Recommendations for a Conversational Chatbot

Uyen Bui

California State University, Long Beach

**Feature at a Glance**

A chatbot is a computer program meant to engage in conversation with human users as if it were another human. The implementation of chatbots in different settings has proved to be promising, as they can, for example, serve as conversation partners for English as a Second Language (ESL) students and provide customer service to several users simultaneously, which lowers company costs. However, chatbots are not without their issues, with the most common ones being lack of ability to truly understand user input and maintain conversation history. As research interest in chatbots has drastically increased since 2016 and chatbots continue to become increasingly present in everyday life, researchers need to investigate and address current and emerging issues. This paper will discuss issues with the conversational aspects of chatbots necessary for completing tasks and provide recommendations. Limitations and future areas of interest will be discussed as well.

**Keywords:** Turing test, artificial intelligence, conversational qualities, privacy, conversation history

**Chatbots Origins and General Usage**

A chatbot is a well-known form of artificial intelligence (AI) system within the field of human-computer interaction (HCI; Adamopoulou & Moussiades, 2020). To be more precise, a chatbot is a computer program specifically created to engage in conversation with human users as if it were another human, typically on the internet (Oxford University Press, n.d.). Alternative names for chatbots include smart bots, interactive agents, digital assistants, and/or artificial conversation entities (Adamopoulou & Moussiades, 2020). The first known chatbot was ELIZA, which imitated the role of a psychotherapist to patients in clinical treatment (Weizenbaum, 1966). At first, patients believed they were actually speaking to another human, but soon discovered that ELIZA did not truly understand their input. Instead, it was merely taking their inputs and repeating it back in the form of standard phrases or questions.

Regardless, researchers and developers still used ELIZA as the basis for other chatbots, such as ALICE (Artificial Intelligence Foundation, 2007; Abu Shawar & Atwell, 2003; Wallace, 2003), which was created to entertain and converse as a human with other human users. Imitating humans and providing entertainment to users are just a few of the many usages of chatbots (Adamopoulou & Moussiades, 2020), as they are now used in a variety of contexts for several purposes. Some of these areas include education, information retrieval, business, and e-commerce (i.e., online shopping; Abu Shawar & Atwell, 2007), with other appeals to chatbots being their assistance with productivity tasks and novelty factor (Adamopoulou & Moussiades, 2020). Chatbots can also quickly respond to specific user questions and therefore serve as a quick form of user support. By handling several users at once, chatbots are able to lower customer service costs for companies, making them a good choice from a business standpoint (Brandtzaeg & Følstad, 2017).

**Chatbot Categorization**

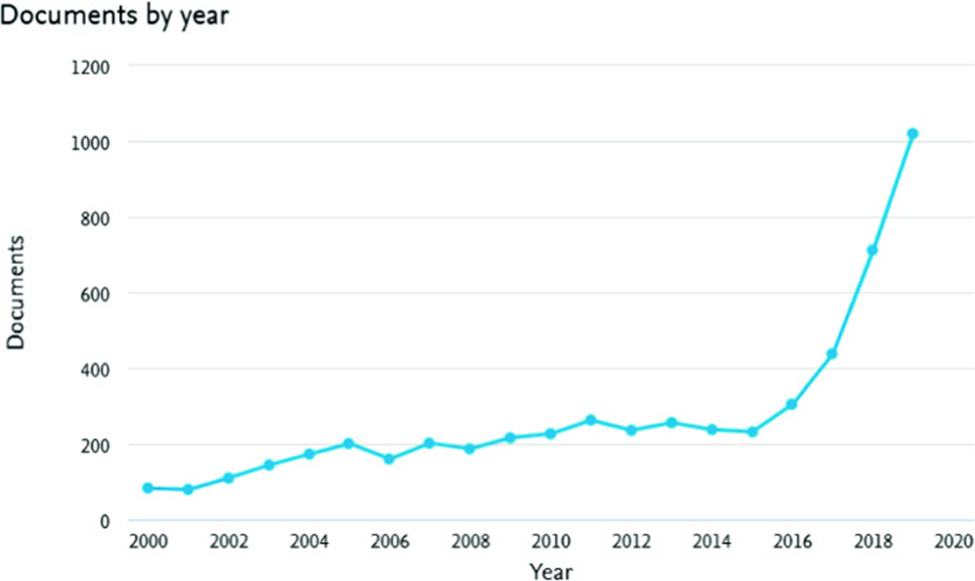
There are many ways to categorize the different types of chatbots more broadly, but for the sake of discussion, this paper will only focus on two types of classification, the first one being according to the type of service provided and how emotionally close the chatbots are to the user (Nimavat & Champaneria, 2017). Following this classification, there are interpersonal, intrapersonal, and inter-agent chatbots. Interpersonal chatbots assist with tasks such as restaurant or flight booking and frequently asked questions (FAQ). They are not close friends with the user but will gather necessary information and pass it on to them. Intrapersonal chatbots, on the other hand, are meant to understand the user as if the chatbots were human themselves and therefore are more similar to friends. Examples of intrapersonal chatbots can be found in chat applications (apps), such as Messenger and Slack. Lastly, inter-agent chatbots act as a means for bots to communicate with one another, which allows developers to integrate bots together (e.g., Alexa-Cortana integration).

