Shrey Kharbanda

4/11

Progression 2 Working Draft

Prof. Zaretsky

The Power of ChatBots: Conversational Agents in AI

Need to pick up your bouquet from the florist for your wife? Ask Siri to remind you. Need help understanding a technical question in your project? Ask ChatGPT to provide a deep explanation in simple words. In 1966, MIT computer scientist Joseph Weisenbaum invented the first chatbot, ELIZA, in the history of Artificial Intelligence. ELIZA was the first program that allowed "some kind of plausible conversation between humans and machines" (Jarow). Chatbots are computer programs that simulate dialogue with humans on the Internet by using Natural Language Processing (NLP)[1]and sentiment analysis to imitate human-like conversations.

Commented [NZ1]: Add a transition here before moving on to the history.

Commented [NZ2]: I might move the material about chatbots in general and then turn to the history of ELIZA.

ELIZA played a significant role in popularizing the concept of chatbots and virtual assistants. While ELIZA was intended to be a tool for exploring the relationship between humans and machines, users treated ELIZA as a confidant, disclosing personal information and intimate things to the chatbot. Its simple yet effective approach to conversation laid the groundwork for more sophisticated chatbots to come. For example, the chatbot Parry was developed by a computer scientist at Stanford's Psychiatry Department in 1972 to counteract the arguments of ELIZA and further provided more insight into the workings of chatbots and NLP in general (Zemčík).

While ELIZA and PARRY engaged in conversation with humans by rephrasing their words entered questions and acting like a "pretty small mirror" (Jarow), chatbot interactions have

been improved since then by training these conversational AI agents on databases consisting of over a billion words. 50 years later, the field of chatbots has been revolutionized by the introduction of powerful chatbots such as Amazon Alexa and ChatGPT, which stands for 'Generative Pre-trained Transformer,' that run on advanced deep learning techniques of unsupervised learning basis. An unsupervised learning basis allows chatbots to generate text by forming an analysis of patterns in a large database of words, without any human interference or instruction. Right before OpenAI introduced ChatGPT in the public domain, there was a lack of human belief in chatbots to provide accurate answers to most complex questions. Customer service chatbots turned out to be frustrating for most people and affected the brand image of most organizations in the market (Press). Launched in 2020 by OpenAI, ChatGPT averages "13 million unique visitors a day" and has surpassed the user base of the most popular social media app TikTok (Cerullo). With this drastic increase in chatbot usage with the new developments in the field of chatbots, it is important to consider how are these new developments in chatbots changing the relationship between humans and computers?

Commented [NZ3]: Good question!

To understand the extent to which human-computer interaction in society has changed, let us consider looking back at this relationship during the time before chatbots were invented. Before chatbots were invented, society relied on traditional methods of communication and customer service. This meant that customers would have to call or visit a business to get answers to their questions or receive any kind of assistance (Onlim). This process was generally time-consuming and frustrating, especially if the business was busy or understaffed. Additionally, before chatbots, businesses had to rely on human customer service representatives to handle customer inquiries. Overall, the interaction between humans and computers was limited and only there was only an indirect use of machines by the final consumers through the service of the

Commented [NZ4]: I would add more about your sense of what was "traditional" — has that always the same? Have modes of communication changed over time?

Commented [NZ5]: I would add more detail here and incorporate some references. Did we "interact" with computers or simply "use" them for different functions?

humans behind the customer service desk or in the form of scrolling, swiping, or button clicks. The human-computer interaction was not personal and hence was restricted. How did this limited interaction suddenly get affected by the invention of ChatGPT?

The human-computer interface was largely personalized with ChatGPT's introduction in society. In essence, ChatGPT "doesn't exactly know anything" (Shankland) since it's trained on multitudes of data and builds patterns based on what it is trained on by humans, what is being told to it by the users and what can be retrieved from similar past responses "to deliver more useful, better dialog" (Shankland). The personalization of contextual information comes from ChatGPT's emphasis on making its responses sound more "plausible" and "authoritative" along with a warning that "they might be entirely wrong" (Shankland). By altering the language style, human-like chatbots can provide greater satisfaction and trust among customers, leading to greater adoption of the chatbot (Jenneboer). This points to a connection between chatbots and human trust in computers.

When it comes to using chatbots in environments such as finance, healthcare, and other fields where the privacy of data is a concern, trust acts as a dominating factor in determining the popularity and acceptance of new technologies. Human intuition may lead us to believe that since chatbots can converse easily, they should be "anthropomorphized" (Przegalinska 788). This derives from the excessive social expectations humans have from computers and in turn, their trust level depends on how human-like is a computer. However, the process of anthropomorphization is not only "about the attribution of superficial human characteristics" but most importantly "a humanlike mind" (Przegalinska 788). People tend to trust human-like technology because of how powerful their competence is compared to human-human responses. In Przegalinska, a research study was conducted that factored chatbot's perceived ability to

Commented [6]: I would modify this so you don't have a question at the end of this paragraph [if you had one at the end of the previous paragraph].

Commented [NZ7R6]: You could also move into this point in the next paragraph - how this limited interaction was then affected by ChatGPT.

