## Title
Friendly-Bot: The Impact of Chatbot Appearance and Relationship Style on User Trust



**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

**Authors**
Bae, Seoyeon
Lee, Yoon Kyung
Hahn, Sowon

**Publication Date**
2023

Peer reviewed

# Friendly-Bot: The Impact of Chatbot Appearance and Relationship Style on User Trust

**Seoyeon Bae (cathy4339@snu.ac.kr)**
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

**Yoon Kyung Lee (yoonlee78@snu.ac.kr)**
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

**Sowon Hahn (swhahn@snu.ac.kr)**
Department of Psychology, Seoul National University
1, Gwanak-ro, Gwanak-gu, Seoul, 08826 Republic of Korea

## Abstract

With AIs becoming prevalent in our daily lives, it is still controversial whether they are a reliable conversational assistant. We, therefore, developed a chatbot, a *Friendly-Bot*, where users can share interpersonal experiences (e.g., friendship and romantic relationships). We then manipulated chatbot appearances into two types: robot and human-looking. We found that chatbot appearance predicted users' trust. Participants preferred the robot-looking chatbot over the human-looking one. Participants also showed higher trust in robot-looking chatbots when conversing about positive interpersonal relationships. Participants showed a different pattern of trust depending on the appearance condition: positive experience led to higher trust in robot-looking condition, whereas the opposite was observed in human-looking condition. Our findings show how chatbot appearance influences rapport-building in AI-assisted interactions (e.g., counseling).

**Keywords:** conversational AI; trust; chatbot appearance; relationship style; human-robot interaction

## Introduction

*Conversational Agents* (CAs, or chatbots) are applied to many domains, such as medical and business. The ratio of enterprises which utilized AI services rose from 10% to 37% between 2015 and 2019, a 270% increase in 4 years (Costello, 2019). Chatbots are used for making restaurant reservations or receiving financial consultation services (Kim et al., 2020; Okuda & Shoda, 2018). Others prefer to use chatbots for personal entertainment (Cho et al., 2022) or counseling (Fitzpatrick et al., 2017). With more large text data and advancements in natural language processing, chatbots like "*Iruda*" can make more natural and longer conversations with users (Cho et al., 2022).

Despite these widespread applications, some people still have difficulty relying on AI. According to the survey, 42% of nearly 2,000 American consumers responded that they did not trust AI (Dujmovic, 2017). Accordingly, factors like chatbots' human likeness and users' characteristics should be considered to study trust in AI.

## Related Works

### Chatbots for Open-domain Conversation

Conversational agents are systems that can talk with people based on natural language understanding (NLU) algorithms (Chaves & Gerosa, 2021). Chatbots are divided into task-oriented and open-domain ones. Task-oriented CAs have specific tasks or goals to be achieved, with a format of question and answer, such as banking and shopping (Bordes et al., 2016).

Meanwhile, open-domain CAs create free chats between bots and users like ordinary human-human conversations. Therefore, open-domain chats must maintain users' interest during conversations. Large Language models such as GPT-3 and developed deep learning models recently improved the performance of open-domain chats performance, enabling more natural and human-like utterances (Brown et al., 2020).

However, conversational agents still have limitations. Repetitive and monotonous dialogues happen as the length of conversations increases (Roller et al., 2021). Ethical issues are another problem. The Korean conversational agent *Iruda* had temporarily terminated its service due to ethical concerns (Choi & Hong, 2021). These limitations can impede users' trust in chatbots. Hence, it is essential to consider ways to enhance user trust in conversational agents.

### Trust in Conversational AI

Trust, belief in another party's action, strength, or ability, is imperative in building relationships. It consists of two aspects - cognitive and affective trust (Johnson & Grayson, 2005). While cognitive trust is obtained by providing reliable information, affective trust is achieved through maintaining relationships and feeling comfort.

People attribute social rules to machines that act like humans (*Computers are Social Actors*; Reeves & Nass, 1996; Nass & Moon, 2000). As trust plays a significant role in interpersonal relationships among humans, so does it in human-robot (or AI) interaction.



In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*. 