The second type of classification is according to the goals the chatbots are meant to achieve (Nimavat & Champaneria, 2017). Informative chatbots are concerned with tasks related to providing information to users. They generally have a pre-determined “bank” of information that is built into them beforehand (e.g., FAQ page, site database), and thus their knowledge is limited. Chat-based or conversational chatbots have a goal of responding appropriately to user input as if they were another human user (e.g., Siri, Alexa) and are programmed with skills, such as cross-questioning, to be able to continue a conversation. Task-based chatbots assist users in performing a specific task, such as booking an airplane flight or making a restaurant reservation. These types of chatbots specialize in asking the user for information and understanding subsequent user input, as that information is needed to carry out the task.

There are common issues across all types of chatbots, regardless of their classification, generally relating to whether the chatbots can pass the Turing Test (i.e., whether an AI machine can “think” as a human; Turing, 1950). Two of these issues, which are present in the aforementioned ALICE chatbot, are the lack of abilities to maintain conversation history and truly understand user input, instead opting to give an answer from a pre-determined knowledge domain (Abu Shawar & Atwell, 2007). However, chatbots are becoming increasingly relevant and important in everyday life, as indicated by a Scopus search, which reveals there has been a large growth in research interest for chatbots since the year 2016 (Figure 1). Therefore, it is important to continue studying how to mitigate these common issues as well as other issues that have been discovered. For the sake of discussion, this paper will only focus on issues relating to conversational qualities of chatbots needed to complete tasks and provide design recommendations.

**Figure 1**

*Number of Chatbot-related Documents from 2000 to 2019*



*Note*. Scopus search results from 2000 to 2019 for the keywords “chatbot,” “conversation agent,” or “conversational interface.”

**Issues Regarding Conversation**

As previously mentioned, common weaknesses of chatbots include not truly understanding user input and failing to maintain a conversation history (Abu Shawar & Atwell, 2007). These weaknesses in conversational intelligence continue to be present in more recent studies, where users criticized a video game chatbot for being entirely scripted and disregarding user input, and a shopping chatbot for not retaining their conversation history across separate chat sessions (Jain et al., 2018). In the case of the shopping chatbot, the user asked it to make a recommendation for shoes to go along with a dress they chose the day prior, which the chatbot failed to carry out to a satisfactory degree.

Other criticisms of the shopping chatbot include continuing to carry out an incorrect task instead of admitting failure and not actively asking questions to narrow down the search space (Jain et al., 2018). For example, a user would ask the chatbot to help them find eye-liner products, and the chatbot would proceed to show socks or a product in a different color than what was requested (e.g., showing brown dresses instead of red dresses). Additionally, the shopping chatbot had trouble understanding negation in requests, where a user would ask for “not [product]” and the chatbot would still search up that particular product.

