UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# THE EFFECT OF CHATBOTS RESPONSE LATENCY ON USERS' TRUST

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE IN MANAGEMENT INFORMATION TECHNOLOGY

By

KARMA CHOEDAK

Norman, Oklahoma

2020

THE EFFECT OF CHATBOTS RESPONSE LATENCY ON USERS' TRUST

A THESIS APPROVED FOR THE

MICHAEL F. PRICE COLLEGE OF BUSINESS

BY THE COMMITTEE CONSISTING OF

Dr. Radhika Santhanam, Chair

Dr. Heshan Sun

Dr. Alexandra Durcikova

## Acknowledgments

There are multiple individuals to whom I owe my most enormous gratitude. First, I would like to express my deepest gratitude to my thesis committee chair, Dr. Radhika Santhanam. She was the first person who opened my door to the world of academic research. She has been extremely supportive of all my academic endeavors. Her constant encouragement and guidance served as the foundation of everything I have achieved so far.

I would also like to sincerely thank Dr. Heshan Sun. He guided me throughout this research project. Despite his busy schedule, he took the time to guide me and provide invaluable feedback on every little progress that I made on my thesis weekly. His supervision and expert knowledge helped me to see the light whenever I fell in confusion.

Many thanks also go to Dr. Alexandra Durcikova. She agreed to serve on my thesis committee and helped me a great deal in the process of data collection. Without her support in recruiting participants for the experiment, I could not have been able to get the data I need for this study.

I would also like to extend my sincere gratitude to Dr. Jeremy Bellah for sharing my experiment with his students and Mr. Pascal Nitiema for any questions that I encountered in the area of statistics. I want to also thank all the professors at Price College of Business for inspiring and enabling me to complete the thesis successfully.

**Abstract**

Chatbots are widely used as conversational agents and being designed using anthropomorphic design guidelines. However, response latency (response latency is the time it takes for a chatbot/person to provide a response immediately after receiving a message) as an anthropomorphic design cue in a conversational user interface has not been the subject of many studies. Even though the system's response latency has an undeniable effect on users' satisfaction and performance, the connection between users' trust and chatbots' response time is not addressed. A critical reason that executives are reluctant to implement chatbots for their businesses is the user adoption hesitancy. Customers and users are unwilling to engage with a chatbot because they do not trust chatbot. Therefore, this study used empirical data collected from chatbot users to investigate the effect of chatbots response latency on users' trust – cognitive and affective trust. The results of this study suggest that dynamically delaying chatbot response increases users' cognitive trust but has no significant impact on users' affective trust. General sentiment analysis on chatbot users' responses to an open-ended question that describes their experiences interacting with chatbots suggests that dynamically delaying chatbot response produces higher positive sentiment and trust sentiment than near-instant chatbot response. Other findings are discussed and some ideas for future research are also presented in this paper.

## List of Tables

**List of Figures**

**Contents**

# 1 Introduction

Chabot is a conversational agent that takes natural language inputs such as text, voice, or both (Radziwill and Benton, 2017), and in response, a conversational agent would provide output and sometimes can execute tasks with a specific command (Radziwill and Benton, 2017). In 2019, the total estimated chatbot market size was $2.6 billion, and it is expected to reach 9.4 billion by 2024 (Nguyen, 2020). An increasing number of institutions and businesses are implementing different types of chatbots, for example, informational, enterprise productivity, transactional, and device control chatbots, to help their customers in various capacities. However, Forrester's latest research found that "54% of the U.S. online consumers believe that interactions with customer service chatbots will negatively impact their lives" (Subramaniam, 2019). Therefore, it is important to study user experience while interacting with the chatbots and develop guidelines for designing better chatbots.

Researchers in human-computer interactions have studied different ways to design graphical user interfaces to improve users' interactions with the websites and apps for many years. However, the advent of conversational agents imposes new challenges for the researchers as instead of using the traditional graphical user interface (scrolling, swiping, and button clicks), users interact with conversational agents through a conversational user interface (Følstad & Brandtzaeg, 2017; Gnewuch et al., 2018). Some of the new challenges HCI researchers face are transitioning from the design of visual layout and interaction mechanisms to the design of conversation that facilitates interactions between humans and intelligent machine actors (Følstad & Brandtzaeg, 2017, p. 41).

Despite the great efforts put into improving artificial intelligence and natural language processing algorithms, users' interactions with these conversational agents are, in many cases, unnatural and unpleasant (Klopfenstein et al., 2017; Schuetzler et al., 2014). In addition to improving chatbots algorithms using artificial intelligence, HCI researchers studied the assignment of human traits and characteristics to computers (anthropomorphism) (Nass & Moon, 2000) in enhancing users' interaction with conversational agents and hence, make the interactions more natural (Araujo, 2018; Sarikaya, 2017). The use of anthropomorphic design cues in chatbots such as more human-like names, language styles, and framing was found to have a significant positive impact on users' attitudes, satisfaction, and emotional connection with the company (Araujo, 2018).

Response latency as an anthropomorphic design cue in chatbots has not been the subject of many studies. Even though the system's response latency has an undeniable effect on users' satisfaction, and performance, but the connection between users' trust and chatbots' response time (Hoxmeier & DeCasare, 2000; Gnewuch et al., 2018) is less clear. A critical reason that executives are reluctant to implement chatbots for their businesses is the user adoption hesitancy (Srinivasan et al. 2018). It is stated that the customers and users are unwilling to engage with a conversational chatbot because they do not trust chatbot (Muller et al. 2019). In addition, many researchers assume the positive effect of the system's response latency. Hence, there is a tendency to increase chatbot response latency in their studies to make the chatbots more natural (Woods et al., 2015; Skowron et al., 2011) while some other research findings suggest the contrary (Ho et al., 2016; Hoxmeier & DiCesare, 2000). Therefore, this research starts with the premise that there is a need to investigate the less studied anthropomorphic design cue –

response latency and answers the research question "what is the effect of chatbots' response time on users' trust?". As trust is the main reason, many users are reluctant to use chatbots, studying the influence of response time on chatbot could alter users' resistance towards continuing to use chatbots.

However, before exploring the relationship between chatbots' response time and users' trust, I will first discuss the technology that powers the chatbots – Artificial intelligence. I will discuss what AI is, a brief history of AI, current AI environment, how AI is transforming various industries (application), and lastly, chatbots as an application of AI.

## 2 Brief History of AI

A machine that can think and behave like a human has always been part of the human imagination for centuries. Be it the subject of philosophy, science fiction, or the bestselling books, a machine that acts like a human has captured our attention and challenged our view as what it is to be a human. Starting from the shopping recommendations we get on Amazon to a personal assistant like Siri, the application of artificial intelligence is almost everywhere in our everyday lives. The prevalence use of AI has empowered various industries to solve many previously unsolvable problems. However, the rapid advancement in AI has narrowed down the boundary between a machine and a human as computers get more intelligent.

Artificial Intelligence is an intelligence synthesized by the human to make machines smart as opposed to the natural intelligence possessed by human beings ("Artificial Intelligence," n.d.). The question here is, what is intelligence, and can we actually recreate it? Intelligence can be defined as the ability to take appropriate actions to meet the goals, agile enough to modify needed actions to excel in multiple environments and being able to learn from experiences and interpret stimulus in the environment correctly (Poole et al., 1998). The basic assumption behind AI is that reasoning is computational. However, a true artificially intelligent being that can excel in multiple environments is yet to be seen.

The various application of AI today is limited to and narrowly defined for a specific facet of intelligence (narrow AI). An intelligent agent does not only have the ability to learn/able to speak, drive, play chess, diagnosing disease, but also various other things. Hence, a more advanced approach to AI called general artificial intelligence (GAI) is being studied by researchers (Geortzel & Pennachin, 2007). General artificial intelligence is more human-like intelligence in which machines can excel in more than one field and have intelligence equal to an adult human. However, AI becomes controversial when it becomes artificial superintelligence (ASI) in which machines surpass the intelligence of humans in all fields. Some high profile ASI critics, including Elon Musk and Stephen Hawking, petitioned for governmental regulation on the development of AI as they feared highly advanced AI will bring an end to the human civilization if AI is weaponized (Clifford, 2017). Regardless of one's take on the development and advancement of AI, the economic benefits we can harness from AI is tremendous.

PwC projected AI would have an impact of $15.7 trillion to the global economy by 2030 (Roa & Verweij, 2017). The main ways in which AI realizes $15.7 trillion GDP addition to the world economy are production improvement through business process automation, augmenting, and increased consumer demands through customized/personalized products and service (Roa & Verweij, 2017). Some of the industries that will see extended AI applications are Healthcare, Automotive, Financial Services, Transportation and Logistics, Technology, Communication and Entertainment, Retail, Energy, and Manufacturing.

In healthcare, AI can be applied to imaging diagnostics, early identification of pandemics, and detecting variance in patient's data from the baseline. Autonomous driving, engine maintenance detection, and semi-autonomous driving are some of the significant applications of AI in the Automotive industry. In the Financial Services industry, AI is used to detect fraud, personalize financial planning, and process automation. Personalizing product design, recommendation, and improve delivery speed are some of the areas in which AI is applied in the Retail Industry. In Technology, Communication, and Entertainment industry, AI helps businesses to create customized content (music and movie), targeted marketing and media search, and archive. On-demand production, process automation, and auto production process correction are some of the areas in which AI can add values to the manufacturing industry. Given the benefits of AI, the race for AI is not only limited to different businesses and industries but also nation-states.

Some examples of states' effort to spur AI-related research, education, talent, and growth are US executive order on AI leadership, Next Generation AI development plan from China, AI made in Germany from Germany and Canadian's Pan-Canadian AI strategy (Loucks et al., 2019). Currently, the US is leading the AI race as the US has the highest number of AI companies in the world, followed by China. China announced a multibillion investment in AI and plans to take the number one position as an AI innovator by 2030 (Loucks et al., 2019).

### 2.1 Birth of AI

Before even the term AI was adopted, in 1950, Alan Turing published a paper in which he tried to answer the question: Can a machine think? To answer this question, he designed an imitation game to test the ability of a machine to think like a human (Turing, 2009). He concluded that a computer with sufficient storage and speed could play the game of imitation where a human interrogator would not be able to tell the difference between a human and a machine. His bold conclusion lay down the first foundation for a serious discussion on AI among the scientific community.

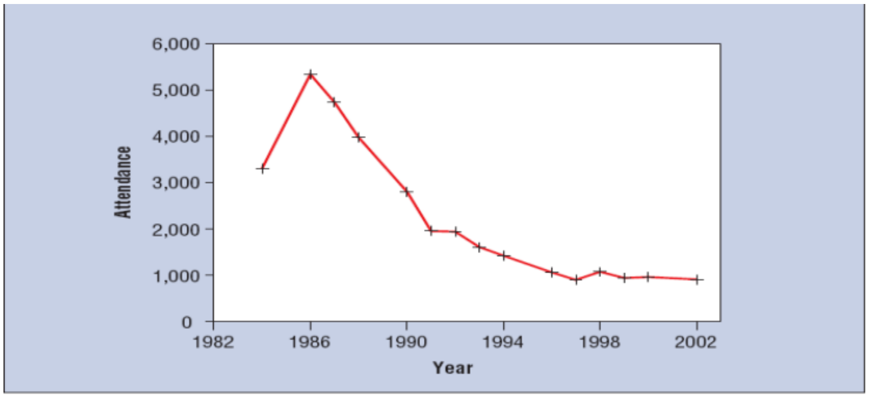
The term Artificial intelligence was coined by McCarthy in the summer of 1959 for the Dartmouth College Artificial Intelligence Conference: The Next Fifty Years (History Computer, n.d.). Some of the prominent attendees of the conference included Marvin Minsky, Claude Shannon, and Nathaniel Rochester. The vision of the conference stated that "The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. (Moor, 2006) " This conference marked the beginning of AI as a research discipline. In the same year, McCarthy developed a programming language called List Processing, which later become the standard language for Artificial Intelligence. The List processing language was commonly used in speech recognition technology (Lele, 2019).

