, 2004, as cited in Söllner et al., 2016). However, the dynamics of user trust over time and its evolution with continued use may also be of importance, warranting further investigation.

Taking into account everything stated above we can observe the impact that trust has towards continuance intention and user engagement across various domains and chatbot uses. This ultimately influences the chatbots adoption and utility, providing another relevant reason behind the decision of evaluating user trust in the FarmBuddy chatbot. Future research should consider addressing the current need for a more longitudinal approach to obtain a comprehensive understanding of the evolution of trust dynamics over time.

2.7 Factors that affect trust

The factors that affect trust have been studied to a certain extent throughout the literature. In this section we will provide a collection of what we have found to be the most prevalent factors affecting trust and their importance in relation to chatbot trust evaluation. This is the basis that was used to determine the factors of trust that were included in the exploratory analysis performed in this study in regard to trust in the FarmBuddy chatbot.

One of the most cited models of trust is Mayer et al. (1995) integrative model of organizational trust, where ability, benevolence, and integrity were found to have an important effect on trust. However, whilst this is consider to be a widely accepted model, it primarily focuses on organizational contexts and may not fully capture the subtle distinctions found related to trust in technology. Furthermore, Rempel et al (1985) divided trust into three components, these being predictability, dependability, and faith (Rheu et al., 2021). Similarly, the framework presented by this model may need further analysis in the context of chatbots as information systems.

In terms of trust in automation systems, Körber (2019) found that familiarity in the system or a similar system has an indirect effect on trust in automation. This may provide valuable insight into possible differences in user trust for the FarmBuddy chatbot, warranting its adequate analysis. Additionally, during an exploratory interview study Følstad et al. (2018) found that, aside from factors concerning the chatbot itself, factors related to the service context as a whole, such as the brand behind the chatbot and the perceived risk whilst using the chatbot, were also reported by the participants to affect trust. This was later confirmed via a multiple regression analysis of trust in chatbots for customer service (Nordheim et al., 2019). Nevertheless, the generalizability of these results across various chatbot application remains uncertain.

In regard to the ease of use of the chatbot, it was not only considered to be one of the most significant factors towards ChatGPT usage (Choudhury & Shamszare, 2023), but also one of the factors, alongside others such as predictability and reputation, that was found to explain variation in trust in chatbots for customer service (Nordheim et al., 2019). Morevover, Følstad et al. (2018) found that the overall appearance or chatbot interface, as well as the use of adequate or correct language, was also considered to influence user trust in chatbots. This finding was later confirmed by Tan, Liew, et al. (2022), who found that variety in chatbot interface did in fact significantly affect perceived trust. Despite these findings, more rigorous experimental validation across diverse chatbot applications is suggested. In 2023 Xu et al. (2023) confirmed the belief found in the literature, regarding the importance of transparency and its effect on user's trust. In this study the degree of transparency was found to be significant in increasing trust, contingent upon the initial level of user trust. For this particular study transparency refers to the AI's knowledge, how they know it, and what they do with this knowledge. The use of transparency in chatbots can also aid in mitigating the so called "cry wolf" effect, which refers to the phenomenon where the threshold set for the alarm trigger is considerably low, in order to detect as many critical events as possible, leading to a large number of false alarms (Yang et al., 2017). Transparency as an important factor explaining user trust has been found extensively throughout the literature (Choudhury & Shamszare, 2023; Mozafari et al., 2022; Przegalinska et al., 2019), which ultimately emphasizes the necessity of including it in this study.

The likeability or empathetic characteristics of the chatbot also seem to have a significant impact on user trust towards chatbots (Ltifi, 2023). Song and Shin (2024) found that the trust that was determined by the eeriness related to the uncanny valley effect was a predictor of the willingness to reuse a chatbot and a possible purchase intention. They found that if the chatbot displayed hyper realistic anthropomorphic features or human likeness, the purchase intention and the willingness to reuse the chatbot were negatively impacted. This importance given to the human likeness of the chatbot in relation to trust was also observed in the exploratory interview study performed by Følstad et al. (2018), as well as the experiment conducted by Bae et al. (2023), which found that higher trust was achieved through the use of a more robot-looking AI chatbot. It must be noted that the human likeness of the chatbots and its likeability are non-linearly correlated, further proving the findings brought up regarding the impact of the uncanny valley effect on trust (Baek & Kim, 2023). Nordheim et al. (2019) was able to replicate the discovery that was revealed through the literature regarding the fact that the users' propensity to trust technology was another significant determinant of trust in chatbots. Furthermore, Generative AI chatbots such as ChatGPT have been shown to have similar factors affecting trust. For instance, Choudhury and Shamszare (2023) included and validated various factors affecting the trust relationship between users and ChatGPT in their study, such as competence, reliability, transparency, and risk. (Söllner et al., 2016) found through the research model proposed in their study that user trust in the information system was significantly affected by factors such as trust in the provider and perceived ease of use. However, it was also acknowledged that the users' trust in the internet as a whole was not found to significantly affect the trust in the information system. This was further expanded by Söllner et al. (2016) to the relationship of trust between users and technology overall.

In conclusion, possessing an appropriate understanding of the factors affecting user trust in chatbots is essential for evaluating user trust in chatbots. This study aims to build on the factors identified in the literature so as to assess user trust in the FarmBuddy chatbot. Moreover, the findings of the study should provide a comprehensive understanding of the factors affecting trust in relation to chatbots as information systems, addressing the existence of this research gap.

2.8 Summary of the literature and gaps

As can be observed in the review of the literature presented above, the importance of evaluating user trust in chatbots has seen a significant increase. However, chatbots designed as information systems, for instance the FarmBuddy chatbot, have not received sufficient attention regarding this topic. The literature review has explored the impact of trust in chatbots in relation to continuance intention, technology adoption, and overall user experience, as well as acknowledging the main factors affecting said trust, such as competence, predictability, and transparency, as shown in *Appendix A*. As well as the importance of evaluating and considering this aspect of user experience during the process of development and evaluation of the chatbot.

Through this study, we intend to fill the gap found in regard to the existing lack of tools for assessing user trust in chatbots which were specifically designed as a form of information system. We aim to design and validate a questionnaire, through the use of partial least squares structural equation modeling (PLS-SEM), that can be utilized in the future to quantitatively assess the user trust and gain insight on the factors affecting user trust towards a chatbot that has been developed with the aforementioned functionality. As well as this, we wish to evaluate user trust on the FarmBuddy chatbot in order to obtain a more comprehensive understanding of the factors affecting trust, and overall user trust, given its importance as seen in this review. According to the literature this also provides significant insight into continuance intention, the utility of the chatbot, and overall user experience whilst using the FarmBuddy chatbot.

3 Methodology

3.1 Population and sample size

In regard to choosing the participants that would take place in the final experiment, these were chosen from a group of consortium members from the EU-FarmBook project given their relationship and interest with the project and future as potential users of the FarmBuddy chatbot, as well as their knowledge regarding the agriculture and forestry sector. Furthermore, this method allows for efficient data collection due to the proximity and availability of the participants, providing what can be considered as a cost-effective means of gathering appropriate responses, considering that the chatbot is not yet operational so no active users can be utilized. However, it is important to acknowledge the risks that arise from the use of this type of sampling, also known as convenience sampling, since it may introduce selection bias, given that the sample may not fully represent the target population, which in this case is potential future users of the chatbot.

To ensure the robustness of our study a power analysis was conducted to determine the appropriate or minimum sample size that should be used both for the pilot test and for the final experiment. Based on the the adjusted R squared obtained in the literature related to the factors used in the questionnaire (Nordheim et al., 2019), we used this number to obtain the minimum sample size that we would need for the PLS-SEM. According to Nordheim et al. (2019) the adjusted R squared was equal to 0.60 which means that if we introduce this in the formula displayed below we obtain a Cohen f2 or effect size of 1.5. If we input this data into the G Power software, specifying 11 predictors for the factors that affect trust, a significance level (alpha) of 0.05, and a desired power of 0.8, we obtain a minimum sample size of 24. The G Power software has been shown to accurately compute sample sizes for given effect sizes, alpha levels, and power values (Faul et al., 2007). It is to be considered that PLS-SEM provides more accurate estimates with small sample sizes, and it should therefore be applied in such instances (Hair et al., 2019). Given that the number obtained is less than the default sample size of 30 that is recommended for the pilot test to detect problems with a sufficient power according to Perneger et al. (2015), 30 was chosen as an appropriate sample size value for the pilot test, which is considered to be valid in relation to the detection of possible technical problems, ambiguous or confusing items, etc. For the final experiment the value of 24 obtained for the power analysis was chosen as the minimum sample size.

$$f^2 = \frac{R^2}{1 - R^2} \tag{1}$$

3.2 Questionnaire development

The development of the questionnaire was primarily based on a thorough review of the existing literature regarding trust and the factors affecting trust, specifically for assessing chatbots but also information systems. The objective behind this review was to utilize the obtained information in order to create a comprehensive instrument that could accurately measure or evaluate the overall trust in the FarmBuddy chatbot and identify the key factors influencing this trust.

The items that were designed based on the information obtained through the literature and can be observed in *Appendix A*. These items were designed to comprehensively cover the factors influencing trust in the FarmBuddy chatbot, in such a way that ensures that all of the relevant dimensions affecting trust that were found in the literature could be included and tested in the pilot test. This was decided so as to check their relevance in relation to chatbots as information systems. In regard to the items that were formulated by the researcher, each item was carefully worded to be clear and concise, facilitating ease of response and attempting to minimize any possible misunderstanding that may arise. This refers to the modification or creation of items that were thought to be necessary, in order to address a part of the construct which was not dealt with via the items found in the literature. Through the items created the following factors or latent variables were being measured based on the literature presented in the previous section titled *Factors that affect trust*:

- **Familiarity**: The latent variable ”Familiarity” in this context refers to the users' previous knowledge, experience, and comfort level in relation to chatbots and other conversational agents. Encompassing aspects such as previous usage, confidence in their ability to use these technologies, and frequency of use.

- **Propensity to Trust**: The construct ”Propensity to Trust” in the context of the questionnaire refers to the respondents' natural inclination or tendency to trust new technology giving valuable insight into potential trust towards this chatbot. This factor is comprised of aspects such as initial distrust, ease of trust, the respondents' willingness to trust under potential doubts. These dimensions are measured in order to evaluate the extent to which respondents are predisposed to trust new technology, influencing their overall trust in technological innovations.

