

Conversations with two chatbots: tutor mike and cleverbot

Mahesh Vanjani, *Texas Southern University, mahesh.vanjani@tsu.edu*

Jamison Posey, *University of Mississippi, jposey@bus.olemiss.edu*

Abstract

Chatbots enable machines to emulate human conversation. Heretofore comparisons of the systems have largely relied on human subjects interacting with the chatbots using a set of uniform prompts that have been predetermined and then provided to the participants. For this experiment, we requested students at a university to conduct a two-minute free flowing conversation with one of two chatbots. Subjects were aware that they are conversing with a chatbot. After conversing with the chatbot the subjects completed a questionnaire and responded based on their experience. The two chatbots used for the experiment had some variation vis-à-vis modus of chatbot development and purpose. The survey results indicate that while one chatbot performed marginally better than the other neither chatbot was perceived to be a human or an acceptable substitute for a conversation with a human.

Keywords: chatbot, Tutor Mike, Cleverbot, artificial intelligence, AI

Introduction

ChatGPT is all the rage. However, chatbots are not new and have been in development for over six decades. A chatbot, also known as a chat bot, virtual assistant, conversational agent, or virtual agent, seeks to imitate human dialogue in order to provide a more intuitive computer interface. In essence, Chatbots are computer programs that mimic human conversation by attempting to provide a natural user interface and human-like conversation flow. Chatbots rely on technology to function and to emulate human dialogue to provide a more intuitive user interface to applications. Technology such as AI (Artificial Intelligence) and NLP (Natural Language Processing) can enhance the ability of chatbots to learn and to better reproduce a more natural and free flowing conversation. In current times a chatbot must be mobile friendly in order to be successful. Figure 1 depicts a conversation between a human and a chatbot via a mobile phone device.

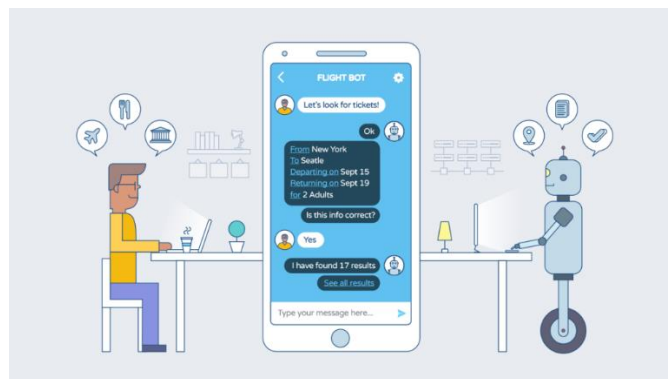


Figure 1: Conversation with a Chatbot

The earliest well known chatbot is ELIZA, also an NLP (Natural Language Processing) program created at the MIT Artificial Intelligence Laboratory. When the original ELIZA chatbot first appeared in the 60's, some people mistook her for human. At the time, the illusion of intelligence worked best if a conversation with ELIZA or any other chatbot was limited to short and simple topic-specific communication. Earlier chatbots were generally text-based. They were programmed to respond to a limited set of relatively simple queries. Responses were generally pre-configured responses to preempted questions. While interactive they were limited to the knowledge that had been added by the chatbot developer and tended to fail when presented with unexpected or complicated queries.

Over time, chatbots have integrated more rules and natural language processing, so end users can engage them in a conversation. Hundreds of these systems have been developed. Chatbots.org is an excellent web-resource to browse chatbots available by category. (<https://www.chatbots.org/>). The website also provides user reviews, quick-start tutorials and guidance by industry and application. Another site (<https://www.personalityforge.com/chatbot-finder.php>) had compiled 7,499 chatbots that were available as of date. However, some are available only as apps via social media platforms and/or via mobile phones. Increasingly, the latest types of chatbots are contextually aware and able to learn as they're exposed to more and more human language. We have since witnessed a quantum leap in the evolution of Chatbots with the emergence of ChatGPT.

ChatGPT is an extremely sophisticated artificial intelligence chatbot developed by AI research company OpenAI. ChatGPT was officially launched on November 30, 2022. Microsoft has formed a strategic partnership with OpenAI's with a \$10 billion investment. Since then, this Artificial Intelligence (AI) technology has been added to Microsoft products including Bing, the corporation's search engine. ChatGPT is a generative AI that can produce content from text to images, have conversations with humans, suggest edits to computer programming code and more. The chatbot can answer questions or assist humans in queries or tasks through its vast training using social media, websites, articles, datasets, books and other forms of text on the internet. Since then, the competition has ramped up. Google Bard is Google's attempt to compete with ChatGPT. Like ChatGPT, Google Bard is an AI chatbot. It is powered by a language model with the intent to allow conversations with users. For now, Google Bard is strictly text-based, but this is expected to change in the very near future.