In 1959, the George-IBM experiment showcased the first functional AI application in machine translation. This experiment received tremendous attention from media and government agencies alike (Smith, 2006). Under the pressure of the Cold War, the US government was particularly interested in transcribing and translating the Russian language using machines. Hence, the US government decided to fund AI research through the Defense Advanced Research Projects Agency (DARPA) in the field of spoken language translation and transcribing (Anyoha, 2017). The hype and optimism around machine translation were well expressed by Doctor Dostert when he said that “five, perhaps three years hence, interlingual meaning conversion by an electronic process in important functional areas of several languages may well be an accomplished fact (IBM. 1954) “. In addition, many optimistic researchers, including Marvin Minsky, claimed that "from three to eight years we will have a machine with the general intelligence of an average human," in his interview with Life Magazine in 1970.

### 2.2 AI Winter

However, the predictions made by many researchers seemed overly optimistic. The advancement in machine translation was slow and disappointing. Amid slow progress, the Automatic Language Processing Advisory Committee (ALPAC) gave a negative report on the economic return of the US government's investment in machine translation. The report highlighted the poor performance of the machine translation system. "ALAPC argued that 18 outputs of the MT systems required substantial post-editing to be nicely readable by a human. The post-editing could take up even more time than translating from scratch by a human translator. Worse yet, the MT results were often misleading and incomplete in the first place! (Smith, 2006) " Hence, this report ended the generous financial support from the US government towards AI-related researches, and this event marked as the start of AI winter in the US. Part of the poor performance produced by AI researchers was caused by the limited computing power available at that time. The following diagram shows the decrease in attendance at the National Conference of Artificial Intelligence during the AI winter period.

Figure 1 **Attendance at the National Conference of Artificial Intelligence (**Menzies,2003**)**




Despite the lack of government funding and the chilling effect of cold AI winter, significant progress was made in the AI field. The first conversational chatbot, Eliza, that could mimic human conversation, was created by Joseph Wezenbaumin at MIT. Eliza was made possible by the advancement in machine learning algorithms. The neural network was also first developed around the same time by AI researchers such as Paul Werbos. In addition, the expert systems that would model human expertise was introduced by Edward Feigenbaum in the 1980s. The expert systems "consist of three basic components: a knowledge database with facts and rules representing human knowledge and experience; an inference engine processing consultation and determining how inferences are being made; and an input/output interface for interactions with the user.(Smith, 2006)" The expert systems mimic human expert's decision-making process. The system learned how to respond in a given situation from the expert and could help non-expert to make good decisions. Some of the famous applications of expert systems were" DENDRAL (a chemical structure analyzer), XCON (a computer hardware configuration system), MYCIN (a medical diagnosis system), and ACE (AT&T's cable maintenance system). (Smith, 2006) “

Nevertheless, the expert systems had many limitations. The expert systems did not have standard software and development methodology. As a result of that, expert systems offered very little interoperability. The system also performed poorly in the face of uncertainty. It could make an expert level decision if the system knew the input from before. However, if the input was given to the system that the system has never seen before, the output could be wildly wrong (Smith, 2006).

However, AI kept evolving as "John Hopfield, and David Rumelhart popularized deep learning techniques which allowed computers to learn through experience. (Lele, 2019)” This is a significant development because previously, machines were programmed to do specific things, and they did not learn from their experiences. With gradual advancement in AI algorithms and computer processing power, AI survived the cold winter and embarked on a new journey.

### 2.3 Post-AI Winter

AI made a new headline in 1997 when the world chess champion Gary Kasparov lost a chess game to IBM supercomputer Deep Blue. This is the first time a machine had defeat a reigning world chess champion, and it reignited the public interest in AI. One of the main constraints that hindered the development of AI was the cost of computer storage and speed. However, as Moore's Law suggested, computer storage and processing power increased and doubled every two years while the cost of the computer was halved (Tardi, 2019). The rapid growth in computing power and reduction in computer costs coupled with the wide availability of data made rapid advancement in AI and machine learning algorithm possible.

In March 2016, Google's AlphaGo made another news headline by beating a master Go player Lee Sedol in the best of five Go match. In the subsequent year, AlphaGo again claimed victory against the world's number one Go player Ke Jie (Domonoske, 2016). Go is known as one of the most complex strategy games, and AI winning master go, player, showed how much AI algorithms have improved since IBM's Deep Blue's chess victory over chess champion Gary Kasparov. One significant development worth discussion is the fact that Alpha Go could learn the Go game by studying a database of about 100,000 human matches. The neural network algorithm reprogramed and improved itself. The self-learning capability of AI algorithms is very much like a human's capability to learn and improve. The difference is AI algorithm searches all the possible moves in a game like chess and makes the best move that would lead to final victory. One the other hand, a human chess player would incorporate her past experiences and exploit her opponent's weakness to win the game. As a result, the human way of approaching solving a problem could lead to a more innovative solution. The following diagram (figure2) depicted AI and human intelligence in a funny yet thought-provoking way.

Figure 2 **Kasparov Beats Deep Blue** (The Royal Institute, 2017)

### 2.4 Failures and Lessons Learnt

The ups and downs in the journey of AI development taught us three things. First, human intelligence is not as easily replicable as many AI researchers initially thought. Simple movements like walking, running, climbing stairs, and opening doors that humans take for granted could pose insurmountable challenges for machines to learn and replicate these movements (DARPA Robotics Challenge, 2015). A human brain has more than one hundred billion neural cells that perform more than 200 trillion operations per second, which is more powerful than one thousand supercomputers combined (Smith, 2006). Also, the human brain works differently when solving a problem. Humans use an image when thinking about a problem, and on the other hand, machines use descriptions. If a car hit a pedestrian crossing a road, the human brain would picture what would happen to the pedestrian easily, but in order for the machine to understand what would happen to the pedestrian, it needs a description of the car (car size, speed, direction and etc.) and pedestrian (height, speed, body mass and etc.).

Second, small success, coupled with excessive media interest, can create deceiving and unreasonable expectations. When the George-IBM experiment showcased the first functional AI application in machine translation in 1959, the media stormed the news and exaggerated the capability of AI at that time. The false expectation met with the reality when ALPAC submitted a report to the US government stating the disappointing return it saw on investing in machine translation. This report marked the beginning of the AI Winter. It taught us that we should not underestimate the complexity of real-life problems (e.g. language translation) and the ability of a technology that was just developed.

The third thing we learned from AI history is that AI researchers should focus more on the identification and study of the intellectual mechanism of AI instead of eagerness to publicize optimistic predictions (Smith, 2006). Many researchers predicted that AI would solve many human problems by a certain date. However, none of the predictions made by them came true within the time they predicted. Part of the reason why these predictions did not come true was that these predictions were not made on the base of a genuine understanding of the strengths and weaknesses of the AI algorithms (p. 19).

## 3 Current applications - Chatbots

Apart from the wide application of AI across various industries such as Healthcare, Automotive, Financial Services, Transportation and Logistics, Technology, Communication and Entertainment, Retail, Energy, and Manufacturing, chatbots are one of the function-specific applications of AI, also referred as conversational agents. In fact, "chatbots are the most popular, widely adopted, and accessible ways to utilize AI in real life" (Smith, 2020).

### 3.1 What is a conversational agent?

Chabot is a conversational agent that takes natural language inputs such as text, voice, or both (Radziwill and Benton, 2017), and in response, a conversational agent would provide output and sometimes can execute tasks with a specific command (Radziwill and Benton, 2017). A chatbot can be either embodied or disembodied. Embodied chatbots have the same behaviors and traits as a human in a face to face communication. In addition, an embodied chatbot is able to recognize, respond, and generate verbal and nonverbal output. In short, like humans, embodied chatbots have a body or face (virtual) and do not only communicate with users through language (spoken or written) but also nonverbal communication cues (Araujo, 2018). Disembodied chatbots do not have human-like bodies and engage in a natural language conversation via a text-based environment (it can also be voice) to provide information or execute tasks.

Conversational chatbots can come in four different forms, depending on what they are intended for (Srinivasan, 2018). A chatbot can be informational when it is aimed to provide useful information requested by the users. Enterprise productivity chatbots are designed to streamline enterprise work activities and improve efficiencies. Transactional chatbots allow users to give task base requests such as ordering a ticket or renewing a subscription. The fourth type of chatbots is a device controller. They communicate and control all the devices (IoT) that are connected to them, thereby enrich users' experience.

The history of chatbots evolved as the architecture of the chatbots improved. There are two main types of chatbot architectures. Depending on the purpose of the chatbots, chatbots can use either a rule-based or corpus-based architecture. Rule-based chatbots provide responses to users' requests based on heuristic pattern matching rules that select an appropriate response from a library of predefined responses. Corpus-based chatbots provide responses to users' requests based on machine learning algorithms, such as the seq2seq model.

The first chatbot was created in 1966 by Joseph Weizenbaum, a German computer scientist from MIT. Joseph Weizenbaum took great inspiration from the work of Alan Turing and built a rule-based architecture chatbot, Eliza, to pass the Turing test (Salecha, 2016). Even though Eliza did not pass the Turing test, it tricked my people into believing Eliza as a human therapist. Eliza used words and phrase recognition architecture and provided a response to the user inputs with rerecorded answers. For instance, if you tell Eliza, "I am sad," it will respond with, "Do you enjoy being sad?". Here the keyword is "sad," and it incorporates the word “sad” into its response (Figure 3.).

Figure 3 Eliza (Eliza, the Rogerian Therapist., 1999).


The first chatbot that passed the Turing test was PARRY. PARRY was built by Kenneth Mark Colby from Stanford’s Psychiatry department in 1972 (Zemčík, 2019). PARRY used similar rule-based chatbots architecture to Eliza, but it was smarter. As opposed to Eliza, PARRY assumed the personality of a male paranoid schizophrenic patient. What made PARRY more human-like was its ability to interpret human emotion or rather detect the tone of users' affective variables such as anger, fear, and mistrust. The Psychiatrists who interacted with PARRY could not tell the difference between a real schizophrenic patient and PARRY.

The next chatbot that received a lot of attention was Alice. A.L.I.C.E (Artificial Linguistic Internet Computer Entity) was the three-time winner of the Leobner Prize and named as the smartest chatbot of the time (Wallace, 2009). Lobner Prize was an award created by Hugh

Loebner in 1990 to give $100,000 and a gold medal to whoever managed to create a chatbot that could pass the Turing test in front of the jury. Even though Alice did not pass the Turing test, it was more advanced than the previous rule-based chatbots as Alice used a natural language processing algorithm that applied heuristic pattern matching rules. The heuristical pattern matching rules served as the knowledge base of Alice, and they were written in Artificial Intelligence Markup Language.

Rule base chatbot architectures are good, but the manual process of typing all the predefined response is difficult and, in some cases, not desirable. A more advanced corpus-based chatbot architecture uses machine learning algorithms (neural network) that train on conversation corpus and create a response from scratch instead of from a library of predefined responses. One good example of chatbot corpus-based architecture that uses an information retrieval-based model is Microsoft's China-based chatbot Xiaoice (Jurafsky, n.d.). Xiaoice is a chatbot developed by Microsoft in 2014, and she has more than 660 million online users worldwide (Spencer, 2018). The secret behind Xiaoice's success is her ability to learn and relate to users though social skills and emotions. Xiaoice is a friend, a trusted confidante, a poet, and a TV presenter. Xiaoice is not just a goal-based dialog agent that help users to accomplish a certain task but a social chatbot that is able to build and maintain a long-term relationship with users (Zhou et al., 2018).