Characteristics of AIs, users' personal traits, and cultural/environmental context can influence peoples' willingness to trust AI (Lee & See, 2004). For example, increased user engagement through the interface led to increased trust in fact-checking AI (Nguyen et al., 2018).

Anthropomorphism, one of the important attributes regarding AIs, has actively been studied in the Human-AI Interaction (HAI) field (Waytz et al., 2014; Foehr & Germelmann, 2020). Users tend to endow non-human things with human-like attributes such as emotions or personalities (Epley et al., 2007). For instance, participants were more willing to trust autonomous vehicles when they had more anthropomorphized attributes such as name, gender, and voice (Waytz et al., 2014). Users showed higher trust and satisfaction with smart speakers (e.g., Alexa) when they had human-like personalities (Foehr & Germelmann, 2020).

## The Role of User Traits in Human-AI Interaction

Studies have also focused on the personal characteristics of AI users, such as personality traits or gender. Neuroticism, one of the personality traits from the Big 5, negatively predicted trust in AI algorithms during the card game (Sharan & Romano, 2020). Users' gender also played a role, indicating that females showed higher trust in autonomous security robots than males (Gallimore et al., 2019).

*Psychological attachment* is crucial for understanding human-AI interactions, as users tend to think of chatbots as friends as time passes (Skjuve et al., 2021). However, *psychological attachment style* has rarely been studied in the HAI field.

Attachment stems from early interactions between caregivers and babies (Bowlby, 1982), substantially shaping life-long interpersonal styles. A study observed different patterns in children's behaviors when being left alone and reuniting with their mothers (Ainsworth et al., 1978). These patterns were categorized into three main types: *Security*, *Anxiety*, and *Avoidance*. Gillath et al. (2021) discovered that one's particular attachment style resulted in different patterns of trust in AI: While secure attachment style raised participants' trust, anxiety attachment decreased it. The study, however, was based on hypothetical scenarios of using AI, without direct interaction (Gillath et al., 2021). In addition, it has been found that attachment style did not last for one's whole life, depending on significant life events or nurturing environments (Weinfield et al., 2000; Weinfield et al., 2004).

Therefore, we designed our study to fill in the research gap. First, we developed a counseling chatbot, *Friendly-Bot*, so that participants can interact with it real-time. Furthermore, we newly defined the "*relationship style*" in designing conversations with the chatbot based on attachment theory from psychology.

## The Present Study

We instructed participants to discuss with the chatbot after recalling certain types of interpersonal experiences. The chatbot was either robot or human-looking. We then measured participants' trust in the chatbot and collected how participants thought of its personality.

Hypotheses of the study are as follows:

**H1.** Participants who converse with the human-looking chatbot will show higher trust than those who interact with the robot-looking chatbot.

**H2.** Participants who make conversations about secure relationships will report higher trust than those who discuss anxious and avoidant relationships.

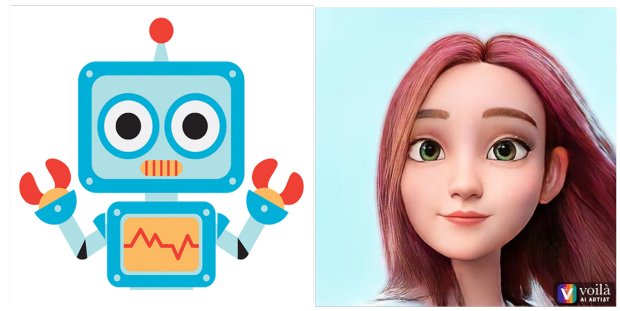


Figure 1: Profile photo of each chatbot appearance. Robot (Left), Digital human (Right)

# Method

## Participants

A total of 26 adults ($M_{age}$ = 22.8, $SD_{age}$ = 2.47) participated in this study. Two of them were excluded from the analysis due to unreliable responses. 18 of 24 were female. All participants' native language was Korean.

## Experiment Design

We conducted a 5*2 between-subject design. The first predictor, *experience style*, was divided into five types - Security, Anxiety, Avoidance, Success, and Failure. Security, Anxiety, and Avoidance indicate an individual's relationship style. Participants in the Security group read guidelines containing examples of secure relationship style and how to use the chatbot and were instructed to recall past secure experiences.