- **Competence**: The factor "Competence" is used to determine the users' perception in relation to the chatbot's ability to perform the tasks that it was designed to do in an effective manner. This includes aspects like confidence in the chatbot's competence, perceived expertise of the chatbot's content, and whether the chatbot is well-equipped for its tasks.

- **Understandability**: The latent variable of "Understandability", which is also often referred to as "ease of use", measures how easily users can interact with and understand the chatbot. It includes aspects such as the ease of use, clarity and understandability of the dialogue, and flexibility of interaction.

- **Appearance**: In regard to the factor of "Appearance", in this particular instance this refers to the users' perception of the chatbot's visual and interface design. It includes aspects such as thoughtful development, visual appearance, interface usability and functionality, language adequacy, etc.

- **Brand**: When it comes to the latent variable "Brand", it refers to the users' perception of the brand that developed the chatbot, which in this case is both *TNO* and the *EU-FarmBook Project*. It is comprised of aspects such as respect from others towards the brand, intention to help users, and brand recognition.

- **Human-Likeness**: The construct of "Human-Likeness" evaluates how human-like the chatbot appears or is perceived by its users. It encompasses aspects such as realism, presence, naturalness, and authenticity.

- **Likeability**: The latent variable of "Likeability" measures users' overall affection towards the chatbot. It includes aspects such as friendliness, comfort in conversation, empathy and understanding in responses, and the accommodating nature of the chatbot.

- **Predictability**: In regard to the factor of "Predictability", it refers to the consistency and reliability of the chatbot's behavior. It includes aspects such as expected behavior, lack of surprises in responses, and overall predictability.

- **Transparency**: The construct of "Transparency" measures how clearly the chatbot communicates its capabilities and workings to users. It includes aspects such as clear communication of functionality, honesty about shortcomings, and clarity of operations.

- **Risk**: The latent variable "Risk" measures users' perception of potential negative consequences that may arise whilst using the chatbot. This comprises aspects such as perceived negative consequences, feelings of safety, need for caution, sense of security, and overall risk perception.

- **Trust**: With the final latent variable of "Trust" the objective is to measure the users' overall trust in the chatbot. It encompasses aspects such as perceived trustworthiness, suspicion towards the chatbot, and general trust in the chatbot.

3.2.1 Likert Scale and Reverse Coding

The use of Likert scale items in questionnaires has been a common practice in a wide variety of research, especially in the context of assessing more subjective concepts such as trust. One of the key advantages of utilizing these types of scales in item design is the ability to convert what is considered to be more subjective opinions or perceptions into quantitative data. Therefore facilitating statistical analysis, which allows for the identification of potential trends, the comparison of demographics, and performing a wide variety of statistical tests which may provide valuable insight. Furthermore, the statistical results obtained can help make an informed decision into which items accurately measure the constructs and which of them appear to measure a different latent variable(Nemoto & Beglar, 2014). In the past, Likert scales have been used to assess various concepts such as attitudes and satisfaction levels in a diverse number of contexts, for instance (Glaser et al., 2006). In relation to trust in the field of chatbots the study presented by Nordheim et al. (2019) utilized a 5-point Likert scale in their attempt to evaluate trust in chatbots. However, in this study we opted for the use of a 7-point Likert scale given its ability to provide more precise and detailed measurements of attitudes and opinions (Joshi et al., 2015). Furthermore, it is considered by Joshi et al. (2015) to reduce ambiguity, as well as enhancing the reliability and validity of the findings. Another main consideration behind this decision is that this increase in the number of responses aligns appropriately with the human cognitive processing capabilities found in the research concerning span of immediate memory, which support this notion given that the human mind is believed to have a span of absolute judgment that can distinguish 7 categories at a time(Joshi et al., 2015).

In regard to the use of a multi-item scale instead of a single item scale, this decision was reached as a result of multi-item questionnaires being considered psycho-metrically advantageous given their higher reliability and validity (Sarstedt & Wilczynski, 2009). Furthermore, multi item questionnaires enable researchers to analyze latent variables that contribute to an overall variable and obtain a more comprehensive understanding of these constructs (Castro et al., 2023). It must be noted that utilizing this type of items does make the design more complex, as well as potentially reducing the response rate (Sarstedt & Wilczynski, 2009). The inclusion of reverse coded items was decided to attempt to control any possible response bias, such as acquiescence bias, that may arise (Paulhus, 1991). Acquiescence bias refers to the tendency that respondents may have towards agreeing with statements regardless of the contents of them. Through the use of reverse coding sufficient insight can be obtained into these types of biases and ensure the reliability and validity of the data being collected.

3.3 Data analysis

3.3.1 Pilot test

In the pilot test, we used a Likert Scale questionnaire as well as open ended questions in order to obtain both qualitative and quantitative information so as to validate the items that will be used for the final experiment as well as to qualitatively assess that these items are understandable and produce no confusion based on the feedback. This validation is necessary since, in the questionnaire, a combination of validated items were used from the literature in combination with some items formulated by the researcher to create a more complete questionnaire for assessing user trust in chatbots designed as information systems. Ultimately, this will result in a validated questionnaire that can be utilized to accurately assess trust and the factors that affect trust for any specific chatbot which has information retrieval as its primary purpose. The qualitative feedback was obtained in order to observe if the participants had any issues regarding functionality or understandability whilst performing the pilot test as well as any valuable insight towards the functionality of the survey itself. In order for this to be valid as mentioned before, the default sample size of 30 was reached, since this is recommended for any pilot test that wishes to detect problems with a sufficient power according to Perneger et al. (2015).

In order to assess the internal consistency reliability, Cronbach's alpha was used given its popularity in the literature. The Cronbach's alpha value was developed in 1951 by Lee Cronbach as a way of describing if all of the items utilized in a test accurately measure the same construct or concept, which pertains to the inter relatedness of the items that are used in the test (Tavakol & Dennick, 2011). Measuring the internal consistency of each of the items is important given that it ensures the reliability of the test results. This occurs since it evaluates whether or not the items are cohesively working together to assess the intended construct, thus providing more robust and trustworthy data. This information was obtained through PLS-PM. Furthermore, aside from Cronbach's alpha values, composite reliability is often used as a popular alternative way of measuring the internal consistency of the items and ensuring that all of the items measure the same construct (Peterson & Kim, 2013). This value was used to provide more robustness to this measurement and given its common use in conjunction with structural equation modeling which is the main focus of this study (Peterson & Kim, 2013).

As mentioned by Slocum-Gori and Zumbo (2011), observing the unidimensionality of the data obtained for each of the items from a survey is a crucial part of construct validity. In order to do so, we decided to use what is known as the ratio of the first-to-second eigenvalues due to it being an established and effective method according to the literature and the study published by Slocum-Gori and Zumbo (2011). Through this method we can observe whether the items from the survey predominantly measure a single construct or latent variable by comparing the variance which is explained by the first factor to that explained by the second factor, consequently enhancing the validity of the scale.

Another common value obtained through PLS-PM is the Average Variance Extracted, which can be used as a measure of convergent validity. Convergent validity is based on the principle that measures of the same or similar constructs should exhibit a high degree of correlation, and is a fundamental aspect of construct validity (Chin & Yao, 2021). Through this approach we aim to measure if the items designed to measure a particular latent variable or construct effectively reflect said construct. Since we assume that these constructs cannot be directly observed, through this approach we can consider a construct to be valid if the items measuring that construct are closely related, which refers to the items sharing a significant proportion of their variance.

Discriminant validity will also be measured via the use of the Fornell-Larcker criterion. This decision was based on the fact that this is the most common method of assessing discriminant validity, which is considered to be crucial in regards to research involving latent variables so as to prevent any problems regarding multicollinearity (Ab Hamid et al., 2017). This refers to the fact that discriminant validity aids in the process of confirming that each latent variable represents a unique construct, rather than overlapping significantly between each other, which could lead to multicollinearity issues as previously mentioned.

3.4 Factors affecting user trust and user trust

In this section, a Partial Least Squares Structural Equation Modeling (PLS-SEM) was conducted using the SEMinR package in R to evaluate factors affecting user trust in the FarmBuddy chatbot and their subsequent impact. Initially, constructs are to be identified as being reflective or formative and utilized as such for the PLS-SEM, and indicator reliability should be ensured by removing items with loadings below 0.4 as recommended by Hair Jr et al. (2021). Internal consistency reliability is checked using Cronbach's alpha and composite reliability (rhoC), both of which should exceed the common minimum threshold of 0.6. Convergent validity is to be verified through the Average Variance Extracted (AVE) values, all above 0.50. Discriminant validity is assessed using the heterotrait-monotrait ratio (HTMT), with values below the 0.90 threshold indicating no issues. Collinearity must be checked using the Variance Inflation Factors (VIF) values, with a common threshold of 5. The path analysis is then analyzed to observe the impact of each of the factors on the overall trust score. The subsequent analysis uses ANOVA to explore the impact of demographic factors on user trust, with assumptions validated through Shapiro-Wilk and Levene's tests. Lastly, a summary of overall trust scores using descriptive statistics is used to observe the intitial trust of the participants towards the FarmBuddy chatbot.

3.5 Ethical considerations

In order to perform the experiment following appropriate ethical considerations, this study obtained approval from our institute's ethics review board. Participants received the necessary ethical information before starting the experiment. They gave their informed consent prior to starting the experiment and were adequately informed that they could terminate their participation at any time without stating a reason. The data was treated anonymously and participants received unique IDs so that the answers could not be traced back to them.

4 Results

4.1 Pilot test

As mentioned before, the pilot test was performed in order to validate the items that will be used for the final experiment as well as to qualitatively assess that these items are understandable and produce no confusion based on the feedback. Using the values that we will mention in this section as well as the feedback from the pilot test, the items that are shown in italics in *Appendix A* were deleted resulting in the final questionnaire that was used in the experiment.