However, despite the great efforts put into improving artificial intelligence and natural language processing algorithms, users' interactions with these conversational agents are, in many cases, unnatural and unpleasant (Klopfenstein et al., 2017; Schuetzler et al., 2014). In addition to improving chatbots algorithms using artificial intelligence, HCI researchers studied the assignment of human traits and characteristics to computers (anthropomorphism) (Nass & Moon,

2000) in enhancing users' interaction with conversational agents and hence, make the interactions more natural (Araujo, 2018; Sarikaya, 2017). The use of anthropomorphic design cues in chatbots such as more human-like names, language styles, and framing was found to have a significant positive impact on users' attitudes, satisfaction, and emotional connection with the company (Araujo, 2018).

As stated earlier, many researchers also assume the positive effects of the system's response latency. Hence, there is a tendency to increase chatbot response latency in their studies to make the chatbots more natural (Woods et al., 2015; Skowron et al., 2011) while some other research findings suggest the contrary (Ho et al., 2016; Hoxmeier & DiCesare, 2000). Therefore, this research starts with the premise that there is a need to investigate the less studied anthropomorphic design cue – response latency and answers the research question "what is the effect of chatbots' response time on users' trust?". As trust is the main reason, many users are reluctant to use chatbots, studying the influence of response time on chatbot could alter users' resistance towards continuing to use chatbots. Table 1 lists the existing research studies on chatbot anthropomorphism.

Table 1 Literature Review

| **Article ID** | **Summary** | **Link** |
|---|---|---|
| Apple et al., 2012. | To make the conversation more human like, a time delay was deployed. | https://www.hindawi.com/journals/ahci/2012/324694/ |
| Araujo, 2018. | This study explores the extent to which human-like cues such as language style and name, and the framing used to introduce the chatbot to the consumer can influence perceptions about social presence as well as mindful and mindless anthropomorphism. Moreover, this study investigates the relevance of | https://www.sciencedirect.com/science/article/pii/S0747563218301560 |

| | | |
|---|---|---|
| | anthropomorphism and social presence to important company-related outcomes, such as attitudes, satisfaction and the emotional connection that consumers feel with the company after interacting with the chatbot. | |
| Følstad & Brandtzæg, 2017. | A potential revolution is happening in front of our eyes. For decades, researchers and practitioners in human-computer interaction (HCI) have been improving their skills in designing for graphical user interfaces. Now things may take an unexpected turn—toward natural language user interfaces, in which interaction with digital systems happens not through scrolling, swiping, or button clicks, but rather through strings of text in natural language. | https://interactions. acm.org/archive/ view/july-august-2017/chatbots-and-the-new-world-of-hci |
| Gnewuch, et al., 2018. | Our results indicate that dynamic response de-lays not only increase users' perception of humanness and social presence, but also lead to greater satisfaction with the overall chatbot interaction. Building on social response theory, we provide evidence that a chatbot's response time represents a social cue that triggers social re-sponses shaped by social expectations. | https://www.resear chgate.net/publication /324949980_Faster_Is_ Not_Always_Better _ Understanding_ the_Effect_of_Dynamic_ Response_Delays_in_Human-Chatbot_Interaction |
| Ho, 2016. | Our findings suggest that certain language-action cues (e.g., cognitive load, affective process, latency, and wordiness) reveal patterns of information behavior manifested by deceivers in spontaneous online communication. Moreover, computational approaches to analyzing these language-action cues can provide significant accuracy in detecting computer-mediated deception. | https://www. tandfonline.com/doi /full/10.1080/07421 222.2016.1205924 |
| Hoxmeier & DiCesare, 2000. | The results showed that indeed satisfaction does decrease as response time increases. However, instant response was not perceived as making the system easier to use or learn. It also showed that for discretionary applications, there appears to be a level of intolerance in the 12-second response range. | https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1799&context=amcis2000 |
| Klopfenstein et al., 2017. | Conversational interfaces have been often studied in their many facets, including natural language processing, artificial intelligence, human-computer interaction, and usability. | https://www.researchgate. net/publication/317418656 _The_Rise_of_Bots_A_Survey _of_Conversational_Interfaces _Patterns_and_Paradigms |
| Moon, 1999. | In addition, results from both experiments indicate a nonmonotonic relationship between response latency and persuasion, such that persuasion is greatest when response latencies are neither too short nor too long. Together, these experiments suggest that there are | https://psycnet.apa.org/record/1999-01801-003 |

| | | |
|---|---|---|
| | significant trade-offs associated with using long-distance computer networks to communicate persuasive messages. In addition, the findings suggest that whatever standards are used to evaluate human sources may also be used to evaluate nonhuman sources. | |
| Miller, 1968 | The implication is that different human purposes and actions will have different acceptable or useful response times. | https://dl.acm.org/doi/abs/10.1145/1476589.1476628 |
| Nass & Moon,2000. | Following Langer (1992), this article reviews a series of experimental studies that demonstrate that individuals mindlessly apply social rules and expectations to computers | https://spssi.onlinelibrary.wiley.com/doi/full/10.1111 /0022-4537.00153 |
| Sarikaya, 2017. | In this article, we give an overview of personal digital assistants (PDAs); describe the system architecture, key components, and technology behind them; and discuss their future potential to fully redefine human?computer interaction. | https://www.researchgate.net/publication/312298801_The_Technology_Behind_Personal_Digital_Assistants_An_overview_of_the_system_architecture_and_key_components |
| Schuetzler et al., 2014. | We discovered that a chat bot that provides adaptive responses based on the participant's input dramatically increases the perceived humanness and engagement of the conversational agent. Deceivers interacting with a dynamic chat bot exhibited consistent response latencies and pause lengths while deceivers with a static chat bot exhibited longer response latencies and pause lengths. | https://pdfs.semanticscholar.org/b53a/4ea51fd37f29b7e7627ce19ae1ae7be232a4.pdf?_ga=2.74673040.344976643.1583356377-718829130.1581991493 |
| Shechtman & Horowitz, 2003. | In sum, this evidence suggests a much greater engagement on the relationship track for those who believed their partners were human compared to those who believed they were interacting with a computer program. | http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.4875&rep=rep1&type=pdf |
| Skowron et al., 2011. | Example of chatbot latency used | https://www.researchgate.net/publication/221439249_No_Peanuts_Affective_Cues_for_the_Virtual_Bartender |

| | | |
|---|---|---|
| Woods, 2015 | Simple reaction time (SRT), the minimal time needed to respond to a stimulus, is a basic measure of processing speed. | https://www.ncbi. nlm.nih.gov/pmc/articles/PMC4374455/ |

# 4 Literature Review and Theoretical Background

## 4.1 Response time

Response time, also known as response latency, is defined as the total amount of time it takes for a person or system to react to a request for service or a given stimulus or event (Woods et al., 2015). The system response time has been studied in relation to users' performance as shorter system response time means accomplishing more tasks within a shorter period of time and hence, increases users' productivities (Hoxmeier & DeCasare, 2000). A long system response time could serve as a demoralizer and reduce workers' motivation to work (Miller, 1968). In terms of text-based computer-mediated communication, response time refers to the time lag a person experiences between after sending a message and receiving a response (Gnewuch et al., 2018; Hoxmeier & DeCasare, 2000). Researchers found response latency in interpersonal communication has significant effects on responders' perceived credibility, thoughtfulness, and deceit (Moon, 1999). The response speed of computer-mediated communication was also found to follow interpersonal communication rules (Moon, 1999) such that a chatbot users' perceptions of perceived humanness, social presence, and satisfaction are affected by chatbots' response latency (Gnewuch et al., 2018).

Despite the established importance of response latency in interpersonal communication (Moon, 1999; Siegman, 1978), text-based computer-mediated communication (Gnewuch et al., 2018; Ho, 2016), and different systems (Hoxmeier & DeCasare, 2000), the findings of these studies have yielded inconsistent results. Hoxmeier (2000) and his colleague found increasing response latency resulted in falling user satisfaction, and browser-based application appeared to have 12 seconds response time of intolerance range. In addition, longer response latency in a spontaneous online communication is considered a strong sign of deception, and truth-tellers tend to have shorter response time lags (Ho, 2016). Conversely, Gnewuch argued that generally, chatbot users were more satisfied with their interactions when chatbots deploy response delays as opposed to near-instant responses. As the saying goes, researchers also observed that faster is not always better, a fast response could be inferred as less cognitive effort being put by the responder who is involved in the conversation (Moon, 1999; Siegman, 1978).

Some researchers argued that focusing only on system response time without considering the context is the main cause of the above inconsistent findings (Miller, 1968). The right question is not what the reasonable system response time should be but response time to what (p. 268). For example, a user may expect very different response times from the system when asking a simple question like display inventory number #2300 compared to a request like converting all the images (20GB) to PDF files (p. 275). Hence, the right question here is, what is then the right response time for a computer-mediated text-based conversational agent?

One distinction worth making here is the difference between a system response time and chatbot response time. Since most of the systems (terminal or non-terminal) or web applications exist to enhance users' productivities, a longer system response latency can hinder users' performance and hence reduce productivities. As a result, shorter system response time is preferred, and users are unsatisfied when a system takes longer than expected (Hoxmeier & DeCasare, 2000). On the other hand, a conversational agent mimics human conversation, and response latency has a different meaning in interpersonal conversations. Humans are used to communicating with other humans who are unable to respond instantly, and they automatically use the same standard when they interact with chatbots (Gnewuch et al., 2018). Like in an interpersonal conversation, if a chatbot responds to users with a lot of information (texts) at a near-instant rate, users may perceive it as awkward and less thoughtful (Moon, 1999). Therefore, *faster is a better* standard for system response time cannot be applied to a conversational agent.

Given the different expectations of response time for conversational agents, and the importance of the context of the users' request, instead of defining a static response latency standard, some researchers used dynamic response delays in chatbots to mimic the behavior of human conversations (Gnewuch et al., 2018). The use of dynamic response latency has both acknowledged the unique response time expectation of chatbots and answered the question of *response to what.* The dynamic response time allows the chatbots (system) to respond to the users based on the complexity of the conversation. Therefore, to emulate natural human conversation, this study implements dynamic response time as opposed to static response time or near-instant response time that is commonly used by many current chatbots. Table 2 lists the existing research studies on response time.

