The conversational user interface: A brief history and futurology of chatbots

Mike McTear[1], Raymond Bond[1]
[1]School of Computing, Ulster University, Jordanstown, Newtownabbey, UK, BT37 0QB

Abstract
The short paper provides a brief overview of the history of conversational user interfaces and the recent re-emergence of the 'chatbot'. Chatbots and smart speakers use natural language processing and understanding ('AI') to converse with humans either using speech or typographical interactions. Human-computer dialogues can be ephemeral or they can be perennial depending on the purpose of the chatbot, whether it be ordering pizza (ephemeral) or a chatbot that coaches a user with mental health and wellbeing content (perhaps more perennial). This paper provides a rapid tour of the history of chatbots from ELIZA to smart speakers, and elicits a series of near future challenges related to conversational user experience, ethics and beyond.

Definitions, background and evolution of the chatbot
Conversational User Interfaces have emerged as a new way to interact with computers and smart devices that is ostensibly more intuitive and more natural than traditional user interfaces. At the most basic level, a conversational interface supports human-computer interaction on a turn-by-turn basis i.e. the human and the computer take turns in an interaction that may be either transactional or conversational. A transactional interaction is used to accomplish some task, such as booking a flight, making a purchase, or asking about football scores, while in a conversational interaction the human and the system engage in "chit-chat" for primarily social purposes. Where the conversation is text-based, the interaction is similar in form to a chat between two human users on a messaging app, except that in this case one of the participants in the interaction is an artificial entity. With voice-based conversations, the interaction is similar to a conversation between two human users on the telephone.

Conversational interfaces have become mainstream since the launch in 2011 of Apple's Siri, a personal assistant that supports spoken interactions on iPhones and other devices, followed by a proliferation of conversational systems that we can talk to: on smart speakers such as Amazon Echo and Google Home, channels such as Facebook Messenger or Telegram, and with social robots, such as Pepper from Softbank Robotics. However, the origins of conversational systems can be traced back to the text-based dialogue systems and chatbots of the 1960s, such as GUS, a dialogue system for booking flights, and ELIZA [1], a chatbot that simulated an interaction with a Rogerian therapist.

Conversational technologies have changed considerably over the past decade. Previously spoken dialogue systems were hand-crafted, with complex rules and algorithms being devised to control the flow of the dialogue (i.e. Dialogue Management), while for their commercial counterparts, known as Voice User Interfaces, a body of best-practice guidelines were accumulated. More recently, machine-learning algorithms, particularly deep neural networks, have dominated the field, made possible by the availability of massive amounts of data and computing power to train the systems. However, there is still much to be learned from examining

insights and approaches from the earlier symbolic approaches, particularly in more advanced systems where a deeper level of understanding and interaction is required. To date most of the focus in conversational technology has been on issues of design and implementation. At the most basic level interaction is controlled completely by the system and the user can only respond by clicking on pre-defined buttons. This type of interaction is typical of applications such as accessing the news, information about a product, or some simple coaching systems. Another more basic interaction is the one-shot queries typical of interactions with smart speakers, where the user asks a question or issues a command and the system responds. These systems require extensive speech recognition and natural language understanding capabilities to be able to process a wide range of user inputs that cannot be predicted in advance, but they do not have the ability to engage in a multi-turn conversation, except more recently the ability to handle subsequent related queries, known as follow-up questions. At the other end of the scale the interaction is open-ended and approximates natural conversations between humans. This type of interaction is typical of advanced chatbots, such as Google's Meena, Facebook's Blender, or Microsoft's XiaoIce. In between these two extremes there are systems that receive restricted input from the user in the form of written text or speech, where the system has to be able to interpret the user's input, decide on its response, and then generate a response. Thus systems differ in terms of what the user can do with the system i.e. what Norman [2] calls their affordances. They also differ in terms of the technological requirements for these affordances. It is important for designers and developers to be aware of these differences in order to be able to provide an optimal user experience and to make best use of the range of technologies and functionalities available. This is particularly important, as there have been questions raised recently about the usability of current conversational systems.

Conversational user experience and usability

In a report by the Nielsen Norman Group [3] on results from extensive usability testing of conversational systems - it was concluded that "usability testing finds that both voice-only and screen-based intelligent assistants work well only for very limited, simple queries that have fairly simple, short answers. Users have difficulty with anything else." While some of these issues probably arise from poor design, it is also the case that users may have unrealistic expectations of conversational systems or that they simply do not know what the systems can do and, more importantly, what they cannot do. For this reason it is important to develop a better understanding of how different types of conversational systems provide different levels of user experience and to appreciate the technological requirements and constraints involved in designing and developing these different types of system.