Participants in the Anxiety group recalled one's actual experiences of anxiety style. The Avoidance group participants followed the same process, with experiences based on avoidance style. Participants of the Success and Failure groups focused on personal achievement, not related to interpersonal relationships. After reading the guidelines, all participants discussed the recalled experiences with the chatbot for about 5 minutes.

The second predictor was the bot's *appearance*, composed of a robot and a digital human ($N_{robot}$ = 11). These were manipulated by a profile photo of the chatbot and its name. Robot appearance included a robot profile photo with the chat-name, *Friendly-Bot*. A profile photo for digital human was presented with the nickname *Minzy*, a typical Korean name. The photo of *Minzy* was created with Voilà AI Artist, an app that makes human photos look like cartoon artwork.

Participants were randomly assigned to each of the ten groups.

## Apparatus

We used a chatbot builder program RASA to create the chatbot. RASA is an open-source hybrid chatbot builder based on machine learning algorithms and supports both task-oriented and open-domain chatbots.

To build a basic chatbot, three main components must be defined: *intents*, *responses*, and *stories*.

*Intents* are user inputs. These are examples that users are expected to enter when chatting with the bot. For instance, if users have to say with whom they hung out during the last vacation, they may enter *friends*, *my friend Tom*, *my little sister*.

Intents can be categorized based on many properties (e.g., address, emotions, location) with a keyword following the expected user input: “*(keyword)*.” These categories are called *entities*. Entities help algorithms recognize which categories users write down and at which stage of the discourse the users are.

*Responses* are the bot’s answers. For example, if users share their recent breakup with their romantic partners, the bot may answer, *I’m sorry to hear that*, or *I got your back*. All these examples can be trained to the chatbot and enable chatbots to respond appropriately and timely. We prepared all responses for every step bots are planned to go through. Besides general forms of responses (letters), the bot also sent images in jpg or gif format.

*Stories* indicate the dialogue flows of the chatbot. One block of the story should contain every conversational step, from the greeting to the termination. We entered as many stories as we wanted, which helped chatbots handle various and unexpected patterns of conversations.

We connected the chatbot to the SLACK platform[1] so that participants engage with the bot through a stable message interface, as realistic as possible. When users initiated the conversation, the bot responded and led throughout the conversation. Participants followed pre-instructed steps with limited answer options (via buttons) to control the conversation flows.

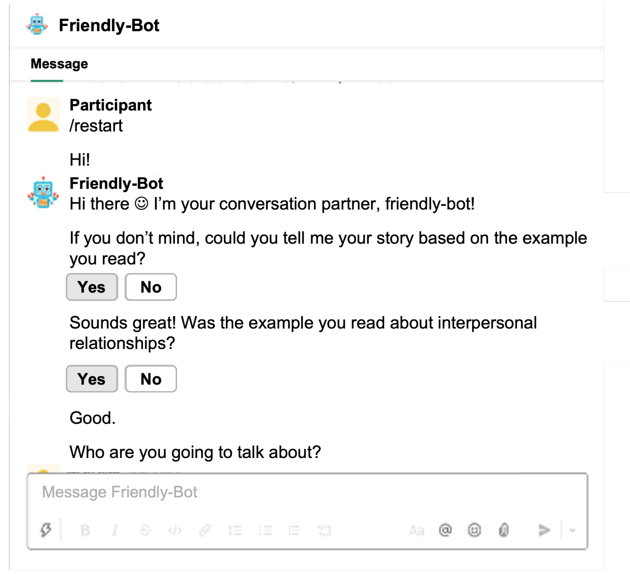


Figure 2: Interface of the *Friendly-Bot*.

## Measurements

**Agreeableness and Neuroticism** We used the Ten Item Personality Inventory to measure participants’ personality traits. The questionnaire is a 10-item measure assessing the Big 5 personality traits using a 7-point Likert scale (Gosling et al., 2003). In this study, items for only two traits - agreeableness and neuroticism - were applied for analysis (e.g., “I see myself as anxious, easily upset,” “I see myself as sympathetic, warm”).