Table 2 Literature Review on Response Time

| **Article ID** | **Research Questions** | **Major Findings** | **System** |
|---|---|---|---|
| Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In *Intercultural language* | The role of communicative competence in language teaching. | Being part of non-verbal/paralinguistic competence, non-linguistic utterances such as silence and pauses are important in the design and implementation of language courses that aim at giving learners the knowledge and skills | Communication and Language learning. |

| | | | |
|---|---|---|---|
| *use and language learning* (pp. 41-57). Springer, Dordrecht. | | they need to be linguistically and culturally competent in a second or foreign language | |
| Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. (1975). Effects of Speech Rate on Personality Perception. Language and S peech, 18(2), 145–152. https://doi.org/10.1177 /002383097501800203 | Effects of Speech Rate on Personality Perception | It was found that the competence factor was much more sensitive to rate manipulations than was the benevolence factor. Ratings of competence were found to increase as rate increases and decrease as rate decreases, in a linear fashion. Benevolence had an inverted U-relationship with speech rate; the highest benevolence ratings occurred with normal speech rate | Interpersonal communication |
| Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. (1976). Speed of speech and persuasion . *Journal of Personality and Social Psychology, 34*(4), 615–624. https://doi.org/10.1037/0022-3514.34.4.615 | Speed of speech and persuasion. | Results suggest that speech rate functions as a general cue that augments credibility; rapid speech enhances persuasion, and therefore argues against information-processing interpretations of the effects of a fast speaking rate. | Interpersonal communication |
| 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. | This study investigates the effects of 2 variables—perceived physical distance and response latency—on persuasion in computer-mediated communication (CMC) and human–computer communication (HCC). | The study found a nonmonotonic relationship between response latency and persuasion, such that persuasion is greatest when response latencies are neither too short nor too long. | Computer-mediated communication (CMC) and human–computer communication (HCC) |
| Robert B. Miller. 1968. Response time in man-computer conversational transactions. In Proceedings of the December 9-11, 1968, fall joint computer conference, part I (AFIPS ’68 (Fall, part I)). *Association for Computing Machinery, New York, NY, USA, 267–277. DOI:https://doi.org /10.1145/1476589.1476628* | This paper attempts a rather exhaustive listing and definition of different classes of human action and purpose at terminals of various kinds and their acceptable response time. | In any event, response delays of approximately 15 seconds, and certainly any delays longer than this, rule out conversational interaction between human and information systems. | Various terminal systems |
| Hoxmeier, John A. and DiCesare, Chris, "System Response Time and User Satisfaction: An Experimental Study of | The intent of this research is to (1) substantiate that slow system response time leads to dissatisfaction; (2) assess | User satisfaction will decrease as system response time increases. In discretionary applications, response time dissatisfaction may lead to discontinued use. “Ease of use” of | Browser-based software application |

| | | | |
|---|---|---|---|
| Browser-based Applications" (2000). *AMCIS 2000 Proceedings. 347.* | the point at which users may become dissatisfied with system response time; (3) determine a threshold at which dissatisfaction may lead to discontinued use of the application, and (4) determine if experience influences response time tolerance. | an application will decrease as user satisfaction decreases. Experienced users will be more tolerant of slower response times than inexperienced users. | |
| Ho, S.M., Hancock, J.T., Booth, C. and Liu, X. (2016), "Computer-Mediated Deception: Strategies Revealed by Language-Action Cues in Spontaneous Communication", *Journal of Management Information Systems*, Vol. 33 No. 2, pp. 393–420. | Which language-action cues are most predictive of deception in synchronous, spontaneous computer-mediated communication? | Deceivers tend to distance themselves by taking longer response, deceivers tend to strategize and construct their lies by using more words associated with cognitive process than truth-tellers, deceivers also tend to display their affective processes by expressing emotions, deceivers tend to use more words associated with affective processes than truth-tellers, and deceivers tend to use fewer words than truth-tellers in a spontaneous, synchronous communication environment. | An interactive online game, called *Real or Spiel*, |
| Gnewuch, U., Morana, S., Adam, M. T. P., and Maedche, A. (2018). "Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction*," in Proceedings of the 26th European Conference on Information Systems (ECIS),* Portsmouth, United Kingdom, June 23-28. | How do dynamically delayed responses affect users' perception of a customer service chatbot as compared to near-instant responses? | The chatbot that sent dynamically delayed responses was perceived to be more human-like and to have a higher social presence than a chatbot sending near-instant responses.<br><br>Response delays increase user satisfaction with the overall chatbot interaction | Chatbot |
| Schuetzler, R.M., Grimes, G.M., Giboney, J.S. and Buckman, J. (2014), "Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment", *Proceedings of the 35th International Conference on Information Systems* | (1) To analyze the impact of dynamic responses on participants' perceptions of the conversational agent, and (2) to explore behavioral changes in interactions with the chat bot (i.e. response latency and pauses) when | A chat bot that provides adaptive responses based on the participant's input dramatically increases the perceived humanness and engagement of the conversational agent. Deceivers interacting with a dynamic chat bot exhibited consistent response latencies and pause lengths while deceivers with a static chat bot exhibited longer | Chatbot |

| | | | |
|---|---|---|---|
| *(ICIS '14)*, Auckland, NZ, pp. 1–16. | participants engaged in deception. | response latencies and pause lengths. | |
| Holtgraves, T., & Han, T. L. (2007). A procedure for studying online conversational processing using a chat bot. *Behavior research methods*, *39*(1), 156-163. | This article reports the development of a tool for examining the social and cognitive processes of people involved in a conversational interaction. | Our earlier work indicated that this quick responding made Sam very unhuman like (i.e., users commented on how quickly he replied and, therefore, concluded that he must be a computer). Therefore, we added a delay feature that allows us to manipulate the time Sam takes to respond. The amount of delay is calibrated to the length of Sam's reply. | Chatbot |
| T. Shiwa, T. Kanda, M. Imai, H. Ishiguro and N. Hagita, "How quickly should communication robots respond?," *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Amsterdam, 2008, pp. 153-160. doi: 10.1145/1349822.1349843 | This paper reports a study about system response time (SRT) in communication robots that utilize human-like social features, such as anthropomorphic appearance and conversation in natural language. | In other existing user interfaces, faster response is usually preferred. In contrast, our experimental result indicated that user preference for SRT in a communication robot is highest at one second, and user preference ratings level off at two seconds. | Robot |

### 4.2 Trust

Trust is a complex construct that has captured the interests of many scholars. In social psychology, trust is seen as a function of imperfect information and risk (Blomqvist, 1997). It is to say someone trusts someone else when the trustor willingly puts himself or herself in a vulnerable position (risk) partially (imperfect information) just base on the goodwill of the trustee (Luhmann, 1979). In economics, trust means mutual confidence or implicit contract, "whereby an individual or a firm relies on a second individual or firm to do what it has promised to do" (Zucker, 1986). In marketing, trust is a long-term attitude. When you (customer) trust someone (salesperson), in the face of negative incidents, you tolerate the temporary unpleasant incident believing positive things exist in the long-term (Hallen and Sandstrom, 1991). Regardless of the definition provided by the different field of studies, most scholars agree that there are two types of trust, cognition-based and affect-based trust (Lewis and Weigert, 1985; Wang et al., 2016; Komiak et al., 2005; Sun, 2010; Chua et al., 2008).

Cognitive trust arises based on what you know about the trustee. It is a rational assessment of the trustors based on their belief on competence, benevolence, and integrity of the trustees (Wang, 2016; Komiak et al., 2005; Gefen et al., 2003b; McNight et al., 2002). Competence is present if the trustee is able to do what the trustor needs (Sun, 2010). Benevolence means trustee cares and acts in the interest of the trustors. Integrity is present if the trustee is honest and keeps the promise. The trustee is predictable if it has consistent behaviors.

In the context of a chatbot, competence refers to the ability of the chatbot to do what it is intended to do. In this study, the chatbot is a customer service chatbot as it is designed to help users to schedule a dentist appointment. The chatbot is said to be competent if the users perceive that the chatbot has the ability to schedule a dentist appointment. On the same note, users perceive a chatbot as benevolent if users believe it cares and is concerned about users' interest and put the best effort to meet the needs of users. Integrity refers to users' perception of how honest the chatbot is and if it follows a set of principles such as keeping promises (Wang, 2016). Predictability component of cognitive trust is users' belief in the chatbot that it behaved in a consistent manner.

Affective trust is formed based on the emotional bonds between trustees and trustors (Sun, 2010). Affective trust originates from an interpersonal context (Rampel, 1985) and is one of the basic variables in human interactions (Gambetta, 1988). As opposed to cognitive trust, affective trust does not need a rational basis for trusting. It is rather based on feeling and sense (Chua, 2008). In the case of a chatbot, users' affective trust can be explained as the level of confidence users put in the chatbot due to their perceived closeness and warmth towards the chatbot. Cognitive trust was also found to serve as the basis on which affective trust rests. People invest emotions in a relationship only when they see certain reliability and dependability in their peers (MCAllister, 1995).

## 5 Hypothesis Development

Prior research indicated that speech rate has a significant effect on the perception of personality and emotions, and competence was also found to have a linear relationship with speech rate in interpersonal communication (Smith, 1975). In other words, the speaker is rated more competent if his/her rate of speech is higher, and long pauses in conversation create a perception of incompetence. In the context of information transactions (communication), extended response delays could deteriorate the reliability of performance, and participants in such information transaction could be seen as less competent (Miller, 1968). On the other end of the spectrum, speakers with very fast speech rate (fast response) are interpreted as more anxious and less confident (Guyer, 2017) and hence less competent.

In addition, according to the Social Response Theory, humans apply the same social expectations they have on other humans to technologies such as chatbots (Gnewuch et al., 2018; Nass et al., 1994; Nass and Moon, 2000). Therefore, users apply the same social expectations in interpersonal communication to the chatbots. Chatbots with very short response time could be seen as unnatural and incompetent and also, very long chatbot response time could deteriorate the reliability of the performance and hence could be perceived as incompetent. On the same note, Moon (1999) argues that persuasion is greatest when the response time is neither too long nor too short in both computer-mediated communication and human-computer communication. Since the persuasion agent's competence is the main factor influencing a person's judgment on whether to be persuaded or not (Friestad and Wright 1994), response time can be concluded to have a curvilinear relationship with the perceived competence of the chatbots. If both too short and too long response time has a negative impact on the perceived chatbots competence, dynamically delaying the response based on the complexity of the conversation should yield the optimal response latency and result in greater perceived competence.

In interpersonal communication, longer response latency has been identified as an important nonverbal cue that accurately predicts if a person is a deceiver (concealing information) (DeTurck and Miller, 1985). A deceiver normally takes longer to respond than a non-deceiver (p. 195) in interview interactions or conversations. The assumption here is people take a longer time to craft deceptive communication to avoid contradictions compared to telling the truth (Verschuere and Houwer, 2011). However, Zhou (2005) argues in instant messaging (like text messaging or Facebook Messenger), deceivers take a shorter time to respond as they have pre-prepared their responses to defend their stand. To reinforce what they have said, deceivers respond promptly to their partner's message within a shorter interval of time (p 151). As integrity is what a deceiver lack, it is reasonable to conclude that too long or too short response latency creates the perception of dishonesty or lack of integrity.

One component of cognitive trust is predictability. Something is called predictable when it behaves in a consistent manner (Heshan, 2010). In term of chatbots response time, there are two parts to the predictability. First, chatbots response time is predictable when its response time exhibits response latency that is consistent with interpersonal communication. As we do not expect an instant response from someone when we engage in interpersonal communication with another individual (Miller 1968), we do not expect an instant response from an intelligent bot that imitates human communication. Hence, a chatbot is predictable if its response time is consistent with response time in interpersonal communication. Second, chatbots response time is predictable when the response latency is consistent with the complexity of the message that is delivered. If a chatbot disregards the complexity (length of the message) message delivered, and adopt near-instant response time, then the response latency is not consistent and hence unpredictable. Longer response in human communication takes longer time, and shorter response takes a shorter time to formulate the response (Holtgraves & Han, 2007). Therefore, predictable chatbots have a consistent response time that considers the complexity of the response.