As previously mentioned, in order to assess the internal consistency reliability Cronbach's alpha was used. According to (Nunnally and Bernstein, 1994) as cited in (Daud et al., 2018), a Cronbach's alpha value (C.alpha) greater than 0.6 is deemed to be acceptable and these values can be observed in the block unidimensionality table 2 shown below, which was obtained through PLS-PM. Furthermore, in exploratory research, as is the case in this questionnaire validation, both Cronbach alpha values and composite reliability values (DG.rho) between 0.6 and 0.7 are acceptable. Taking this into account we can state that all factors meet these criterion and therefore both internal consistency and composite reliability are met (Ahmad et al., 2016).

In regards to the unidimensionality of the factors, which refers to making sure that these items measure a single trait, we can observe that the 1st eigenvalue captures significant variance in all of the factors as well as being significantly larger than the 2nd eigenvalues. This concept appears to be met for all of the factors, indicating that each factor represents a single underlying construct.

Below we have displayed the tables for both the original and final tables related to block unidimensionality. It must be acknowledged that some items were previously deleted from the original table based on the qualitative feedback from the pilot test. During this process of item deletion, items were deleted based on the results to the PLS-PM one at a time to observe how this affected the table. The items were chosen based on both the data from the PLS-PM as well as personal judgement regarding aspects such as possible correlation to other items as well as similarity between the items which can affect their unidimensionality. Furthermore, the feedback from the survey was also used to assess if items from a problematic factor may produce confusion or understandability issues, for instance the deletion of one item due to confusion related to the word "flexible" that was brought up by multiple participants. One notable deletion is that of the factor of "Human-likeness", the deletion of this construct was determined due to the confusion created by the items used for this latent variable. In future research, it would be valuable to recognize this and design clear items to assess this potentially important factor. In the case of this study, given the limited time and the fact that the pilot test could not be conducted again we decided to completely eradicate this factor.

	Mode	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
Familiarity	A	5	0.736	0.827	2.46	0.858
Propensity to Trust	A	5	0.884	0.916	3.44	0.759
Competence	A	6	0.754	0.832	2.84	1.143
Understandability	A	5	0.598	0.757	2.03	1.086
Appearance	A	6	0.706	0.805	2.57	1.361
Brand	A	6	0.782	0.849	2.98	1.281
Human-likeness	A	4	0.573	0.754	1.84	1.036
Likeability	A	4	0.610	0.773	1.90	0.957
Predictability	A	4	0.661	0.797	1.99	0.957
Transparency	A	5	0.779	0.850	2.67	1.323
Risk	A	4	0.831	0.888	2.66	0.676

Table 1: Original Table of Factors with Corresponding Block Unidimensionality Values

	Mode	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
Familiarity	A	5	0.736	0.827	2.46	0.858
Propensity to Trust	A	5	0.884	0.916	3.44	0.759
Competence	A	4	0.807	0.874	2.54	0.651
Understandability	A	3	0.668	0.820	1.81	0.702
Appearance	A	3	0.784	0.874	2.10	0.512
Brand	A	3	0.740	0.854	1.99	0.715
Likeability	A	4	0.741	0.841	2.32	0.918
Predictability	A	4	0.661	0.797	1.99	0.957
Transparency	A	2	0.840	0.926	1.72	0.275
Risk	A	4	0.831	0.888	2.66	0.676

Table 2: Final Table of Factors with Corresponding Block Unidimensionality Values

One of the most common ways of evaluating the convergent validity of an instrument was presented by Fornell and Larcker (1981). Through this method convergent validity is established when a latent construct accounts for at least 50 percent of the variance in its associated indicators, which can be determined using the Average Variance Extracted (AVE) values. This means that the AVE should be 0.50 or higher. However, if the AVE value is slightly below 0.50, it can still be considered acceptable if the composite reliability (CR) is above 0.60. Given that the composite reliability for all factors is indeed over 0.60 and that the AVE is near or above 0.5, we can conclude that convergent validity is met.

	Type	R2	Block_Communality	Mean_Redundancy	AVE
Familiarity	Endogenous	1	0.4928897	0.4928897	0.4928897
Propensity to Trust	Endogenous	1	0.6877667	0.6877667	0.6877667
Competence	Endogenous	1	0.6344523	0.6344523	0.6344523
Understandability	Endogenous	1	0.6021669	0.6021669	0.6021669
Appearance	Endogenous	1	0.6984187	0.6984187	0.6984187
Brand	Endogenous	1	0.6633083	0.6633083	0.6633083
Likeability	Endogenous	1	0.5794030	0.5794030	0.5794030
Predictability	Endogenous	1	0.4972005	0.4972005	0.4972005
Transparency	Endogenous	1	0.8624179	0.8624179	0.8624179
Risk	Endogenous	1	0.6653021	0.6653021	0.6653021

Table 3: Table of AVE

As mentioned before, discriminant validity was assessed by comparing the square root of the AVE of each construct to the correlation between constructs, which is also known as the Fornell-Larcker criterion. After observing the results from the Fornell-Larcker criterion, where we checked that the square root of the Average Variance Extracted (AVE) for each factor is greater than the correlation between that factor and any other factor. The results demonstrated that all 10 factors possess discriminant validity.

Comparison	Square root of AVE (Familiarity)	Correlation
Familiarity vs Propensity to Trust	0.702061	0.1723
Familiarity vs Competence	0.702061	0.2014
Familiarity vs Understandability	0.702061	0.2986
Familiarity vs Appearance	0.702061	-0.1894
Familiarity vs Brand	0.702061	0.2345
Familiarity vs Likeability	0.702061	0.1123
Familiarity vs Predictability	0.702061	-0.0567
Familiarity vs Transparency	0.702061	0.3210
Familiarity vs Risk	0.702061	0.1456

Table 4: Comparisons of Factor 1 with Factors 2 through 10

	Type	Square root of AVE
Familiarity	Endogenous	0.702061
Propensity to Trust	Endogenous	0.829317
Competence	Endogenous	0.7965251
Understandability	Endogenous	0.7759941
Appearance	Endogenous	0.8357145
Brand	Endogenous	0.8144374
Likeability	Endogenous	0.7611853
Predictability	Endogenous	0.7051245
Transparency	Endogenous	0.9286646
Risk	Endogenous	0.8156605

Table 5: Table of Square root of AVE

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7	Factor8	Factor9	Factor10
Factor1	1.0000									
Factor2	0.1723	1.0000								
Factor3	0.2014	0.3506	1.0000							
Factor4	0.2986	0.2205	0.4252	1.0000						
Factor5	-0.1894	0.1486	0.0807	0.1143	1.0000					
Factor6	0.3339	0.2386	0.1783	-0.0277	0.3895	1.0000				
Factor7	-0.0811	0.3088	0.4837	0.1859	0.3772	0.3675	1.0000			
Factor8	0.1222	0.0233	0.1455	0.0229	0.2716	0.2654	0.3052	1.0000		
Factor9	-0.2023	0.3170	0.4674	0.1029	0.5451	0.4903	0.7041	0.2972	1.0000	
Factor10	-0.0205	-0.4173	-0.5300	-0.2039	-0.1746	-0.3432	-0.6114	-0.2788	-0.4551	1.0000

Table 6: Correlations Between Factors

Notably, via the qualitative assessment we were able to observe that various participants from the pilot test commented positively on the use of an integrated link to the chatbot in various places throughout the survey to allow for tasks to be performed when necessary. This allowed for the participants to comfortably go back and forward from the survey to the chatbot making the overall experience more pleasant.

Taking into consideration the aforementioned, we can state that the pilot test effectively refined the final questionnaire by validating the items used based on both qualitative and quantitative data. The internal consistency and composite reliability, was satisfactory, indicating a reliable measurement of the constructs or factors affecting user trust. The unidimensionality and convergent validity of the factors were also accurately confirmed, and discriminant validity was achieved based on the results obtained following the Fornell-Larcker criterion. Although the factor of "Human-likeness" was removed due to various issues regarding misinterpretation and clarity, the refined questionnaire is robust for assessing the other key constructs. Future research should aim to design clearer items to provide a more comprehensive evaluation of potentially important factors, such as that of "Human-likeness".

4.2 Factors affecting user trust for the FarmBuddy chatbot

The main objective of this section is to assess which of the factors obtained from the literature have the most impact in regards to user trust for the FarmBuddy chatbot. In order to do so a Partial Least Squares Sequential Modeling or PLS-SEM was performed following the recommended criteria provided by Hair Jr et al. (2021). This was performed via the use of the SEMinR package in R. As previously mentioned, the data that was analyzed was obtained through a survey that was given to consortium members or potential future users of the FarmBuddy chatbot.

Firstly, the constructs for the measurement model must be defined. In this particular case, based on the analysis of each construct, we can conclude that all of the factors used in the questionnaire are reflective constructs. This refers to the fact that each of the items or set of items are reflective of an underlying latent construct, each construct is expected to be correlated, the items are interchangeable, and any changes that may occur to the latent construct should cause changes to the items to which it corresponds (Hanafiah, 2020). In regard to the structural model, we are aiming to observe the paths and relationships between each of the factors and user trust in order to assess their impact on this dependent variable of user trust.

Before we are able to observe the model itself and the relationships between the paths used in the model, we must first perform a few procedures to ensure the robustness and reliability of the PLS-SEM. According to Hair Jr et al. (2021), the first step towards this goal is to ensure indicator reliability. This is done via the analysis of the bivariate correlation between indicator and construct, which is obtained by squaring the indicator loadings. According to (Hair, Hult, Ringle, & Sarstedt, 2022, as cited in Hair Jr et al., 2021), any value below 0.4 needs to be removed from the measurement model. After the necessary eliminations the following table was obtained, were we can observe that all of the values exceed the minimum cutoff of 0.4 presented in the literature.

