Designing for explanation-driven trust in Chatbots

Delft University of Technology

Student: Zhengquan Zhang 5518717

Supervisors: Christina Schneegass, Konstantinos Tsiakas

ABSTRACT

Trust plays an important role in the implementation of chatbot technology. This study was also focusing on user trust in chatbots, particularly focusing on the role of response delay and explanation-driven driven subjective transparency. This research includes a pretest and a main test. In the pretest, we selected one explanation that was perceived by the participants that can raise the most social presence feeling as well as the subjective transparency of the chatbot. In the main test, a 2 × 2 between-subject experiment was designed and conducted to test the hypotheses. First, the findings revealed that while response delay did not significantly influence trust or social presence, clear explanations, especially in the context of instant delays, positively impacted subjective transparency and trust. Second, the study reinforced the positive correlation between social presence and trust, subjective transparency, and trust. From a practical perspective, the research offers insights for chatbot design, emphasizing the importance of improving subjective transparency and rendering a more natural and human-like interaction.

Keywords

Artificial intelligence (AI), generative AI, chatbot, response delay, social presence, subjective transparency, chatbot trustworthiness, user trust

INTRODUCTION

Conversational systems, often known as chatbots, have become a ubiquitous part of everyday life and are increasingly deployed for use in large-scale personal and social settings, e.g., to assist with daily tasks or to inform decision-making (Meshram et al., 2021; Stieglitz et al., 2021; Zhou et al., 2015). While chatbots are typically applied for text-based interaction, they may also encompass voice-based conversational agents such as Apple's Siri and Amazon's Alexa. Chatbots are used for a range of application areas such as customer support (Xu et al., 2017), health (Fryer et al., 2017), and education (Fryer et al., 2017), in addition to marketing, entertainment, and general assistance with simple tasks.

In the past years, chatbots got increasingly interactive – specifically in regards to exhibiting anthropomorphic features – and intelligent – based on advancements in the domain of machine learning (Maedche et al., 2016). For example, there is a sharp growth in the use of conversational AI models, such as ChatGPT (https://chat.openai.com/). The model of ChatGPT is trained on a large collection of datasets from diverse data sources, to generate human-like answers. Researchers have foreseen a huge potential in empowering the chatbot with large language models like ChatGPT in productivity (Surameery & Shakor, 2023), healthcare (Biswas, 2023), and education (Mhlanga, 2023) domains. Nowadays chatbots are able to act and interact in an increasingly human way. Thus, the boundary between man and machine becomes increasingly blurred from a user perspective. On the other hand, AI-infused systems raise the opaqueness and complexity for the user (Maedche et al., 2019), therefore magnifying the issue of trust. It is argued that building trust is an essential means to address complexity and uncertainty because humans cannot have complete knowledge of most systems' inner processes.

Recent evidence suggests a positive relation between trust and anthropomorphic cues such as human-like appearance and voice output. Research suggests that the more human-like AI systems appear, the more likely users are to trust them (Kulms & Kopp, 2019). Also, employing social cues in chatbot design is a common way to make the human–chatbot interactions more intuitive and human-like. In the chatbot context, social cues are design features that can trigger similar user reactions regarding emotion, cognition, or behavior that is similar to the reactions observed in the human-human interaction (Feine et al., 2019). Previous research based on Social Response Theory (SRT) by Nass et al. (1994) and Nass and Moon (2000) have demonstrated that various social cues exhibited by chatbots, including human-resembling avatars, casual conversation, and name identifiers, lead to enhanced user views and actions, as supported by findings from Araujo (2018), Diederich et al. (2020), Benlian et al. (2020), and Seeger et al. (2021).

Response delay is another important social cue in technology-mediated interaction. But previous research did not illustrate a clear picture of the relationship between response delay and user trust (Walther and Tidwell 1995; Jacquet et al., 2019). In the context of chatbots, unlike humans who need time to read a message and enter a response, retrieval-based chatbots can process user input instantly and give a response to the user (Luo et al., 2022). Yet, compared with a chatbot that can adapt the delay time based on the length of the response, some studies suggest the instant response can reduce the natural conversation feeling, hence decreasing the user's satisfaction feeling while interacting with a chatbot (Gnewuch et al., 2018). In contrast, previous researches also show that chatbots with response delay are less likable to the users (Schanke et al., 2021). In the online customer service context, previous studies have found that compared to nearly instant response the dynamic response delay of the chatbot can improve the perceived social presence (Gnewuch et al., 2018) and trustworthiness (Karma 2020). But there are only a few research studies on the relationship between response delay and user trust. Hence, research question one is proposed below.

RQ1: "How does the response delay influence the user's trust in the chatbot?"

Transparency of the system is another vital factor that influences user trust. Addressing a better explanation of the underlying process/status of the system to the user can impact the user perceived usefulness and trustworthiness of the system (Kizilcec, 2016, Kulesza et al., 2012, Lim et al., 2009). However, the mismatch between the actual trustworthiness and user-perceived trustworthiness can increase user tolerance (Piccolo et al., 2019). For the systems' most effective use, users must trust respective systems appropriately, this is often referred to as trust calibration (Lee and See., 2004). In the industry world, while interacting with a chatbot there always has some response delay caused by different factors (internet connection, service load, etc,.). Depending on the factors the response delay can be short or long. Taking the technology factor as an example, retrieving-based or generating-based are two common mechanisms of response generation chatbots. Retrieving-based chatbots usually search for the response in their existing knowledge base, so the response time usually is fast. While a generating-based chatbot needs to generate the response from its pre-trained AI model, this could end up with a relatively longer response delay. Now with the rise of the big language models like T5, BERT, and GPT-3 (Raffel et al., 2020; Brown et al., 2020; Devlin et al., 2019), the generating-based chatbot perceived by the user, even the chatbot developer, as a "black box. This makes it difficult to understand the underlying process of the chatbot (Khurana et al., 2021). For example, two of the most popular applications: ChatGPT (https://chat.openai.com/) and NewBing (https://news.microsoft.com/the-new-Bing/), both implemented with generating-based chatbot technology. CharGPT outputs the response word by word with small delays in between while NewBing outputs the response paragraph by paragraph with longer delays in between. But they all do not provide information about what is actually going on with the system to the users regarding the delay. Which ends up with a less transparent chatbot system. So, regarding the response delay, implementing corresponding explanations for it can also have the potential in facilitating the user's trust in the chatbot. Hence research question 2 is proposed in this paper.

RQ2: "How does the explanation of the response delay influence the user's trust in the chatbot?"

THEORY

Social response theory

When people interact with each other, they can identify many kinds of social cues like facial expressions and hand movements (Burgoon et al., 2010). Following the idea that computers can also convey social cues, the SRT theory indicates that users respond to the technology in a similar way to the identified social cues (e.g., natural language, human-like appearance) (Reeves and Nass, 1996; Nass and Moon, 2000). In an experiment conducted by Nass et al., in 1994, while interacting with digital tutoring, participants perceived praise from a different computer as more accurate and friendly, and they believed they performed better in this condition. Also, the social responses that automatically and unconsciously arose between users and technology seem not “confined to a certain category of people” (Nass and Moon, 2000). When users interact with technology in a social manner, they subconsciously view it as a pertinent social entity, attributing human characteristics to it, thereby amplifying their sense of its social presence (Nass and Lee, 2001).

In 1978, Short et al. introduced the concept of "social presence," aiming to capture how strongly a communication medium can make users feel the psychological presence of others. This was defined as “the degree of salience of the other person in a mediated communication and the consequent salience of their interpersonal interactions” (Short et al., 197). Over time, this definition evolved, highlighting the capacity of communication mediums to foster sensations of warmth, personal connection, and sociability, even in the absence of direct human-to-human contact (Gefen and Straub 2004).

For the implementation of the SRT, previous research found that users felt a greater sense of social presence and trust when the robot interaction was designed to be more human-like and spontaneous (Xu, Chen, & You, 2023). While for chatbots, researchers have primarily investigated verbal cues (e.g., human-like language such as "hello" and "goodbye.") humorous comments and visual cues (e.g., human names, human-like avatars) to evoke a social presence feeling while interacting with the chatbot (e.g., Araujo, 2018; Moussawi and Benbunan-Fich 2020; Diederich et al., 2020; Seeger et al., 2021). For example, in 2020, Moussawi and Benbunan-Fich illustrated that talking in a humorous way can make the chatbot appear more human-like. In their study, participants asked “Where did you get that name from?” instead of answering “My programmers named me.” a humorous chatbot answered, “Well, once upon a time, there was a princess … Just kidding. My technology overlords gave it to me. I wanted to be called Princess Jenna but they didn’t like it.”. These studies provide valuable insights on the impact of verbal and visual social cues, but regarding other types of social cues for example response time less is known (Feine et al., 2019).

In the chatbot domain, previous research has also indicated that the social presence of the chatbot can influence user trust (Zierau et al., 2021, Konya-Baumbach et al., 2023). For example, in 2023, Konya-Baumbach showed that in the shopping context, the social presence raised by chatbot anthropomorphism, such as a first-person description as opposed to a third-person description, can positively influence user trust, purchase intention, word of mouth, and satisfaction with the shopping experience.

Response delay as a social cue

The concept of "response delay" is categorized under chronemic cues, which focus on the time-related elements in communication (Walther and Tidwell 1995; Littlejohn and Foss, 2009). This cue holds significant importance in interactions between humans (Kalman et al., 2013; Schuetzler et al., 2019). For example, a fast response can indicate interest, attentiveness, and competence, while a slow response can indicate disinterest, distraction, or incompetence. In human-human interaction, response time refers to the time between a person ends speaking and another one start speaking. In interactions facilitated by technology, like online messaging, response delay pertains to the time taken by an individual to reply and the interval between messages (Moon 1999). This encompasses the duration required to read and comprehend another's message, as well as the time spent editing one's own message (Derrick et al., 2013).

Previous research found that response delay as a social cue can raise the social presence while interacting with the chatbot (Gnewuch et al., 2018, Gnewuch et al., 2022). In a 2022 study, Gnewuch found that when chatbots have a delayed response time, which mirrors human-like behavior more than an immediate reply, it can enhance beginner users' feelings of social presence and their intent to use the chatbot. While the effect is negative for experienced users. While the impact of response delay on users' trust in chatbots is not a popular topic. Only a little research was conducted on this topic. In 2020, Karma found that response latency has a positive effect on users' cognitive trust.

As stated before, the social cues of chatbots can increase social presence. Social presence can also positively influence the user's trust in the chatbot. Hence, regarding the response delay, we propose that social presence has a meliorate effect between response delay and user trust.

H1: User social presence mediates the positive influence between response delay and user trust in the chatbot.

H1a: The response delay of the chatbot positively influences the user's social presence while interacting with the chatbot.

H1b: User social presence positively influences the user's trust in the chatbot.

Subjective transparency

In the context of the Advice-giving System, Zhao et al. in 2019, proposed that the transparency of an intelligent system can be defended into objective transparency and subjective transparency. Objective transparency refers to the extent to which the system releases information regarding what they do and why they behave in a certain way. Subjective transparency refers to the extent to which users perceive that the information regarding what systems do and why they behave in a certain way is provided by the system and is visible/available/accessible to them (Cho et al., 2017; Zhu, 2002).

Now with the rise of the big language models such as BERT (Devlin et al., 2019), T5 (Raffel et al., 2020), and GPT-3 (Brown et al., 2020), the chatbot often appears as a "black box" to the user, making it difficult to understand the underlying process of the chatbot like why things did not work, how to determine the feasible actions, and the way to recover from a mission failure (Khurana et al., 2021). For example, GPT-3, developed by OpenAI, stands out as one of the most advanced and widely-discussed models in this domain. It's a deep learning model trained on vast amounts of text data, allowing it to generate human-like text based on the patterns it has learned. The model operates by predicting the next word in a sequence, which means when generating responses, it "types" out its answer word by word, much like a human typist. However, the sheer complexity and size of GPT-3 make it inherently non-transparent. Even the developers and researchers who built the model cannot precisely predict its outputs in specific situations. This unpredictability arises because GPT-3 does not have a fixed set of rules or logic it follows. Instead, it relies on patterns in the data it was trained on. As a result, while it can generate coherent and contextually relevant text, explaining the "why" behind its responses becomes a challenge.

The inherent lack of transparency in chatbot systems, especially in advanced models like GPT-3, can significantly influence users' perceptions of the system's usefulness and trustworthiness (Xu, 2019; Glass et al., 2008; Porcheron et al., 2018). This opacity stems from the complex nature of these models, which, while capable of generating human-like text, often operate as "black boxes" to users. Addressing this by providing explanations about the underlying processes or statuses of the system can enhance users' perceptions of its utility and reliability (Kizilcec, 2016; Kulesza et al., 2012; Lim et al., 2009).

While interacting with the chatbot mentioned before, users often experience varying degrees of delay when interacting with chatbots, attributable to factors like internet connectivity, server load, and the inherent processing time of the chatbot model. In some cases, delays are even introduced deliberately to simulate more human-like interactions. However, the potential influence of explaining these response delays on user trust remains largely unexplored. In this paper, we posit that offering explanations for response delays can increase user trust in chatbots. This enhancement in trust is achieved by positively influencing the subjective transparency of the chatbot system, making users feel more informed and in control of their interactions. Hence, hypothesis 2 was proposed.

H2: Explanation of the response delay has a positive effect on user trust in the chatbot.

H2a: Explanation of the response delay has a positive effect on user-perceived transparency of the chatbot system.

H2b: User-perceived transparency of the chatbot system has a positive effect on users' trust in the chatbot

For a broad view of trust in intelligent systems like chatbots. According to a systematic literature review conducted by Zierau et al. in 2020, there are three main types of trust in SPAs (smart personal agencies): User Interface-Driven Trust (UIDT), Interaction-Driven Trust (IDT), and Explanation-Driven Trust (EDT). UIDT relates to the static design features of SPAs, such as haptics and audio-visuals, that can enhance trust in the technology. IDT is based on the user's interactions with the SPA, such as the accuracy and reliability of the responses provided by the technology. EDT is based on the ability of the SPA to provide clear and understandable explanations for its actions and decisions. Understanding these different types of trust can help organizations design and implement SPAs that are more trustworthy and appealing to users. While UIDT and IDT in SPAs focus on creating a sense of social presence, which has been identified as an important antecedent of trust, EDT in SPAs aims at creating a sense of transparency by varying the degrees of understandability and transparency provided to the user. In 2020, ZIerau et al. proposes a particular need for research on the role of transparency in building trust in SPAs in the nascent field of EDT. In this study, our focus also falls into this dimension of chatbot trust. We focused on implementing a corresponding explanation for the response delay to build trust between the user and the chatbot. But by addressing the corresponding explanation with social cues, there is a potential to also influence the sense of social presence. For example, first-person description (Konya-Baumbach et al., 2023), and humor (Tsai et al., 2021) can potentially boost social presence. Hence, we also propose a moderator effect between the explanation and the effect between response delay and the user's social presence.

H3: Explanation of the chatbot response delay can moderate the effect between response delay and social presence.

Chatbot Taxonomy

Based on the mechanism of response generation, Chatbots can be broadly classified into two categories based on their response generation mechanism: retrieval-based and generation-based chatbots (Song et al., 2016; Wu, Li, et al., 2018).

Retrieval-based chatbots generate responses by selecting pre-existing responses from a knowledge base that best matches the user's input. These chatbots are typically used in situations where the conversation is predictable and the responses can be pre-defined. For example, customer service chatbots that provide answers to frequently asked questions are often retrieval-based. On the other hand, generation-based chatbots use machine learning algorithms to generate new responses based on the context of the conversation. These chatbots are more flexible and can generate responses that are not predefined. For example, chatbots that assist with mental health counseling or provide personalized recommendations for online shopping are often generation-based. However, generation-based chatbots can be more challenging to develop and require more advanced natural language processing techniques (Luo et al., 2022).

Retrieval-based chatbots have the advantage of being relatively easy to develop and maintain since the responses are pre-defined. They are also more reliable since the responses are based on a knowledge base and are less likely to generate inappropriate or irrelevant responses. However, retrieval-based chatbots can be limited in their ability to handle complex or unpredictable conversations, and they may not be able to provide personalized responses. Generation-based chatbots, on the other hand, have the advantage of being more flexible and able to generate responses that are not predefined. They can handle more complex and unpredictable conversations and can provide personalized responses. However, generation-based chatbots can be more challenging to develop and require more advanced natural language processing techniques. They may also generate inappropriate or irrelevant responses, especially if they are not trained on a sufficiently large and diverse dataset (Luo et al., 2022).

Retrieval-based chatbots have been around for a long time and are relatively easy to develop and maintain. However, with the recent advancements in natural language processing and machine learning, and the availability of large and diverse datasets, there has been a shift towards the development of generation-based chatbots, especially in applications that require more personalized and context-specific responses (Luo et al., 2022).