**Self-Esteem** We used Rosenberg Self-Esteem Scale to measure participants’ self-esteem levels. The questionnaire contains ten items using a 4-point Likert scale (Rosenberg, 1979). (e.g., “I feel that I have a number of good qualities”)

**Familiarity with AI** Whether participants have experienced using any artificial intelligence and to what extent they are familiar with AI in their daily lives were asked. These items were measured with ‘yes or no’ and a 5-point Likert scale.

**Trust in Chatbot** To measure participants’ trust in the *Friendly-Bot*, we referred to the Trust in AI scale applied by Gillath et al. (2021). As the original one was developed for various kinds of AI, we revised the scale to focus on trust in the chatbot. The scale includes items such as “Did you feel emotionally safe during the conversation?” using a 7-point Likert scale.

**Personality of Chatbot** Participants were asked to express the personality (characters) of the *Friendly-Bot* at the end of the survey. They were required to answer for the item using adjectives such as *friendly*, *irritable*.

[1] https://slack.com/

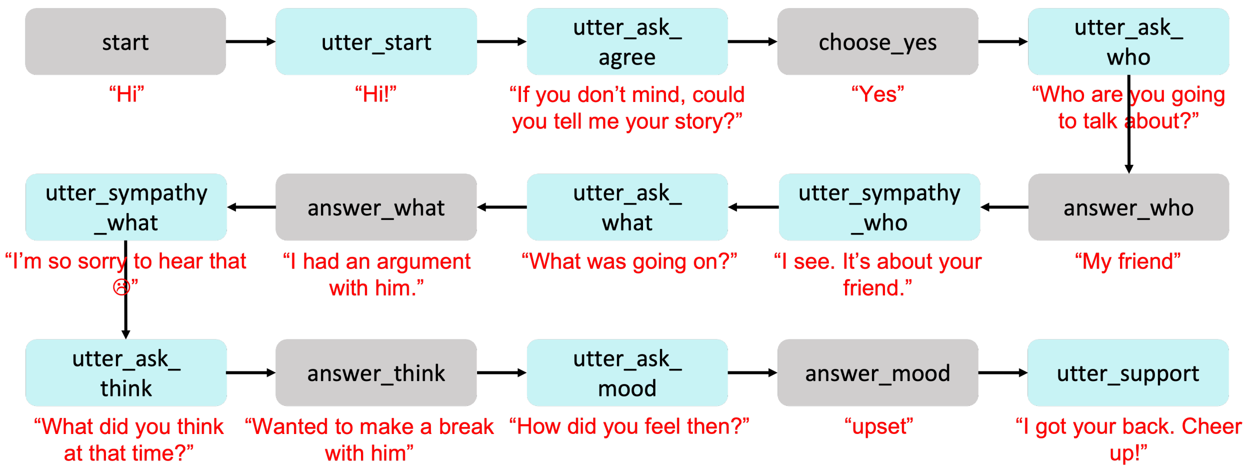


Figure 3: Dialogue flow of the chatbot.

## Procedure

After providing informed consent, participants took the first survey to measure their agreeableness, neuroticism, and self-esteem. Then, they read the chatbot guidelines that differed by experience style groups. After recalling an experience related to a particular experience style, participants made conversations with the *Friendly-Bot* about the experience. After that, they completed the post-survey measuring their familiarity with AI, trust in the *Friendly-Bot*, the bot's personality, and demographic information.

## Result

When the appearance style was put in regression as the only predictor controlling other factors (experience styles, personality traits, and familiarity with AI), it significantly influenced trust in the *Friendly-Bot* ($F(1, 22) = 4.11$, $\beta = .40$, $SE = 1.96$, $p = .055$). Participants' experience styles did not significantly affect their trust scores.