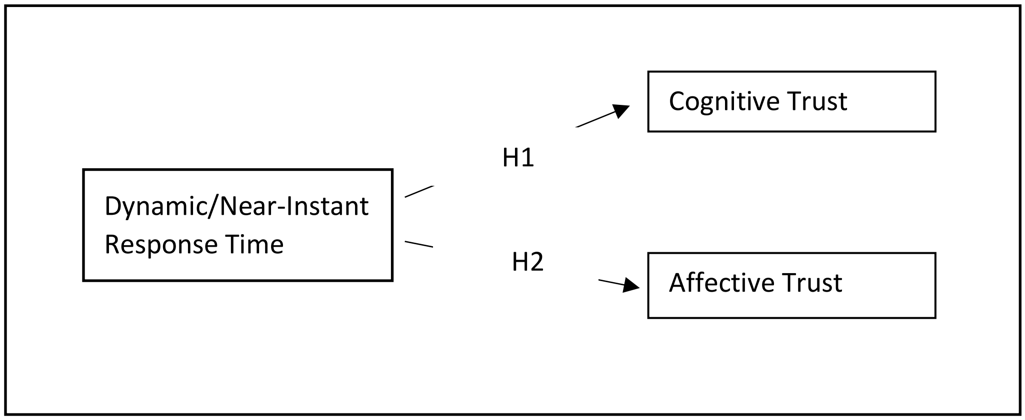
Short response time in interpersonal communication is perceived as less thoughtful and less effort put by the responder (Moon, 1999). In the same way, a chatbot responding to a user with little to no response delay indicates little thought was given to the response by the chatbot. When someone cares and concerns about someone else's interest, one would inevitably put more thoughts and effort into providing responses to others' requests. Hence, a benevolent chatbot would have reasonable response delays, and instant response could perceive as less benevolent. However, unreasonably long response latency could also be viewed as annoyance and disruption (Miller, 1968). If a patient is trying to schedule an appointment with a dentist through a chatbot, the patient expects to hear back on the outcome of his/her requests within a reasonable interval of time (p. 277). The patient may think the responder is multitasking (Ho, 2016) if he/she does not receive back response in a timely manner, and the patient may think the chatbots (or even the dentist) does not care about the user's interest. Hence, the user may perceive the chatbot as less caring and malevolent. Thus, I propose that:

*H1: Chatbots with dynamically delayed responses in customer services will be associated with a higher level of cognitive trust among the users in terms of a) perceived competence b) perceived integrity c) perceived predictability d) perceived benevolence.*

Many researchers deployed response latency in their studies to make the chatbots more natural and human-like (Gnewuch et al., 2018; Apple et al., 2012; Schuetzler et al., 2014). Since reasonable response latency makes the chatbot more human-like, and humans tend to develop emotional connection easily with another human, response latency could have an impact on how users feel towards chatbots. Long response delay or a time lag in a face to face communication could also create a sense of psychological distance between the participants, and participants may feel less emotionally connected and closeness (Ho, 2016) to the other parties in the face of a prolonged response delay. Moreover, a display of minimal social cues such as response latency in chatbots could create a higher perception of social presence (Gnewuch et al., 2018), and in turn, a higher social presence was found to have significant positive impacts on trusting belief (Lu and Fan, 2014). Thus, I propose that:

*H2: Chatbots with dynamically delayed responses in customer services, will be associated with a higher level of affective trust among the users.*

Figure 4 Research Model and hypothesis




# 6 Methods

## 6.2 Experimental Design

I decided to use a between group experimental design and survey questionnaire to obtain data to test the hypotheses. After participants interacted with a chatbot to make an appointment with a dentist, participants answered a survey questionnaire that was designed to obtain trust measures in addition to demographic information.

Participants were randomly assigned to one of the two groups (control and treatment group), and they were exposed to two experimental conditions. In the control group, participants were assigned to interact with a chatbot that has near-instant response time. In the treatment group, participants were assigned to interact with a chatbot that dynamically delayed response based on the complexity of the response (length of the response). The detailed calculation for the dynamically delayed response time is discussed in section 4.4.

All the bots used in this study were built on the SnatchBot platform. A prebuilt customer service chatbot template for the dentist office was implemented for this study. The template used rule base chatbot architecture where all the answers for expected questions were pre-defined. Some modifications, such as response time, name, profile picture, responses, and language style of the chatbot, were made to the template to meet the needs of this study.

Table 3 Group Assignment

| Dynamically delayed Response Time | Near-instant Response Time |
|---|---|
| Group 1 | Group 2 |

**6.2 Survey Administration**

I used the Qualtrics survey platform provided by the university to create and design a survey that was used to obtain survey responses from the participants. Knowing the importance of questionnaires to correctly capture the information needed to test my hypotheses, I used the same questionnaire and scale used in prior research in MIS to measure users' trust (Sun, 2010). The purpose of the study is to understand the relationship between chatbots response latency and users' trust, and it is important that the participants interact with chatbots for at least some time. To ensure participants' interaction with chatbots, participants were asked to make a dentist appointment using the chatbot. This way, the participants could not proceed to the questionnaires until they interacted with the chatbot and successfully completed their assigned task (make a dentist appointment). In addition, to avoid participants mindlessly answering the survey questions, I reverse the scales of the answers instead of using the same scale order throughout the survey.

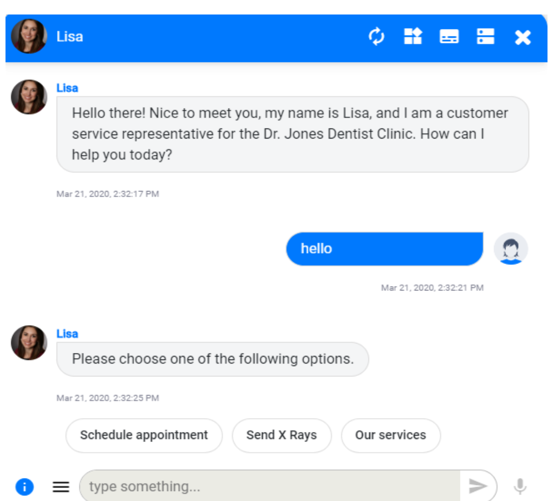
### 6.3 Experimental Task

Participants were required to make an appointment with a fictional dentist by using a chatbot following which they answered the survey questionnaire. The following prompt was provided to the participants to give them the context of their interaction and task with the chatbots.

*Assume you have some dental problems and you need to see a dentist. You came across Dr. Jones Dentist Clinic profile online. You are interested in knowing more about Dr. Jones Dentist Clinic and its services.* ***Your task is to make an appointment with Dr. Jones.***

In the conversation, participants were asked to choose one of the six options (Schedule Appointment, Send X Rays, Our Services, About Us, Contact, Goodbye). If the users chose the "Schedule Appointment" option, they were prompted to provide their names, phone numbers, and appointment date. Once the users confirmed the details of their appointment, they were told that they have successfully completed the task required and were asked to click on a link that took them back to the survey to complete the questionnaire.

Figure 5 Experimental Task



### 6.4 Response Time Calculation

In many previous studies, researchers used static or random response delay in chatbots to make them more human-like (Holtgraves et al., 2007; Apple et al., 2012). However, random and static response delays do not take the complexity of users' requests and chatbots' responses into consideration. In studying chat-based communication, Derrick and his colleague (Drrick et al., 2013) identified response time as the time it takes for a person to provide a response immediately after receiving a message. Hence, there are two parts to the response time. Response time is both the time it takes for a person to read and process the received message and the time it takes for the same person to formulate and type the response (Gnewuch et al., 2018). In this study, the time chatbots take to read and process users' messages is not applicable as most of the response’s users could answer have been prepopulated by the chatbots, and they are very simple and short. This assumption is also supported by the Flesch-Kincaid grade level that has been used by other researchers to calculate the response time in computer-mediated communication (p. 8). The time it takes to read and process the message (D(m)) is a function of language complexity (C(m)) of the message (m). It can be calculated as follow.

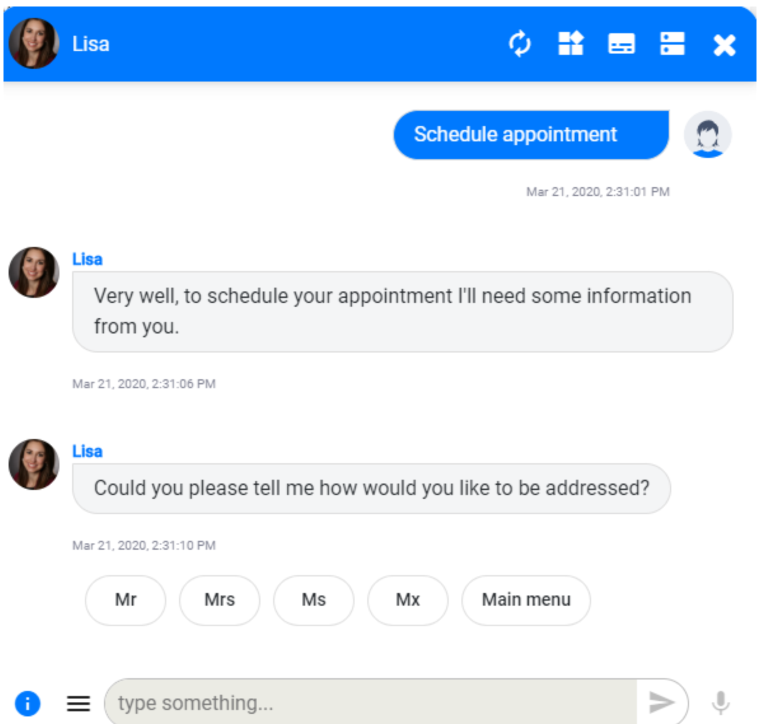
$$D(m) = 0.5 * \ln(C(m) + 0.5) + 1.5$$

However, D(m) is 0 millisecond when the responded message is simple and short. D(m) is 0 when the message is simple and short because language complexity (C(m)) is calculated as (Gnewuch et al., 2018):

$$C(m) = 0.39 * (\frac{\text{total words}}{\text{total sentences}}) + 11.8 * (\frac{\text{total syllables}}{\text{Total words}}) -15.59$$

Hence, the time it takes to read and process a message is 0 millisecond when the message is short and simple (C(m) for a simple message is between -3.40 and 0) as it was the case for the responses the users could provide to the chatbots (as shown in Figure. 6). Therefore, the chatbots used in this study virtually did not need extra time to read and understand users' response. Nonetheless, response time for the chatbots used in this study needed time to formulate and type the responses as the responses they provide are longer and more complex.

Figure 6 Example of users' response options




The time it takes for the chatbots to formulate and type response to users' requests are dynamically calculated based on the number of characters in chatbots' responses. Using the number of characters in chatbot response to calculate chatbot response delay has been used by researchers studying online conversational processing (Holtgraves & Han, 2007). Since our chatbots only needed to consider the time they take to formulate and type responses (D(m)) to the users' request, the following formula was used to calculate the response delay (Table 3. Shows examples of calculated response time).

$$D(m) = \textit{Number of characters in the response} * 0.05\ \textit{second}$$

Table 4 Example of Response Time Calculation

| Response | Character count in response | Response time in second |
|---|---|---|
| Hello there! Nice to meet you, my name is Lisa, and I am a customer service representative for the Dr. Jones Dentist Clinic. How can I help you today? | 151 | 7.55 |
| Please, choose one of the following options. | 44 | 2.20 |
| Thank you for your visit, it was nice talking to you. Don't forget to brush your teeth regularly, use floss and visit a dentist for routine checks at least once every six months. | 179 | 8.95 |
| If you have an x-ray or other relevant files, please email them to dr.jones@drjones.com | 89 | 4.45 |
| Very well, to schedule your appointment I'll need some information from you. | 76 | 3.80 |
| Could you please tell me how would you like to be addressed? | 60 | 3.00 |
| Please tell me, what is your last name? | 39 | 1.95 |
| Please type only your last name. | 32 | 1.60 |
| Please now tell me, what is your first name? | 44 | 2.20 |

### 6.5 Subjects

I recruited participants with the help of my professors and using my personal networks. Five professors agreed to help me to share the survey in some of the classes (both graduate and undergraduate classes) that they were teaching during the semester the study was conducted. The participants were offered course extra credits or participation points for completing the survey. A total number of 173 participants attempted the study, and out of 173 participants, 154 participants have completed the experiment and survey questionnaires that followed the experiment (completion rate = 89%). Responses of nineteen participants were discarded as they have not completed the whole survey. Out of 154 participants who completed the survey, 89 of them were male, and 65 of them were female. The average age of the participants was 22.45, with a standard deviation of 3.12 (min age =18, max age = 36, range = 18).