	FAM	PRO	COMP	UND	APP	BRAND	PRED	TRANS	RISK	TRUST
F1_1	0.682	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F1_3	0.770	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F1_4	0.669	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F1_5	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRO1_1	0.000	0.944	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRO1_2	0.000	0.644	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRO1_3	0.000	0.832	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRO1_4	0.000	0.774	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C1_1	0.000	0.000	0.813	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C1_2	0.000	0.000	0.621	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C1_3	0.000	0.000	0.841	0.000	0.000	0.000	0.000	0.000	0.000	0.000
C1_4	0.000	0.000	0.798	0.000	0.000	0.000	0.000	0.000	0.000	0.000
U1_1	0.000	0.000	0.000	0.472	0.000	0.000	0.000	0.000	0.000	0.000
U1_2	0.000	0.000	0.000	0.838	0.000	0.000	0.000	0.000	0.000	0.000
U1_3	0.000	0.000	0.000	0.634	0.000	0.000	0.000	0.000	0.000	0.000
A1_1	0.000	0.000	0.000	0.000	0.866	0.000	0.000	0.000	0.000	0.000
A1_2	0.000	0.000	0.000	0.000	0.647	0.000	0.000	0.000	0.000	0.000
B1_1	0.000	0.000	0.000	0.000	0.000	0.772	0.000	0.000	0.000	0.000
B1_2	0.000	0.000	0.000	0.000	0.000	0.755	0.000	0.000	0.000	0.000
PRE1_1	0.000	0.000	0.000	0.000	0.000	0.000	0.822	0.000	0.000	0.000
PRE1_2	0.000	0.000	0.000	0.000	0.000	0.000	0.490	0.000	0.000	0.000
PRE1_4	0.000	0.000	0.000	0.000	0.000	0.000	0.731	0.000	0.000	0.000
T1_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.936	0.000	0.000
T1_2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.889	0.000	0.000
R1_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.846	0.000
R1_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.597	0.000
R1_4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.438	0.000
COMP1_1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.909
COMP1_3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.907

Table 7: Summary of Loadings Squared

Similarly to what we performed in the questionnaire validation, Hair Jr et al. (2021) suggests that internal consistency reliability is examined. Hair Jr et al. (2021) defines internal consistency reliability as "the extent to which indicators measuring the same construct are associated with each other". As previously done, we will perform this evaluation by checking the Cronbach's alpha and the composite reliability (rhoC). We can clearly observe in table 8 that all of the factors or constructs exhibit an adequate or high internal consistency reliability, given that the alpha and rhoC values exceed the threshold of 0.6 used in exploratory research. Furthermore, the table also displays the Average Variance Extracted (AVE), which relates to the convergent validity of the model. As previously mentioned, the minimum acceptable AVE is 0.50, which is used to show that the factor explains 50 percent or more of the variance of the items composing said factor (Hair et al., 2022, as cited in Hair Jr et al., 2021). In this particular case, we can observe that all of the factors demonstrate good to excellent convergent validity, as indicated by the AVE values surpassing the aforementioned threshold.

	Alpha	rhoC	AVE
Familiarity	0.915	0.883	0.655
Propensity to Trust	0.946	0.940	0.799
Competence	0.901	0.930	0.768
Understandability	0.726	0.845	0.648
Appearance	0.694	0.861	0.756
Brand	0.690	0.866	0.764
Likeability	0.774	0.864	0.681
Predictability	0.906	0.954	0.913
Transparency	0.698	0.832	0.627
Risk	0.899	0.952	0.908

Table 8: Summary of Reliability Metrics

Contrary to the pilot test, in order to assess the discriminant validity of the model we used a different approach, which some consider to be a superior alternative for PLS-SEM. This is known as the heterotrait-monotrait ratio (HTMT) of correlations (Henseler et al., 2015, as cited in Hair Jr et al., 2021). According to this, discriminant validity issues arise when HTMT values are above the threshold of 0.90 for structural models with conceptually similar constructs (Henseler et al., 2015, as cited in Hair Jr et al., 2021). Taking this into consideration, given that all of the HTMT values are below the threshold, this indicates that there are no issues regarding discriminant validity according to the HTMT criterion. Consequently, the constructs in the model are then considered to be distinct from each other.

	FAM	PRO	COMP	UND	APP	BRAND	PRED	TRANS	RISK	TRUST
FAM	.									
PRO	0.620	.								
COMP	0.149	0.110	.							
UND	0.340	0.333	0.511	.						
APP	0.305	0.361	0.221	0.596	.					
BRAND	0.404	0.183	0.419	0.523	0.525	.				
PRED	0.273	0.141	0.733	0.773	0.198	0.207	.			
TRANS	0.156	0.205	0.683	0.396	0.203	0.271	0.606	.		
RISK	0.314	0.272	0.415	0.257	0.430	0.445	0.382	0.249	.	
TRUST	0.105	0.160	0.699	0.403	0.648	0.765	0.283	0.484	0.710	.

Table 9: Discriminant Validity Using HTMT

For the final recommended step in the evaluation of the model, we need to consider any potential issues regarding collinearity. This is examined through the use of the variance inflation factor (VIF). The common threshold set for the VIF values according to Hair Jr et al. (2021) is 5, with greater values indicating potential issues regarding collinearity. In the bar graph shown in Figure 1 we can observe that non of the factors exceed the threshold, proving minimal multicollinearity.

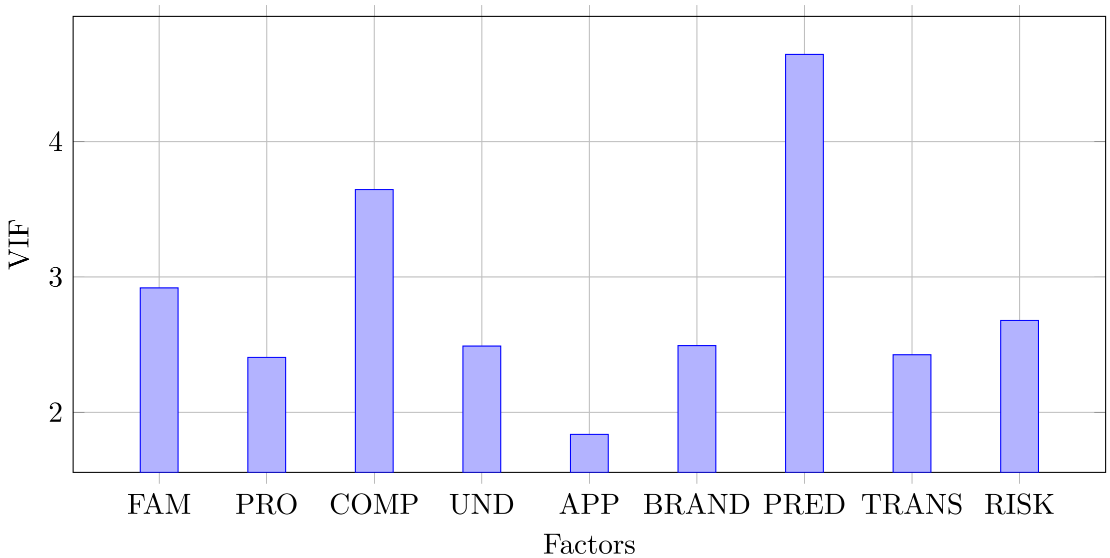

Figure 1: Variance Inflation Factors (VIF)

After completing the preliminary validations of the Partial Least Squares Structural Equation Modeling (PLS-SEM), including assessments of indicator reliability, internal consistency, convergent validity, discriminant validity, and collinearity, we are now positioned to examine the results of our analysis.

In order to assess the factors that impact user trust in relation to the FarmBuddy chatbot we can take a look at the path analysis results obtained through PLS-SEM. The plot of the model can be observed in *Appendix B*. Nevertheless, the values for the paths are also presented in the table below for clarity. If we take a look at the R^2 value of 0.816 as well as the adjusted R^2 of 0.666 we can observe that the model explains a substantial proportion of the variance found in user trust, consequently indicating a strong overall fit. This suggests that the model has overall good explanatory power.

When we take a look at the path coefficients for each of the factors affecting trust we can observe that the three most impactful factors on user trust for the FarmBuddy chatbot are Appearance with a path coefficient of 0.387, followed by Risk (0.335), and finally Brand (0.308). This suggests that these factors are particularly crucial in fostering user trust in the FarmBuddy chatbot. To a lesser extent, we can observe that both Transparency (0.293) and Competence (0.265) also contribute positively towards user trust. On the other hand, the data shows that the factors of Familiarity and Predictability have a minimal positive or even negligible impact (0.123 and -0.012, respectively), whilst Propensity to Trust, Understandability, and Predictability show negative but relatively minimal effects on user trust (-0.111, -0.058, and -0.012, respectively). Considering the findings from the data presented below, we can highlight the relative importance of the factors of Appearance, Brand, Risk, Transparency, and Competence in building user trust for the FarmBuddy chatbot, while the lower or negative impacts of some other factors suggest areas where improvements may be needed.

Unfortunately, considering the small sample size used in this particular study, we need to consider the inherent limitation in the robustness and generalizability of the findings presented. Specifically, one of the constrains that arises from this small sample size is our ability to perform reliable bootstrapping to assess the statistical significance of the path coefficients, which is often recommended whilst performing PLS-SEM by Hair Jr et al. (2021). Even though Hair Jr et al. (2021) recommend this common technique for significance testing it requires a sufficiently large sample size to produce reliable estimates, and taking this into account we cannot provide reliable p-values or confidence intervals for the paths in our model. Ultimately, this means that even though we can interpret the results as we have done previously, considering the magnitude and direction of the relationship. We cannot provide whether or not these relationships are statistically significant. Therefore, in future research we suggest utilizing a larger sample size, which would be beneficial for overcoming these limitations presented. The application of bootstrapping and other techniques in order to perform a more rigorous test for the significance of the path coefficients, would be possible with larger sample sizes, leading to more robust conclusions regarding the relationships in the model presented.

Path	**Path Coefficient**
Familiarity	0.123
Propensity to Trust	-0.111
Competence	0.265
Understandability	-0.058
Appearance	0.387
Brand	0.308
Predictability	-0.012
Transparency	0.293
Risk	0.335
R^2	**0.816**
Adjusted R^2	**0.666**

Table 10: Summary of Path Coefficients and Model Fit Indices

4.3 User trust in the FarmBuddy chatbot

In order to assess the user trust of the participants towards the FarmBuddy we will first observe the differences that arise in relation to user trust based on demographics factors. To ensure the validity of ANOVA, it is essential to first check the necessary assumptions for this statistical test. In order to do so, we will be following the paper presented by Emerson (2022) were they state the conditions that should be met before conducting this statistical test.