In ESL (English as a Second Language) settings, chatbots have been highly requested by teachers to allow students an opportunity to practice with an English-speaking conversational partner (Atwell, 1999). While these ESL chatbots have been found to have proper grammar and a wide range of vocabulary, they are lacking in semantics (Coniam, 2014). Essentially, the responses formed by chatbots are grammatically correct and are adequately diverse in vocabulary, but do not make sense. As chatbots can be conversational partners for learning another language or assistants in completing various tasks, it is important that the responses are coherent to ensure a positive user experience.

In mental health contexts, chatbots are perceived as “cold” (i.e., emotionally distant) in comparison to an actual health professional (Nadarzynski et al., 2019). This perception may be due to chatbots being unable to truly understand the more emotional issues of patients, an ability which would be required for them to appear empathetic. Users are also concerned with the privacy of their conversations with these chatbots, as they are sharing their private health information and talking about sensitive subjects, as well as the accuracy of the more detailed health information that is provided by the chatbots. Generally, users are comfortable with chatbots providing commonly known information and doing simple tasks (e.g., booking appointments), but they prefer that a health professional inform them on more detailed health information.

**Recommendations to Improve Conversational Qualities**

In order to decrease the “coldness” of mental health chatbots and increase friendliness in general, designers could consider having the chatbot disclose information about itself to the users or patients. When both the chatbot and user are disclosing information about themselves, there is an increase in perceived intimacy between the two parties (Lee et al., 2020). Additionally, regardless of the amount of disclosure (i.e., low or high), users felt the chatbot was more trustworthy and therefore felt more comfortable discussing sensitive subjects. Consequently, users were more inclined to continue using the chatbot. These reactions were in comparison to chatbots that did not disclose any information about themselves. However, users did feel that some answers from low disclosure chatbots were generic, indicating that chatbots should disclose a large amount of information to ensure the highest feelings of intimacy, trust, and comfort from the user. Other studies have also found that when chatbots ask follow-up questions pertaining to the topic, they appear more human-like and encourage users to provide more involved responses (Schuetzler et al., 2014). Therefore, chatbots should not only disclose information about themselves but also actively ask questions to the user about the topic at hand.

While users would like some retention of conversation history across different sessions (Jain et al., 2018), patients who use a mental health chatbot are concerned with privacy of their disclosed information (Nadarzynski et al., 2019). To address privacy concerns while maintaining conversation history, designers can follow ethics originally outlined in mental health settings (Kretzschmar et al., 2019). The first ethic is concerning privacy and transparency, where user data obtained during a chatbot conversation should be kept private and protected through anonymization. Companies should consider including a privacy policy with each chatbot that explicitly states whether conversation data will be shared. Secondly, users should be informed of the potential benefits and risks as well as the capabilities of the chatbot, potentially in a chatbot introduction or tutorial. Reminders of capabilities can also help to mitigate the aforementioned issue of chatbots not admitting failure (Jain et al., 2018), as chatbots can say they do not know something and remind the user what they are capable of doing. Lastly, in the event that users run into a major issue with the chatbot, there should be emergency external contact information available, or monitoring done by designers. In fact, the design should have already been carefully reviewed before implementation of the chatbot, but otherwise there needs to be a “safety net” in case issues arise during operation (Kretzschmar et al., 2019).

The common chatbot issue of not being able to truly understand user input (Abu Shawar & Atwell, 2007) still needs to be addressed, as some chatbots will follow a “script” and disregard input (Jain et al., 2018). Designers should therefore work on chatbot responses in conversations to be grammatically and semantically adequate (i.e., ensure they are “human-like”) as these qualities are the aim of chatbots (Oxford University Press, n.d.) and have been received positively with Pandorabots (Jain et al., 2018). In particular, users liked how the chatbot spoke naturally and appeared to understand their input, as they responded appropriately in most of their conversations. Chatbots should also be able to actively ask questions about the topic at hand based on previous responses, rather than passively gathering information and responding based on their pre-determined knowledge base (Nuruzzaman & Hussain, 2018). When chatbots ask questions, they will be able to narrow the search space for users and better provide assistance, which is useful in contexts such as shopping (Jain et al., 2018).