Chatbot Evaluation

This study was conducted prior to the release of and wide media coverage and exponential popularity of ChatGPT. This paper focusses on the subject's conversation experience with the following two chatbots: Tutor Mike and Cleverbot.

Tutor Mike

https://www.rong-chang.com/tutor_mike.htm

The "Tutor Mike" chatbot has been created by Professor Ron Lee to help English learners practice English. It has no self-learning ability and must be "taught" or "trained". The goal of this chatbot is different from chatbots such as Alexa or Siri. These chatbots are mainly used for finding information or performing tasks. Tutor Mike, as named, is an English language tutor with the persona of a young unmarried American male. He is knowledgeable in grammar, history and geography and is capable of correcting a student's grammatical errors. The Tutor Mike chatbot won the second place in the 2018 Loebner Prize Contest.

The Loebner Prize is the oldest Turing Test contest, started in 1991 by Hugh Loebner and the Cambridge Centre for Behavioral studies. The Loebner Prize was awarded to the creators of the first bot that could pass

an extended Turing Test involving textual, visual, and auditory components. The Turing Test is a method of inquiry in artificial intelligence (AI) for determining whether a computer is capable of thinking like a human being. According to Professor Lee, “Tutor Mike” did so well in the competition because he created the Final Patterns script based on his Final Patterns theory. “Tutor Mike is written in AIML created by Dr. Richard Wallace. Figure 2 illustrates a sample screen for chatbot Tutor Mike. Figure 3 is an image of the chatbot’s persona/avatar.



Figure 2: Tutor Mike Sample Screen



Figure 3: Tutor Mike Sample Persona/Avatar

Cleverbot

<https://www.cleverbot.com/>

Rollo Carpenter, a British scientist specializing in building and researching artificial intelligence systems, initially created the Cleverbot chatbot in 2006. Cleverbot is an AI-powered chatbot and was designed to be a tool to improve conversational skills. The chatbot can mimic natural language learn from its conversations

with users. Cleverbot is a unique AI chatbot and stands out from other similar AI chatbots due to its advanced artificial intelligence. It is powered by an AI algorithm that is constantly learning from conversations it has with people. This allows it to understand the context of conversations and respond with more natural-sounding responses. It also has the ability to remember past conversations, which allows it to provide more personalized responses. Additionally, Cleverbot is able to recognize and respond to slang, humor, and sarcasm, which sets it apart from other AI chatbots. This makes it more engaging and lifelike, making it a popular choice for businesses looking to provide customers with a more human-like experience.

Over the years, the Cleverbot chatbot has been continuously updated and improved. It has been integrated into Discord allowing users of this communication platform to speak with Cleverbot without visiting the website. With over a quarter of a billion users, Discord is one of the most popular VoIP and instant messaging social platforms with millions of regular users. In recent years,

Cleverbot has become more sophisticated, incorporating features such as natural language processing, sentiment analysis, and machine learning. It is now capable of understanding complex conversations and responding in a more human-like manner. This chatbot is multilingual and can understand and respond in several different languages, including English, Spanish, French, German, Italian, and Dutch. It can also learn and adapt to new languages over time, making it an even more versatile tool for communication.

However, while the chatbot has been known to fool some users into thinking they are talking to a human it has not officially passed the Turing Test. This is largely because the Turing Test requires a chatbot to fool a panel of judges, rather than just individual users. Figure 4 illustrates a sample screen for chatbot Cleverbot. Figure 5 is an image of the chatbot's persona/avatar.