According to Emerson (2022), the first consideration is that the dependent variable is continuous. Strictly speaking, Likert scale data is ordinal, which means that this assumption is not perfectly met. However, in the case of Likert scales with sufficient enough points (5 or more), the data can be treated as approximately interval for the purpose of ANOVA, with Norman (2010) proving that even with a slight violation to these assumptions parametric statistics still provide robust results. The third assumption postulated by Emerson (2022) refers to the independence between the groups. Given that the participants chose a single response per demographic factor this assumption is also met.

For the second and fourth assumptions according to Emerson (2022), we can see that Table 11 provides a comprehensive overview of the statistical tests conducted to assess the normality and variance of user trust in the FarmBuddy chatbot. The analysis of normality of residuals was performed via Q-Q plot as well as the Shapiro-Wilk test. The findings show that all demographic variables display p-values above the significance level of 0.05, which is commonly used in the literature. This suggests that the residuals for these variables meet the assumption, since they follow a normal distribution. The Q-Q plot further proves this finding evidenced by the p-values from the Shapiro-Wilk test. Levene's Test for equality of variance is the most widely used method for evaluating the fourth assumption of equal variances presented by Emerson (2022). The p-values above 0.05 indicate that the variances across the different groups used in the study are homogenous meeting the aforementioned assumption. Ultimately, these findings support the necessary assumptions of normality and equal variance depicted by (Emerson, 2022), validating the robustness of the ANOVA results in examining user trust in the chatbot.

The next step is to take a look at the p-values for the demographic values in regard to how they affect user trust. The findings presented in table 11 suggest that the demographic variables do not significantly influence user trust in relation to the FarmBuddy chatbot, given the notably high p-values obtained. All of the p-values obtained are significantly above the common significance threshold of 0.05. This indicates that variations in user trust are not strongly associated with the participant's gender, age, education level, level of experience, their attendance to the focus group, the domain where they work in, their occupation, or their country of residence. These results demonstrate that user trust towards the FarmBuddy chatbot is likely to be consistent even across the different demographic groups used in this study.

Variable	p-value
GENDER	0.945
AGE	0.892
EDUCATION	0.929
EXPERIENCE	0.782
FOCUS	0.788
COMPOSITE_DOMAIN	0.613
COMPOSITE_OCCUPATION	0.650
COUNTRY	0.238
Shapiro-Wilk p-value	
GENDER	0.108
AGE	0.830
EDUCATION	0.238
EXPERIENCE	0.243
FOCUS	0.179
COMPOSITE_DOMAIN	0.156
COMPOSITE_OCCUPATION	0.466
COUNTRY	0.112
Levene's Test p-value	
GENDER	0.233
AGE	0.712
EDUCATION	0.542
EXPERIENCE	0.827
FOCUS	0.707
COMPOSITE_DOMAIN	0.754
COMPOSITE_OCCUPATION	0.572
COUNTRY	0.127

Table 11: Summary of ANOVA p-values, Shapiro-Wilk p-values, and Levene's Test p-values

In order to accurately evaluate the overall trust score of the participants towards the FarmBuddy chatbot we can observe the statistical summary displayed in table 12, which provides a comprehensive and detailed overview of the distribution of overall trust scores.

The minimum value of 1.5 suggests that at least one of the users rated their trust in the chatbot quite low. However, as one can observe there is also at least one participant who rated their trust in the chatbot at the highest possible level. If we observe figure 2 we can observe that the distribution is slightly skewed to the left, which is also represented by the mean (4.514) being marginally lower than the median (4.800). This ultimately means that most of the users tend to exhibit higher trust scores and that the bulk of the scores are clustered towards the higher values present in the trust scale. However, this also indicates that the fewer instances that exhibit low trust scores extend further into the lower section of the scale.

Taking this into account, we can state that, overall, these statistics suggest what can be categorized as a moderate to high level of trust towards the FarmBuddy chatbot, considering that both the mean and median are above the midpoint of possible range of trust values (3.5). However, we can also clearly infer that there is plenty of room for improving this score with future iterations of the chatbot.

Statistic	**Value**
Minimum	1.500
1st Quartile (25th Percentile)	3.000
Median	4.800
Mean	4.514
3rd Quartile (75th Percentile)	5.500
Maximum	7.000

Table 12: Summary Statistics of Overall Trust Scores

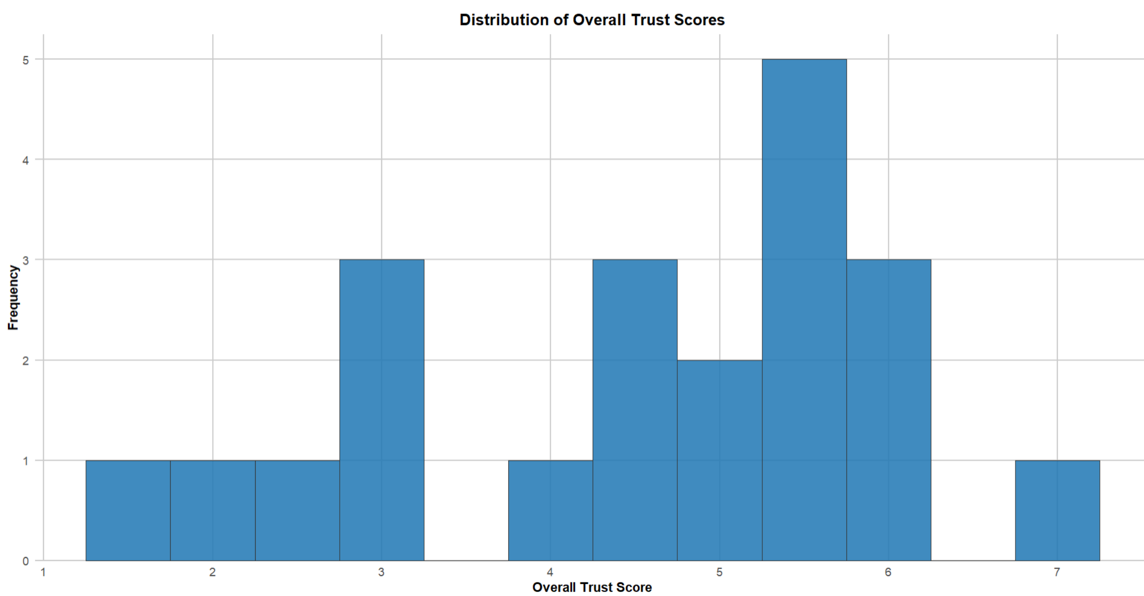

Figure 2: Distribution of Overall Trust Scores

5 Discussion

The main objective of this study was to evaluate the factors that influence user trust towards the FarmBuddy chatbot, as well as to explore to which extent do the users perceive the chatbot as being trustworthy. In order to accurately answer this question, we validated and utilized a questionnaire, which was composed of the main factors affecting trust found in the literature, to gather quantitative data. Moreover, this validated questionnaire was also one of the most important results from this study, which was obtained through a pilot test and the subsequent PLS-PM analysis of the data obtained from this test. In regard to which extent the users perceive the chatbot as being trustworthy, an ANOVA was performed to analyze the influence of demographic factors on the overall trust score, as well as an analysis of this score by itself. Finally, we will touch on the implications of the literature review performed regarding trust as a measure of user experience and likelihood for future interactions. In this following section, we will be delving deeper into the potential practical and theoretical considerations that these findings present.

5.1 User trust questionnaire

The development and validation of this questionnaire holds several significant practical implications for future research and for the development of other chatbots in similar contexts, specifically considering the importance of trust evaluation and analysis.

First of all, this questionnaire serves as a robust foundation for analyzing the factors affecting user trust and overall user trust evaluation in different chatbots. It can be utilized by researchers or chatbot developers to conduct studies across various chatbots identifying the differences or similarities that arise in these varying contexts. Furthermore, using this approach of breaking down trust into latent variables or factors allows for a more nuanced understanding of what drives trust in their specific context, leading to improvements in chatbot development. The findings that they obtain can be used to prioritize and allocate more resources to the areas which are most influential to user trust.

The importance of using benchmarking has been previously observed as a tool to enhance organizational performance(Alosani et al., 2016) but it can also be used for comparing results accross various studies. In order to achieve this goal effectively, this validated questionnaire can serve as a benchmarking tool in the future if the limitations are treated appropriately.

A user centered approach to the development of the chatbot can be taken through this questionnaire, to better meet the expectations of the users and enhance their trust. Potentially, the findings can also affect the marketing of the chatbot. For instance, if the trustworthiness of the brand needs to be emphasized.

Finally, various participants from the pilot test commented positively on the use of an integrated link to the chatbot in various places throughout the survey to allow for tasks to be performed when necessary. Given the positive feedback, future similar research should also consider implementing this practice. Overall, the validated questionnaire, which was designed through a thorough literature review and statistical analysis, can be a beneficial tool for researchers or developers in the field of chatbots. By using this questionnaire, future chatbot development and research can achieve more effective and reliable results.

5.2 Factors that affect trust

After performing the PLS-SEM, we were able to obtain the most influential factors affecting user trust in the FarmBuddy chatbot, which were Appearance, Risk, Brand, Transparency, and Competence, in that order. On the other hand, Familiarity, Predictability, Propensity to Trust, Understandability, and Predictability seemed to have minimal impacts on the overall trust score.

In contrast to previous studies evaluating factors affecting trust in chatbots (Nordheim et al., 2019), which had Competence (coined as Expertise in their study) as the most impactful factor explaining trust. The findings of the PLS-SEM suggest that this was not the case for FarmBuddy, with Competence, whilst still important, falling to fifth place in terms of relevance. We believe that this shift has occurred given the high functionality of generative artificial intelligence chatbots that have appeared after the release of ChatGPT. Since the paper by Nordheim et al. (2019) was published in 2019 we can infer that this increase in base functionality of chatbots, has increased the importance of what were previously seen as less important factors, according to the literature. This would explain why the more technical aspects, such as Understandability and Predictability, have also seen a significant decrease in their influence on user trust. Furthermore, as a result, less technical factors, such as Appearance, Brand, and Risk, are coming to the forefront. For instance, a similar phenomenon was observed in the shift in consumer trust and purchase intentions after the release of the smartphone. Whereas before the introduction of the smartphone, technical aspects such as functionality, and call quality were very influential. This changed especially after the release of the iPhone, with less technical factors gaining prominence, such as design, and brand image given the high quality of functionality of the products (Ivanov et al., 2021).