Commented [NZ8]: Try to use ACTIVE voice construction **[apply throughout essay].**

Commented [NZ9]: Add more here and try to be more precise. How exactly does this generate "human trust" and what does this trust mean/entail?

Commented [NZ10]: Does that necessarily increase trust?

Commented [NZ11]: Avoid generalizing [unless there is evidence of this from your source?].

provide assistance, the user's prior experience with chatbots, and the user's perception of the chatbot's ethical behavior to figure out how much trust governs the human-computer interactions. The results showed how even though the users knew that they were having a conversation with an artificial being, they kept on provoking the chatbot to say something human. As chatbots become more advanced and sophisticated, the dividing line between human and computer interactions has started to become blurred, making it more difficult to distinguish between automated and human interactions in the online world. Will computers replace human-human interactions by a majority one day in the next few years? Do innovations such as chatbots have the potential to disrupt our communication norms and practices, influencing the way we interact with one another and perceive technology?

Disruptive innovation, according to Jill Lepore of the New Yorker, is the gradual takeover of a particular innovation in society that grows progressively and looks at innovation as a competitive tool to transform societal norms and practices. Despite affecting our day-to-day lives, the disruption caused by these innovations is “predictive” (Lepore) and can be mitigated in advance. While innovations such as chatbots have the potential to disrupt our communication norms and practices, they cannot be substituted for human connection and understanding. Innovation, in the form of disruptive and progressive change, serves both as “a chronicle of the past” and as “a model for the future” (Lepore). Essentially, no one can ever know the future of disruptive change through innovations. These innovations would, therefore, predictably leave behind a profound impact on the way we interact with society and its elements and how it shapes our communication norms and practices. However, can this impact be profound enough to simulate real-life experiences?

Commented [NZ12]: In what way does this tell us something about trust?

Commented [NZ13]: These are good questions in general BUT see my earlier point about the structure of using questions at the end of so many paragraphs. *Also, you should EXPLORE these points through your analysis and development of your argument further in what follows.

Commented [NZ14]: Add a clearer transition into this point.

Commented [NZ15]: Always capitalize magazine or journal titles.

Commented [16]: Instead of having this question here, maybe start the next paragraph with a version of these themes - related to the simulation of real-life or what real -life experiences really mean.

In his essay "The Experience Machine," Robert Nozick explores how while technologies and artificial intelligence have the potential to simulate experiences, they cannot replace the depth and authenticity of real-world experiences. According to Nozick, the question of "what are we" is significant to the progress of the human mind and society. This essay proposes a thought-provoking philosophical concept of a machine that can help provide any human being with any experience they desire by enabling them to feel pleasure, success, and happiness, but the only condition for this experiment is that the participants physically cannot do anything in the real world to partake in those experiences. Innovations created by humans, for humans, are always going to be artificial and artificial innovations cannot account for making human life meaningful, but at the same time, they can enhance our understanding of the world and our place in it. Nozick suggests that real-world experiences offer more than just the experience itself. They provide a sense of authenticity and reality while innovations "limit[s] us to a man-made reality" (Nozick 44) such that there is no deep connection to the understanding of our own selves and our place in this world. This limited view of reality and authenticity can be attributed to the reinforcement of biases and inequalities in society through the process of innovation.

Surprisingly, since innovations are the consequences of human actions, humans themselves are responsible for reinforcing biases and inequalities in society. According to Paola Andrea Russo et al., people's self-perceived reputation influences their fairness towards humans and artificial intelligence individually. Innovation is generally meant to create new technologies, products, and services that aim to improve human lives. However, since these innovations are developed by humans of certain demographics or perspectives, they tend to shadow a limited perspective that may reinforce biases and inequalities in society. For instance, in the article "Could a Conversational AI Identify Offensive Language?" by da Silva et al., the authors argue

Commented [17]: ADD A CLEARER TRANSITION INTO THIS MATERIAL - WORK ON THE LINE OF ARGUMENT
[from conference: how profound is the effect of innovation. does it replace the "real" of experiences?]

Commented [18]: FROM CONFERENCE: is human-human better than human-computer, in that sense?

Commented [19]: maybe a way to think through trust, experience, and the real?

that AI can reinforce bias in society by using whatever it's trained on, which was originally fed into it by humans. Modern conversational AI models that detect offensive language have "acceptable accuracy," but for example, "tweets in African American English were considered abusive in higher rate than the Standard American English" (da Silva et al.). This shows how only certain groups that are often the target of abuse in real life can be easily discriminated against and segregated by such chatbots.

Moreover, the limited view of reality and authenticity can directly influence people's perceptions of AI and innovation in general. People may view AI as "fake" or "inauthentic" (Russo et al.), leading to distrust in the idea of innovation. This, in turn, reinforces biases and inequalities in society as people may be more likely to rely on their own perceptions and biases rather than accepting new technologies that could lead to progress This brings out a fair warning about the future of perceiving these biases. If chatbots are trained on biased datasets, will these biases continue to perpetuate in the field of innovation? And is relying on our own perceptions and biases the best way to reduce biases and inequalities within innovation in society?