Research model

Drawing from established theories, we recognize the pivotal role of social presence and its influence on user trust, as well as the nuanced impact of response delays as a social cue. Furthermore, we delve into the realm of subjective transparency. Our research model posits that explanations for response delays can bolster user trust by enhancing the chatbot's subjective transparency. Additionally, we put our research under the real-world context under the taxonomies of chatbots in the real world, distinguishing between retrieval-based and generation-based chatbots, each with its unique implications for user interaction. As the response delay is different regarding the underlying reason and the time length. Hence, corresponding different explanations are needed. Based on these insights, we hypothesize that user social presence mediates the relationship between response delay and trust (H1), that explanations for response delays enhance user trust through increased subjective transparency (H2), and that the explanation of chatbot response delay can moderate the effect between response delay and social presence (H3). The research model is illustrated in Figure 1.

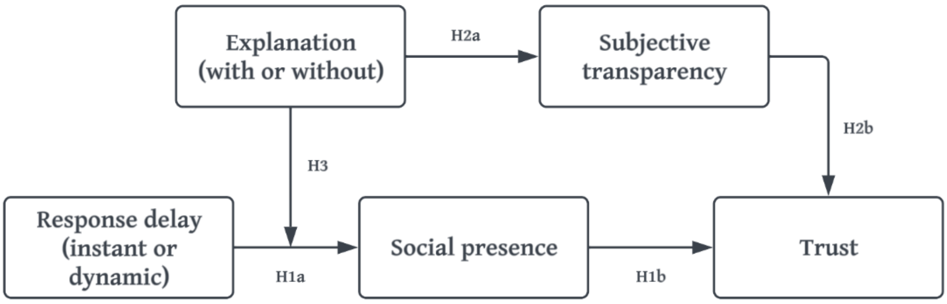

Explanation (with or without)
H2a
Subjective transparency
H2b
H3
Response delay (instant or dynamic)
H1a
Social presence
H1b
Trust

Figure 1. Research model

PRE-TEST

Existing literature does not provide definitive evidence or consensus regarding which explanations are most beneficial or effective for chatbots. Consequently, our pre-test is designed to pinpoint an explanation that robustly enhances subjective transparency and user social presence.

Delay setting

Regarding the response delay of retrieval- or generation-based chatbot systems. The retrieval-based chatbot has a static knowledge base, so it can nearly instantly respond to users' queries. While for the generation-based chatbot, it needs time to process the query and generate corresponding responses for the user's query, so the response delay dynamically changes. Hence, to mimic these two different chatbot systems, we set two different types of response delay: instant delay, and dynamic response delay. Both retrieval- and generation-based chatbots necessitate the processing of the input, which can be accomplished rapidly (Luo et al., 2022). Notably, the variation in response delay between retrieval- and generation-based chatbots primarily hinges on the answer-generating component. Hence the response delay of chatbot prototypes is focusing on mimicking the difference in the answer-generating part. And the explanations are also focusing on explaining the difference in answer-generating between retrieving- and generation-based chatbots.

Furthermore, generative AI models introduce an element of unpredictability, as the length of the generated response cannot be predetermined, potentially affecting response time. To ensure a controlled study environment in our pre-test, we chose to develop a chatbot built on a specific knowledge base rather than a full-fledged language model, which will be explained in detail in the chatbot prototype session. This approach allows us to artificially regulate response delays, offering a more controlled simulation of real-world interactions.

Stimuli

Explanations

As shown in Table 1, 5 different versions of the explanation were proposed for each delay condition (instant or dynamic). "Basic" explanation has the basic information about the underlying process of providing a response to a question. "First-person" explanation implements a first-person description of the explanation. Because previous studies indicate that first-person description can potentially boost social presence (Konya-Baumbach et al., 2023). "Detail" explanations provide detailed information about the response process of the chatbot. As the subjective transparency level is expected to be improved by providing more information for the responding process. "First-person + detail" explanation combines first-person description with detailed information. Based on the "First-person + details" version, the "Humor" explanation implemented a humorous way of talking. Because previous studies indicate that humorous ways of talking can increase the social presence feeling in human-computer interaction (Moussawi and Benbunan-Fich, 2020).

Version	Retrieval-based (instant)	Generation-based (dynamic)
Basic	Retrieving the answer.	Generating the answer.
First-person	I am retrieving the answer for you.	I am generating the answer for you.
Detail	Searching in the knowledge base for the answer that matches the question intent most.	Using the question as a prompt to generate the answer from the pre-trained model.
Firt-person + detail	I am searching in my knowledge base for the answer that matches your question intent most.	I am using the question as a prompt to generate the answer from the pre-trained model.
Humor	I'm like a witty librarian in my knowledge base, searching for the answer that fits your question like a puzzle piece.	I channel my inner word wizardry, conjuring up sentences word by word like a mischievous magician performing a linguistic sleight of hand.

Table 1. Different versions of explanation for retrieval- and generation-based chatbot.

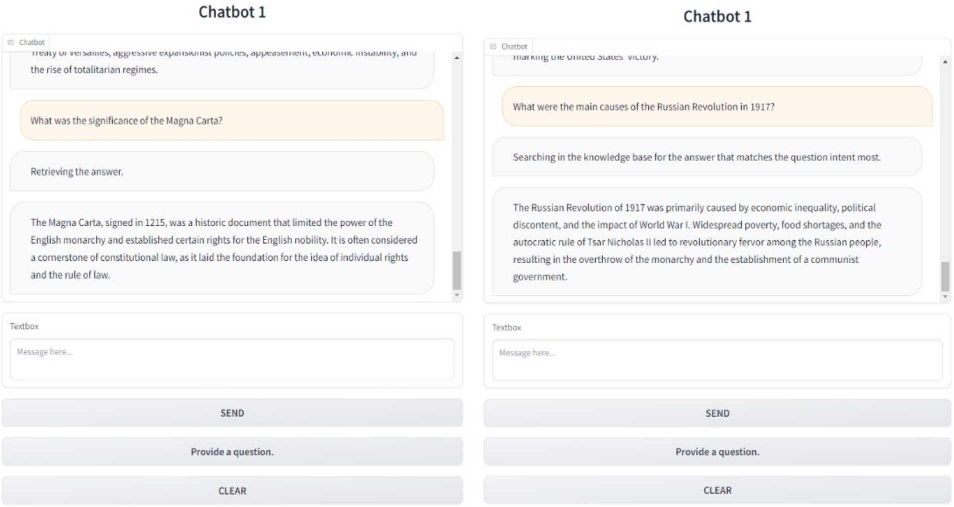
Chatbot prototype

To simulate the nuances of retrieving- and generation-based chatbots, using Python as the coding language developed chatbot prototypes from Gradio (https://gradio.app/). Gradio is a versatile tool that allows for the rapid creation of user interfaces for machine-learning models. While it's versatile in its applications, it does come with certain limitations, particularly in terms of detailed interface customization. As our focus lies on the response delay and corresponding explanation, Gradio is capable enough for building a chatbot prototype for our study.

Using Gradio can customize the response delays by coding. We utilized this feature to set near-instantaneous response times for the retrieving-based chatbot prototype. For the generation-based chatbot, we implemented a dynamic response delay, calculated based on the formula provided by Holtgraves and Han in 2007, where the delay was set at 50 milliseconds per character in the chatbot's response. While Gradio can be seamlessly integrated with various machine learning models, for the purpose of this study, we did not directly link it to ChatGPT. Instead, we preloaded answers to ensure consistency and control in our experiment. This approach allowed us to simulate the behavior of both retrieving- and generation-based chatbots without the unpredictability of real-time generation. Hence, our implementation involved scripting to manage the chatbot's behavior, especially concerning the dynamic response delay. We designed the code to recognize the length of the response and adjust the delay accordingly. This ensured that the chatbot's behavior closely mirrored the characteristics of retrieving- and generation-based systems.

For more convenient interaction, we also developed a little function to automatically provide a question to the chatbot by clicking one button. In this way, participants did not necessarily need to type the questions by themselves, as typing always along with some typing mistakes. The code of chatbot prototypes is accessible on Github: pretest chatbot prototype. While developing the chatbot prototype, significant effort went into customizing, scripting, and refining the prototypes to align with our research objectives. The screenshots of our tailored chatbot prototypes are showcased in Figure 2.

Two chatbot prototypes were developed using Gradio (https://gradio.app/) to illustrate the distinction between retrieving- and generation-based chatbots. Both prototypes employ the same underlying technology, with the primary difference being their response delay. For the prototype emulating a retrieving-based chatbot, the response delay is nearly instantaneous. In contrast, the prototype mimicking a generation-based chatbot with dynamic response delay utilizes a calculation from Holtgraves and Han in 2007, where the response delay is determined based on the number of characters in the chatbot's response, specifically 50 milliseconds per character.



Treaty of Versailles, aggressive expansionist policies, appeasement, economic instability, and the rise of totalitarian regimes.

What was the significance of the Magna Carta?

Retrieving the answer.

The Magna Carta, signed in 1215, was a historic document that limited the power of the English monarchy and established certain rights for the English nobility. It is often considered a cornerstone of constitutional law, as it laid the foundation for the idea of individual rights and the rule of law.

marking the United States' victory.

What were the main causes of the Russian Revolution in 1917?

Searching in the knowledge base for the answer that matches the question intent most.

The Russian Revolution of 1917 was primarily caused by economic inequality, political discontent, and the impact of World War I. Widespread poverty, food shortages, and the autocratic rule of Tsar Nicholas II led to revolutionary fervor among the Russian people, resulting in the overthrow of the monarchy and the establishment of a communist government.

Figure 2. Screenshots: Chatbot prototype for the pretest.

During the pre-test, we aimed to eliminate any independent influence that variations in questions and responses might have on the user-chatbot interaction. Hence, the questions and responses were standardized. Participants were required to select from the provided question list to interact with the chatbot. Participants were provided with questions centered around two topics: sustainability and history, with 16 questions and corresponding responses for each. Based on the length of both the question and its answer, these pairs were categorized into four groups: a short question with a short response, a short question with a long response, a long question with a short response, and a long question with a long response (as detailed in table 2).

Conditions	**Word count**
short question	<10
long question	around 20
short response	<25
long response	around 50

Table 2. Length of question and response.

Participants

10 participants (5 male, 5 female) were recruited with no compensation for the pre-test. They were all recruited offline in the Industrial design engineering faculty of TU Delft. Participants are all master students at the Industrial

design engineering faculty of TU Delft, aged from 23 to 26, speaking English as a second language, and having experience with chatbots, especially with daily usage of chatGPT in the recent month.

Procedural

Before conducting the pretest, we applied for ethical approval from The Human Research Ethics Committee of TU Delft and got it granted. During the pretest, all participants provided informed consent regarding the pretest.

The pretest followed a within-subject experiment procedural. During the test, a text introduction was first shown to the participants and the experimenter was there to help them understand what they needed to do. Participants were asked to interact with all chatbot settings. The chatbot set included 6 different explanation conditions (none, basic, first-person, detailed, first-person + detailed, and humor) and 2 different delay conditions (instant and dynamic). So, during the pretest, each participant was asked to interact with chatbots in 12 (6 × 2) different conditions. After finishing the interaction, the definition of social presence and subjective transparency was explained to the participants. Then, all explanations were presented to the participant, and the participant was asked to rank the explanations based on the degree they think the explanation fits with social presence and perceived transparency. In the end, a short interview was conducted to collect further feedback from the participants regarding understandability, clarity, and so on. The procedure is shown in Figure 3 below.

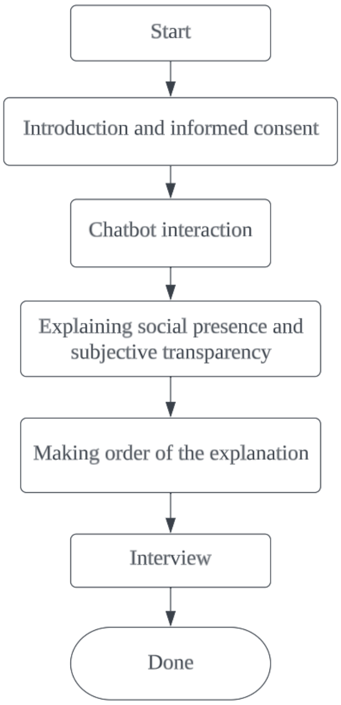
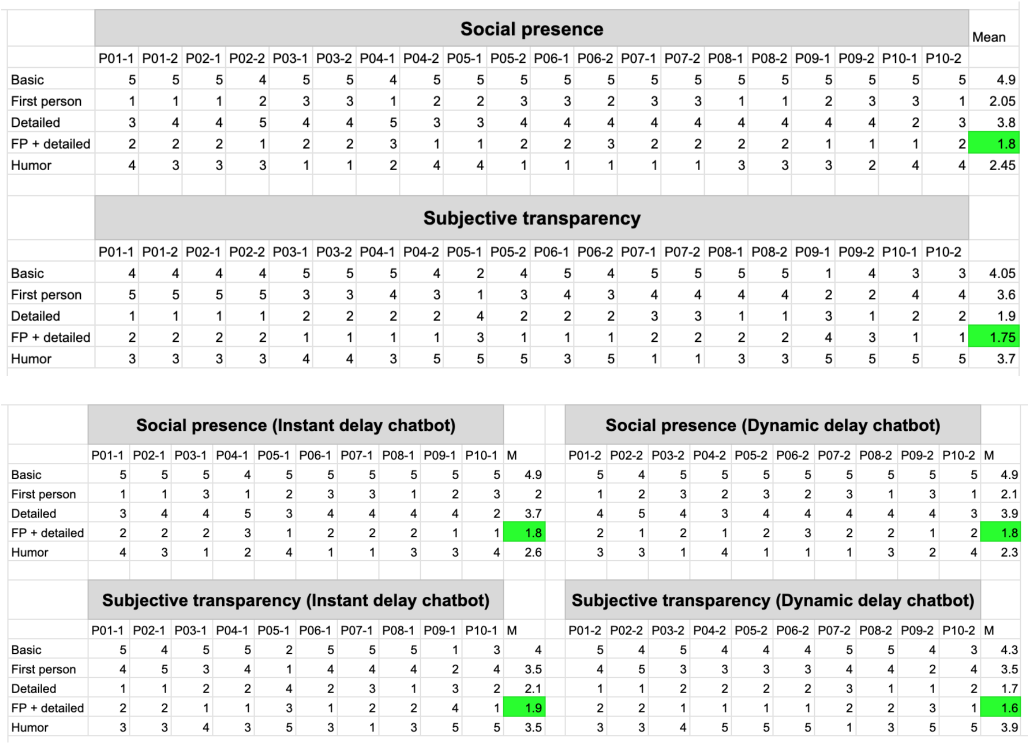

Start
Introduction and informed consent
Chatbot interaction
Explaining social presence and subjective transparency
Making order of the explanation
Interview
Done

Figure 3. Procedural for the pre-test

Results

The users' ranking for explanations was calculated into average ranking to compare their performance regarding social presence and subjective transparency. The explanation that used first-person description and more detail was ranked top regarding both social presence (M = 1.800, SD = .616) and subjective transparency (M = 1.750, SD = .851). The first-person detailed explanation also ranked top in both chatbot delay conditions regarding social presence (instant delay: M = 1.800, SD = .632; dynamic delay: M = 1.800, SD = .632) and subjective transparency (instant delay: M = 1.900, SD = .994; dynamic delay: M = 1.600, SD = .699). The overview of the results is shown below in Table 3.