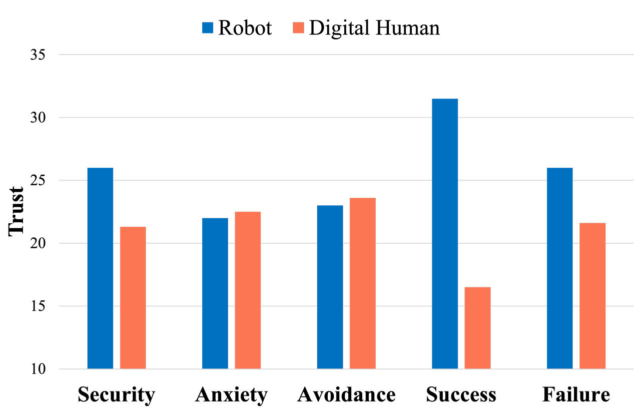


Figure 4: Trust scores of each group.

Furthermore, the mean trust scores of participants who talked with the robot-appearance chatbot ($M_{robot} = 25.36$, $SD_{robot} = 5.71$) were generally higher than those who talked with the digital human-appearance chatbot ($M_{human} = 21.38$, $SD_{human} = 3.86$), which was marginally significant ($t(22) = 2.03$, $p = .055$).

Participants who recalled secure relationship experience ($M_{sec} = 26$, $SD_{sec} = 5.66$) and success ($M_{suc} = 31.5$, $SD_{suc} = 4.95$) reported higher trust in the *Friendly-Bot* than those who recollected anxious ($M_{anx} = 22$, $SD_{anx} = 6.08$), avoidant relationship experiences ($M_{avo} = 23$, $SD_{avo} = 8.49$), and failure ($M_{fail} = 26$, $SD_{fail} = 1.41$); This pattern was only observed in the robot-looking condition. In human-looking condition, on the contrary, anxious ($M_{anx} = 22.5$, $SD_{anx} = .70$), avoidant relationships ($M_{avo} = 23.6$, $SD_{avo} = 5.69$), and failure ($M_{fail} = 21.6$, $SD_{fail} = 2.08$) induced higher trust than secure relationship ($M_{sec} = 21.3$, $SD_{sec} = 4.16$) and success ($M_{suc} = 16.5$, $SD_{suc} = 2.12$).

Participants expressed the personality of the *Friendly-Bot*, usually with adjectives such as "friendly" "kind," "warm," and "empathetic," regardless of experience styles they remembered.

## Discussion

In this study, we investigated if the participants' relationship styles and the bot's appearances affected their evaluation of the chatbot. We, therefore, developed the counseling bot, the *Friendly-Bot*, and constructed the conversational flow for relationship styles.

We found that the chatbot's appearance (robot vs. digital human) significantly predicted users' willingness to trust the bot. We observed a different pattern from what we initially hypothesized; those who made conversations with the robot-looking chatbot showed higher trust than those in the human-looking chatbot group. This pattern may be because people felt uncomfortable with a human-like appearance of the chatbot, also known as the *uncanny valley* (Mori et al., 2012). Additionally, the mismatch between the bot's name *Minzy* and its identity may have caused participants to feel awkward throughout the conversations.

Participants who recollected positive experiences (i.e., secure relationships and success) reported higher trust scores than those who did negative ones (i.e., anxious and avoidant relationships, and failure), which was observed only in the robot-looking condition. This pattern partially supports our second hypothesis. Participants who talked about their negative experiences to *Minzy* trusted her more than those in positive conditions. It seems that the human-like appearance of the chatbot created a safe and comfortable environment like counseling, which may have led to more self-disclosing behaviors (Go & Sundar, 2019).

Finally, participants described the bot's personality as social (e.g., *kind*, *friendly*, *warm-hearted*, and *empathetic*) regardless of relationship styles they recalled or the bot's appearance. This is because the conversation topic was mainly about personal experiences. This can be backed up by prior research that social chatbots facilitate more self-disclosing and rapport building between counselors and clients (Lee et al., 2020; Ta et al., 2020; Skjuve et al., 2021).

The findings of the study are as follows. First, it suggests that the visual attributes of chatbots influence users' trust. Second, relationship style is a reliable predictor for trust in CAs. Lastly, it suggests that social personalities are crucial in positive interactions between humans and chatbots.