Perhaps, there is no way to end this perpetuating cycle of reinforcement of biases and inequalities because chatbots and other conversational agents of AI provide humans with a very accessible platform to seek advice and information. Before this cycle of modern innovation, humans heavily relied on their own perceptions and biases to seek advice or information. However, innovation was finally introduced as a concept in society because humans were not trying to reinforce biases. They have, in fact, always had a tendency to seek out more interactions and information that validates their existing beliefs and ideas, or, in other words, the tendency for social validation. Profit-based chatbots and other innovations today have been trained by humans to maximize human validation so that they could enhance the user experience

Commented [20]: I would reflect on how you might want to elaborate further on the connections between bias, innovation, inequality, etc....

Commented [21]: Consider converting this into sentences/analytical claims....

Commented [22]: Perhaps you can take these concerns about innovation and bias --> and bring it to your conclusion in terms of broader implications.
--> we have to look to ourselves for really addressing issues like bias nad inequality?

Commented [23]: Try to rephrase.

and increase their profits. In essence, humans have been projecting onto societal reinforcements such as innovation to view this theory of innovation as a mirage of human validation and self-perception, or a confirmation of their existing beliefs and ideas. Hence, innovation acts, as a mirage for human validation, or a tool for human society to validate their own beliefs and is not really changing our existing ideas as much as reinforcing what we already believe.

Commented [NZ24]: Interesting!

Works Cited

Commented [NZ25]: Great sources!

Cerullo, Megan. "ChatGPT is Growing Faster than TikTok." Moneywatch, CBS News, 1 February 2023, 5:49 PM. https://www.cbsnews.com/news/chatgpt-growing-faster-than-tiktok/.

Chandra, Shalini, et al. "To Be or Not to Be . . . Human? Theorizing the Role of Human-Like

Competencies in Conversational Artificial Intelligence Agents." *Journal of Management Information Systems,* vol. 39, no. 4, 2022, pp. 969-1005, doi: 10.1080/07421222.2022.2127441.

Chowdhary, K.R. "Natural Language Processing." *Fundamentals of Artificial Intelligence, Springer, New Delhi*, 2020, https://doi.org/10.1007/978-81-322-3972-7_19.

da Silva, Daniela A., et al. "Could a Conversational AI Identify Offensive Language?" Information, vol. 12, no. 10, 2021, pp. 418. ProQuest, http://proxy.library.nyu.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Fcould-conversational-ai-identify-offensive%2Fdocview%2F2584397915%2Fse-2, doi:https://doi.org/10.3390/info12100418.

Jarow, Oshan. "How the first chatbot predicted the dangers of AI more than 50 years ago." *Vox*, 5 Mar. 2023, https://www.vox.com/22314501/chatbot-eliza-predicted-dangers-artificial-intelligence-ai.

Jenneboer, Liss, et al. "The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review." *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 1, Jan. 2022, pp. 212–29. Crossref, https://doi.org/10.3390/jtaer17010011.

Lepore, Jill. "The Disruption Machine: What the Gospel of Innovation Gets Wrong." *The New Yorker*, 23 June 2014.

Nozick, Robert. "The Experience Machine." *Anarchy, State, and Utopia, Basic Books*, 1974, pp. 42-45.

Onlim. "The History of Chatbots: From ELIZA to Alexa." *Onlim*, 2022, https://onlim.com/en/the-history-of-chatbots/. Accessed 11 Apr. 2023.

Press, Gil. "One Negative Chatbot Experience Drives Away 30% Of Customers." *Forbes*, 1 Feb. 2023, www.forbes.com/sites/gilpress/2023/02/01/one-negative-chatbot-experience-drives-away-30-of-customers/?sh=147f4cf82510.

Przegalinska, Aleksandra, et al. "In bot we trust: A new methodology of chatbot performance measures." *Business Horizons*, vol. 62, no. 6, Nov.-Dec. 2019, pp. 785-797, doi: 10.1016/j.bushor.2019.08.005.

Russo, Paola Andrea, et al. "How self-perceived reputation affects fairness towards humans and artificial intelligence." *Computers in Human Behavior*, vol. 124, 2021, article no. 106920, ISSN 0747-5632, doi: 10.1016/j.chb.2021.106920.

Shankland, Stephen. "Why we're all obsessed with the mind-blowing ChatGPT AI chatbot." *CNET*, 19 Feb. 2023, https://www.cnet.com/tech/computing/why-were-all-obsessed-with-the-mind-blowing-chatgpt-ai-chatbot/.

Wilson, H. James, et al. "What Do We Do About the Biases in AI?" *Harvard Business Review*, 15 Oct. 2019, https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai.

Zemčík, Tomáš. "A Brief History of Chatbots." *DEStech Transactions on Computer Science and Engineering*, 2019, doi: 10.12783/dtcse/aicae2019/31439.

Still exploring 2 body paragraph notions:

1. *Is putting a price tag on ChatGPT similar to putting a price tag on human interaction?*
2. *What is the right balance between human and machine?*

[1] See Chowdhary for more information on Natural Language Processing.