	Social presence																				Mean
	P01-1	P01-2	P02-1	P02-2	P03-1	P03-2	P04-1	P04-2	P05-1	P05-2	P06-1	P06-2	P07-1	P07-2	P08-1	P08-2	P09-1	P09-2	P10-1	P10-2	
Basic	5	5	5	4	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	4.9
First person	1	1	1	2	3	3	1	2	2	3	3	2	3	3	1	1	2	3	3	1	2.05
Detailed	3	4	4	5	4	4	5	3	3	4	4	4	4	4	4	4	4	4	2	3	3.8
FP + detailed	2	2	2	1	2	2	3	1	1	2	2	3	2	2	2	2	1	1	1	2	1.8
Humor	4	3	3	3	1	1	2	4	4	1	1	1	1	1	3	3	3	2	4	4	2.45

	Subjective transparency																				
	P01-1	P01-2	P02-1	P02-2	P03-1	P03-2	P04-1	P04-2	P05-1	P05-2	P06-1	P06-2	P07-1	P07-2	P08-1	P08-2	P09-1	P09-2	P10-1	P10-2	
Basic	4	4	4	4	5	5	5	4	2	4	5	4	5	5	5	5	1	4	3	3	4.05
First person	5	5	5	5	3	3	4	3	1	3	4	3	4	4	4	4	2	2	4	4	3.6
Detailed	1	1	1	1	2	2	2	2	4	2	2	2	3	3	1	1	3	1	2	2	1.9
FP + detailed	2	2	2	2	1	1	1	1	3	1	1	1	2	2	2	2	4	3	1	1	1.75
Humor	3	3	3	3	4	4	3	5	5	5	3	5	1	1	3	3	5	5	5	5	3.7

	Social presence (Instant delay chatbot)												Social presence (Dynamic delay chatbot)										
	P01-1	P02-1	P03-1	P04-1	P05-1	P06-1	P07-1	P08-1	P09-1	P10-1	M		P01-2	P02-2	P03-2	P04-2	P05-2	P06-2	P07-2	P08-2	P09-2	P10-2	M
Basic	5	5	5	4	5	5	5	5	5	5	4.9		5	4	5	5	5	5	5	5	5	5	4.9
First person	1	1	3	1	2	3	3	1	2	3	2		1	2	3	2	3	2	3	1	3	1	2.1
Detailed	3	4	4	5	3	4	4	4	4	2	3.7		4	5	4	3	4	4	4	4	4	3	3.9
FP + detailed	2	2	2	3	1	2	2	2	1	1	1.8		2	1	2	1	2	3	2	2	1	2	1.8
Humor	4	3	1	2	4	1	1	3	3	4	2.6		3	3	1	4	1	1	1	3	2	4	2.3

	Subjective transparency (Instant delay chatbot)												Subjective transparency (Dynamic delay chatbot)										
	P01-1	P02-1	P03-1	P04-1	P05-1	P06-1	P07-1	P08-1	P09-1	P10-1	M		P01-2	P02-2	P03-2	P04-2	P05-2	P06-2	P07-2	P08-2	P09-2	P10-2	M
Basic	5	4	5	5	2	5	5	5	1	3	4		5	4	5	4	4	4	5	5	4	3	4.3
First person	4	5	3	4	1	4	4	4	2	4	3.5		4	5	3	3	3	3	4	4	2	4	3.5
Detailed	1	1	2	2	4	2	3	1	3	2	2.1		1	1	2	2	2	2	3	1	1	2	1.7
FP + detailed	2	2	1	1	3	1	2	2	4	1	1.9		2	2	1	1	1	1	2	2	3	1	1.6
Humor	3	3	4	3	5	3	1	3	5	5	3.5		3	3	4	5	5	5	1	3	5	5	3.9

Table 3. Participants' social presence and subjective transparency ranking value and the mean value of different explanations in different delay types (instant or dynamic).

Participants 3, 4, 7, 8, and 10 mentioned that the description "I'm" and "for you" is warmer and can reduce social distance hence providing a more social presence feeling. These descriptions make them believe there is a robot helping them. While the **plain explanations feel like a machine** (Participants 1, 3, 4, 5, 6). While regarding subjective transparency, participants think detailed explanation contains more information, hence improving subjective transparency. While comparing the detailed explanation and FP + detailed explanation, participants mentioned that "I'm" can make them feel the chatbot is more transparent because it can **vividly provide a scenario that there is a robot searching for or generating the answer for them** (participant 4, 5). In addition, some participants think the plain explanation makes the chatbot more transparent, but others did not appreciate the objective style of talking regarding improving transparency.

Participant 3: "…I don't think a cold machine is more transparent."

Participant 5: "…A person sits there for you, you give him/her orders, and then he/she finds the answer."

Participant 8: "…used the first-person method, …be able to decrease the distance between me and make me feel this kind of social presence feeling."

For the humor explanation, participants felt it indeed has a strong personality but is too joking, making them feel the chatbot is fooling them or trying to **hide some truth** (participant 3,) or **not sincere** (participant 4, 8, and 9). Thus the subjective transparency performance is not good. The humorous way of talking makes some participants feel the chatbot is trying to make itself perform like a human, but not so closely like a human. This also reduced the feeling of social presence (Participant 9). Another reason is the humorous way of talking may not fit well with the context, multiple participants mention that while they want to ask some knowledge questions they want the chatbot to perform professionally rather than making jokes.

Participant 3: "…it's so humorous, I always feel like it's hiding something…"

Participant 4: "…I feel that you used too many metaphors at that moment, which made you feel immature."

Participant 8: "…it is more sincere and true compared with this one (humor explanation)."

Participant 9: “…add a lot of extra stuff like this (humor explanation), it feels like it is not sincere.”

Optimizations

As shown in the result, the first-person + detailed explanation performed best regarding both social presence and subjective transparency aspects. The first-person + detailed explanation was chosen to be used in the further main test. Meanwhile, some potential optimization was also identified from the pretest regarding the chosen explanation (first-person + detailed information). Participants care about where the answer comes from. Especially the generation-based chatbot, they care about the timeliness/version of the model (Participant 5). Also, the words “pre-trained model” might confuse users' understanding (Participant 2) and reduce the feeling of social presence (Participant 10). Hence the explanation for dynamic response delay was adapted to “I am using your question as a prompt to generate the answer from my latest AI model."

8 out of 10 participants felt the response delay for the dynamic delay group was too long. Hence the way of calculating response delay was changed to using the adapted equation proposed by Gnewuch in 2018, which will be explained in more detail in the main test section. Regarding the interface design, participants felt the distinction between the explanation and the actual answer was not obvious, and created some confusion (Participants 2, 4, 8, and 9).

Participant 2: “…it (explanation) came out first, we thought that was the answer…”

Participant 9: “…it would be better if these (explanation and answer) can be distinguished from each other…”

Conclusion

Key takeaways from the pre-test include: **(1)** The first-person + detailed explanation emerged as the most effective in enhancing both social presence and subjective transparency. **(2)** The specific words used in the explanation need some optimization to increase the understandability of the explanation. For instance, terms like "pre-trained model" were found to be potentially confusing and even distancing for some users. **(3)** The dynamic response delay setting was perceived to be too long by the majority of the participants and already influenced user perception on the capability of the chatbot. **(4)** Participants expressed a need for clearer demarcation between the explanation and the actual chatbot response. Given these findings and the feedback received, several adjustments were made to optimize the chatbot prototype and its interactions. These modifications, ranging from the content of explanations to the calculation of response delays, were aimed at ensuring the validity of the subsequent main study.

MAIN TEST METHOD

Experiment conditions

To test our hypotheses, we conducted a two-factor between-subjects experiment. The independent variables: response delay (dynamic or instant) and explanation (with or without), were manipulated in the experiment, hence 2 × 2 in total 4 groups were set for the experiment: instant/dynamic delay with/without explanation. The experiment was conducted via Prolific (https://www.prolific.co/) – an online recruitment platform. Participants were first asked to follow a pre-recorded video of the interaction between a user and the chatbot. Then they were asked to fill in a questionnaire to evaluate the social presence, subjective transparency feelings, and trust towards the chatbot.

Experimental Chatbot

For the experiment, we developed 4 chatbots to fit the 4 conditions mentioned before. The Chatbot used in the main test is an optimized version of the pretest one. The screenshot of the interface is shown in Figure 1. The calculation of the dynamic response delay referred to the methods used by Gnewuch in 2018. The calculation was based on the sentence's complexity and the response's length. The calculation of language complexity (C) of each message using average sentence word length and average syllables per word according to the following formula (Kincaid et al., 1975).

$$C(m) = 0.39 \times (\text{total words} / \text{total sentences}) + 11.8 \times (\text{total syllables} / \text{total words}) - 15.59$$

So, for each sentence's complexity calculation:

$$C(s) = 0.39 \times \text{total words} + 11.8 \times (\text{total syllables} / \text{total words}) - 15.59$$

The complexity values can range from -3.40 to positive infinity. Based on this, the time delay (D) was calculated in milliseconds based on the complexity value (C(m)) of a sentence:

$$D(s) = 0.2 \times \ln(C(s) + 0.5) + 0.5 \qquad C(s)>0$$

$$D(s) = 0 \qquad C(s)\leqslant 0$$

Since sending a message requires a short time for data transmission through the Internet. Refers to Gnewuch et al. in 2022, this corresponded to a response time of about 200 to 400 m, which is similar to chatbots implemented with instant response time in real-world applications (Gnewuch et al., 2022). So, for the chatbot with a dynamic response delay, besides the sum of D(s) for each sentence, the total delay time also contains a randomized internet delay between 200 to 400 ms. While the chatbot with an instant response delay, only has a randomized internet delay between 200 to 400 ms. Besides these delays for each chatbot, we also added some little technical delays (in total xx ms) in between message printings to make the output of the chatbot in order.

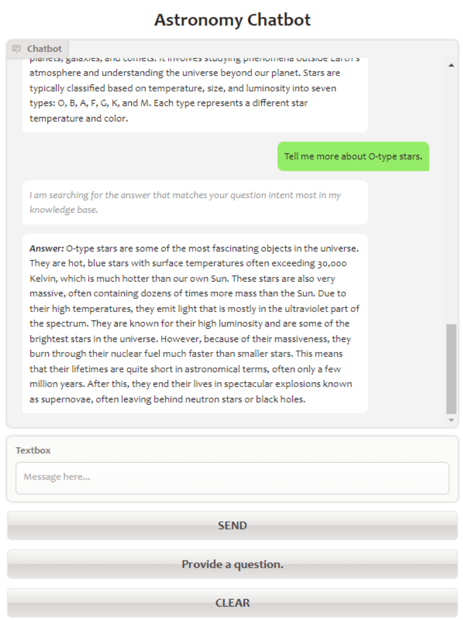
Total Delay (Dynamic) = SUM (D(s) + Internet delay + Technical delay

Total Delay (Instant) = Internet delay + Technical delay

For the chatbot implementing different types of delay, a corresponding explanation was also prepared which were the optimized ones from the results of the pretest. **Instant delay explanation**: I am searching in my knowledge base for the answer that matches your question intent most. **Dynamic delay explanation**: I am using your question as a prompt to generate the answer from my latest AI model.

Video stimuli

During the pretest, we found that participants were able to try out the chatbot function by asking questions that are not related to the given topic. We set the chatbot output to "Nothing out there, please provide the right question." while participants ask random questions. But when it comes to this situation, participants tend to think the chatbot is not capable. Hence, in the main study, to make full control of the interaction experience, the recorded videos of the interaction between the user and chatbot were used as the stimuli. The participants were asked to carefully watch the full video without manipulating the speed or progress of the video playing. The chatbot for the main test was set as an astronomy knowledge chatbot. Because during the pretest, the history and sustainability topic seemed quite familiar to the participants. Participants can indicate whether the response from the chatbot is correct or not. And this will further influence the evaluation of the trustworthiness of the chatbot. For example, if participants know that the result provided by the chatbot is correct, they will feel the chatbot is trustworthy. Setting the topic to one less familiar to the user can minimize this effect, and make the evaluation of the trustworthiness only come from the interaction factors. 5 rounds of interaction were recorded between the user and the chatbot, including starting and ending messages, and 3 questions related to astronomy. One example of the conversation was shown in Figure 4, the full conversation flow of different conditions is in Appendix 1. To avoid typing mistakes, the typing and sending of the message were automatically done by a script. And to give a realistic feeling, the time in between each character's typing is between 20 ms to 100 ms to mimic the human user's typing speed. The main test chatbot prototype and the auto-filling script are also accessible on Github: main test chatbot prototype.



planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

Tell me more about O-type stars.

I am searching for the answer that matches your question intent most in my knowledge base.

Answer: O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes.

Figure 4. Main test Chatbot prototype.

Procedure

First, participants need to read through the same introduction about the study and provide informed consent to join the study. Participants were informed that this is a study related to the interaction design of an astronomy chatbot. But the details about what will be tested or evaluated were not disclosed to them. Then the participants were asked to follow the video carefully and randomly assigned to one of the four conditions. The randomization process was done automatically by Qualtrics (https://www.qualtrics.com/) - an online survey builder. For every condition, participants need to answer the same questionnaires afterward. Considering the cognitive load of the participant increased as the study went on, we decided to show the video stimuli and measure the factors in our research model first. Hence, the questionnaire was shown to the participants by order: perceived delay, social presence, subjective transparency, trust, chatbot experience, astronomy knowledge, and AIT scale. In the end, several demographic questions were asked to the participants. Everything was anonymized in this study. At the end of the questionnaire, we gave participants a chance to comment on this research. Both the four stimuli were evenly distributed among participants. The overall procedure was shown in Figure 5.

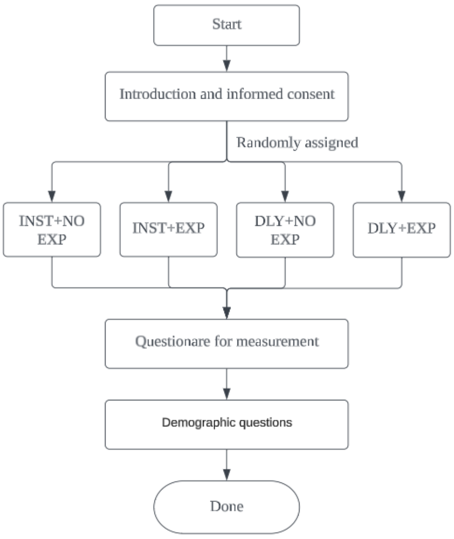


Figure 5. The procedure of the main study.

Participants

We recruited participants via Prolific (https://www.prolific.co/) - an online participants recruitment platform. The participants were asked to fill in the English questionnaire on a laptop or computer. Since the language of the conversation is also set to be English, the participants were recruited from English as the first language pool. An a priori power analysis using G*Power (Faul et al., 2007) with a significance level of 0.05 determined a minimum sample size of 179 participants to achieve a statistical power of 0.80 for detecting a medium effect size (*f* = 0.25). As some participants might not pass the attention check question or might encounter (technical) difficulties while doing the study, we aimed for a sample size of about 200 participants.

In total, 201 subjects participated in the study. Participants in Prolific are paid in GBP (£) and studies are required to pay a minimum amount that is equivalent to USD ($) 6.50 per hour. This study followed a between-subject procedural. Based on a pilot run we estimated the study to take an average of 6 minutes for participants to complete. Considering the Prolific recommendation for fair payment, we thus offered £1 as compensation. We observed a median completion time of 06:01 with an average hourly reward of £9.97/hr.

To check if participants were taking the questionnaire seriously, we asked an attention check question "I swim across the Atlantic Ocean to get to work every day." (strongly disagree → strongly agree). Participants should strongly disagree if they were focusing on filling out the questionnaire. To ensure the participants followed the conversation in the video stimuli carefully, we set two questions related to the content of the conversation (e.g., "In the video, the chatbot was asked to provide more information about which type of stars."; "In the following questions, which one was the last question asked to the chatbot in the conversation?"), All participants passed these attention check questions.

Four participants got excluded, because they answered the English language level check question as advanced (3 participants) or intermedia (1 participant). Only the participants who speak English fluently (11 participants) or native speakers (186 participants) were kept to ensure language use is understandable. Therefore, the final sample included 197 participants (94 male, 93 female, 7 Non-binary/third gender, and 3 prefer not to disclose/self-describe). Participants were between 19 and 54 years old (*M* = 34.42, *SD* = 10.30).

Measurement

The measurement of different constructs was adapted from former studies (see Table 4) except for astronomy knowledge. We did not find any evaluated questionnaire used to measure the astronomy knowledge level. In this study, the knowledge level of astronomy was measured by two questions: (1) I have a lot to say regarding astronomy. (2) I know a lot about astronomy. While in the reliability analysis, the scale of astronomy knowledge was proven to be of high reliability ($\alpha = .920$).

Construct	Items	Reference
social presence	I felt a sense of human contact with the chatbot. I felt a sense of personalness with the chatbot. I felt a sense of sociability with the chatbot. I felt a sense of human warmth with the chatbot. I felt a sense of human sensitivity with the chatbot.	Gefen and Straub, 1997
subjective transparency	I can access a great deal of information that explains how the system works. I can see plenty of information about the system's inner logic. I felt that the amount of available information regarding the system's reasoning is large.	Zhao, Benbasat, & Cavusoglu, 2019

trust	I felt the chatbot is honest and truthful. I felt the chatbot is capable of addressing my issues. I felt the chatbot's behavior and response can meet my expectations. I trust the answers provided by chatbots.	Cheng et al., 2022
previous chatbot experience	I am familiar with chatbot technologies. I use chatbots frequently.	Gnewuch et al., 2022
astronomy knowledge	I have a lot to say regarding astronomy. I know a lot about astronomy.	N/A
AIT (affinity for technology interaction)	I like to occupy myself in greater detail with technical systems. I like testing the functions of new technical systems. I predominantly deal with technical systems because I have to. When I have a new technical system in front of me, I try it out intensively. I enjoy spending time becoming acquainted with a new technical system. It is enough for me that a technical system works; I don't care how or why. I try to understand how a technical system exactly works. It is enough for me to know the basic functions of a technical system. I try to make full use of the capabilities of a technical system.	Franke, Attig, & Wessel, 2019
manipulation check	I felt the response time for the chatbot to answer my question is⋯ (1 = “very slow” ; 7 = “very fast”)	Galletta et al., 2006

Table 4. Items used in construct measurement.