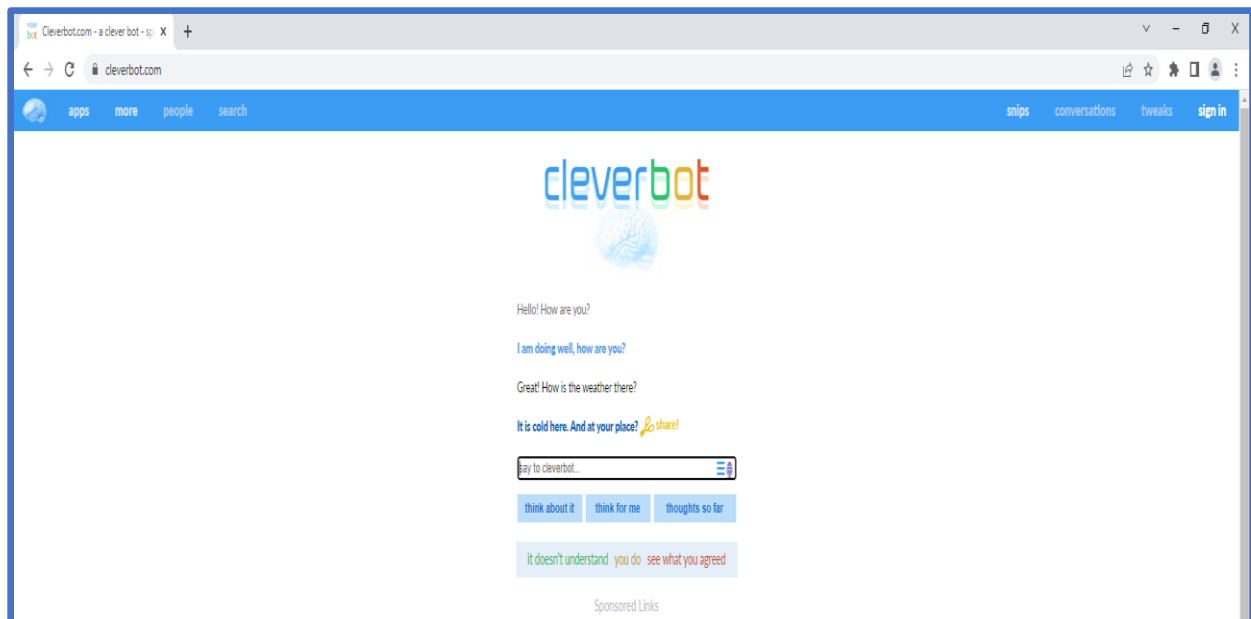


Figure 4: Cleverbot Sample Screen



Figure 5: Cleverbot Persona/Avatar

Purpose

Chatbots have been in development for a while and, over the past six decades, a multitude of chatbots have been developed and released for public use. Few if any studies have been conducted on satisfaction with conversating with a chatbot using a free-flowing natural conversation. Most studies require participants to use pre-selected prompts and the evaluation is based on the whole conversation transcript based on the prompts provided to participants conversing with the chatbot. The purpose of this study was to evaluate how two chatbots performed when engaged in a free-flowing conversation.

Subjects and Task Description

The chatbot evaluation survey results are based on a questionnaire completed by 141 students from a university in the southern region of the United States. Students were assigned to one of the two chatbots and asked to conduct a two-minute free flowing conversation with the chatbot assigned. Once they were done conversing with the chatbot, students completed a questionnaire and responded based on their experience. We had 76 students who conversed with chatbot Tutor Mike and 65 students who conversed with the Cleverbot chatbot. Table 1 below lists the nine questions on the questionnaire. The response requested was based on a Likert scale ranging from 1 (totally disagree) to 4 (neutral) to 7 (totally agree). The appendix includes a copy of the questionnaire provided to the students.

Table 1: Survey Questions

Survey Questions	
Q1	The chatbot responses seemed natural (what you expect a human might say).
Q2	The chatbot responses were too fast to seem like a human was typing on the other end.
Q3	The chatbot seemed friendly (did not argue with you or seem rude).
Q4	The chatbot seemed knowledgeable (it seemed to know the topic we discussed).
Q5	If I didn't know that I was using a chatbot, I would have thought I was chatting with a real person.
Q6	I would use this chatbot again.
Q7	Chatting with the system made me less lonely.
Q8	Chatting with this system was better than talking with a friend.
Q9	I trust the chatbot did not record or share the information I typed.

Analysis of Results and Discussion

Survey response results are reported in Table 2. The table includes the mean response for each of the nine questions for each of the two chatbots and the overall mean value for both chatbots. Independent Samples t Test was used to compare the means of the two chatbot groups to determine whether there is any statistical evidence that the associated population means are significantly different. Any significant differences in the mean responses between the two chatbots are noted in the t Stat column.