As time progresses and more people adopt smart speakers and chatbots, it is likely that human users may become more 'computerised' in their interaction with the computer and the computer will become more 'humanised'. In other words, the user's expectations may be appropriately calibrated to the machine's natural language capabilities. If this is the case, users will create a slightly different set of natural language heuristics for interacting with machines. In a sense, this is how people are already using smart speakers. Humans can learn to use a 'machine lingua' that is most effective when talking to a smart speaker, and even children may quickly learn that being more polite does not increase the likelihood of getting a better or more efficient response from a smart speaker - perhaps having negative unintended consequences by

promoting ‘command language’ towards the computer (as smart speakers are ‘politeness agnostic’). Human users can learn to simplify utterances that are comprised of fewer words to form clearer commands that are more explicit. In this sense, they improve their elocution akin to how we may adjust to converse with young children or people who speak English as a second or third language. Nevertheless, perhaps machines do not need to adapt to humans but humans need to adapt to machines. A compromise like this is probably best termed as semi-natural language communication. And, perhaps the ‘semi’ is more efficient for all actors involved. And of course, a secondary effect might be that human-human conversation may also become more ‘computerised’ depending on the frequency of human-computer conversations that take place in the real world, for example, supposedly a future world might be comprised of more human-computer conversations than human-human conversations (but hopefully not).

Recently within the HCI community various researchers have started to examine the issue of user experience with conversational systems and there have been a number of conferences, workshops and publications addressing this issue. In some cases interactions between humans and conversational systems are recorded and examined in detail. This work will become increasingly important as conversational systems become more prevalent, particularly in new areas of application such as education, healthcare, and well-being support. This area of ‘conversational UX’ is of particular importance since most usability scales, metrics and UX best practices [4] are based on research into conventional graphical user interfaces that use the WIMP model (windows, icons, menus and pointer), which is obviously very different to a natural language interface.

Are conversational user interfaces a step backwards, or forwards?
As already mentioned, the history of conversational systems links back to the 1960s with the creation of ELIZA but the idea of ‘chatbots’ has recently re-merged for a plethora of reasons. Moreover, in the history of computing, the ability of a computer to engage in a conversation has been integral to the concept of Artificial Intelligence, given that it is generally agreed that if a user cannot distinguish between engaging in a computer and a human conversation, then the computer is considered to be truly intelligent (Turing test [5]). And whilst we have still yet to convincingly pass this Turing test, chatbots are now being used as useful mediums for human-computer interactions. Others may see this as a step backwards in history as opposed to a step forward. Specifically, that the graphical user interface and the WIMP’s model emerged to replace character (command) user interfaces (e.g. MS-DOS) – so why step backwards? This is an interesting argument but chatbots are not perfectly analogous to character user interfaces. Moreover, character user interfaces have remerged in popularity in recent years, especially amongst technical software developers and are still the preferred interfaces in specific industries such as pathology laboratories due to the efficiency of shortcuts and commands elicited from the keyboard. User experience specialists may also consider the WIMP’s model as superior given that there is arguably less ambiguity as to what the user’s goal is when compared to a natural language interface which can be a minefield in terms of the number of misinterpretations and errors that a computer can make when determining user goals and intents from natural language (especially given that even humans make numerous errors when interpreting natural language).

The future

The history and future of chatbots is an interesting field that needs attention. There are many other futuristic aspects of chatbots that need to be considered. This includes the ethical aspects of chatbots, for example, current and future chatbots may deceive users into believing that they are interacting with a human, e.g. for fraudulent means. The 'computers as social actors' [6] theory indicates that humans already treat computers as social actors, so how much more so will chatbots be treated that have natural language channels and a convincing persona? And treating computers as social actors and as sentient beings might be problematic when considering the lack of empathy, emotional intelligence and true understanding that a computer has (in light of John Searle's Chinese room argument [7]). Philosophically, the more AI a chatbot has - the more questions arise regarding machine consciousness and mind. The distant future, might even abandon the use of chatbots given the amount of user effort required to type or use vocal chords to converse. Perhaps, the distant future is more akin to sci-fi with computational mind reading or more cognitive based interactions (e.g. brain computer interfaces etc.). Regardless, it seems that the era of chatbots is approaching and is challenging the conventional graphical paradigm of human-computer interaction. The level of user adoption and acceptance of chatbots will dictate the near future prevalence of conversational user interfaces.

References

[1] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM. 1966 Jan 1;9(1):36-45.

[2] Norman DA. Affordance, conventions, and design. interactions. 1999 May 1;6(3):38-43.

[3] Budiu, R. (2018) The User Experience of Chatbots. California: Nielsen Norman Group. Available from: https://www.nngroup.com/articles/chatbots/ [Accessed March 2019].

[4] Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M. Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. InProceedings of the 31st European Conference on Cognitive Ergonomics 2019 Sep 10 (pp. 207-214).

[5] Saygin AP, Cicekli I, Akman V. Turing test: 50 years later. Minds and machines. 2000 Nov 1;10(4):463-518.

[6] Reeves B, Nass CI. The media equation: How people treat computers, television, and new media like real people and places. Cambridge university press; 1996.

[7] Searle JR. The Chinese room revisited. Behavioral and brain sciences. 1982 Jun;5(2):345-8.