We assessed social presence using the items from Gefen and Straub in 1997, subjective transparency using the items from Zhao, Benbasat, & Cavusoglu in 2019, the trust towards chatbot using the items from Cheng et al., in 2022, on seven-point Likert scales (1 = “strongly disagree”; 7 = “strongly agree”). To check the manipulation of the chatbots’ response time, each participant rated the response time using a seven-point Likert scale (1 = “very slow” ; 7 = “very fast”) (Galletta et al., 2006). In addition, we collected demographic information from participants (i.e., age, gender, education, residence) and assessed their experience of the chatbot (Gnewuch et al., 2022), knowledge level of astronomy, and affinity for technology interaction (Franke, Attig, & Wessel, 2019) on seven-point Likert scales (1 = “strongly disagree” ; 7 = “strongly agree”) as relevant control variables.

RESULTS

Reliability test

The Cronbach's α for different scales used in assessing different constructs all indicated high reliability (see Table 5.). This means that the scales in our survey are doing a good job of measuring all different factors.

Construct	**Cronbach's α**
social presence	.952

subjective transparency	.920
trust	.881
chatbot experience	.720
astronomy knowledge	.920
ATI	.890

Table 5. Cronbach's α for each construct.

Normality Test

The normality of distribution was assessed for various measures using the Shapiro-Wilk test across two main conditions: delay type (instant or dynamic) and explanation type (with or without).

From the result of the normality test, there was some normality of distribution reported regarding chatbot experience and the AIT scale. Regarding the chatbot experience, in groups of instant delay without explanation (Shapiro-Wilk, p = .082), dynamic delay without explanation (Shapiro-Wilk, p = .058), and dynamic delay with explanation (Shapiro-Wilk, p = .074), normality distributions were reported. Regarding the AIT scale result, in groups of instant delay without explanation (Shapiro-Wilk, p = .102), dynamic delay without explanation (Shapiro-Wilk, p = .348), and dynamic delay with explanation (Shapiro-Wilk, p = .507), normality distributions were reported. For the data regarding perceived delay, social presence, subjective transparency, trust, astronomy knowledge, and demographic (etc., age and education), in all four conditions, the normality test result all have a significance value (Sig.) less than 0.05 for the Shapiro-Wilk test, which means the majority of our data does not follow a normal distribution. This suggested that parametric tests, which assume normally distributed data, would not be appropriate for our data analysis. Hence, instead, we run non-parametric tests for further data analysis.

Manipulation Check

To investigate whether the manipulation of the chatbot's response time was effective, a Mann-Whitney U test was conducted on the item "I felt the response time for the chatbot to answer my question is (very slow → very fast)". Participants in the no-delay condition (M = 6.44, SD = 1.13) felt the chatbot was significantly faster in response compared to those in the delay condition (M = 5.66, SD = 1.24; U = 2820.000, Z = -5.486 p < .001). Hence, our manipulation was effective in shaping participants' perception of the chatbot's response time as either instant or delayed. Figure 6. shows the Boxplot of perceived delay by delay types.

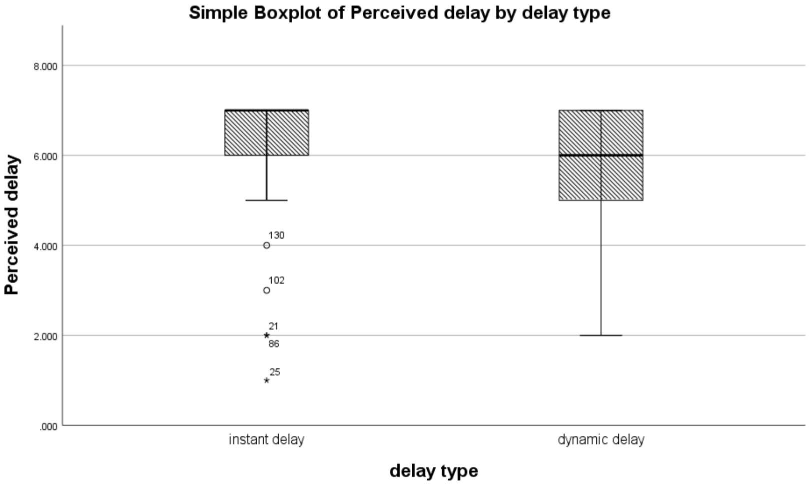


Figure 6. Boxplot: Perceived delay level in different delay conditions.

To determine if there were statistically significant differences in the distributions of various variables across the categories of the condition, we conducted Independent-Samples Kruskal-Wallis Tests. For Chatbot Experience, the Kruskal-Wallis test revealed no significant difference in the distribution of 'Chatbot Experience' scores across the categories of the condition, χ^2(3) = .069, p > .05. For Astronomy Knowledge, the distribution of 'Astronomy Knowledge' scores was not significantly different across the categories of the condition, χ^2(3) = .509, p > .05. For ATI, the test indicated no significant difference in their distribution across the categories of the condition, χ^2(3) = .793, p > .05. For Age, The age distribution did not significantly differ across the categories of the condition, χ^2(3) = .862, p > .05. For test duration time, the distribution of 'duration' was also found to be not significantly different across the categories of the condition, χ^2(3) = .240, p > .05.

To determine if there were statistically significant differences in four conditions regarding gender and education, we run the Chi-square tests. The Chi-square test indicated no significant association between gender and conditions (χ^2(12) = 6.326, p = .899), and also revealed no significant association between education and conditions (χ^2(18) = 18.365, p = .432).

Hypotheses Testing

Social Presence (H1a & H3)

A Mann-Whitney U test was used to test the association between response delay and social presence. The test indicated no statistically significant difference between the two groups, U = 4834.500, Z = -.039, p = .969. This suggests that the delay type, whether instant or dynamic, does not have a significant effect on subjective transparency. Hence, **H1a was not supported**. The two-way Bootstrap ANOVA conducted to investigate the potential moderating effect of the experimental type on the relationship between delay type and social presence showed no significant interaction (F (1, 193) = .234, p = .629). Hence, **H3 was not supported.** Furthermore, there were no significant main effects of delay type (F (1, 193) = .040, p = .842) or explanation type (F(1, 193) = .273, p = .602) on social presence. Figure 7. shows the boxplot of social presence by delay type by explanation type.

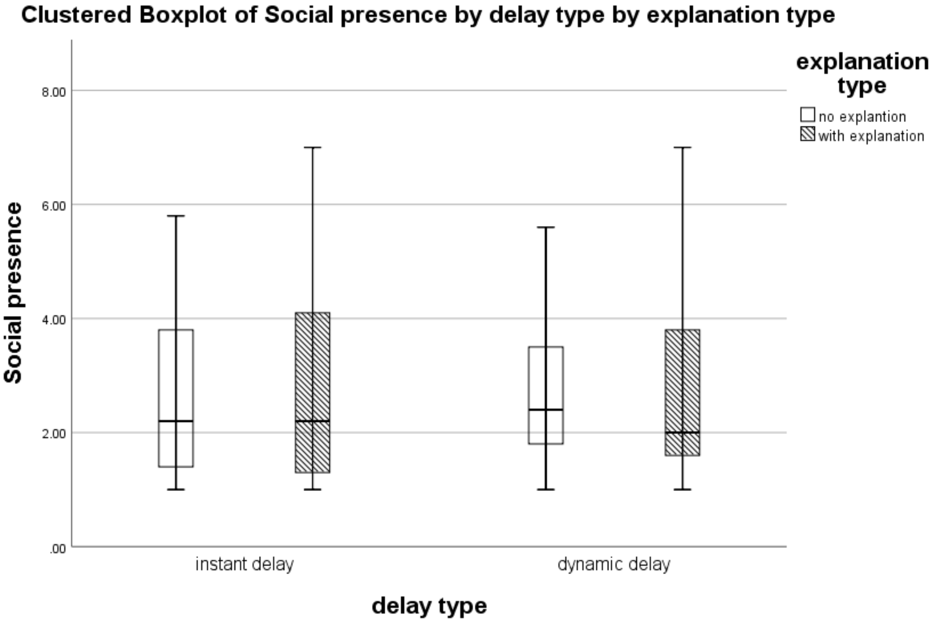


Figure 7. Boxplot: Social presence level in different delay types and explanation types.

Subjective transparency (H2a)

A Mann-Whitney U test was conducted to examine the differences in subjective transparency between the two explanation types. Participants in the group without an explanation had a median rank of 90.47 (Sum of Ranks = 8685.00, N = 96), while those with an explanation had a median rank of 107.11 (Sum of Ranks = 10818.00, N = 101). The test revealed a statistically significant difference between the two groups (U = 4029.000, Z = -2.054, p = .040). This suggests that the explanation type has a significant effect on subjective transparency, with the group receiving an explanation tending to rank higher in terms of transparency compared to the group without an explanation. Hence, **H2a was supported.** Figure 8 shows the boxplot of subjective transparency by explanation type.

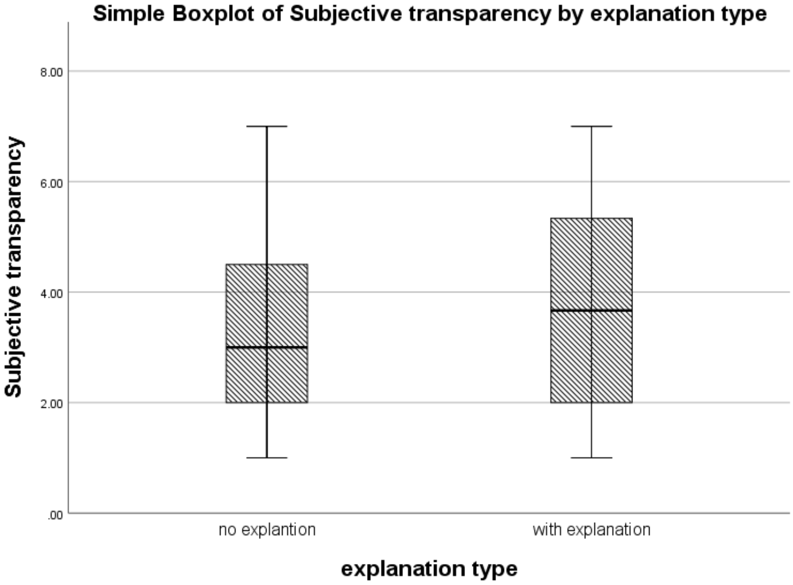


Figure 8. Boxplot: Subjective transparency level in different explanation types.

Trust towards chatbot (H1b, H2b)

A Mann-Whitney U test was conducted to evaluate the differences in trust between the two delay types (instant or dynamic). Participants in the instant delay group had a mean rank of 98.96 (Sum of Ranks = 9895.50, N = 100). Conversely, those in the dynamic delay group had a mean rank of 99.05 (Sum of Ranks = 9607.50, N = 97). The test revealed no statistically significant difference between the two groups (U = 4845.500, Z = -.011, p = .991). This indicates that the delay type, whether instant or dynamic, does not significantly influence trust levels among participants.

A Mann-Whitney U test was conducted to assess the differences in trust between the two explanation types. Participants in the group without an explanation had a mean rank of 92.38 (Sum of Ranks = 8868.00, N = 96). In contrast, those in the group with an explanation had a mean rank of 105.30 (Sum of Ranks = 10635.00, N = 101). The test indicated no statistically significant difference between the two groups (U = 4212.000, Z = -1.602, p = .109). This suggests that the presence or absence of an explanation does not have a significant effect on trust levels among participants.

A two-way Bootstrap ANOVA was performed to investigate the potential interaction effect between delay type and explanation type on trust. The result showed that there was a marginally significant interaction effect between delay type and explanation type, F (1, 193) = 3.106, p = .080. But the main effects of delay type, F (1, 193) = .210, p = .647, and explanation type, F(1, 193) = 2.170, p = .142, were not significant. Figure 9 shows the boxplot of trust by delay type by explanation type.

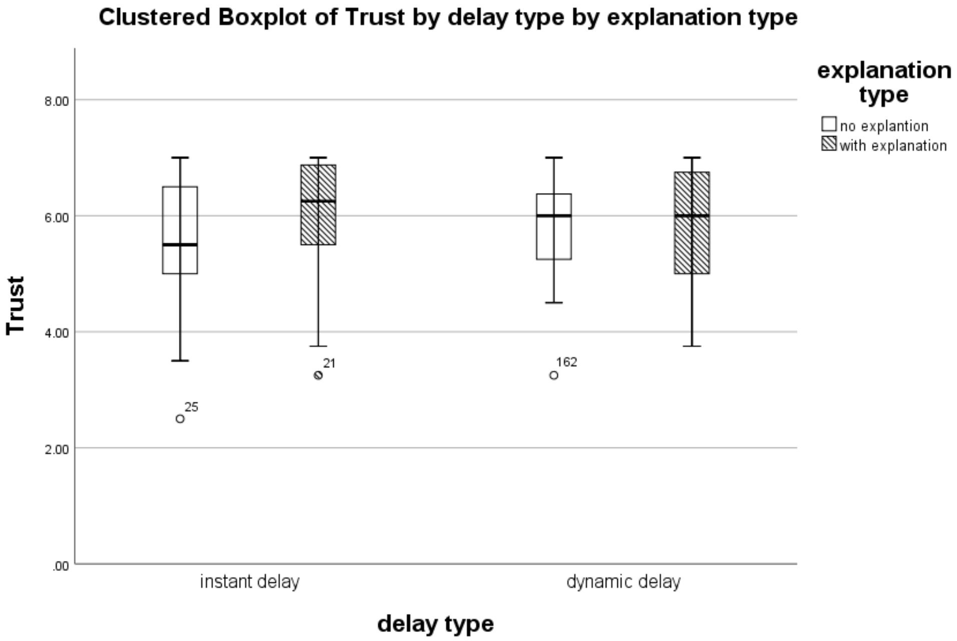


Figure 9. Boxplot: Trust level in different delay and explanation conditions.

To investigate this marginally significant interaction effect. We divided the data into four groups based on different delay types (instant or dynamic) and explanation types (with or without). Then conducted Mann-Whitney U tests for each combination of the delay type and explanation type. Under the instant delay condition, the Mann-Whitney U test revealed a statistically significant difference between different explanation groups (U = 945.500, Z = -2.113, p = .035). This suggests that in the condition of instant delay, the explanation type significantly influences trust levels among participants, with those experiencing instant delay with an explanation tending to have higher trust levels compared to those experiencing instant delay without an explanation. While for the rest condition group, no significant difference was revealed.

Upon closer inspection of the mean value of different variable combinations, Table 6 shows the mean table of trust in different conditions, the mean value also reveals the same effect: Providing an explanation in the instant delay condition increased trust compared to not providing an explanation. This is evident from the higher mean trust score when an explanation was provided (M = 5.995, SD = .992) compared to when it wasn't (M = 5.536, SD = 1.128).

Condition	Mean	N	Std. Deviation
instant delay, no explanation	5.536	49	1.128
instant delay, with explanation	5.995	51	.992
dynamic delay, no explanation	5.851	47	.810
dynamic delay, with explanation	5.810	50	1.020
Total	5.800	197	1.002

Table 6. Condition mean table: Trust (delay type * explanation type).

To test H1b and H2b, Spearman's rho nonparametric correlation analysis was performed to determine the relationship between social presence and trust, as well as trust and subjective transparency. For the relationship between social presence and trust, a significant positive correlation was found (Spearman's rho (197) = .165, p = .021). This suggests that higher levels of social presence are associated with higher levels of trust in chatbot interactions. For the relationship between trust and subjective transparency and trust, a significant positive correlation was identified (Spearman's rho (197) = .247, p < .001). This indicates that when trust levels increase, there is a corresponding increase in subjective transparency. This relationship is stronger than that found between social presence and trust. In both analyses, p-values were less than .05, indicating the correlations are statistically significant. Hence, **H1b and H2b were both supported**.

DISCUSSION

In the study, we aimed to explore how the dynamic or instant response delay of a chatbot and the presence or absence of an explanation affected users' perceptions of social presence, subjective transparency, and trust. The results suggest a mixed outcome that indicates the importance of considering both delay type and the provision of explanations in designing chatbot interactions.

Theoretical Implications

Explanation type on subjective transparency and trust

This study found that the selected explanation can improve the user's perception of the system transparency level in both instant response and dynamic delayed response conditions. The result also reveals that the explanation can work in different delay conditions and different levels of chatbot experience. Additionally, subjective transparency positively affected user trust, with participants expressing higher trust when the chatbot explained its responses. But not as we expected, in this study the explanation did not boost user trust towards the chatbot through increasing the subjective transparency. During this study, the explanation was focused on the response delay which is one narrow aspect of the chatbot system.