**Applications in Summary**

A large increase in research interest in chatbots since the year 2016 indicates that chatbots are becoming increasingly relevant in everyday life (*Scopus - Document Search*, n.d.). They are present in the areas of education, information retrieval, business, and online shopping (Abu Shawar & Atwell, 2007), and can be categorized according to either the type of service provided with emotional closeness or the goal the chatbot is meant to achieve (Nimavat & Champaneria, 2017). Regardless of the classification type, chatbots are meant to converse with humans as if they were another human (Oxford University Press, n.d.), and therefore designers should focus on improving conversational qualities related to tasks. As a starting point, designers should ensure that chatbots are able to truly understand user input rather than answering from a fixed knowledge bank and maintain conversation history across chat sessions, as these two issues are the most common issues to this day (Abu Shawar & Atwell, 2007).

In other contexts, such as online shopping, designers should have chatbots actively ask questions to reduce the search space and determine what exactly the user wants (Nuruzzaman & Hussain, 2018; Jain et al., 2018). The chatbot capabilities should also be clearly communicated in the beginning of usage to ensure users are aware of what services are available (Kretzschmar et al., 2019). In the case that chatbots are not able to perform a service, they should admit failure to prevent user frustration (Jain et al., 2018) and restate their capabilities. Additionally, while maintaining conversation history is beneficial, there are privacy concerns in mental health settings as users are disclosing private health information (Nadarzynski et al., 2019). To help address these concerns, user data should be protected via anonymization of chat information, and a privacy policy stating whether information will be shared should be included with the chatbot (Kretzschmar et al., 2019). As a safety net for users in case of failure, the chatbot should also be monitored by experts or provide external contact information.

Research in ESL settings has revealed that chatbots have grammatically correct responses and diverse vocabulary, but do not have proper semantics (Coniam, 2014). Designers should ensure that chatbots have both proper grammar and semantics, not only to help ESL students properly improve their conversational skills but also ensure chatbot responses are coherent in general. In mental health settings, chatbots tend to feel “cold” in comparison to a human health professional (Nadarzynski et al., 2019), but this issue can be addressed by ensuring the chatbot also discloses information about itself, as when chatbots do so, users feel higher levels of intimacy and trust (Lee et al., 2020). The chatbot can also ask follow-up questions to increase user engagement (Schuetzler et al., 2014). Users will then feel more comfortable disclosing information about themselves and feel more inclined to continue using the chatbot.

**Limitations and Future Areas of Consideration**

As this paper only focused on some conversational qualities of chatbots, future areas of consideration include other aspects (conversational or not) in which chatbots are still lacking. Recent studies reveal that chatbots are still unable to recognize grammatical errors in sentences (Nuruzzaman & Hussain, 2018), cannot determine user personas (Nimavat & Champaneria, 2017) and emotions (Nuruzzaman & Hussain, 2018) from conversations, and cannot differentiate the structures of various languages or be multilingual. Therefore, further research is needed to determine how to extract such user information (e.g., personality, emotions) from chatbot conversations and implement that information into their responses to ensure the user experience feels more personalized and natural (Nimavat & Champaneria, 2017). The other issues regarding grammar and different languages would be useful in ESL settings, where students from various language backgrounds will be learning proper English language structure.

However, making a chatbot appear more “human” in the overall chat experience is also dependent on current technological capabilities. Programmers must continue to work on natural language processing (NLP) to ensure chatbots are flexible and able to understand user inputs written in different styles (e.g., “What’s the temperature?” and “Could you check the current temperature?” should elicit same answer; Rahman et al., 2017). Additionally, researchers have suggested that chatbots having a particular personality would make them appear more human (Nimavat & Champaneria, 2017), and that personality would likely correspond to their services. Soon, though, chatbots may be able handle more complex conversational tasks, such as monetary payments, given current advancements in machine learning (Rahman et al., 2017). Other issues to further investigate could include development, implementation, and maintenance costs of chatbots, which will most likely vary depending on the chatbot tasks and goals.