This brings forward some practical considerations that might be considered during the development of FarmBuddy and similar chatbots. Once a certain level of functionality or competence is reached, the main factors to be considered in order to increase the overall trust towards the chatbot seem to be the Appearance, Risk, Brand, Transparency of it. This shift in trust factors highlights the importance of prioritizing non-technical aspects, for chatbot developers, once a baseline level of technical competence is achieved. Focusing on aspects such as ensuring that the appearance of the chatbot is user friendly and modern. The interface should be appealing and easy to navigate, given that "interactive products with engaging usable interfaces positively influence user experience and performance" (Granić, 2017, p. 43). Addressing the possible perceived risk is also influential, which can be addressed, for example, through clear communication of data privacy and security measures to ensure a risk free interaction. Furthermore, building and reinforcing brand trust is also vital given that this ensures a certain level of quality to the information being presented by the chatbot. This might be particularly important for the FarmBuddy chatbot since potential future users will expect quality information to be obtained from reputable sources gathered through the EU-FarmBook knowledge base. Finally, appropriately informing the users of the shortcomings and functionality of the chatbot also seems to impact user trust and should be another objective to include during the development process.

Taking this into consideration, we can observe the important contribution of our study towards the existing literature by highlighting the fact that trust factors are dynamic in the context of technological advancements. It indicates the possible dependence of user trust in chatbots on traditionally secondary factors, and underscores the necessity of continually researching trust models to fit the technological landscape present at the time.

5.3 User trust toward the FarmBuddy chatbot

Given the importance of fostering a positive overall user trust in the FarmBuddy chatbot, we analysed this value through the questionnaire validated in this paper. The findings of the analysis indicated a generally positive perception of user trust since the mean and median scores were above the 3.5 midpoint of the trust scale used, which was a 7-point Likert scale going from strongly disagree to strongly agree. Subsequently, the participants or potential future users of the FarmBuddy chatbot generally deem the FarmBuddy chatbot as being trustworthy. Considering the already positive level of trust, further enhancements should focus on building on existing strengths of the chatbot. These enhancements should be informed by the factors impacting user trust as discussed in the previous section. This should be possible since there is still plenty of room for improvement as shown by the overall trust score. On the other hand, given the presence of users that rated the chatbot lower in terms of overall trust score, further development should aim to mitigate and enhance these trust scores by identifying the issues faced by the users through feedback or a qualitative analysis of the logs gathered during the interaction. Potentially, developing a mechanism that allows for consistent user feedback and suggestions might aid in this process, as well as allow for regular assessment, which we have seen to be crucial due to the constant technological advancements.

The non-significant influence of demographic factors on user trust towards the FarmBuddy chatbot has several implications. The first implication is that, if the sample used is considered to be a good representation of potential future users of the chatbot, we should maintain a certain level of consistency across all of the user groups to ensure reliable functionalities. However, once the chatbot becomes active and establishes an active user base, we can repeat this experiment with these participants, via random sampling, to ensure the robustness and repeatability of the findings of the study.

In conclusion, even though the FarmBuddy chatbot possesses an positive initial level of trust perception among potential future users, the path towards future improvement is clear and important. By reiterating assessment and focusing on user feedback we can maintain high user trust scores influencing user experience and continuance intention. Therefore, during future development, the establishment of a feedback mechanism might be beneficial to maintaining or even enhancing this user trust score towards the FarmBuddy chatbot.

5.4 User trust as a measure of user experience and continuance intention

In this paper's literature review we were able to observe the significant impact that trust has towards user experience, impacting both the pragmatic and hedonic qualities of user experience (Hassenzahl, 2018, as cited in Haugeland et al., 2022). Moreover, in order to foster long-term interactions between users and chatbots, trust was also considered to be pivotal. Therefore, a high level of trust results in an increase of overall user experience, as well as sustained user relationships increasing the chatbot's potential implementation, adoption and overall utility (Przegalinska et al., 2019).

Taking this into consideration, if we wish to develop a chatbot that provides a beneficial user experience and that fosters continuance intention between the user and the chatbot, we need to consider the findings presented before to enhance the user trust in the FarmBuddy chatbot. Aligning with the results of this study, two notable factors to consider are Transparency and Risk due to the fact that both of these were impactful in regards to user experience and continuance intention. For instance, chatbots associated to reputable brands or organizations, which offer transparency, tend to have higher trust, leading to better overall user experience (White et al., 2022). The findings on trust as a measure of user experience and continuance intention offer a comprehensive understanding of the multidimensional nature of user trust in this context and its vital role accross a wide variety of applications. By focusing on the aforementioned areas, we can create a more trustworthy chatbot that also encourages an overall better user experience and long-term engagement and loyalty to both the brand and the chatbot itself. This approach is key in maintaining high levels of trust and ensuring success for the FarmBuddy chatbot, or any other similar chatbot, over the long term.

5.5 Limitations

While this study provides valuable insights into the factors affecting trust and overall user trust toward the FarmBuddy chatbot, there are several limitations that must be acknowledged.

As previously mentioned, due to limited time and availability the sample size is relatively small which did not allow for a wide variety of statistical analysis and limited the ability to reliably bootstrap the PLS-SEM model. Furthermore, the sample population itself is composed of consortium members given that the chatbot is not yet publicly available, this may bias the findings and limit their generalizability. The other main limitation of the study is the removal of the factor of "Human-likeness" due to issues with misinterpretation and clarity. This is a limitation since it does not consider a potentially important factor, which might have provided valuable insights into the factors affecting trust.

Future research should consider including the aforementioned factor in the analysis, increasing the sample size, and making use of active users through random sampling techniques to achieve more generalizable results.

6 Conclusion

Taking into account everything stated above, this study provides vital insights into the factors influencing user trust in the FarmBuddy chatbot and evaluates the extent to which users perceive the chatbot as being trustworthy. This research was conducted by using a validated questionnaire, developed through a pilot test and subsequent PLS-PM analysis, in order to gather the necessary quantitative data. The findings highlight the influence of non-technical factors, such as Appearance, Risk, Brand, and Transparency, in relation to their influence on trust. On the other hand, out of the technical aspects, only Competence appeared to have relative importance, even though this factor has seen a significant decrease in importance. This shift may suggest a response to advancements in chatbot technology by the factors affecting user trust, indicating the need for ongoing research to adapt to the technological landscape.

User trust towards the FarmBuddy chatbot was found to be overall positive with a mean and median overall trust score (4.514 and 4.800) significantly above the midpoint of the 7-point Likert scale. This finding suggest a strong foundation with room for improvement and further development. Moreover, demographic factors were found to be of non-significant influence on trust, implying the need for consistency across user groups, or a re-evaluation with a broader more representative user base in the future.

Furthermore, user trust was identified as a critical component of user experience and continuance intention highlighting the importance of maintaining high trust scores to ensure overall positive interactions and user experience, as well as fostering a sustained relationship with the users and long-term engagement.

Overall, despite the limitations the study offers valuable contributions to the literature, as well as other researchers and developers related to this increasingly important and prevalent field. Future research should aim to address the limitations presented before to ensure robust findings and aid in the process of maintaining and enhancing user trust in chatbots.

Acknowledgments and Disclosure of Funding

First of all, I would like to express my deepest gratitude to my supervisors, Daan, Liv, and Stephan, for their invaluable guidance, support, and encouragement throughout the whole process of writing this thesis. Your insights, patience, and belief in my work has allowed for all of this to happen as smoothly as possible. I have learned so much from each of you, and I am incredibly grateful for this opportunity that you have given me to have worked under your supervision.

I also want to extend my sincere thanks to TNO and Leiden University for providing me with the resources and environment necessary to conduct this research appropriately. The support from both institutions has been vital to the successful completion of this project.

I appreciate you all for making this journey possible.

Appendix A.

The structure of this table is based on *Appendix 1* (Nordheim et al., 2019). The deleted items are highlighted in red. The user answers each item by selecting one of the options from the 7-point Likert scale ranging from *Strongly Agree* to *Strongly Disagree*.

The structure of this table is based on *Appendix 1* (Nordheim et al., 2019). The deleted items are highlighted in red. The user answers each item by selecting one of the options from the 7-point Likert scale ranging from *Strongly Agree* to *Strongly Disagree*.

Table 13: Questionnaire items used in the study and their original sources

Questionnaire		
Factors	**Questionnaire Items**	**Source**
Familiarity	I am familiar with the way chatbots and/or other conversational agents work	How familiar are you with the way chatbots and/or other conversational agents work? (Pollmann, 2021)
	I have never used a chatbot or a conversational interface before	Have you used a chatbot or a conversational interface before? (Pollmann, 2021) [Reverse coded]
	I am confident using a chatbot and/or conversational interface	How confident do you feel using a chatbot and/or conversational interface? (Pollmann, 2021)
	I do not often use chatbots and/or other conversational interfaces	How often do you use chatbots and/or other conversational interfaces? (Pollmann, 2021) [Reverse coded]
	I am unfamiliar with chatbots and/or other conversational agents	How familiar are you with chatbots and/or other conversational agents? (Pollmann, 2021) [Reverse coded]
Propensity to Trust	Even under doubt I will choose to trust new technology	Even under doubt I will choose to trust new technology (Nordheim, 2018)
	I generally distrust new technology until they give me a reason to trust	I generally trust new technology until they give me a reason not to (Nordheim, 2018) [Reverse coded]
	It is easy for me to trust new technology	It is easy for me to trust new technology (Nordheim, 2018)
	My typical approach is to distrust new technology	My typical approach is to trust new technology (Nordheim, 2018) [Reverse coded]
	My tendency to trust new technology is high	My tendency to trust new technology is high (Nordheim, 2018)