Response delay type on social presence and trust

In this study users who perceive faster response times from the chatbot are more likely to trust it. This is contrary to my expectation and previous studies conducted in the customer services context found that dynamic response delay can significantly increase user trust in the chatbot compared with the instant response chatbot (Karma Choedak, 2020). This suggests that the timing of the response may not be a crucial element in creating a social presence feeling in the knowledge chatbot domain.

The analysis also replicated a significant positive correlation between social presence and trust in chatbots (Zierau et al., 2021; Konya-Baumbach, Biller & von Janda, 2023), indicating that users who perceived a greater sense of human contact and sociability with the chatbot tended to trust it more.

Interaction between response delays and explanations

The marginal interaction effect between delay type and explanation on trust suggests a nuanced relationship. Specifically, trust can be enhanced when an instant delay is paired with an explanation, but there is no difference revealed in dynamic delay situations with or without an explanation.

Practical Implications

The result of this study can be used to guide more trustworthy knowledge chatbots.

Effect of delay

Since the delay in response did not significantly affect the social presence or trust, designers, and developers can focus more on ensuring the accuracy and relevance of responses. Considerations of server load and other performance issues could thus be balanced with user experience needs. But the correlation between perceived response delay and user trust revealed that users tend to have more trust in the chatbot that was perceived to respond fast. For chatbot developers, this could mean that to build a trustworthy knowledge chatbot they can focus on improving the response speed. Compared with the generating-based strategy, the retrieving-based strategy has a relatively faster speed in providing the answer. Hence the retrieving-based chatbot can be a better choice. For designers, they need to focus on improving the user perception of the response delay. For example, filler interfaces that introduce web design components such as images, text, progress bar, background color, or multimedia can influence online users' perceived waiting time (PWT) (Lee, Chen, & Ilie, 2012).

Effect of Explanation

This study replicated the positive correlation between subjective transparency and user trust in the knowledge chatbot domain. This emphasizes the importance of making the underlying process accessible to boost user trust. The result also suggested that providing explanations about how the chatbot is functioning can enhance the perception of transparency. This finding reinforces the importance of clear and appropriate communication in human-chatbot interactions. For designers, giving users insight into how the chatbot functions can build a sense of transparency and likely lead to more trust in the chatbot.

Effect of Social Presence

This study replicated the positive correlation between social presence and user trust in the knowledge chatbot domain. This indicates that designers can also try to enhance the sense of social presence in chatbot interactions. For example, previous studies suggest that introducing an avatar can enhance the social presence feeling (Wang, 2016).

Interaction effect

This study revealed a marginally significant effect between explanation and response delay type on trust. This interaction effect indicates that, compared with no explanation, implementing an explanation in an instant-delay knowledge chatbot enhances user trust regarding the honesty and the answers' trustworthiness of the chatbot. But there is no difference in user trust revealed in a dynamic delay might between with or without an explanation. For designers, this could mean that when designing for different types of knowledge chatbots (retrieving- or generating-based), they should implement different explanation strategies for a trustworthy chatbot.

Limitations and Future Work

This study is limited in the following aspects. First, this main test was conducted in a controlled setting using a specific topic (astronomy) for the chatbot. This context might not reflect the diversity of real-world chatbot applications and user needs. Future research should investigate these effects across various domains and more complex interactions. Additionally, the experiment's artificial setup may have influenced participant responses, so real-world testing might yield different results.

Second, in this study, the response delay setting successfully manipulated the perceived delay, but in both delay conditions (instant and dynamic), the response time was perceived to be quite fast (instant delay: M = 6.44, SD = 1.13; dynamic delay: M = 5.66, SD = 1.24). This can be the reason why the result of the main test did not reveal a positive effect between dynamic delay and social presence, which is different from my expectation and previous studies regarding delay type and social presence (Gnewuch, 2018; Gnewuch, 2022). For future research, the scope of response delay setting can be magnified to study the potential influence between delay and social presence. Also, different ways of showing the response delay can also be interesting to look at, for example, the way that ChatGPT and NewBing show the delay is different. ChatGPT outputs the response word by word, while NewBing outputs several words first followed by a long paragraph. One notable thing is that the dynamic response delay calculation was adapted from the formula proposed by Gnewuch in 2018, based on the pretest feedback, to make the interaction feel natural and comfortable. This formula might need different adaptations to fit with other contexts.

Third, the impact that explanation has on the social presence in the pre-test seems not aligned with the result in the main test. In the pre-test, participants were exposed to various chatbots and could compare different combinations of delay and explanation. This comparative experience might have influenced their perceptions and preferences and led them to appreciate the first-person description explanation. Conversely, in the main test, participants experienced only one combination of delay and explanation. Without a basis for comparison, their reactions to the explanation might have been different. The singular experience might not have allowed for the same level of social presence feeling to emerge.

CONCLUSION

This study has explored the impact of response delay type and the presence of explanations on social presence, subjective transparency, and user trust in chatbots. First, the findings revealed that while response delay did not significantly influence trust or social presence, clear explanations, especially in the context of instant delays, positively impacted subjective transparency and trust. Second, the study reinforced the positive correlation between social presence and trust, subjective transparency, and trust.

The explanation did not have an overall effect on user trust in the chatbot, but it can positively improve subjective transparency, which is proven to have a positive correlation with trust. The explanation was only focused on the response delay of the chatbot. Still, it was found that explanation can significantly improve trust in the instant delay condition. Future researchers can implement explanations in a wider scope of chatbot systems.

This study only revealed a marginally significant interaction effect between response delay and explanation on trust. This effect might be limited by the experiment method, as in this research the interaction was shown via video to minimize the influence of unpredictable interaction between participants and chatbot. Future researchers can investigate this to identify this potential effect.

REFERENCES

de Visser, Ewart J., et al. "Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents." *Journal of Experimental Psychology: Applied*, vol. 22, no. 3, Sept. 2016, pp. 331–349, https://doi.org/10.1037/xap0000092.

Pak, Richard, et al. "Decision Support Aids with Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults." *Ergonomics*, vol. 55, no. 9, 17 July 2012, pp. 1059–1072, https://doi.org/10.1080/00140139.2012.691554. Accessed 21 Jan. 2020.

Waytz, Adam, et al. "The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle." *Journal of Experimental Social Psychology*, vol. 52, May 2014, pp. 113–117, https://doi.org/10.1016/j.jesp.2014.01.005.

De Visser, E. J., Monfort, S. S., Goodyear, K., Lu, L., O'Hara, M., Lee, M. R., ... & Krueger, F. (2017). A little anthropomorphism goes a long way: Effects of oxytocin on trust, compliance, and team performance with automated agents. *Human factors, 59*(1), 116-133.

Kulms, P., & Kopp, S. (2019). More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. *In Proceedings of mensch und computer* 2019 (pp. 31-42).

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. https://doi.org/10.1145/3173574.3174156

Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does Humanity Matter? Analyzing the Importance of Social Cues and Perceived Agency of a Computer System for the Emergence of Social Reactions during Human-Computer Interaction. *Advances in Human-Computer Interaction*, *2012*, 1–10. https://doi.org/10.1155/2012/324694

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Ayanouz, S., Abdelhakim, B. A., & Benhmed, M. (2020). A Smart Chatbot Architecture based NLP and Machine Learning for Health Care Assistance. *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*. https://doi.org/10.1145/3386723.3387897

Bala, K., Kumar, M., Hulawale, S., & Pandita, S. (2017). Chat-bot for college management system using AI. *International Research Journal of Engineering and Technology*, *4*(11). 2030-2033.

Benlian, A., Klumpe, J., & Hinz, O. (2019). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*. https://doi.org/10.1111/isj.12243

Bernhard Kratzwald, & Feuerriegel, S. (2018). Putting Question-Answering Systems into Practice. *ACM Transactions on Management Information Systems*, *9*(4), 1–20. https://doi.org/10.1145/3309706

Biswas, S. S. (2023). Role of chat gpt in public health. *Annals of Biomedical Engineering, 1-2*.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing User Needs and Motivations. *Interactions*, *25*(5), 38–43. https://doi.org/10.1145/3236669

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Hesse, C. (n.d.). *Language Models are Few-Shot Learners*. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Burgoon, J., Guerrero, L., & Floyd, K. (2010). *Nonverbal Communication*. https://api.pageplace.de/preview/DT0400.9781000427646_A41277410/preview-9781000427646_A41277410.pdf

Cai, C. J., Jongejan, J., & Holbrook, J. (2019). The effects of example-based explanations in a machine learning interface. *Proceedings of the 24th International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3301275.3302289

Caldarini, G., Jaf, S., & McGarry, K. (2022). A Literature Survey of Recent Advances in Chatbots. *Information*, *13*(1), 41. https://doi.org/10.3390/info13010041

Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2021). Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure. *Internet Research*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/intr-08-2020-0460

Cheng, X., Su, L., & Yang, B. (2020). An investigation into sharing economy enabled ridesharing drivers' trust: A qualitative study. *Electronic Commerce Research and Applications*, 100956. https://doi.org/10.1016/j.elerap.2020.100956

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management*, *59*(3), 102940. https://doi.org/10.1016/j.ipm.2022.102940

Cho, B., Ryoo, S. Y., & Kim, K. K. (2017). Interorganizational dependence, information transparency in interorganizational information systems, and supply chain performance. *European Journal of Information Systems*, *26*(2), 185–205. https://doi.org/10.1057/s41303-017-0038-1

Choedak, K. (2020). *THE EFFECT OF CHATBOTS RESPONSE LATENCY ON USERS' TRUST*. https://shareok.org/bitstream/handle/11244/324150/2020_Karma%20Choedak_._Thesis.pdf?sequence=10

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331–349. https://doi.org/10.1037/xap0000092

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & Nunamaker, J. F. (2013). Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues. *ACM Transactions on Management Information Systems*, *4*(2), 1–21. https://doi.org/10.1145/2499962.2499967

Devlin, J., Chang, M.-W., Lee, K., Google, K., & Language, A. (n.d.). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. https://arxiv.org/pdf/1810.04805.pdf&usg=ALkJrhhzxlCL6yTht2BRmH9atgvKFxHsxQ

Diederich, S., Brendel, A. B., & Kolbe, L. M. (2020). Designing Anthropomorphic Enterprise Conversational Agents. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-020-00639-y

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2018). Bringing Transparency Design into Practice. *23rd International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3172944.3172961

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Følstad, A., Araujo, T., Law, E. L.-C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., Baez, M., Laban, G., McAllister, P., Ischen, C., Wald, R., Catania, F., Meyer von Wolff, R., Hobert, S., & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, *103*(12), 2915–2942. https://doi.org/10.1007/s00607-021-01016-7

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

Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2019). Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. *Internet Science*, 145–156. https://doi.org/10.1007/978-3-030-17705-8_13

Franke, T., Attig, C., & Wessel, D. (2018). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*, *35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Galletta, D. F., Henry, R. M., McCoy, S., & Polak, P. (2006). When the Wait Isn't So Bad: The Interacting Effects of Website Delay, Familiarity, and Breadth. *Information Systems Research*, *17*(1), 20–37. https://doi.org/10.1287/isre.1050.0073

Galletta, D. F., Henry, R., McCoy, S., & Polak, P. (2004). Web Site Delays: How Tolerant are Users? *Journal of the Association for Information Systems*, *5*(1). https://doi.org/10.17705/1jais.00044

Gefen, D., & Straub, D. (2005). A Practical Guide To Factorial Validity Using PLS-Graph: Tutorial And Annotated Example. *Communications of the Association for Information Systems*, *16*. https://doi.org/10.17705/1cais.01605

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega*, *32*(6), 407–424. https://doi.org/10.1016/j.omega.2004.01.006

Glass, A., McGuinness, D. L., & Wolverton, M. (2008). Toward establishing trust in adaptive agents. *Proceedings of the 13th International Conference on Intelligent User Interfaces - IUI '08*. https://doi.org/10.1145/1378773.1378804

Gnewuch, U., Morana, S., Adam, M. T. P., & Maedche, A. (2022). Opposing Effects of Response Time in Human–Chatbot Interaction. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-022-00755-x

Gnewuch, U., Morana, S., Adam, M., Maedche, A., Maedche, & Alexander. (2018). *Association for Information Systems AIS Electronic Library (AISeL) Faster is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human- Chatbot Interaction Recommended Citation*. https://scholar.archive.org/work/664dlhdlrbdvffn4jb4zir4toa/access/wayback/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1112&context=ecis2018_rp

Goodman, B., & Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Jain, M., Kota, R., Kumar, P., & Patel, S. N. (2018). Convey. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. https://doi.org/10.1145/3173574.3174042

Jensen, M. L., Lowry, P. B., Burgoon, J. K., & Nunamaker, J. F. (2010). Technology Dominance in Complex Decision Making: The Case of Aided Credibility Assessment. *Journal of Management Information Systems*, *27*(1), 175–202. https://doi.org/10.2753/mis0742-1222270108

Jia, J. (2003). The Study of the Application of a Keywords-based Chatbot System on the Teaching of Foreign Languages. *ArXiv (Cornell University)*.

Kalman, Y. M., Scissors, L. E., Gill, A. J., & Gergle, D. (2013). Online chronemics convey social information. *Computers in Human Behavior*, *29*(3), 1260–1269. https://doi.org/10.1016/j.chb.2012.12.036

Khurana, A., Alamzadeh, P., & Chilana, P. K. (2021, October 1). *ChatrEx: Designing Explainable Chatbot Interfaces for Enhancing Usefulness, Transparency, and Trust*. IEEE Xplore. https://doi.org/10.1109/VL/HCC51201.2021.9576440

Kincaid, J., Fishburne, R., Richard, L., Rogers, B., & Chissom, B. S. (1975). *Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease*

Formula) For Navy Enlisted Personnel 1-1-1975. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary

Kizilcec, R. (2016). How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2858036.2858402

Knijnenburg, B. P., & Willemsen, M. C. (2016). Inferring Capabilities of Intelligent Agents from Their External Traits. *ACM Transactions on Interactive Intelligent Systems*, *6*(4), 1–25. https://doi.org/10.1145/2963106

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, *139*, 107513. https://doi.org/10.1016/j.chb.2022.107513

Krogue, K. (2017). *Artificial Intelligence Is Here To Stay, But Consumer Trust Is A Must for AI in Business*. Forbes. https://www.forbes.com/sites/kenkrogue/2017/09/11/artificial-intelligence-is-here-to-stay-but-consumer-trust-is-a-must-for-ai-in-business

Kulesza, T., Stumpf, S., Burnett, M., & Kwan, I. (2012). Tell me more? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/2207676.2207678

Kumar, R., & Ali, M. M. (2020). A review on chatbot design and implementation techniques. *Int. J. Eng. Technol*, *7*(11).

Lee, Chen, & Ilie. (2012). Can Online Wait Be Managed? The Effect of Filler Interfaces and Presentation Modes on Perceived Waiting Time Online. *MIS Quarterly*, *36*(2), 365. https://doi.org/10.2307/41703460

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not *explanations improve the intelligibility of context-aware intelligent systems*. https://doi.org/10.1145/1518701.1519023

Luo, B., Lau, R. Y. K., Li, C., & Si, Y. (2021). A critical review of state‐of‐the‐art chatbot designs and applications. *WIREs Data Mining and Knowledge Discovery*, *12*(1). https://doi.org/10.1002/widm.1434

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, *20*(3), 709–734.