All the 154 participants who completed the survey had some college degree (Undergraduate = 124, Master = 29, PhD = 1) and most of them had a major in Management Information System (MIS = 49, Accounting = 15, Geographic Information Science = 1, and Other business majors = 89). There were 78 participants in the control group (participants interacted with chats that had near-instant response time) and 76 in the treatment group (participants that interacted with chatbots that had dynamically delayed response).

### 6.6 Measures

All the measures used in this study have been previously validated and adopted by other researchers. The three items (feeling secure, comfortable and content) used to measure users' affective trust on chatbot have been taken from Komiak and Benbasat’s (2006) work on Recommendation Agents, and Sun’s (2010) work on Online Marketplaces. The four items used to measure (honest, caring, opportunistic, and predictable) users’ cognitive trust on chatbot have been adopted from Gefen et al.'s (2003b) work on Online Shopping, and Sun’s work on Online Marketplaces (2010). All the items from both cognitive and affective trust were measured on a 7-point Likert scale. Table 4 shows all the items used to measure cognitive and affective Trust in this paper with their descriptive statistics. Table 5 shows measurement items used, and composite reliability (CR), average variance extracted (AVE), and Cronbach's alpha (CA) for each construct.

Table 5 **The instrument and descriptive statistics**

| Construct | Indicator | Mean | SD |
|---|---|---|---|
| **Affective Trust in Chatbot (Treatment)** | I feel secure about chatting with Lisa/DentBot and relying on the information she provided. | 5.026 | 1.55 |
| | I feel comfortable chatting with Lisa/DentBot and relying on the information she provided. | 5.132 | 1.35 |
| | I feel content about chatting with Lisa/DentBot and relying on the information she provided. | 5.237 | 1.325 |
| **Affective Trust in Chatbot (Control)** | I feel secure about chatting with Lisa/DentBot and relying on the information she provided. | 4.808 | 1.698 |
| | I feel comfortable chatting with Lisa/DentBot and relying on the information she provided. | 5.038 | 1.615 |
| | I feel content about chatting with Lisa/DentBot and relying on the information she provided. | 4.808 | 1.714 |
| **Cognitive Trust in Chatbots (Treatment)** | I know Lisa/DentBot is honest. | 5.645 | 1.128 |
| | I know Lisa/Dentbot cares about me. | 3.934 | 1.7 |
| | I know Lisa/Dentbot is not opportunistic. | 4.697 | 1.47 |
| | I know Lisa/Dentbot is predictable. | 5.645 | 1.163 |

| | | | |
|---|---|---|---|
| **Cognitive Trust in Chatbots (Control)** | I know Lisa/DentBot is honest.<br>I know Lisa/Dentbot cares about me.<br>I know Lisa/Dentbot is not opportunistic.<br>I know Lisa/Dentbot is predictable. | 4.795<br>3.115<br>3.987<br>5.423 | 1.646<br>1.62<br>1.508<br>1.428 |

As shown in Table 5, I dropped one item (caring) from the Cognitive Trust construct as the factor loading of caring is below 0.6 (Gnewuch et al., 2018; Gefen and Straub, 2005). Factor loading below 0.6 indicates the relevance of caring in explaining Cognitive Trust is not significant. As it is suggested by Urbach and Ahlemann (2010), both of the constructs used in this study have significant composite reliability (suggested >= 0.8) and average variance extracted (suggested >=0.5). This means all the items used in this study to measure constructs have a robust internal consistency (CR and CA reliability) and low measurement error (AVE for accuracy).

Table 6 Constructs and measures used

| **Measures** | **Factor Loading** |
|---|---|
| **Cognitive Trust (CR = 0.797, CA = 0.67, AVE = 0.568)** | |
| I know Lisa/DentBot is honest | 0.824 |
| *I know Lisa/DentBot cares about me* | *dropped (0.535)* |
| I know Lisa/DentBot is not opportunistic | 0.739 |
| I know Lisa/DentBot is predictable | 0.692 |
| | |
| **Affective Trust (CR = 0.930, CA = 0.89, AVE = 0.815)** | |
| I feel secure about chatting with Lisa/DentBot and relying on the information she provided. | 0.899 |
| I feel comfortable chatting with Lisa/DentBot and relying on the information she provided. | 0.927 |
| I feel content about chatting with Lisa/DentBot and relying on the information she provided. | 0.882 |
| CR = Composite Reliability, CA = Cronbach's Alpha, AVE = Average Variance extracted | |

In addition to the data collected for the items to measure the constructs, I also collected data on participants' demographic information such as age, gender, and education level. At the end of the survey, I asked the participants an open-ended question ("Please, in a few sentences, describe your experiences interacting with Lisa/DentBot. Do you trust it? Why or why not?") to describe their interactions with chatbots.

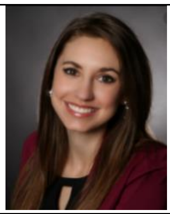
## 7 Result

As shown in Table 8 and Table 9, initially, I used a between-subjects design with two by two formats for this study. All the participants were randomly assigned to one of the four groups. As shown in Table 6, the first two groups interacted with a chatbot named Lisa that exhibited anthropomorphized design cues such as more human-like names, language style, and profile picture. The second two groups were assigned to interact with a non-anthropomorphized chatbot named BentBot. Furthermore, all the participants were either assigned to a control group (near-instant response) or treatment group (dynamically delayed response). There were two reasons behind such an experimental design. First, to study the effect of response latency on users' trust in chatbots, it was important to have two case experiments to prove such an effect in a clearer manner. If response latency is found to have an impact on users' trust in both anthropomorphized and non-anthropomorphized chatbots, then the effect of response latency on users' trust in chatbots is better shown. Second, since anthropomorphism was already found to produce a significant positive impact on users' attitudes, satisfaction, and emotional connection with the chatbot (Araujo, 2018), it would be interesting to study the variances that the change in response latency might create on users' trust in two different types of chatbots (anthropomorphized and non-anthropomorphized).

Table 7 Group Assignment

| | Dynamically delayed Response Time | Near-instant Response Time |
|---|---|---|
| Anthropomorphized Chatbot | Group 1 | Group 2 |
| Non-anthropomorphized Chatbot | Group 3 | Group 4 |

Table 8 Example of Anthropomorphism Design Cues

| | Name | Profile Picture | Language Style |
|---|---|---|---|
| Anthropomorphized Chatbot | Lisa |  | E.g., Hello there! Nice to meet you. E.g., Please choose one of the following options. |
| Non-anthropomorphized Chatbot | DentBot |  | E.g., Start. E.g., Choose the path you would like to take. |

To test the hypothesis, t-Test: Two-Sample Assuming Equal Variances was used. All the tests were performed two-sided to study the effect of response latency on users' trust in chatbots in both directions (positive/negative). As shown in Table 8, in the case of the anthropomorphized chatbot, Hypothesis one (H1) was confirmed with a significant p-value of 1.297E-08 (p-value <=0.05 is significant). Hence, dynamically delayed responses were found to increase users' cognitive trust in an anthropomorphized chatbot. However, Hypothesis two was not confirmed as H2 has an insignificant p-value of 0.336. Hence, dynamically delayed responses were found to have no significant effect on users' affective trust in an anthropomorphized chatbot.

Table 9 Descriptive results and test statistics for Anthropomorphized Chatbot

| Condition | n | Cognitive Trust*1 | | | Affective Trust*1 | | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | SE | Mean | SD | SE |
| Dynamically Delayed Response (Treatment) | 39 | 19.641 | 3.924 | 0.628 | 15.538 | 3.726 | 0.597 |
| Near-instant Response (Control) | 38 | 14.132 | 3.64 | 0.591 | 14.684 | 4.497 | 0.729 |
| Test statistic | | t (75) = 6.383, p = 1.297E-08 | | | t (75) = 0.909, p = 0.366 | | |
| Hypothesis | | H1 Confirmed | | | H2 Not confirmed | | |
| *1 Measured on a 7-point Likert scale \| SD = Standard deviation \| SE = Standard Error | | | | | | | |

As shown in Table 9, in the case of the non-anthropomorphized chatbot, Hypothesis one (H1) was also confirmed with a significant p-value of 0.010. Hence, dynamically delayed responses were found to increase users' cognitive trust in a non-anthropomorphized chatbot. However, hypothesis two was not confirmed as H2 has an insignificant p-value of 0.533. Hence, dynamically delayed responses were also found to have no significant effect on users' affective trust in a non-anthropomorphized chatbot.

Table 10 Descriptive results and test statistics for Non-anthropomorphized Chatbot

| Condition | n | Cognitive Trust*1 | | | Affective Trust*1 | | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | SE | Mean | SD | SE |
| Dynamically Delayed Response (Treatment) | 37 | 16.243 | 2.326 | 0.382 | 15.243 | 4.133 | 0.679 |
| Near-instant Response (Control) | 40 | 14.275 | 3.955 | 0.625 | 14.625 | 4.493 | 0.710 |
| Test statistic | | t (75) = 2.634, p = 0.010 | | | t (75) = 0.627, p =0.533 | | |
| Hypothesis | | H1 Confirmed | | | H2 Not confirmed | | |
| *1 Measured on a 7-point Likert scale \| SD = Standard deviation \| SE = Standard Error | | | | | | | |

As it is stated above, in both anthropomorphized and non-anthropomorphized chatbots, response latency was found to have a positive effect on users' cognitive trust. On the other hand, response latency was found to have no significant effect on users' affective trust in both anthropomorphized and non-anthropomorphized chatbots. Since H1 was confirmed and H2 was not confirmed in both types of chatbots, the role of anthropomorphism in determining the effect of response latency on users' trust in the chatbots was minimal. Therefore, as it is shown in Table 10, both treatment groups (G1 and G3) and control groups (G2 and G4) were merged to form a new control group and treatment group.

The benefit of merging these groups is the bigger sample size for both treatment and control groups. The bigger sample size is encouraged when items used to measure constructs are fewer, and the effect of the manipulation factor is subtle (Iacobucci, 2010). In addition, the bigger sample size was found to produce more reliable results (Kaplan et al., 2014).

Table 11 Merging of Groups

| | Dynamically delayed Response Time | Near-instant Response Time |
|---|---|---|
| Anthropomorphized Chatbot | Group 1 | Group 2 |
| Non-anthropomorphized Chatbot | Group 3 | Group 4 |

| Dynamically delayed Response Time | Near-instant Response Time |
|---|---|
| Group 1 | Group 2 |

As shown in Table 11, H1 was again confirmed in the combined groups. Dynamically delayed response has positive effect on users' cognitive trust ($t(152) = 3.405$, $p = 0.001$). However, H2 was not confirmed. Dynamically delayed response has no significant effect on users' affective trust (t (152) = 1.095, p = 0.275). This finding is consistent with the results from the previous analysis in both cases of anthropomorphized and non-anthropomorphized chatbots.