Table 2: Survey Response Data Analysis

Survey Question Number	Tutor Mike		Cleverbot		Tutor Mike versus Cleverbot t Stat	All Responses	
	N = 76		N = 65			N = 141	
	Mean	Standard Deviation	Mean	Standard Deviation			Mean
Q1	3.83	1.45	3.89	1.39	-0.26453	3.86	1.42
Q2	5.99	1.57	4.23	1.78	6.10270	5.18	1.88
Q3	5.14	1.65	3.43	1.70	6.07187	4.35	1.88
Q4	4.14	1.54	3.98	1.66	0.58994	4.07	1.59
Q5	2.70	1.80	2.77	1.98	-0.24310	2.73	1.91
Q6	2.93	1.63	3.62	1.84	-2.12243	3.25	1.73
Q7	2.64	1.65	2.83	1.87	-0.62245	2.73	1.75
Q8	1.51	0.86	1.55	1.17	-0.22814	1.53	1.02
Q9	3.24	2.03	2.75	1.84	1.49246	3.01	1.95

Q1: The chatbot responses seemed natural (what you expect a human might say).

For the question related to the chatbot responses seeming natural and human like the average response for both chatbots was higher than the median value with no statistically significant difference in the average response. The overall mean for the entire sample set was also above the median value. While no definitive conclusions can be drawn vis-à-vis any difference between the two chatbots the response mean values suggest that at least half of the students surveyed felt like they were communicating with a human.

Q2: The chatbot responses were too fast to seem like a human was typing on the other end.

For the question related to the chatbot responses being too fast to appear that a human was responding the data indicates that both the Tutor Mike and Cleverbot chatbots were perceived to be artificial or non-human. Mean values for each chatbot and the overall sample were higher than the median value. However, there is a statistically significant difference in the perception of the sample populations using the chatbots. Tutor Mike was more strongly perceived to be non-human as opposed to Cleverbot. This is not surprising given the purpose of the two chatbots and how they were developed and configured. Cleverbot is an AI based conversational chatbot and can learn from conversations whereas Tutor Mike cannot. Clearly both chatbots responded too fast to fool the users into thinking a human was responding to them.

Q3: The chatbot seemed friendly (did not argue with you or seem rude).

Tutor Mike was perceived to be friendlier than Cleverbot. The mean response for this question for Tutor Mike was much higher than the mean and statistically significantly higher than the mean response for Cleverbot. In fact, the mean response for Cleverbot was marginally lower than the mean implying that a slight majority of users did not perceive this chatbot to be friendly. The overall sample mean was above the median value implying that, in general, users perceived that the chatbots were somewhat friendly.

Q4: The chatbot seemed knowledgeable (it seemed to know the topic we discussed).

Both chatbots performed similarly vis-à-vis being knowledgeable. The mean response for both chatbots and the overall sample was over the median with no statistically significant variation between the two chatbots. We can infer that more than half the survey respondents felt that the chatbots were knowledgeable. Students were not provided with any prompts and were simply asked to engage in a natural conversation with the chatbot assigned to them. Consequently, their individual perception of a chatbot being knowledgeable will be a function of their specific conversation or user interaction with the assigned chatbot. Also, the chatbots differ in how they were developed and intended use. For example, Tutor Mike chatbot was specifically created to operate as an English tutor so is not expected to respond knowledgeably in other areas such as Physics or Chemistry. Cleverbot is configured to learn from prior conversations. Given that, it is interesting to note that for this experiment, Tutor Mike had a slightly higher mean for this question so was perceived to be more knowledgeable.

Q5: If I didn't know that I was using a chatbot, I would have thought I was chatting with a real person.

Overall, for both chatbots, the students did not perceive that either conversed like a real person. This result is in line with their responses to a related question (question number 2) which asked them if they thought they were communicating with a human. The mean response is very similar for both chatbots with no statistically significant difference ergo students generally concurred that they would not have thought that they were chatting with a real person. It is important to note that participants has been informed that they would be conversing with a chatbot and not a real person.

Q6: I would use this chatbot again.

We had mixed results with regard to a preference for using the chatbot again. Based on the mean value of responses, the students who were assigned the Tutor Mike chatbot would not like to use the chatbot again. However, for users assigned to Cleverbot more than half would use the chatbot again. This difference was statistically significant so we can conclude that, at least for this sample population, Cleverbot fared better in that there is more likelihood of users being inclined to repeat their experience in conversing with the chatbot.