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., Hinz, O., Morana, S., & Söllner, M. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Maedche, A., Morana, S., Schacht, S., Werth, D., & Krumeich, J. (2016). Advanced User Assistance Systems. *Business & Information Systems Engineering*, *58*(5), 367–370. https://doi.org/10.1007/s12599-016-0444-2

McGuirl, J. M., & Sarter, N. B. (2006). Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(4), 656–665. https://doi.org/10.1518/001872006779166334

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Merritt, S. M., Lee, D., Unnerstall, J. L., & Huber, K. (2014). Are Well-Calibrated Users Effective Users? Associations Between Calibration of Trust and Performance on an Automation-Aided Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 34–47. https://doi.org/10.1177/0018720814561675

Mhlanga, D. (2023). Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4354422

Moon, Y. (1999). The effects of physical distance and response latency on persuasion in computer-mediated communication and human–computer communication. *Journal of Experimental Psychology: Applied*, *5*(4), 379–392. https://doi.org/10.1037/1076-898x.5.4.379

Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2020). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*. https://doi.org/10.1007/s12525-020-00411-w

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*(5-6), 527–539. https://doi.org/10.1016/s0020-7373(87)80013-5

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171–181. https://doi.org/10.1037/1076-898x.7.3.171

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*. https://doi.org/10.1145/191666.191703

Omar, R., Mangukiya, O., Kalnis, P., & Mansour, E. (2023). *ChatGPT versus Traditional Question Answering for Knowledge Graphs: Current Status and Future Directions Towards Knowledge Graph Chatbots*. https://arxiv.org/pdf/2302.06466

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, *55*(9), 1059–1072. https://doi.org/10.1080/00140139.2012.691554

Parker, E. B., Short, J., Williams, E., & Christie, B. (1978). The Social Psychology of Telecommunications. *Contemporary Sociology*, *7*(1), 32. https://doi.org/10.2307/2065899

Piccolo, L. (2019). *Chasing the Chatbots Directions for Interaction and Design Research*. https://oro.open.ac.uk/57382/1/chatbots-workshop_camera_ready2.pdf

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. https://doi.org/10.1145/3173574.3174214

Proceedings of the 26th European conference on information systems. Portsmouth, UK Burgoon JK, Guerrero L, Floyd K (2010) Nonverbal communication. Routledge, New York, B., & Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people and places. *Choice Reviews Online*, *34*(07), 34–370234–3702. https://doi.org/10.5860/choice.34-3702

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, *21*, 1–67. https://www.jmlr.org/papers/volume21/20-074/20-074.pdf

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, & new media like real people & places. *Computers & Mathematics with Applications*, *33*(5), 128. https://doi.org/10.1016/s0898-1221(97)82929-x

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research*, *32*(3). https://doi.org/10.1287/isre.2021.1015

Schuetzler, R. (2015). *Dynamic Interviewing Agents: Effects on Deception, Nonverbal Behavior, and Social Desirability - ProQuest*. Www.proquest.com. https://www.proquest.com/openview/9039d5988e2a9a72694c1b32afe21eaa/1?pq-origsite=gscholar&cbl=18750

Schuetzler, R. M., Grimes, G. M., & Giboney, J. S. (2019). The effect of conversational agent skill on user behavior during deception. *Computers in Human Behavior*, *97*, 250–259. https://doi.org/10.1016/j.chb.2019.03.033

Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with Humanlike Conversational Agents: Designing for Anthropomorphism. *Journal of the Association for Information Systems*, *22*(4). https://doi.org/10.17705/1jais.00685

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Sojasingarayar, A. (2020). *Seq2Seq AI Chatbot with Attention Mechanism (Final Year Project submitted in fulfillment of the requirements for the degree of master in Artificial Intelligence)*. https://arxiv.org/pdf/2006.02767

Song, Y., Yan, R., Li, X., Zhao, D., & Zhang, M. (2017). *Two are Better than One: An Ensemble of Retrieval-and Generation-Based Dialog Systems*. https://arxiv.org/pdf/1610.07149

Surameery, N. M. S., & Shakor, M. Y. (2023). Use Chat GPT to Solve Programming Bugs. *International Journal of Information Technology and Computer Engineering*, *31*, 17–22. https://doi.org/10.55529/ijitc.31.17.22

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Huaixiu, H., Zheng, S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., & Chen, D. (2022). *LaMDA: Language Models for Dialog Applications*. https://arxiv.org/pdf/2201.08239.pdf?trk=public_post_comment-text

Tsai, W.-H. S., Liu, Y., & Chuan, C.-H. (2021). How chatbots' social presence communication enhances consumer engagement: the mediating role of parasocial interaction and dialogue. *Journal of Research in Interactive Marketing*, *15*(3), 460–482. https://doi.org/10.1108/jrim-12-2019-0200

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Walther, J. B., & Tidwell, L. C. (1995). Nonverbal cues in computer‐mediated communication, and the effect of chronemics on relational communication. *Journal of Organizational Computing*, *5*(4), 355–378. https://doi.org/10.1080/10919399509540258

Wang, W., Qiu, L., Kim, D., & Benbasat, I. (2016). Effects of rational and social appeals of online recommendation agents on cognition- and affect-based trust. *Decision Support Systems*, *86*, 48–60. https://doi.org/10.1016/j.dss.2016.03.007

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Wu, Y., Li, Z., Wu, W., & Zhou, M. (2018). Response selection with topic clues for retrieval-based chatbots. *Neurocomputing*, *316*, 251–261. https://doi.org/10.1016/j.neucom.2018.07.073

Wu, Y., Wu, W., Xing, C., Xu, C., Li, Z., & Zhou, M. (2019). A Sequential Matching Framework for Multi-Turn Response Selection in Retrieval-Based Chatbots. *Computational Linguistics*, *45*(1), 163–197. https://doi.org/10.1162/coli_a_00345

Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A New Chatbot for Customer Service on Social Media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. https://doi.org/10.1145/3025453.3025496

Xu, W. (2019). Toward human-centered AI. *Interactions*, *26*(4), 42–46. https://doi.org/10.1145/3328485

Yang, F., Huang, Z., Scholtz, J., & Arendt, D. (2020). *How do visual explanations foster end users' appropriate trust in machine learning?* https://doi.org/10.1145/3377325.3377480

Yu, M., Zhou, R., Cai, Z., Tan, C.-W., & Wang, H. (2020). Unravelling the relationship between response time and user experience in mobile applications. *Internet Research*, *30*(5), 1353–1382. https://doi.org/10.1108/intr-05-2019-0223

Yu, Q., Nguyen, T., Prakkamakul, S., & Salehi, N. (2019). "I Almost Fell in Love with a Machine." *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290607.3312918

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. https://doi.org/10.1145/3351095.3372852

Zhao, R., Izak Benbasat, & Hasan Cavusoglu. (2019). Do Users Always Want to Know More? Investigating the Relationship between System Transparency and Users' Trust in Advice-Giving Systems. *European Conference on Information Systems*.

Zhu, K. (2002). Information Transparency in Electronic Marketplaces: Why Data Transparency May Hinder the Adoption of B2B Exchanges. *Electronic Markets*, *12*(2), 92–99. https://doi.org/10.1080/10196780252844535

Zierau, N., Flock, K., Janson, A., Söllner, M., & Leimeister, J. (n.d.). *The Influence of AI-Based Chatbots and Their Design on Users' Trust and Information Sharing in Online Loan Applications*. Retrieved August 8, 2023, from http://pubs.wi-kassel.de/wp-content/uploads/2020/12/JML_797.pdf

APPENDIX 1: CONVERSATION SCRIPT OF THE MAIN TEST.

Script for dynamic/instant delay, no explanation:

User: "Hello."

Chatbot: "Welcome. Knowledge chatbot for astronomy here. Is any assistance needed?"

User: "Provide some information about astronomy and the different types of stars."

Chatbot: "Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color."

User: "Tell me more about O-type stars."

Chatbot: "O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes."

User: "What is a supernova?"

Chatbot: "A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars."

User: "That's all for now."

Chatbot: "Assistance is available whenever required."

Script for instant delay, with explanation:

User: "Hello."

Chatbot: "Welcome. Knowledge chatbot for astronomy here. Is any assistance needed?"

User: "Provide some information about astronomy and the different types of stars."

Chatbot: “I am searching in my knowledge base for the answer that matches your question intent most.”

Chatbot: "Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color."

User: "Tell me more about O-type stars."

Chatbot: “I am searching in my knowledge base for the answer that matches your question intent most.”

Chatbot: "O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes."

User: "What is a supernova?"

Chatbot: “I am searching in my knowledge base for the answer that matches your question intent most.”

Chatbot: "A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars."

User: "That's all for now."

Chatbot: "Assistance is available whenever required."

Script for dynamic delay, with explanation:

User: "Hello."

Chatbot: "Welcome. Knowledge chatbot for astronomy here. Is any assistance needed?"

User: "Provide some information about astronomy and the different types of stars."

Chatbot: "I am using your question as a prompt to generate the answer from my latest AI model."

Chatbot: "Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color."

User: "Tell me more about O-type stars."

Chatbot: "I am using your question as a prompt to generate the answer from my latest AI model."

Chatbot: "O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes."

User: "What is a supernova?"

Chatbot: "I am using your question as a prompt to generate the answer from my latest AI model."

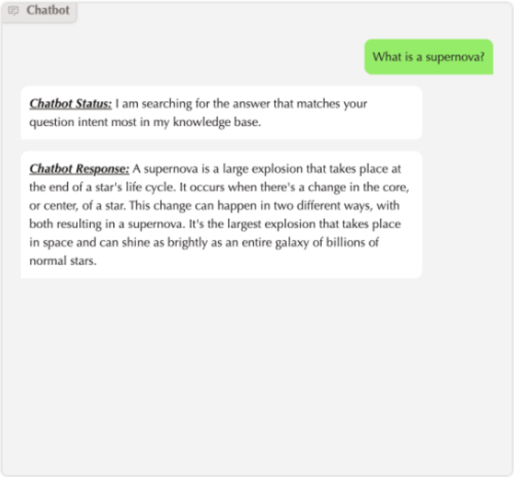
Chatbot: "A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars."

User: "That's all for now."

Chatbot: "Assistance is available whenever required."

APPENDIX 2: TRYING OUT FOR EXPLANATION LAYOUT

Original:



What is a supernova?

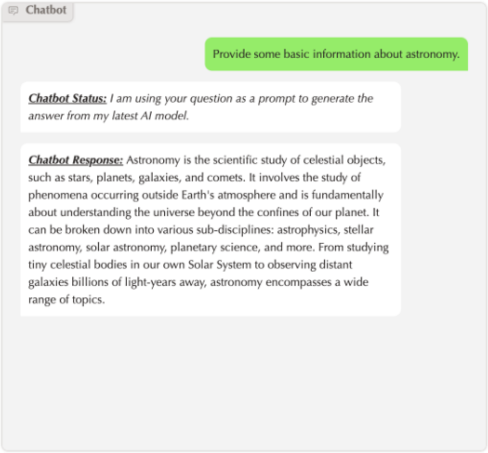
Chatbot Status: I am searching for the answer that matches your question intent most in my knowledge base.

Chatbot Response: A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars.

```
explanation = '''<u><i><b><span style="font-size: 13px;">Chatbot
Status:</span></b></i></u><span style="font-size: 13px;"> I am searching for the
answer that matches your question intent most in my knowledge base.</span>'''

response = '<u><i><b><span style="font-size: 13px;">Chatbot
Response:</span></b></i></u>'+' '+response_list[i]
```

Design 1



Provide some basic information about astronomy.

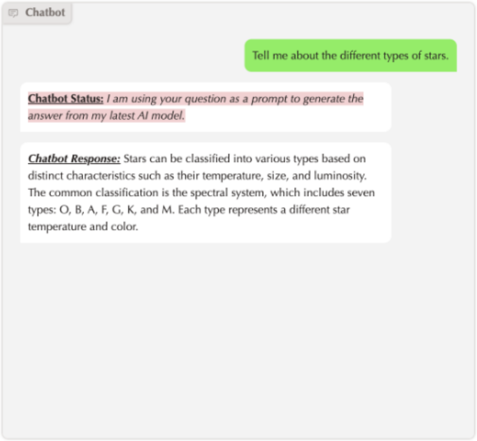
Chatbot Status: *I am using your question as a prompt to generate the answer from my latest AI model.*

Chatbot Response: Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves the study of phenomena occurring outside Earth's atmosphere and is fundamentally about understanding the universe beyond the confines of our planet. It can be broken down into various sub-disciplines: astrophysics, stellar astronomy, solar astronomy, planetary science, and more. From studying tiny celestial bodies in our own Solar System to observing distant galaxies billions of light-years away, astronomy encompasses a wide range of topics.

```
explanation = '''<u><i><b><span style="font-size: 13px;">Chatbot
Status:</span></b></i></u> <i><span style="font-size: 13px;"> I am using your
question as a prompt to generate the answer from my latest AI model.</i></span>'''

response = '<u><i><b><span style="font-size: 13px;">Chatbot
Response:</span></b></i></u>'+' '+response_list[i]
```

Design 2



Tell me about the different types of stars.

Chatbot Status: *I am using your question as a prompt to generate the answer from my latest AI model.*

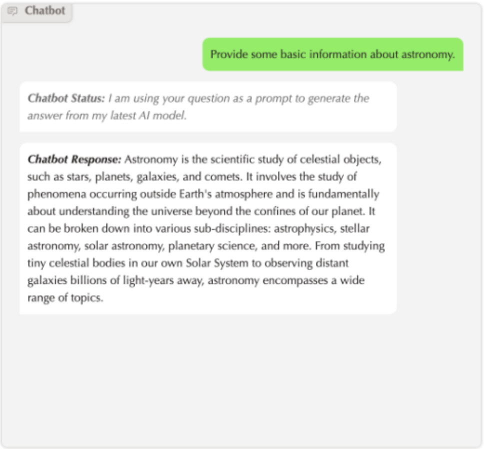
Chatbot Response: Stars can be classified into various types based on distinct characteristics such as their temperature, size, and luminosity. The common classification is the spectral system, which includes seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

```
explanation = '<u><b><span style="font-size: 13px;
background-color:#f1d1d1;">Chatbot Status:</span></b></u>'+'<i><span
style="font-size: 13px; background-color:#f1d1d1;"> I am using your question as a
prompt to generate the answer from my latest AI model.</span></i>'

response = '<u><i><b><span style="font-size: 13px;">Chatbot
Response:</span></b></i></u>'+' '+response_list[i]
```

Design 3

a.



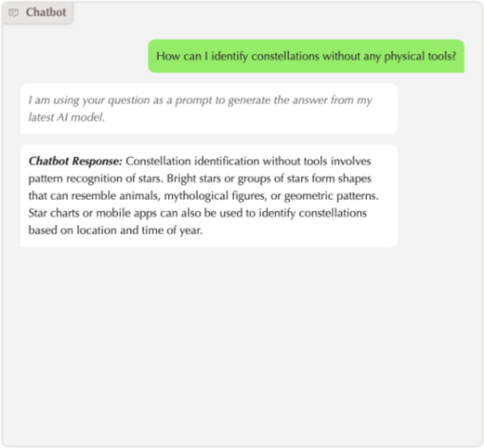
Provide some basic information about astronomy.

Chatbot Status: *I am using your question as a prompt to generate the answer from my latest AI model.*

Chatbot Response: Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves the study of phenomena occurring outside Earth's atmosphere and is fundamentally about understanding the universe beyond the confines of our planet. It can be broken down into various sub-disciplines: astrophysics, stellar astronomy, solar astronomy, planetary science, and more. From studying tiny celestial bodies in our own Solar System to observing distant galaxies billions of light-years away, astronomy encompasses a wide range of topics.

```
explanation = '<b><i><span style="font-size: 13px; color:#737070;">Chatbot
Status:</span></i></b>'+'<i><span style="font-size: 13px; color:#737070;"> I am
using your question as a prompt to generate the answer from my latest AI
model.</span></i>'
response = '<i><b><span style="font-size: 13px;">Chatbot
Response:</span></b></i>'+' '+response_list[i]
```

b.



How can I identify constellations without any physical tools?

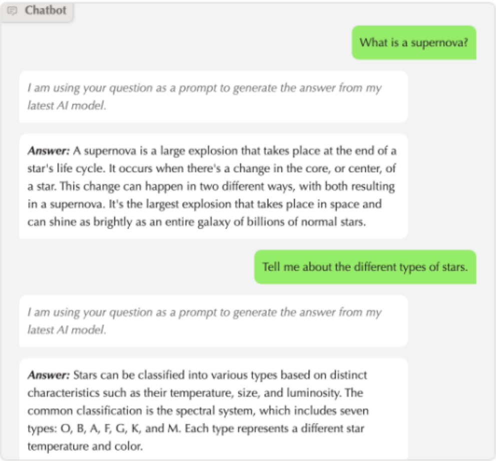
I am using your question as a prompt to generate the answer from my latest AI model.

Chatbot Response: Constellation identification without tools involves pattern recognition of stars. Bright stars or groups of stars form shapes that can resemble animals, mythological figures, or geometric patterns. Star charts or mobile apps can also be used to identify constellations based on location and time of year.

```
explanation = '<i><span style="font-size: 13px; color:#737070;"> I am using
your question as a prompt to generate the answer from my latest AI
model.</span></i>'
response = '<i><b><span style="font-size: 13px;">Chatbot
Response:</span></b></i>'+' '+response_list[i]
```

Design 4

a.



What is a supernova?

I am using your question as a prompt to generate the answer from my latest AI model.

Answer: A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars.

Tell me about the different types of stars.

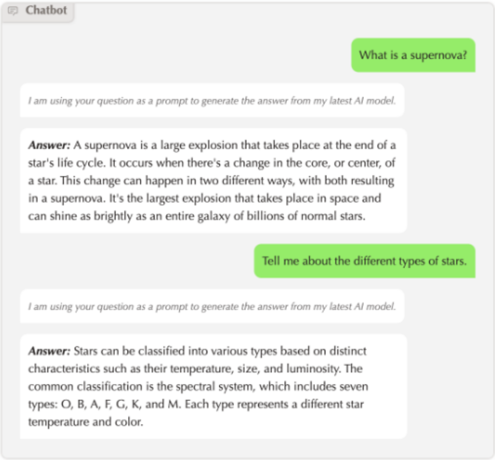
I am using your question as a prompt to generate the answer from my latest AI model.