References

Abu Shawar, B., & Atwell, E. (2003). Using dialogue corpora to train a chatbot. In D. Archer, P.

Rayson, A. Wilson, & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference (CL2003)* (pp. 681–690).

Abu Shawar, B., & Atwell, E. (2007). Chatbots: Are they really useful? *Journal for Language*

*Technology and Computational Linguistics*, *22*(1), 29–49.

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *Artificial*

*Intelligence Applications and Innovations*, *584*, 373–383. <https://doi.org/10.1007/978-3-030-49186-4_31>

Artificial Intelligence Foundation. (2007). The A. L. I. C. E. Artificial Intelligence Foundation.

<https://alicebot.org/>

Atwell, E. (1999). *The language machine*. London: British Council.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In I. Kompatsiaris, J. Cave,

A. Satsiou, G. Carle, A. Passani, E. Kontopoulos, S. Diplaris, & D. McMillan (Eds.), *Internet Science* (pp. 377–392). Springer International Publishing. <https://doi.org/10.1007/978-3-319-70284-1_30>

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text*

*& Talk*, *34*, 545–567. <https://doi.org/10.1515/text-2014-0018>

Jain, M., Kumar, P., Kota, R., & Patel, S. N. (2018). Evaluating and informing the design of

chatbots. *Proceedings of the 2018 Designing Interactive Systems Conference*, 895–906. <https://doi.org/10.1145/3196709.3196735>

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People’s

Advisory Group. (2019). Can your phone be your therapist? Young people’s ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, *11*, 1178222619829083. <https://doi.org/10.1177/1178222619829083>

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. <https://doi.org/10.1145/3313831.3376175>

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial

intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, *5*, 2055207619871808. <https://doi.org/10.1177/2055207619871808>

Nimavat, K., & Champaneria, T. (2017). Chatbots: An overview types, architecture, tools and

future possibilities. *International Journal for Scientific Research & Development*, *5*(7), 1019–1026.

Nuruzzaman, M., & Hussain, O. K. (2018). A survey on chatbot implementation in customer

service industry through deep neural networks. *2018 IEEE 15th International Conference on E-Business Engineering (ICEBE)*, 54–61. <https://doi.org/10.1109/ICEBE.2018.00019>

Oxford University Press. (n.d.). Definition of chatbot. In *Lexico.com*. Retrieved March 22, 2021,

from <https://www.lexico.com/en/definition/chatbot>

Rahman, A. M., Mamun, A. A., & Islam, A. (2017). Programming challenges of chatbot: Current

and future prospective. *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 75–78. <https://doi.org/10.1109/R10-HTC.2017.8288910>

Schuetzler, R. M., Grimes, M., Giboney, J. S., & Buckman, J. (2014). Facilitating natural

conversational agent interactions: Lessons from a deception experiment. In *Proceedings of the International Conference on Information Systems* (ICIS’14; online). Retrieved from http://aisel.aisnet.org/icis2014/proceedings/HCI/9/

*Scopus—Document search*. (n.d.). Retrieved March 22, 2021, from

<https://www.scopus.com/search/form.uri?display=basic#basic>

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460.

<https://doi.org/10.1093/mind/LIX.236.433>

Wallace, R. (2003). *The elements of AIML style*. A.L.I.C.E. Artificial Intelligence Foundation,

Inc.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language

communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. <https://doi.org/10.1145/365153.365168>

**Uyen Bui** is a student in the Master of Science in Human Factors program at California State University, Long Beach (CSULB). She received her Bachelor of Arts in Psychology at CSULB. Her research interests currently lie in the areas of usability testing, accessibility, and user behavior surrounding passwords. Uyen can be reached at [uyen.bui@student.csulb.edu](mailto:uyen.bui@student.csulb.edu).