There are some limitations in this study. First, the relationship styles participants remembered did not significantly affect their trust level. It seems that the way of priming relationship style did not persist, and another method should be considered. Second, a future study should be investigated with a larger sample. In addition, the female photo of the human-looking condition might cause biased results in the chatbot's personality traits. Future studies should include different genders. Finally, privacy and social bias should be regarded when designing conversations with social chatbots for ethical concerns (Choi & Hong, 2021).

## References


Ainsworth, M. D. S., Blehar, M., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Psychology Press.

Bordes, A., Boureau, Y. L., & Weston, J. (2016). Learning end-to-end goal-oriented dialog. *arXiv preprint arXiv:1605.07683*.

Bowlby, J. (1982). Attachment and loss: retrospect and prospect. *American Journal of Orthopsychiatry*, *52*(4), 664-678.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729-758.

Cho, W. I., Kim, S., Choi, E., & Jeong, Y. (2022). Assessing How Users Display Self-Disclosure and Authenticity in Conversation with Human-Like Agents: A Case Study of Luda Lee. *Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022*, 145-152.

Choi, S. S., & Hong, A. R. (2021). Identifying issue changes of AI chatbot 'Iruda' case and its implications. *Electronics and Telecommunications Trends*, *36*(2), 93-101.

Costello, K. (2019, January 21). Gartner survey shows 37 percent of organizations have implemented AI in some form. *Gartner*. https://www.gartner.com/en/newsroom/press-releases/2019-01-21-gartner-survey-shows-37-percent-of-organizations-have

Dujmovic, J. (2017, March 30). Opinion: What's holding back artificial intelligence? Americans don't trust it. *MarketWatch*. https://www.marketwatch.com/story/whats-holding-back-artificial-intelligence-americans-dont-trust-it-2017-03-30

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Mental Health*, *4*(2), Article e7785.

Foehr, J., & Germelmann, C. C. (2020). Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. *Journal of the Association for Consumer Research*, *5*(2), 181-205.

Gallimore, D., Lyons, J. B., Vo, T., Mahoney, S., & Wynne, K. T. (2019). Trusting robocop: Gender-based effects on trust of an autonomous robot. *Frontiers in Psychology*, *10*, Article 482.

Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., & Spaulding, R. (2021). Attachment and trust in artificial intelligence. *Computers in Human Behavior*, *115*, 106607.

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, *97*, 304-316.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality, 37*, 504-528.

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500-507.

Kim, H., Jung, S., & Ryu, G. (2020). A study on the restaurant recommendation service app based on AI chatbot using personalization information. *International Journal of Advanced Culture Technology*, *8*(4), 263-270.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80.

Lee, Y. C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-12.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

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

Nguyen, A. T., Kharosekar, A., Krishnan, S., Krishnan, S., Tate, E., Wallace, B. C., & Lease, M. (2018). Believe it or not: Designing a human-ai partnership for mixed-initiative fact-checking. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, 189-199.

Okuda, T., & Shoda, S. (2018). AI-based chatbot service for financial industry. *Fujitsu Scientific and Technical Journal*, *54*(2), 4-8.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.

Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., Xu, J., Ott, M., Shuster, K., Smith, E. M., Boreau, Y. L., & Weston, J. (2021). Recipes for Building an Open-Domain Chatbot. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 300-325.

Rosenberg, M. (1979). *Conceiving the Self*. New York: Basic Books.

Sharan, N. N., & Romano, D. M. (2020). The effects of personality and locus of control on trust in humans versus artificial intelligence. *Heliyon*, *6*(8), Article e04572.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *149*, 102601.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., Decero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: thematic analysis. *Journal of Medical Internet Research*, *22*(3), Article e16235.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113-117.

Weinfield, N. S., Sroufe, L. A., & Egeland, B. (2000). Attachment from infancy to early adulthood in a high-risk sample: Continuity, discontinuity, and their correlates. *Child Development*, *71*(3), 695-702.

Weinfield, N. S., Whaley, G. J., & Egeland, B. (2004). Continuity, discontinuity, and coherence in attachment from infancy to late adolescence: Sequelae of organization and disorganization. *Attachment & Human Development*, *6*(1), 73-97.