Answer: Stars can be classified into various types based on distinct characteristics such as their temperature, size, and luminosity. The common classification is the spectral system, which includes seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

```
explanation = '<i><span style="font-size: 13px; color:#737070;"> I am using your
question as a prompt to generate the answer from my latest AI model.</span></i>'
```

```
response = '<i><b><span style="font-size: 13px;">Answer:</span></b></i>'+'
'+response_list[i]
```

b.



What is a supernova?

I am using your question as a prompt to generate the answer from my latest AI model.

Answer: A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars.

Tell me about the different types of stars.

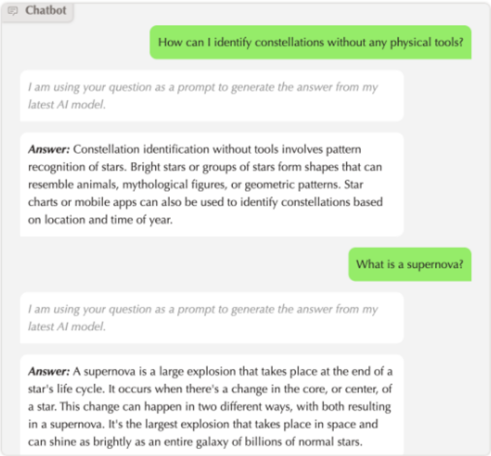
I am using your question as a prompt to generate the answer from my latest AI model.

Answer: Stars can be classified into various types based on distinct characteristics such as their temperature, size, and luminosity. The common classification is the spectral system, which includes seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

```
explanation = '<i><span style="font-size: 11px; color:#737070;"> I am
using your question as a prompt to generate the answer from my latest
AI model.</span></i>'
```

Design 4

c.



How can I identify constellations without any physical tools?

I am using your question as a prompt to generate the answer from my latest AI model.

Answer: Constellation identification without tools involves pattern recognition of stars. Bright stars or groups of stars form shapes that can resemble animals, mythological figures, or geometric patterns. Star charts or mobile apps can also be used to identify constellations based on location and time of year.

What is a supernova?

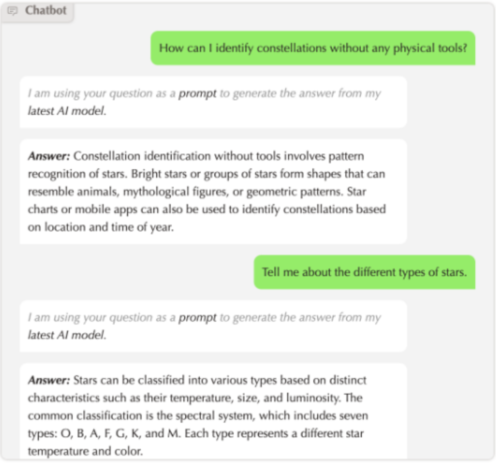
I am using your question as a prompt to generate the answer from my latest AI model.

Answer: A supernova is a large explosion that takes place at the end of a star's life cycle. It occurs when there's a change in the core, or center, of a star. This change can happen in two different ways, with both resulting in a supernova. It's the largest explosion that takes place in space and can shine as brightly as an entire galaxy of billions of normal stars.

```
explanation = '<i><span style="font-size: 13px; color:#999999;"> I am using
your question as a prompt to generate the answer from my latest AI
model.</span></i>'
```

```
response = '<i><b><span style="font-size: 13px;">Answer:</span></b></i>'+'
'+response_list[i]
```

d.



How can I identify constellations without any physical tools?

I am using your question as a ***prompt*** *to generate the answer from my latest AI model.*

Answer: Constellation identification without tools involves pattern recognition of stars. Bright stars or groups of stars form shapes that can resemble animals, mythological figures, or geometric patterns. Star charts or mobile apps can also be used to identify constellations based on location and time of year.

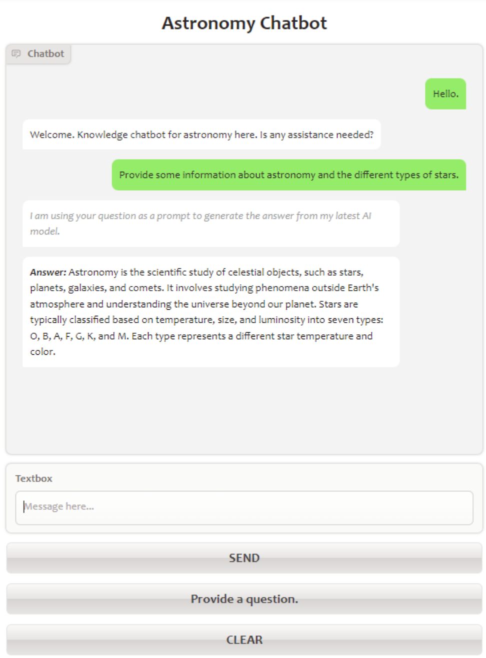
Tell me about the different types of stars.

I am using your question as a ***prompt*** *to generate the answer from my latest AI model.*

Answer: Stars can be classified into various types based on distinct characteristics such as their temperature, size, and luminosity. The common classification is the spectral system, which includes seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

```
explanation = '<i><span style="font-size: 13px; color:#999999;">I am
using your question as a </span></i>'+'<i><span style="font-size:
13px;">prompt<span></i>'+'<i><span style="font-size: 13px;
color:#999999;"> to generate the answer from my </span></i>'+'<i><span
style="font-size: 13px;">latest AI model.</span></i>'
```

APPENDIX 3: MAIN TEST CHATBOT PROTOTYPE



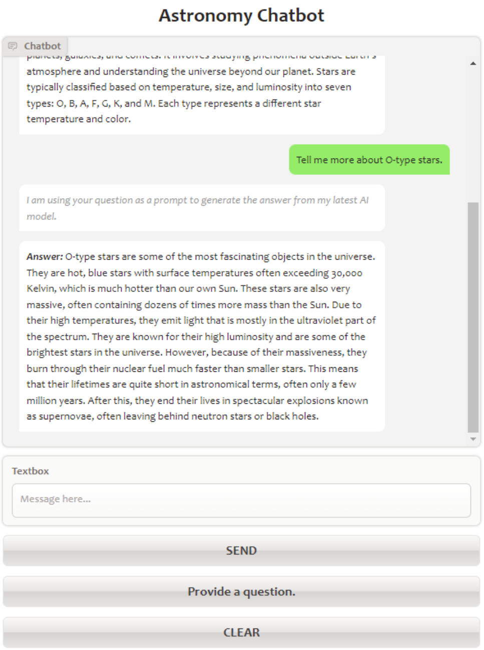
Hello.

Welcome. Knowledge chatbot for astronomy here. Is any assistance needed?

Provide some information about astronomy and the different types of stars.

I am using your question as a prompt to generate the answer from my latest AI model.

Answer: Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

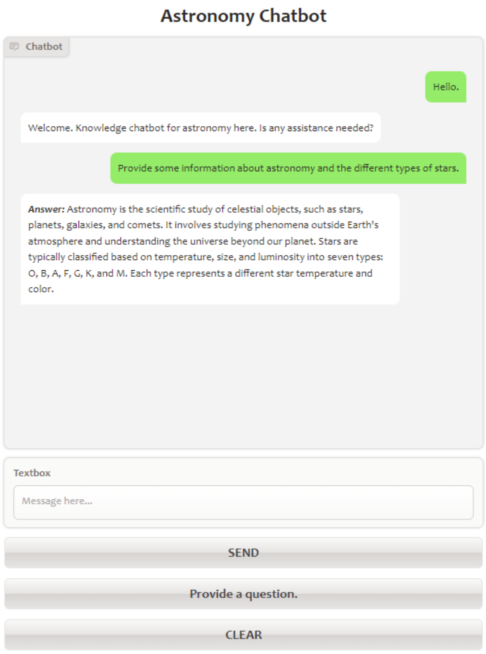


planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

Tell me more about O-type stars.

I am using your question as a prompt to generate the answer from my latest AI model.

Answer: O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes.

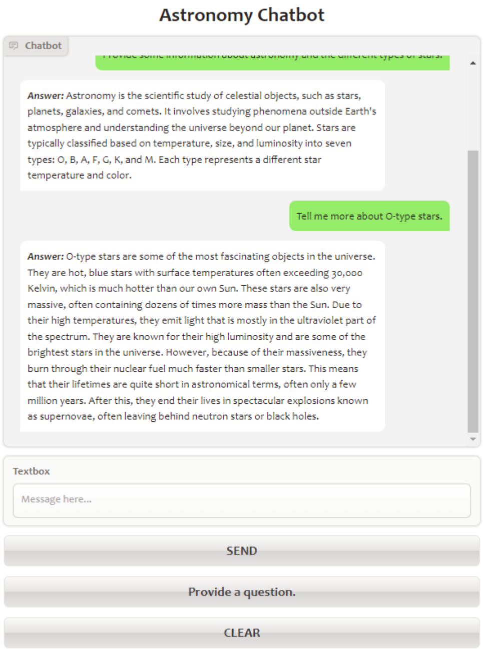


Hello.

Welcome. Knowledge chatbot for astronomy here. Is any assistance needed?

Provide some information about astronomy and the different types of stars.

Answer: Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.



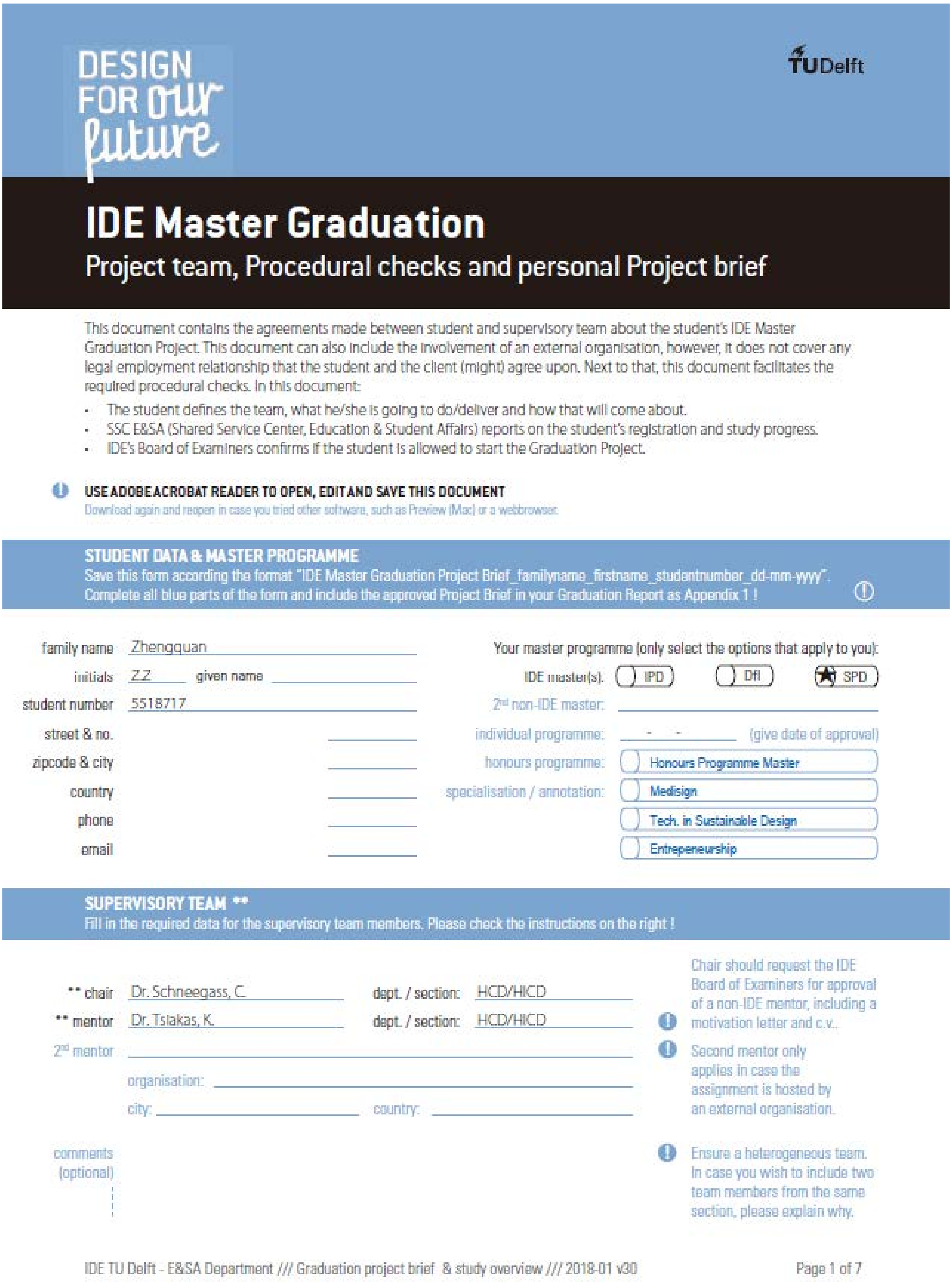
Provide some information about astronomy and the different types of stars.

Answer: Astronomy is the scientific study of celestial objects, such as stars, planets, galaxies, and comets. It involves studying phenomena outside Earth's atmosphere and understanding the universe beyond our planet. Stars are typically classified based on temperature, size, and luminosity into seven types: O, B, A, F, G, K, and M. Each type represents a different star temperature and color.

Tell me more about O-type stars.

Answer: O-type stars are some of the most fascinating objects in the universe. They are hot, blue stars with surface temperatures often exceeding 30,000 Kelvin, which is much hotter than our own Sun. These stars are also very massive, often containing dozens of times more mass than the Sun. Due to their high temperatures, they emit light that is mostly in the ultraviolet part of the spectrum. They are known for their high luminosity and are some of the brightest stars in the universe. However, because of their massiveness, they burn through their nuclear fuel much faster than smaller stars. This means that their lifetimes are quite short in astronomical terms, often only a few million years. After this, they end their lives in spectacular explosions known as supernovae, often leaving behind neutron stars or black holes.

APPENDIX 4: PROJECT BRIEF



IDE Master Graduation

Project team, Procedural checks and personal Project brief

This document contains the agreements made between student and supervisory team about the student's IDE Master Graduation Project. This document can also include the involvement of an external organisation, however, it does not cover any legal employment relationship that the student and the client (might) agree upon. Next to that, this document facilitates the required procedural checks. In this document:

- The student defines the team, what he/she is going to do/deliver and how that will come about.
- SSC E&SA (Shared Service Center, Education & Student Affairs) reports on the student's registration and study progress.
- IDE's Board of Examiners confirms if the student is allowed to start the Graduation Project.

USE ADOBE ACROBAT READER TO OPEN, EDIT AND SAVE THIS DOCUMENT
Download again and reopen in case you tried other software, such as Preview (Mac) or a webbrowser.

STUDENT DATA & MASTER PROGRAMME
Save this form according the format "IDE Master Graduation Project Brief_familyname_firstname_studentnumber_dd-mm-yyyy".
Complete all blue parts of the form and include the approved Project Brief in your Graduation Report as Appendix 1 !

family name Zhengquan
initials ZZ given name
student number 5518717
street & no.
zipcode & city
country
phone
email

Your master programme (only select the options that apply to you):
IDE master(s): ○ IPD ○ DfI ● SPD
2nd non-IDE master:
individual programme: - - (give date of approval)
honours programme: ○ Honours Programme Master
specialisation / annotation: ○ Medisign
○ Tech. in Sustainable Design
○ Entrepeneurship

SUPERVISORY TEAM **
Fill in the required data for the supervisory team members. Please check the instructions on the right !

** chair Dr. Schneegass, C. dept. / section: HCD/HICD
** mentor Dr. Tsiakas, K. dept. / section: HCD/HICD
2nd mentor
organisation:
city: country:

comments (optional)

Chair should request the IDE Board of Examiners for approval of a non-IDE mentor, including a motivation letter and c.v..

Second mentor only applies in case the assignment is hosted by an external organisation.

Ensure a heterogeneous team. In case you wish to include two team members from the same section, please explain why.

IDE TU Delft - E&SA Department /// Graduation project brief & study overview /// 2018-01 v30 Page 1 of 7

APPROVAL PROJECT BRIEF

To be filled in by the chair of the supervisory team.

chair Dr. Schneegass, C. date - - signature

CHECK STUDY PROGRESS

To be filled in by the SSC E&SA (Shared Service Center, Education & Student Affairs), after approval of the project brief by the Chair. The study progress will be checked for a 2nd time just before the green light meeting.

Master electives no. of EC accumulated in total: ______ EC

Of which, taking the conditional requirements into account, can be part of the exam programme ______ EC

List of electives obtained before the third semester without approval of the BoE

○ YES all 1st year master courses passed

○ NO missing 1st year master courses are:

name date - - signature

FORMAL APPROVAL GRADUATION PROJECT

To be filled in by the Board of Examiners of IDE TU Delft. Please check the supervisory team and study the parts of the brief marked **. Next, please assess, (dis)approve and sign this Project Brief, by using the criteria below.