Q7: Chatting with the system made me less lonely.

For both chatbots the respondents did not feel less lonely as a result of conversing with the chatbot. There was no statistically significant difference in the results. This result is in line with responses to earlier questions where respondents did not believe they were chatting with a human. If they do not believe they are conversing with a human, they are unlikely to feel less lonely or feel any sense of companionship when conversing with either of the two chatbots.

Q8: Chatting with this system was better than talking with a friend.

Again, the results for the question related to chatting with a friend are close to similarly slated questions. Respondents for both chatbots did not perceive that conversing with the chatbot was better than talking with a friend. This result makes sense given that users were not convinced that they were speaking with a human and/or having a natural flowing conversation as they would with another person.

Q9: I trust the chatbot did not record or share the information I typed.

In general respondents felt that the chatbot was recording their conversation and/or sharing the recorded conversation. There was no statistically significant difference in the perception of trust that the chatbots were not recording or sharing conversations. The mean values were closer to the median for Tutor Mike than Cleverbot but respondents using either chatbot were generally convinced that their conversations were being recorded and/or shared. Given that students knew they were conversing with a chatbot and, it is now more commonly known that such systems learn from past conversations and interaction, it is not surprising that most respondents felt that their conversation with chatbot would be recorded and shared.

In summary, based on the mean responses with a statistically significant difference, the users perceived Tutor Mike to be less human than Cleverbot. They also thought that Tutor Mike was friendlier. Finally, users assigned to Tutor Mike were less averse to having another conversation with the chatbot as opposed to users assigned to having a conversation with the Cleverbot chatbot.

The results of this experiment are interesting vis-à-vis future chatbot development. User perception is key to the success of a chatbot. Cleverbot was designed to learn from previous conversations whereas Tutor Mke was not. However, based on the results of this experiment, we find that while the audience thought Cleverbot conversed more like a human they also found Tutor Mike friendlier.

In general, neither chatbot users were significantly interested in having another conversation with their assigned chatbot. Perhaps, the fact that they knew they were conversing with a chatbot caused the lack of interest in a future conversation. While we have experienced success with chatbot use for customer service type usage, clearly a lot more research and development is needed to create a chatbot that can serve as a substitute for a conversation with another human.

Conclusion

For this experiment, the survey results indicate that while the Tutor Mike chatbot performed marginally better than Cleverbot neither chatbot was perceived to be human or to be an acceptable substitute for a conversation with a human.

There were some limitations with this study. We used a sample of convenience and all subjects were college students. The subjects had been made aware that they would be conversing with a chatbot. There were no preset prompts and students were asked to have a two-minute free-flowing conversation with a chatbot. They were assigned one of two chatbots but the chatbots were different in how they were developed and intent. This means different users could end up with a different experience based on the content and flow of their conversation with the assigned chatbot. Regardless, Tutor Mike that was never configured to learn from previous conversations seems to have marginally outperformed Cleverbot which is an AI based conversational chatbot specifically configured to learn from prior conversations.

Future research in this area will likely show dramatically different results given the quantum leap in chatbot evolution and development with the advent of ChatGPT. Chatbots are more human-like than ever before so it stands to reason to speculate that users can be fooled into thinking that they are conversing with a human.