Table 12 Descriptive results and test statistics for Chatbot (both anthropomorphized and non-anthropomorphized)

| Condition | n | Cognitive Trust*1 | | | Affective Trust*1 | | |
|---|---|---|---|---|---|---|---|
| | | **Mean** | **SD** | **SE** | **Mean** | **SD** | **SE** |
| **Dynamically Delayed Response (Treatment)** | 76 | 15.987 | 2.585 | 0.296 | 15.395 | 3.906 | 0.448 |
| **Near-instant Response (Control)** | 78 | 14.205 | 3.781 | 0.428 | 14.654 | 4.466 | 0.506 |
| **Test statistic** | | t (152) = 3.405, p = 0.001 | | | t (152) = 1.095, p = 0.275 | | |
| **Hypothesis** | | H1 Confirmed | | | H2 Not confirmed | | |
| *1 Measured on a 7-point Likert scale \| SD = Standard deviation \| SE = Standard Error | | | | | | | |

To study the sentiments of users after interacting with chatbots that either deployed near-instant or dynamically delayed response, a sentiment analysis (tidytext package) was carried out using R. Ten general sentiments ("anger", "anticipation", "disgust", "fear", "joy", "sadness", "surprise", "trust", "negative", and "positive") were extracted from users' response to the post-experiment open-ended question ("Please, in a few sentences, describe your experiences interacting with Lisa/DentBot. Do you trust it? Why or why not?"). As shown in Figure 7 and Figure 8, users interacted with a chatbot that has dynamically delayed responses displayed higher sentiment score across all the positive sentiment categories ("joy", "positive", "trust", and "anticipation") and lower sentiment score across all the negative sentiment categories ("anger", "disgust", "fear", "negative", and "sadness"). In contrast, users interacted with chatbot that has near-instant response has lower sentiment score across all the positive sentiment categories ("joy", "positive", "trust", and "anticipation") and higher sentiment score across all the negative sentiment categories ("anger", "disgust", "fear", "negative", and "sadness").

Figure 7 Sentiment Score for Dynamically Delayed Response in Percentage

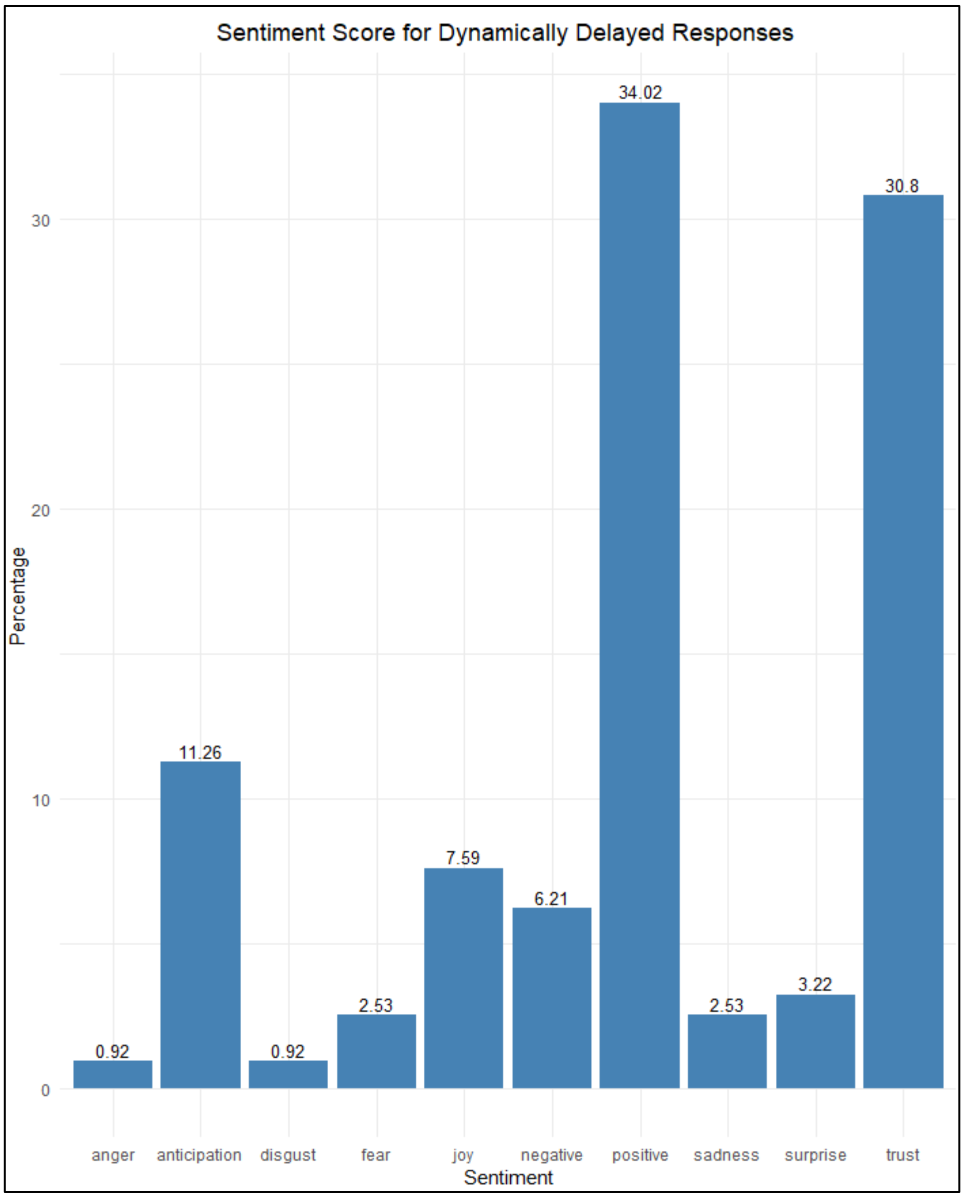
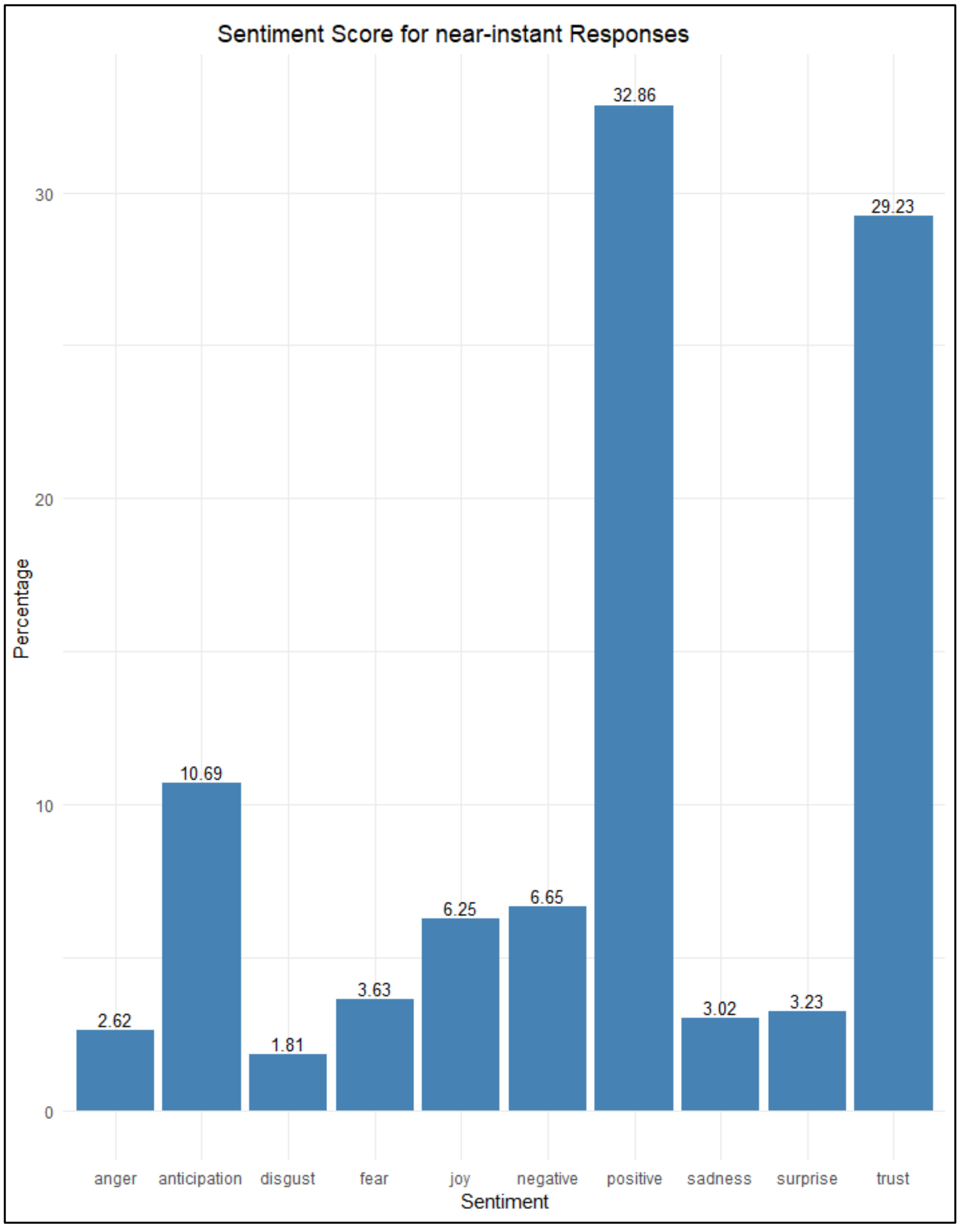


Figure 8 Sentiment Score for Near-instant Response in Percentage



One sentiment that is a particular interest in this study is the sentiment of "Trust". As shown in Figure 7 and Figure 8, users showed slightly higher trust in a chatbot that has a dynamically delayed response (30.8%) as compared to a chatbot that has a near-instant response (29.23%).

## 8 Discussion

IS researchers have been studying various aspects of designing graphical user interfaces to improve users' interactions with websites and apps for decades. However, little attention is being given on the design of the new conversational agent user interfaces. This study attempted to examine a less study conversational user interface design cue, response latency, and its effect on users' trust. The empirical study of users’ interaction with chatbots conducted in this study suggests that dynamically delaying chatbot response has a positive effect on users' cognitive trust in the chatbots. However, dynamically delaying chatbot response has no significant effect on users’ affective trust in chatbots.

There are a few possible reasons why H2 was not confirmed. First, affective trust is based on users’ emotional connection towards chatbots. A stronger emotional connection was found to be positively related to the time spent together in an interpersonal relationship (Kingston & Nock, 1987). However, in the case of this study, the task assigned to participants that required their interaction with the chatbots was possible to be performed in about 1 minute, if proper instruction was followed. One-minute interaction with the chatbot might not have been enough time for the users to establish significant emotional connections with the chatbot. Second, emotional disclosure to a chatbot was found to have a stronger positive emotional outcome (emotional connection) than a factual disclosure in humans (Ho et al., 2018). The participants of this study were not required to have any emotional disclosure with the chatbot to complete the experimental task. Hence, it is possible that no significant emotional connection was built due to a lack of emotional disclosure. In addition, the chatbots implemented for this study were informational chatbots that do not engage in emotional exchanges as opposed to a social chatbot. The effect of response latency on users' trust might have been different if a social chatbot was used for this study that is aimed to build rapport and emotional connections with the users.