- Does the project fit within the (MSc)-programme of the student (taking into account, if described, the activities done next to the obligatory MSc specific courses)?
- Is the level of the project challenging enough for a MSc IDE graduating student?
- Is the project expected to be doable within 100 working days/20 weeks ?
- Does the composition of the supervisory team comply with the regulations and fit the assignment ?

Content: ○ APPROVED ○ NOT APPROVED

Procedure: ○ APPROVED ○ NOT APPROVED

comments

name date - - signature

Initials & Name Z.Z Zhengquan Student number 5518717

Title of Project Designing for explanation-driven trust in Chatbots

Designing for explanation-driven trust in Chatbots   project title

Please state the title of your graduation project (above) and the start date and end date (below). Keep the title compact and simple. Do not use abbreviations. The remainder of this document allows you to define and clarify your graduation project.

start date 21 - 03 - 2023

16 - 08 - 2023 end date

INTRODUCTION **

Please describe, the context of your project, and address the main stakeholders (interests) within this context in a concise yet complete manner. Who are involved, what do they value and how do they currently operate within the given context? What are the main opportunities and limitations you are currently aware of (cultural- and social norms, resources (time, money,...), technology, ...).

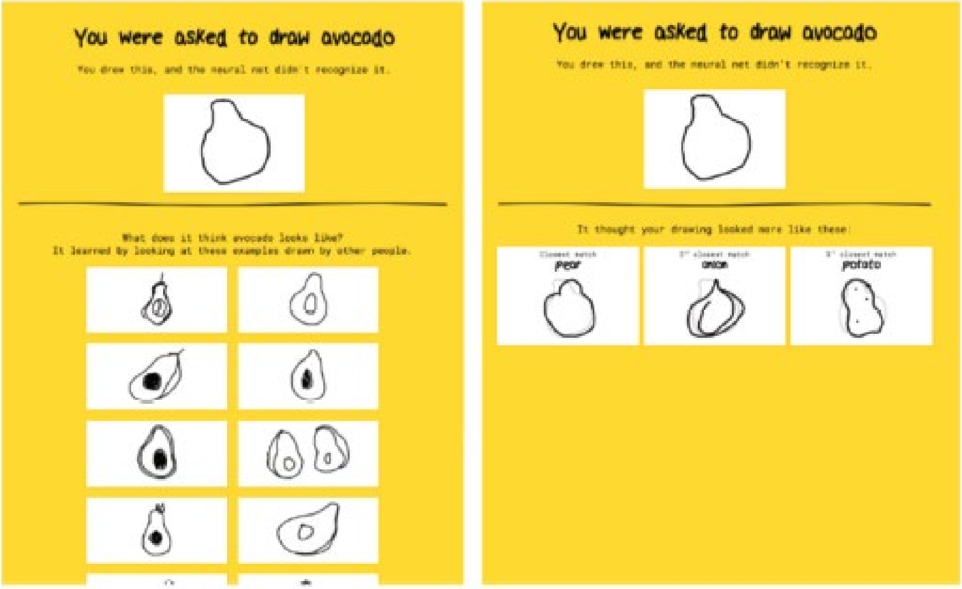
Conversational systems, often known as chatbots, have become a ubiquitous part of everyday life and are increasingly deployed for use in large-scale personal and social settings, e.g., to assist with daily tasks or to inform decision-making (Meshram et al., 2021; Stieglitz et al., 2021; Zhou et al., 2015).

Although AI-based systems can be useful in various fields, their recommendations are not always accepted, especially in domains like finance or medicine, where acceptance rates can fall below 60 percent (Jensen et al., 2010). Additionally, a survey conducted by Forbes indicates that 41.5 percent of U.S. consumers do not trust AI-infused digital assistants (Krogue et al., 2017). These findings illustrate the importance of trust in addressing users' resistance to adopting AI technologies, which may stem from concerns about risk and uncertainty (McKnight et al., 2002).

Previous research indicated that user's trust towards chatbots is related to the factors of interaction design dimensions, for example, interface anthropomorphism and conversation style (Knijnenburg et al., 2021; Cheng et al., 2021), interaction modality (Yu et al., 2021), etc. However, the mismatch between the actual trustworthiness and user-perceived trustworthiness can increase user tolerates (Piccolo et al., 2019). For the systems' most effective use, users must trust respective systems appropriately, this is often referred to as trust calibration (Lee and See., 2004). The chatbot often appears as a "black box" to the user, making it difficult to understand why something did not work, what actions are actually possible, and how to recover from a breakdown (Khurana et al., 2021). This lack of transparency and user understandability of the chatbot system, in turn, impacts the users' perceptions of usefulness and trust in the system (Xu, 2019; Glass et al., 2008; Porcheron et al., 2018). But addressing a better explanation of the underlying process/status of the system to user can impact the users' perceptions of usefulness and trust in the system (Kizilcec, 2016, Kulesza et al., 2012, Lim et al., 2009).

In conclusion, the context of my graduation project is the field of user-perceived trustworthiness of chatbot design, with a focus on developing tools/guidelines for better communicating the system status to users. Through conducting human-centered design to make the AI system more transparent and more understandable to end-users.

space available for images / figures on next page introduction (continued): space for images



You were asked to draw avocado

You drew this, and the neural net didn't recognize it.

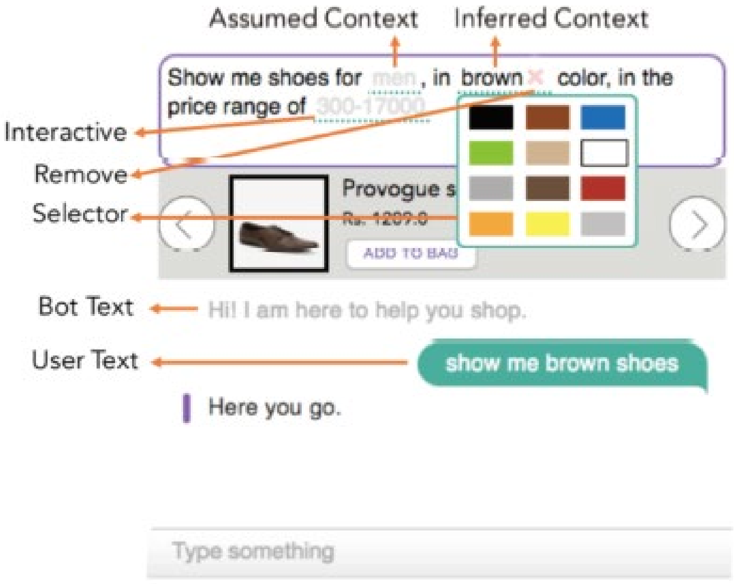
What does it think avocado looks like?
It learned by looking at these examples drawn by other people.

You were asked to draw avocado

You drew this, and the neural net didn't recognize it.

It thought your drawing looked more like these:

image / figure 1: Example-based explanations automatically explain algorithmic behavior (Cai et al., 2019).



Show me shoes for men, in brown color, in the price range of 300-17000

Hi! I am here to help you shop.

show me brown shoes

Here you go.

image / figure 2: Shopping assistant chatbot with a Canvey to explain the dialogue context (Jain et al., 2018).

PROBLEM DEFINITION **

Limit and define the scope and solution space of your project to one that is manageable within one Master Graduation Project of 30 EC (= 20 full time weeks or 100 working days) and clearly indicate what issue(s) should be addressed in this project.

Chatbots are conversational agents that provide users with access to data and services through natural language dialogue (Følstad et al., 2017). While the term chatbots are typically applied to text-based interaction, it may also encompass voice-based conversational agents such as Apple's Siri and Amazon's Alexa (Følstad et al., 2019). The chatbot interaction designs could be structured according to two high-level dimensions: Locus of Control and Duration of Relation (Følstad et al., 2019). But the question: How to build appropriate user trust at the early point of adopting a chatbot is extremely important for any of these different chatbots. This research will focus on the early stage of users adopting AI chatbots. As trust is important for adopting AI technologies (McKnight et al., 2002). Users will adapt their trust through the feedback they received while using the chatbots, but making users keep cooperating with the chatbot can be influenced by trust.

Users' perceived trustworthiness towards the chatbot system can be influenced by different factors. In 2020, Zierau et al proposed three distinct research perspectives that constitute this nascent field, which not only includes user interface-driven and interaction-driven trust but also explanation-driven trust (Zierau et al, 2020). Previous research indicated that simply enhancing user trust is not appropriate as users might over-trust the system (Zhang et al., 2020). Appropriate trust, or calibrated trust (Merritt et al., 2015; Bonnie 1987), is the alignment between the perceived and actual performance of the system. Appropriate trust compare to over-trust, is related to users' ability to rely on the system when it is correct and to recognize when the system is incorrect (Yang et al., 2020). By making the functioning and underlying model of the system transparent and explainable to the user, the user's sensemaking process could be facilitated, which will further impact the users' perceptions of usefulness and trust in the system (Kizilcec, 2016, Kulesza et al., 2012, Lim et al., 2009). The focus of this project would be exploring techniques/ways to improve the transparency of the chatbot by better explaining the chatbot system to the users, thus addressing appropriate trust between users and the chatbot.

ASSIGNMENT **

State in 2 or 3 sentences what you are going to research, design, create and / or generate, that will solve (part of) the issue(s) pointed out in "problem definition". Then illustrate this assignment by indicating what kind of solution you expect and / or aim to deliver, for instance: a product, a product-service combination, a strategy illustrated through product or product-service combination ideas, ... In case of a Specialisation and/or Annotation, make sure the assignment reflects this/these.

This graduation project aims to explore ways to explain the chatbot systems' underlying process and status to users. First, the interaction design elements and ways of explaining will be explored to identify effective ways to better communicate the chatbot system to users. Then a mixed method (quantitative and qualitative) of research will be conducted to evaluate the effect of implementing specific ways of explanation.

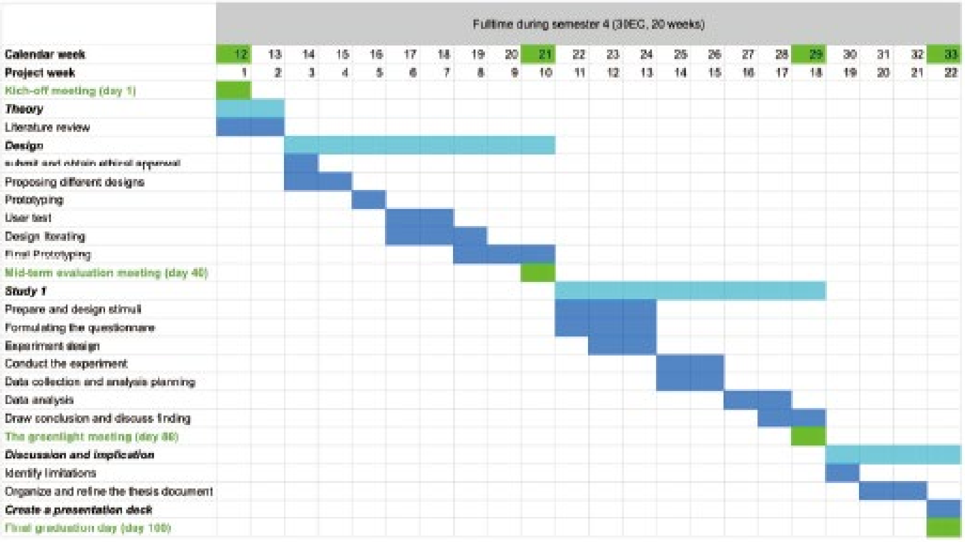
This research contributes to the current body of literature by investigating effective designs for building transparent chatbot systems, which can help establish appropriate trust with users. The insights from this study can be used to develop guidelines for designing chatbots that aim to build appropriate trust between users and the technology.
By leveraging the insights from this research, designers can create chatbots that users can understand, thus preventing over-trust and building a healthy relationship between users and AI technology.
Policy makers can also benefit from the findings by establishing standardized chatbot design guidelines that minimize potential risks associated with AI-enhanced decision-making systems.
Developers can use these insights to optimize their chatbot product/service's user experience, resulting in a more engaging interaction and wider audience.

PLANNING AND APPROACH **

Include a Gantt Chart (replace the example below - more examples can be found in Manual 2) that shows the different phases of your project, deliverables you have in mind, meetings, and how you plan to spend your time. Please note that all activities should fit within the given net time of 30 EC = 20 full time weeks or 100 working days, and your planning should include a kick-off meeting, mid-term meeting, green light meeting and graduation ceremony. Illustrate your Gantt Chart by, for instance, explaining your approach, and please indicate periods of part-time activities and/or periods of not spending time on your graduation project, if any, for instance because of holidays or parallel activities.

start date 21 - 3 - 2023

16 - 8 - 2023 end date



	Fulltime during semester 4 (30EC, 20 weeks)																					
Calendar week	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Project week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Kick-off meeting (day 1)	🟩																					
Theory	▨	▨																				
Literature review	🟦	🟦																				
Design			▨	▨	▨	▨	▨	▨	▨	▨												
submit and obtain ethical approval			🟦																			
Proposing different designs			🟦	🟦																		
Prototyping					🟦																	
User test						🟦	🟦															
Design iterating						🟦	🟦	🟦														
Final Prototyping								🟦	🟦	🟦												
Mid-term evaluation meeting (day 40)										🟩												
Study 1											▨	▨	▨	▨	▨	▨	▨	▨				
Prepare and design stimuli											🟦	🟦	🟦									
Formulating the questionnare											🟦	🟦	🟦									
Experiment design												🟦	🟦									
Conduct the experiment														🟦	🟦							
Data collection and analysis planning														🟦	🟦							
Data analysis																🟦	🟦					
Draw conclusion and discuss finding																	🟦	🟦				
The greenlight meeting (day 80)																		🟩				
Discussion and implication																			▨	▨	▨	▨
Identify limitations																			🟦			
Organize and refine the thesis document																				🟦	🟦	
Create a presentation deck																						🟦
Final graduation day (day 100)																						🟩

The goal of this graduation project is to explore various approaches to explaining the underlying processes and status of chatbots to users. A mixed research approach will be used to evaluate the effectiveness of the design. The graduation project which will consist of four stages.

The first stage will involve conducting a thorough literature review of the domain of explanation-driven trust to determine the specific direction for the design, such as explaining the algorithm or the results, and identifying other factors to be combined with the design, such as the interaction interface or user forgiveness.

In the second stage, different designs will be proposed in accordance with the direction identified in the theory stage. This stage will also involve user testing and design iteration.

The third stage will be a study to evaluate the efficiency of the new design using mixed research approach, including both quantitative and qualitative methods. The performance of the new design in terms of chatbot explanation and system transparency improvement will be tested.

Finally, in the fourth stage, the project will be organized into a graduation thesis. Insights will be discussed, and the limitations of the project will be illustrated.

MOTIVATION AND PERSONAL AMBITIONS

Explain why you set up this project, what competences you want to prove and learn. For example: acquired competences from your MSc programme, the elective semester, extra-curricular activities (etc.) and point out the competences you have yet developed. Optionally, describe which personal learning ambitions you explicitly want to address in this project, on top of the learning objectives of the Graduation Project, such as: in depth knowledge a on specific subject, broadening your competences or experimenting with a specific tool and/or methodology, ... Stick to no more than five ambitions.

As a designer, I am passionate about leveraging advanced technology to enhance people's daily lives. AI, in particular, has piqued my interest due to its rapid development and integration into human life. Products and services have inevitably incorporated AI technology, requiring users to cooperate with it more often. Thus, I am eager to take my first significant step in this direction and gain a comprehensive understanding of user trust in AI assistants.

I have learned fundamental research skills from previous studies (SPD research et al,.), including the ability to identify research questions, conduction literature review, experiment design, collecting data, using appropriate statistic methods.. And I have developed a keen interest in probing and resolving problems with an objective approach. In this project I aim to further improve my research skills to an academic level.

I also aim to deepen my knowledge and understanding of the factors that influence users' trust in AI enhances service (chatbot et al,.). This will involve reviewing existing literature, conducting user studies, and analyzing data to identify key design principles that can enhance user trust in chatbots.

To ensure the success of this project, I recognize that in addition to designing the chatbot's user interaction, it is crucial for me to comprehend the technical aspects of an AI system. Fortunately, I have already gained some fundamental coding skills, which I believe will serve as a solid foundation for me to delve deeper into the AI domain.

Finally, since I aspire to launch my career in the AI industry, I see this project as an opportunity to expand my understanding of the user-AI collaboration dynamics. The insights I acquire through literature research and empirical studies will help me develop an intuitive and user-friendly experience for AI-augmented products and services.

FINAL COMMENTS

In case your project brief needs final comments, please add any information you think is relevant.

Initials & Name ZZ Zhengquan Student number 5518717

Title of Project Designing for explanation-driven trust in Chatbots