References

- Best Chatbot: Testing ALICE, Cleverbot, Rose, and more. (2017). Inverse.
<https://www.inverse.com/article/37615-best-chatbot>.
- Bram, U. (2015). The Reverse Turing Test: Pretending to Be a Chatbot Is Harder Than You Think. Vice.
<https://www.vice.com/en/article/ypw7vm/the-reverse-turing-test-pretending-to-be-a-chatbot-is-harder-than-you-think>
- Brandtzaeg P. and Følstad, A. (2017). Why people use chatbots. In: Kompatsiaris I. et al. (eds) Internet Science. INSCI 2017. Lecture Notes in Computer Science, vol 10673. Springer, Cham.
- Cleverbot - Everything you need to know. (n.d.). Botpenguin.com. <https://botpenguin.com/cleverbot/>
- Cleverbot. (2014). Cleverbot. <https://www.cleverbot.com/>
- Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22(5), 811-817.
- ESL Robot - English Tutor. (n.d.). [Www.rong-Chang.com](http://www.rong-chang.com). Retrieved October 18, 2022, from https://www.rong-chang.com/tutor_mike.htm
- Eze, E. S. (2019). How Chatbots Can Help You Increase Conversion. Medium.
<https://blog.markgrowth.com/how-chat-bots-can-help-you-increase-conversion-6561ba0b8ab0>
- Frąckiewicz, M. (2023). Cleverbot: An Introduction to the Original AI Chatbot. TS2 SPACE.
<https://ts2.space/en/cleverbot-an-introduction-to-the-original-ai-chatbot/#:~:text=Cleverbot%20is%20an%20AI%20powered>
- IBM. (2023). What is a chatbot? | IBM. [Www.ibm.com](http://www.ibm.com). <https://www.ibm.com/topics/chatbots>
- Kerry A., Ellis R., and Bull S. (2009). Conversational agents in E-Learning. In: Allen T., Ellis R., Petridis M. (eds) *Applications and Innovations in Intelligent Systems XVI*. SGAI 2008. Springer, London.
- Lee, M. C. (2018). Chatbot “Tutor Mike” Won Second Place in 2018 Loebner Prize Contest. Medium.
<https://medium.com/@mikeclee/chatbot-tutor-mike-won-second-place-in-2018-loebner-prize-contest-c801c2b264ee>
- Lortie, C. and Guitton, M. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One*, 6(9).
- McMillan, M. (2023). What is Google Bard? Everything you need to know about ChatGPT rival. Tom’s Guide. <https://www.tomsguide.com/news/google-bard-ai>
- Moloney, C. (2018). How to win a Turing Test (the Loebner prize). Medium.
<https://chatbotsmagazine.com/how-to-win-a-turing-test-the-loebner-prize-3ac2752250f1>

Issues in Information Systems

Volume 24, Issue 3, pp. 24-34, 2023

Personality Forge AI Chatbot Platform - Chatbot Finder. (n.d.). Wwww.personalityforge.com. Retrieved May 25, 2023, from <https://www.personalityforge.com/chatbot-finder.php>

Regalbuto, G. (2023). What is ChatGPT? Fox News. <https://www.foxnews.com/tech/what-is-chatgpt>

Shawar, B. and Atwell, E. (2007a). Chatbots: Are they really useful? LDV-Forum 2007, 22(1), 29-49.

Shawar, B. and Atwell, E. (2007b). Different measurements to evaluate a chatbot system. NAACL-HLT-Dialog '07 Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies, 89-96.

The Loebner Prize. (n.d.). Wwww.ocf.berkeley.edu.
<https://www.ocf.berkeley.edu/~arihuang/academic/research/loebner.html>

Vanjani, M., Posey, J. and Aiken, M. (2019). An Evaluation of Three Online Chatbots. Southwestern Business Administration Journal: Volume 18: Issue 1, Article 3.

Appendix

Chatbot Evaluation Questionnaire

The purpose of this study is to evaluate the differences among four online chatbots. Your instructor will assign you to a group that will evaluate one of them. Please open this chatbot: www.one-of-the-two-chatbot-links-here.com

[Assign one of the following two chatbots]

1. Tutor Mike https://www.eslfast.com/robot/tutor_mike.htm
2. Cleverbot <https://www.cleverbot.com/>

Chat with the system for 2 minutes. Try asking it some questions and engaging in a natural conversation. Take a screen shot of any part of your “conversation” and add the image at the bottom of the questionnaire below.

Based on your experience with the chatbot please answer the following questions using the scale: **1 (totally disagree), 4 (neutral), 7 (totally agree)**

1. **The chatbot responses seemed natural (what you expect a human might say).**
1 2 3 4 5 6 7
2. **The chatbot responses were too fast to seem like a human was typing on the other end.**
1 2 3 4 5 6 7
3. **The chatbot seemed friendly (did not argue with you or seem rude).**
1 2 3 4 5 6 7
4. **The chatbot seemed knowledgeable (it seemed to know the topic we discussed).**
1 2 3 4 5 6 7
5. **If I didn't know that I was using a chatbot, I would have thought I was chatting with a real person.**
1 2 3 4 5 6 7
6. **I would use this chatbot again.**
1 2 3 4 5 6 7
7. **Chatting with the system made me less lonely.**
1 2 3 4 5 6 7
8. **Chatting with this system was better than talking with a friend.**
1 2 3 4 5 6 7
9. **I trust the chatbot did not record or share the information I typed.**
1 2 3 4 5 6 7