Questionnaire (continued)		
Factors	**Questionnaire Items**	**Source**
Competence	I feel very sure about the chatbot's competence	I feel very sure about the chatbots competence (Nordheim, 2018)
	The content of the chatbot reflects lack of expertise	The content of the chatbot reflects expertise (Nordheim, 2018) [Reverse coded]
	The chatbot is well equipped for the task it is set to do	The chatbot is well equipped for the task it is set to do (Nordheim, 2018)
	The chatbot does not appear to be knowledgeable	The chatbot appears knowledgeable (Nordheim, 2018)[Reverse coded]
	The chatbot seemed to have a correct interpretation of my questions or requests	*Formulated by the researcher*
	I experienced difficulty getting my question or request answered	*I experienced to get my question answered (Nordheim, 2018) [Reverse coded]*
Understandability	The chatbot is easy to use	The chatbot is easy to use (Nordheim, 2018)
	My dialogue with the chatbot was confusing and unclear	My dialogue with the chatbot was clear and understandable (Nordheim, 2018) [Reverse coded]
	This chatbot is flexible to interact with	*This chatbot is flexible to interact with (Gao et al., 2018; Nordheim, 2018)*
	It was difficult for me to learn how to use this chatbot	*It was easy for me to learn how to use this chatbot [Reverse coded]*
	I feel it is difficult to get the chatbot to do what I want it to do	I feel it is easy to get the chatbot to do what I want it to do (Gao et al., 2018; Nordheim, 2018) [Reverse coded]
Appearance	*I think the chatbot was thoughtfully developed*	*Formulated by the researcher based on feedback from Følstad et al. (2018, p. 10)*
	The visual appearance of the chatbot made it harder for me to use	The visual appearance of the chatbot made it easier for me to use (Essop et al., 2023) [Reverse coded]
	I find the chatbot's interface usable and functional in providing answers to my questions	I find the chatbot's interface usable and functional in providing answers to my questions (Essop et al., 2023)
	The chatbot uses inadequate and incorrect language	*Formulated by the researcher based on feedback from Følstad et al. (2018, p. 10) [Reverse coded]*
	The chatbot has a nice appearance	The chatbot has a nice appearance (Nordheim, 2018)

Questionnaire (continued)		
Factors	**Questionnaire Items**	**Source**
	I felt confused navigating the different areas of the chatbot	*I felt confident navigating the different areas of the chatbot (Essop et al., 2023) [Reverse coded]*
Brand	The chatbot and/or brand that developed the chatbot is respected by others	The chatbot is respected by others (Nordheim, 2018)
	I have heard others talking negatively about this chatbot and/or brand that developed the chatbot	I have heard others talking positively about this chatbot (Nordheim, 2018) [Reverse coded]
	I believe that the brand that developed the chatbot had the users' best interest in mind	*Formulated by the researcher*
	The chatbot and/or brand that developed the chatbot has a bad reputation	*The chatbot has a good reputation (Nordheim, 2018) [Reverse coded]*
	I believe that the brand that developed the chatbot genuinely aimed to help users with their queries	Formulated by the researcher
	The chatbot and/or brand that developed the chatbot is not well known by others	*The chatbot is well known by others (Nordheim, 2018) [Reverse coded]*
Human-Likeness	*The chatbot is humanlike*	*The chatbot is humanlike (Nordheim, 2018)*
	The chatbot is unrealistic	*The chatbot is realistic (Nordheim, 2018) [Reverse coded]*
	The chatbot is present	*The chatbot is present (Nordheim, 2018)*
	The chatbot is unnatural	*The chatbot is natural (Nordheim, 2018) [Reverse coded]*
	The chatbot is authentic	*The chatbot is authentic (Nordheim, 2018)*
Likeability	I like this chatbot	I like this chatbot (Nordheim, 2018)
	The chatbot is not friendly	*The chatbot is friendly (Nordheim, 2018)[Reverse coded]*
	The chatbot is comfortable to talk to	The chatbot is comfortable to talk to (Nordheim, 2018)
	The chatbot did not show empathy and understanding in its responses	Formulated by the researcher [Reverse coded]
	The chatbot is accommodating	The chatbot is accommodating (Nordheim, 2018)
Predictability	The chatbot behaved as predicted	The chatbot behaved as predicted (Nordheim, 2018)

Questionnaire (continued)		
Factors	**Questionnaire Items**	**Source**
	There were surprises in how the chatbot answered me	There were no surprises in how the chatbot answered me (Nordheim, 2018) [Reverse coded]
	I think it is predictable that the chatbot has the type of content it does	I think it is predictable that the chatbot has the type of content is does (Nordheim, 2018)
	The chatbot behaved unpredictably	The chatbot behaved predictable (Nordheim, 2018)[Reverse coded]
	The content of the chatbot was as expected	*The content of the chatbot was as expected (Nordheim, 2018)*
Transparency	The chatbot clearly communicated what it can do, and how it can help	Formulated by the researcher based on feedback from Følstad et al. (2018, p. 10)
	I am poorly informed how the chatbot works	*I am well informed how the system works (Roesler et al., 2022)[Reverse coded]*
	I understand how the chatbot works	*I understand how the system works (Roesler et al., 2022)*
	The chatbot was honest towards its shortcomings	Formulated by the researcher based on feedback from Følstad et al. (2018, p. 10)
	The way the chatbot works is unclear to me	*The way the system works is clear to me (Roesler et al., 2022)[Reverse coded]*
Risk	I think there could be negative consequences when I use this chatbot	I think there could be negative consequences when I use this chatbot (Nordheim, 2018)
	I feel it is safe to talk to this chatbot	I feel it is unsure to talk to this chatbot (Nordheim, 2018) [Reverse coded]
	I feel I must be cautious when I use this chatbot	I feel I must be cautious when I use this chatbot (Nordheim, 2018)
	I feel secure when I interact with the chatbot	I feel vulnerable when I interact with this chatbot (Nordheim, 2018) [Reverse coded]
	I feel there is risk involved by talking to this chatbot	*I feel there is risk involved by talking to this chatbot (Nordheim, 2018)*
Trust	I experience this chatbot as trustworthy	I experience this chatbot as trustworthy (Nordheim et al., 2019)
	I'm suspicious of this chatbot	I'm suspicious of this chatbot (Nordheim et al., 2019)[Reverse coded]

Questionnaire (continued)		
Factors	**Questionnaire Items**	**Source**
	I trust this chatbot	I trust this chatbot (Nordheim et al., 2019)

Appendix B.

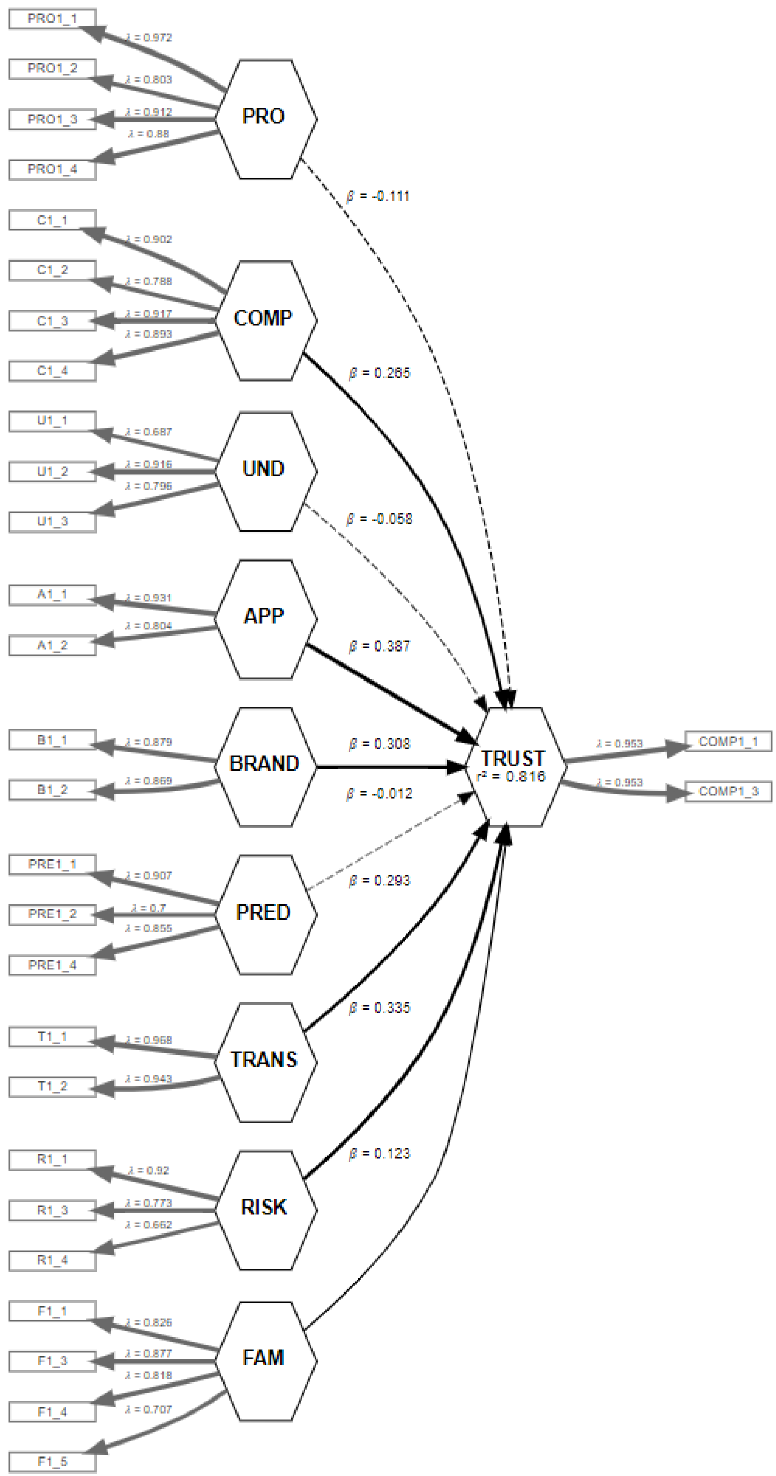

Figure 3: PLS-SEM model

References

(n.d.). https://eufarmbook.eu/en/about

Ab Hamid, M. R., Sami, W., & Sidek, M. M. (2017). Discriminant validity assessment: Use of fornell & larcker criterion versus htmt criterion. *Journal of physics: Conference series*, *890*(1), 012163.

Acharya, N., Sassenberg, A.-M., & Soar, J. (2022). Consumers' behavioural intentions to reuse recommender systems: Assessing the effects of trust propensity, trusting beliefs and perceived usefulness. *Journal of Theoretical and Applied Electronic Commerce Research*, *18*(1), 55–78.

Ahmad, S., Zulkurnain, N., & Khairushalimi, F. (2016). Assessing the validity and reliability of a measurement model in structural equation modeling (sem). *British Journal of Mathematics & Computer Science*, *15*(3), 1–8.