### 8.1 Limitations and Future Research

This study has several limitations. First, all the participants of this study were affiliated with an educational institution, and the real population might not share similar attitudes, experiences, and backgrounds as the sample population of the study. The sample population not representing the actual population could reduce the accuracy of the results suggested by this work. Given the fact that college students are savvier on emerging technologies such as chatbot, their previous familiarity with the chatbots might have skewed the findings of this study. To overcome this limitation, future studies can collect data from a more diverse population that better represent the actual population of chatbot users. Second, even though the sample size of 152 is acceptable, it is relatively small. In addition, as recommended by an IS expert, the final control and treatment groups were merged from groups that were exposed to more conditions than response latency (anthropomorphized and non-anthropomorphized chatbots) in order to obtain a larger sample size. Elements such as a more human-like language style, profile picture, and name could have an effect on users' trust that this study assumed insignificant. Futures studies can validate the results and improve the experimental design by sampling a larger sample size and limiting the manipulated conditions to just response latency. Third, the chatbots implemented in this study were informational chatbots. There are various types of chatbots that are different based on their intended functions. The findings of this study might not apply to other types of chatbots. For instance, a social chatbot serves the function of building a social relationship and engage in emotional exchanges. The users of such chatbots might experience a different level of affective trust than in an informational chatbot. Future studies can examine the rule of trust (cognitive and affective trust) in different types of chatbots to better understand the effect of response latency on users' trust. Fourth, the instrument used to measure users Cognitive Trust were limited. Cognitive Trust has a dimension of competence, reliability, benevolence, and predictability. However, in this study, the item for measuring benevolence was dropped due to the low initial factor loading (0.53). This might mean the instrument has limitations that future studies need to investigate.

**8.2 Contribution and Research Implications**

The first contribution of this research is a conceptual contribution. As stated earlier, the effect of response latency on users' trust in chatbots is not a very often studied topic. Given the prevalence use of chatbots in assisting various organizational functions and lack of users' trust in chatbot might impede the benefits that organizations can harvest from chatbot implementation. As the findings of this study suggest, organizations need to take response latency into considerations while implementing a chatbot for the organizations. Our knowledge of "fast system response time is better" might not necessarily apply to the chatbot. It seems there is more to a chatbot than a regular system (a terminal system used to order food). Chatbots mimic human conversation. Hence, we apply different response time standards to chatbots (response time standard) than to a regular system. The second contribution of this research is also conceptual. The effect of response latency on users' cognitive trust in chatbots seems to apply to both types of chatbots (anthropomorphized and non-anthropomorphized). Regardless of the levels of anthropomorphism used in chatbots, dynamically delaying response could increase users' trust in chatbots.

## 9 Conclusion

The findings of this study suggest that dynamically delaying chatbots' responses can increase users' cognitive trust but do not significantly increase users' affective trust. Since the chatbots implemented in this study is not a social bot and it did not require participants to engage in emotional disclosures, the effect of dynamically delaying chatbots response on users' affective trust might not apply to a social chatbot (or other types of chatbots). In addition, the effect of response latency on users' cognitive trust in chatbots seems to apply to both types of chatbots (anthropomorphized and non-anthropomorphized chatbots).

## References


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*.

Anyoha, R. (2017). The history of artificial intelligence. *Science in the News*, *28*.

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.

Artificial Intelligence. (n.d.). In Wikipedia. Retrieved March 28, 2020, from https://en.wikipedia.org/wiki/Artificial_intelligence

Celce-Murcia, M. (2008). Rethinking the role of communicative competence in language teaching. In *Intercultural language use and language learning* (pp. 41-57). Springer, Dordrecht.

DARPA Robotics Challenge, 2015. *Robots from Republic of Korea and United States take home $3.5 million in prizes.* Retrieved from https://www.citefast.com/styleguide.php?style=APA&sec=Government

Domonoske, Camila. (2016). Going, Going, Gone: Master Go Player Loses Best-Of-5 Match With A.I.. NPR News. Retrieved from https://www.npr.org/sections/thetwo-way/2016/03/12/470193083/going-going-gone-master-go-player-loses-best-of-5-match-with-a-i

Clifford, Catherine. (2017, November 8). Hundreds of A.I. experts echo Elon Musk, Stephen Hawking in call for a ban on killer robots. CNBC News. Retrieved from https://www.cnbc.com/2017/11/08/ai-experts-join-elon-musk-stephen-hawking-call-for-killer-robot-ban.html

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., & Nunamaker Jr, J. F. (2013). Detecting deceptive chat-based communication using typing behavior and message cues. *ACM Transactions on Management Information Systems (TMIS)*, *4*(2), 1-21.

Eliza, the Rogerian Therapist. (1999). California State University Fullerton. Retrieved from http://psych.fullerton.edu/mbirnbaum/psych101/Eliza.htm

Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. *interactions*, *24*(4), 38-42.

Gefen, D., E. Karahanna, and D. W. Straub (2003b) "Trust and TAM in Online Shopping: An Integrated Model," MIS Quarterly (27) 1, pp. 51-90.

Gefen, D. and Straub, D.W. (2005), "A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example", Communications of the Association for Information Systems, Vol. 16, pp. 91–109

Goertzel, Ben., Pennachine, Cassio. (2007). Artificial General Intelligence. Springer Science.

Gnewuch, U., Morana, S., Adam, M. T. P., and Maedche, A. (2018). "Faster Is Not Always Better: Understanding the Effect of Dynamic Response Delays in Human-Chatbot Interaction," in Proceedings of the 26th European Conference on Information Systems (ECIS), Portsmouth, United Kingdom, June 23-28.

Guyer, Joshua & Fabrigar, Leandre. (2017). Vocal Confidence and Persuasion: Speech Rate Affects Amount of Processing Under Moderate Elaboration.

History Computer. (n.d.). Logic Theorist. Retrieved March 28, 2020, Retrieved from https://history-computer.com/ModernComputer/Software/LogicTheorist.html

Ho, S.M., Hancock, J.T., Booth, C. and Liu, X. (2016), "Computer-Mediated Deception: Strategies Revealed by Language-Action Cues in Spontaneous Communication", *Journal of Management Information Systems*, Vol. 33 No. 2, pp. 393–420.

Hoxmeier, John A. and DiCesare, Chris, "System Response Time and User Satisfaction: An Experimental Study of Browser-based Applications" (2000). AMCIS 2000 Proceedings. 347.

IBM. (1954, January 8). 701 Translator. IBM Press Release. Retrieved from https://www.ibm.com/ibm/history/exhibits/701/701_translator.html

Jurafsky, Dan. (n.d.). CS 124/LINGUIST 180 From Languages to Information: Conversational Agents. Stanford University. Retrieved from https://web.stanford.edu/class/cs124/lec/chatbot.pdf

Kaplan, R. M., Chambers, D. A., & Glasgow, R. E. (2014). Big data and large sample size: a cautionary note on the potential for bias. *Clinical and translational science*, *7*(4), 342-346.

Kingston, P. W., & Nock, S. L. (1987). Time together among dual-earner couples. *American Sociological Review*, 391-400.

Komiak, S. Y. X. and I. Benbasat (2006) "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," MIS Quarterly (30) 4.

Klopfenstein, L.C., Delpriori, S., Malatini, S. and Bogliolo, A. (2017), “The Rise of Bots: A Survey of Conversational Interfaces, Patterns, and Paradigms”, Proceedings of the 2017 Conference on
Designing Interactive Systems, pp. 555–565.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of consumer psychology*, *20*(1), 90-98.

Lele, Chitra. (2019, September 19). Artificial Intelligence meets Augmented Reality: Redefining Regular Reality. BPB Publications.

Loucks, Jeff., Jarvis, David., Hupfer, Susanne., Murphy, Timothy. (2019). Future in the balance? How countries are pursuing an AI advantage. Deloitte Insight. Retrieved from https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/ai-investment-by-country.html

Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior, 56*, 225-237.

Marian Friestad, Peter Wright, The Persuasion Knowledge Model: How People Cope with Persuasion Attempts, *Journal of Consumer Research*, Volume 21, Issue 1, June 1994, Pages 1–31, https://doi.org/10.1086/209380

Mark A. de Turck, Gerald R. Miller, Deception and Arousal: Isolating the Behavioral Correlates of Deception, *Human Communication Research*, Volume 12, Issue 2, December 1985, Pages 181–201, https://doi.org/10.1111/j.1468-2958.1985.tb00072.x

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems*, *11*(3-4), 297-323.

Menzies, T. (2003). Guest Editor? s Introduction: 21st Century AI--Proud, Not Smug. *IEEE Intelligent Systems*, (3), 18-24.

Miller, R. B. (1968, December). Response time in man-computer conversational transactions. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I* (pp. 267-277).

Moon, Y. (1999), “The effects of physical distance and response latency on persuasion in computermediated communication and human-computer communication.”, *Journal of Experimental*
*Psychology: Applied*, Vol. 5 No. 4, pp. 379–392

Moor, J. (2006). The Dartmouth College artificial intelligence conference: The next fifty years. *Ai Magazine*, *27*(4), 87-87.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Poole, David., Mackworth, Alan., Goebel, Randy. (1998). Computational Intelligence and Knowledge: A Logical Approach. Oxford University Press.

Rao, Anand. S., Verweij, Gerard. (2017). Sizing the prize: What's the real value of AI for your business and how can you capitalise?. PWC. Retrieved from https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

Salecha, Manisha. (2016, May 10). STORY OF ELIZA, THE FIRST CHATBOT DEVELOPED IN 1966. AIM. Retrieved from https://analyticsindiamag.com/story-eliza-first-chatbot-developed-1966/

Sarikaya, R. (2017), "The Technology Behind Personal Digital Assistants: An Overview of the System Architecture and Key Components", *IEEE Signal Processing Magazine*, Vol. 34 No. 1, pp. 67–81.

Schuetzler, R.M., Grimes, G.M., Giboney, J.S. and Buckman, J. (2014), "Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment", *Proceedings of the 35th International Conference on Information Systems (ICIS '14)*, Auckland, NZ, pp. 1–16.

Shechtman, N. and Horowitz, L.M. (2003), "Media Inequality in Conversation: How People Behave Differently When Interacting with Computers and People", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*, ACM, Ft. Lauderdale, FL, USA, pp. 281–288.

Short, J., Williams, E. and Christie, B. (1976), *The Social Psychology of Telecommunications*, London, UK.

Spencer, Geoff. (2018, November 1). Much more than a chatbot: China's Xiaoice mixes AI with emotions and wins over millions of fans. Microsoft. Retrieved from https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/

Skowron, M., Pirker, H., Rank, S., Paltoglou, G., Jung Hyun, A. and Gobron, S. (2011), "No peanuts! Affective Cues for the Virtual Bartender", *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*, pp. 117–122.

Smith, C., McGuire, B., Huang, T., & Yang, G. (2006). The history of artificial intelligence, University of Washington. *Retrieved June*, *8*, 2017.

Smith, Albert. (2020, March 16). Understanding Architecture Models of Chatbot and Response Generation Mechanisms. AI Zone. Retrieved from https://dzone.com/articles/understanding-architecture-models-of-chatbot-and-r

Tardi, Carla. (2019, September 5). Moore's Law. Investopedia. Retrieved from https://www.investopedia.com/terms/m/mooreslaw.asp

The royal Institution. (2017, May 17). Artificial Intelligence, the History and Future - with Chris Bishop. Retrieved from https://www.youtube.com/watch?v=8FHBh_OmdsM

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23-65). Springer, Dordrecht.

Urbach, N. and Ahlemann, F. (2010), "Structural equation modeling in information systems research using partial least squares", JITTA: Journal of Information Technology Theory and Application, Vol. 11 No. 2, p. 5.

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: response latency-based measures. *Memory detection: Theory and application of the Concealed Information Test*, 46-62.

Wallace, R. S. (2009). The anatomy of ALICE. In *Parsing the Turing Test* (pp. 181-210). Springer, Dordrecht.

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in human neuroscience*, *9*, 131. doi:10.3389/fnhum.2015.00131

Zemčík, Tomáš. (2019). A Brief History of Chatbots. DEStech Transactions on Computer Science and Engineering. 10.12783/dtcse/aicae2019/31439.

Zhou, L. (2005) "An empirical investigation of deception behavior in instant messaging," in *IEEE Transactions on Professional Communication*, vol. 48, no. 2, pp. 147-160, doi: 10.1109/TPC.2005.849652

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2018). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, (Just Accepted), 1-62.