Alosani, M. S., Al-Dhaafri, H. S., & Yusoff, R. Z. B. (2016). Mechanism of benchmarking and its impact on organizational performance. *International Journal of Business and Management*, *11*(10), 172–183.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., et al. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC medical education*, *23*(1), 689.

Asfoura, E., Kassem, G., Alhuthaifi, B., & Belhaj, F. (2023). Developing chatbot conversational systems & the future generation enterprise systems. *International Journal of Interactive Mobile Technologies*, *17*(10).

Bae, S., Lee, Y. K., & Hahn, S. (2023). Friendly-bot: The impact of chatbot appearance and relationship style on user trust. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Baek, T. H., & Kim, M. (2023). Is chatgpt scary good? how user motivations affect creepiness and trust in generative artificial intelligence. *Telematics and Informatics*, *83*, 102030.

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: Does it exist. *Proceedings of conference of human computer interaction, workshop on human animated character interaction.*

Castro, M. S., Bahli, B., Ferreira, J. J., & Figueiredo, R. (2023). Comparing single-item and multi-item trust scales: Insights for assessing trust in project leaders. *Behavioral Sciences*, *13*(9), 786.

Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., Fraternali, P., Quarteroni, S., Ceri, S., Bozzon, A., Brambilla, M., Della Valle, E., et al. (2013). An introduction to information retrieval. *Web information retrieval*, 3–11.

Chin, C.-L., & Yao, G. (2021). Convergent validity. In *Encyclopedia of quality of life and well-being research* (pp. 1–2). Springer.

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of chatgpt: Survey analysis. *Journal of Medical Internet Research*, *25*, e47184.

Daqar, M. A. A., & Smoudy, A. K. (2019). The role of artificial intelligence on enhancing customer experience. *International Review of Management and Marketing*, *9*(4), 22.

Daud, K. A. M., Khidzir, N. Z., Ismail, A. R., & Abdullah, F. A. (2018). Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at pengkalan datu river. *International Journal of Development and sustainability*, *7*(3), 1026–1037.

El Bakkouri, B., Raki, S., & Belgnaoui, T. (2022). The role of chatbots in enhancing customer experience: Literature review. *Procedia Computer Science*, *203*, 432–437.

Emerson, R. W. (2022). Anova assumptions.

Essel, H. B., Vlachopoulos, D., Tachie-Menson, A., Johnson, E. E., & Baah, P. K. (2022). The impact of a virtual teaching assistant (chatbot) on students' learning in ghanaian higher education. *International Journal of Educational Technology in Higher Education*, *19*(1), 57.

Essop, L., Singh, A., & Wing, J. (2023). Developing a comprehensive evaluation questionnaire for university faq administration chatbots. *2023 Conference on Information Communications Technology and Society (ICTAS)*, 1–7.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–191.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: Findings from a questionnaire study. *Quality and User Experience*, *5*(1), 3.

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What makes users trust a chatbot for customer service? an exploratory interview study. *Internet Science: 5th International Conference, INSCI 2018, St. Petersburg, Russia, October 24–26, 2018, Proceedings 5*, 194–208.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39–50.

Funke, N., Stadler, K., Vakkuri, H., Wagner, A., Lunkenheimer, M., & Kracklauer, A. (2023). "your conversational partner is a chatbot": An experimental study on the influence of chatbot disclosure and service outcome on trust and customer retention in the fashion industry. *Journal of Applied Interdisciplinary Research*, (1), 10–27.

Gao, M., Kortum, P., & Oswald, F. (2018). Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proceedings of the human factors and ergonomics society annual meeting*, *62*(1), 1414–1418.

Glaser, A. Y., Hall, C. B., Uribe S, J. I., & Fried, M. P. (2006). Medical students' attitudes toward the use of an endoscopic sinus surgery simulator as a training tool. *American journal of rhinology*, *20*(2), 177–179.

Granić, A. (2017). Technology in use: The importance of good interface design. *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)*, 43–49.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of pls-sem. *European business review*, *31*(1), 2–24.

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial least squares structural equation modeling (pls-sem) using r: A workbook*. Springer Nature.

Hanafiah, M. H. (2020). Formative vs. reflective measurement model: Guidelines for structural equation modeling research. *International Journal of Analysis and Applications*, *18*(5), 876–889.

Haugeland, I. K. F., Følstad, A., Taylor, C., & Bjørkli, C. A. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, *161*, 102788.

Ivanov, A. V., Akiba, E. A. G., & Konov, K. K. (2021). The smartphone consumer decision-makingprocess of university students in sweden: The case of iphone.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, *7*(4), 396–403.

Kim, M., & Chang, B. (2020). The effect of service quality on the reuse intention of a chatbot: Focusing on user satisfaction, reliability, and immersion. *International Journal of Contents*, *16*(4), 1–15.

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics 20*, 13–30.

Kvale, K., Freddi, E., Hodnebrog, S., Sell, O. A., & Følstad, A. (2020). Understanding the user experience of customer service chatbots: What can we learn from customer satisfaction surveys? *International Workshop on Chatbot Research and Design*, 205–218.

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? a study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, *17*(1), 39–71.

Lin, H.-C., Chang, T.-Y., & Kuo, S.-H. (2018). Effects of social influence and system characteristics on traceable agriculture product reuse intention of elderly people: Integrating trust and attitude using the technology acceptance model. *Journal of research in education sciences*, *63*(3).

Liu, W., Jiang, M., Li, W., & Mou, J. (2024). How does the anthropomorphism of ai chatbots facilitate users' reuse intention in online health consultation services? the moderating role of disease severity. *Technological Forecasting and Social Change*, *203*, 123407.

Ltifi, M. (2023). Trust in the chatbot: A semi-human relationship. *Future Business Journal*, *9*(1), 109.

Matook, S., Brown, S. A., & Rolf, J. (2015). Forming an intention to act on recommendations given via online social networks. *European Journal of Information Systems*, *24*, 76–92.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, *20*(3), 709–734.

McKnight, D. H. (2005). Trust in information technology. *The Blackwell encyclopedia of management*, *7*, 329–331.

Mozafari, N., Weiger, W. H., & Hammerschmidt, M. (2022). Trust me, i'm a bot–repercussions of chatbot disclosure in different service frontline settings. *Journal of Service Management*, *33*(2), 221–245.

Nemoto, T., & Beglar, D. (2014). Likert-scale questionnaires. *JALT 2013 conference proceedings*, *108*(1), 1–6.

Nguyen, D. M., Chiu, Y.-T. H., & Le, H. D. (2021). Determinants of continuance intention towards banks' chatbot services in vietnam: A necessity for sustainable development. *Sustainability*, *13*(14), 7625.

Nordheim, C. B. (2018). *Trust in chatbots for customer service–findings from a questionnaire study* [Master's thesis].

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service—findings from a questionnaire study. *Interacting with Computers*, *31*(3), 317–335.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, *15*, 625–632.

Paulhus, D. L. (1991). Measurement and control of response bias. *Academic Press*.

Perneger, T. V., Courvoisier, D. S., Hudelson, P. M., & Gayet-Ageron, A. (2015). Sample size for pre-tests of questionnaires. *Quality of life Research*, *24*, 147–151.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of applied psychology*, *98*(1), 194.

Pollmann, N. (2021). *Testing a usability scale for chatbots: The effect of familiarity on satisfaction ratings* [B.S. thesis]. University of Twente.

Przegalinska, A., Ciechanowski, L., Stroz, A., Gloor, P., & Mazurek, G. (2019). In bot we trust: A new methodology of chatbot performance measures. *Business Horizons*, *62*(6), 785–797.

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human–Computer Interaction*, *37*(1), 81–96.

Roesler, E., Rieger, T., & Manzey, D. (2022). Trust towards human vs. automated agents: Using a multidimensional trust questionnaire to assess the role of performance, utility, purpose, and transparency. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 2047–2051.

Salah, M., Alhalbusi, H., Ismail, M. M., & Abdelfattah, F. (2023). Chatting with chatgpt: Decoding the mind of chatbot users and unveiling the intricate connections between user perception, trust and stereotype perception on self-esteem and psychological well-being. *Current Psychology*, 1–16.

Sarstedt, M., & Wilczynski, P. (2009). More for less? a comparison of single-item and multi-item measures. *Die Betriebswirtschaft*, *69*(2), 211.

Shahzad, M. F., Xu, S., An, X., & Javed, I. (2024). Assessing the impact of ai-chatbot service quality on user e-brand loyalty through chatbot user trust, experience and electronic word of mouth. *Journal of Retailing and Consumer Services*, *79*, 103867.

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social indicators research*, *102*, 443–461.

Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, *25*(3), 274–287.

Song, S. W., & Shin, M. (2024). Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: The moderating role of avatar familiarity. *International Journal of Human–Computer Interaction*, *40*(2), 441–456.

Tan, S.-M., Liew, T. W., et al. (2022). Multi-chatbot or single-chatbot? the effects of m-commerce chatbot interface on source credibility, social presence, trust, and purchase intention. *Human Behavior and Emerging Technologies*, *2022*.

Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International journal of medical education*, *2*, 53.

Vatanasombut, B., Igbaria, M., Stylianou, A. C., & Rodgers, W. (2008). Information systems continuance intention of web-based applications customers: The case of online banking. *Information & management*, *45*(7), 419–428.

White, B. K., Martin, A., & White, J. A. (2022). User experience of covid-19 chatbots: Scoping review. *Journal of medical Internet research*, *24*(12), e35903.

Wijaya, H. D., Gunawan, W., Avrizal, R., & Arif, S. M. (2020). Designing chatbot for college information management. *IJISCS (International Journal of Information System and Computer Science)*, *4*(1), 8–13.

Xu, Y., Bradford, N., Garg, R., et al. (2023). Transparency enhances positive perceptions of social artificial intelligence. *Human Behavior and Emerging Technologies*, *2023*.

Yan, M., Filieri, R., & Gorton, M. (2021). Continuance intention of online technologies: A systematic literature review. *International Journal of Information Management*, *58*, 102315.

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating effects of user experience and system transparency on trust in automation. *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, 408–416.

Zwass, V. (2024). Information system. encyclopedia britannica.