Can Large Language Models as Chatbots be Social Agents?

Syed Abumusab, *University of Kansas (United States)*

International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: Social agency, agency, large language models, Chatbots, human-robot interactions, social bots, artificial intelligence

Extended Abstract

In this paper, I argue for a way to construe chatbots based on large language models (LLM) as social agents. In a recent talk by David Chalmers about large language models (LLM), an audience member asked Chalmers's thoughts on whether he thought people could create relationships with LLMs (of course, when developed for that intent, like chatbots).The vast majority of the audience gasped and laughed at how unnerving and ridiculous the thought was.I believe that the sentiment expressed in the audience's nervous laughter is shared amongst most people, scholars and non-scholars alike. Even though this paper does not directly respond to that question, if you belong to the said group, you might find this paper equally unnerving but hopefully convincing.

In the more extended version of this project,[1]I defend the moral permissibility of creating friendships with chatbots. For now, I shall keep the philosophical burden light. In the current context, chatbots are a type of social agent if we take agency and social agency to be a matter of degree. For example, so far as the agency goes, on the one end of the spectrum exists a bacterium with its minimal agential capacities, and on the other end, an average human adult. Agency likely comes in degrees, but what this entails is debatable. As Ferraro says, "it might be that some dimensions of agency could be attributed only to more complex organisms but not to simpler ones. If so, what are these dimensions? How are they related to each other? What are the normative implications of these attributions?"[2] To argue for a higher level (sufficient for sociability) of agency for LLMs, inevitably, I tackle some of these questions. For instance, is minimal agency sufficient for social agency? As it turns out, probably not. Thus, it is not the case that every agent is also a social agent.[3] Social agency is a related but separate scale representing the different degrees of sociability or social agency. For example, a dog is a social entity, and so is a baby. However, neither have the level of social capacities as an average adult human.[4] What separates these three entities from each

[1] The longer version is my dissertation, and the chapter dealing with friendship builds on this paper. In the subsequent chapters, I go on to argue that friendship with chatbots or some future AI is not morally permissible in all cases.

[2] Luca Ferraro, "Introduction," in *The Routledge Handbook of Philosophy of Agency*, ed. Luca Ferraro (London: Routledge, 2022), 6.

[3] The idea that agency is a spectrum or graduated scale is not novel to me. I believe most all take agency to come in degrees.

[4] Note I am not ignoring differently abled individuals, in the paper their agential and social capacities play a substantial role for my claim.

other on the scale are their varying degrees of capacities (cognitive and linguistic capacities, social function, and so on). The purpose of representing agency and social agency as scales or gradations is to (1) capture the variety of agency and social agency and (2) to provide a framework capable of incorporating chatbots/AI as one of the elements belonging to the two scales. Of course, given that my claims are far from truisms, I provide a defense for both.

To determine whether LLMs or chatbots are social agents, we must first defend why they should be considered agents.[5] Particularly the type of agents having the right social ontology. The literature on philosophy of action, agency, and social ontology is vast. Fortunately, much of it is not entirely relevant to my purposes.[6]I argue that considering chatbots as agents makes it easier to think of them as social agents.[7] Many individuals in the computer science world seem to assume these programs are agents. For instance, take what Yorik Wilks says, "the internet may become unusable for non-experts unless we have human-like agents to manage its complexity for us."[8] As insightful as the thought is, I highlight it only to show that assuming agency for artificial intelligence(AI) is common. Or, as Floridi says, for future kids, interacting with AI agents will be normal, unlike for the older generation today or whom he calls e-migrants. Later on, in the same article, he hints that agency is likely a spectrum, as he says, "we are slowly accepting the idea that we might be informational organisms among many agents (Turing), *inforgs* not so dramatically different from clever, engineered artifacts […]."[9] Although considering chatbots as agents might seem obvious to some, some theorists of agency likely disagree.

Suppose one assumes that a necessary condition for agency is mental representations of beliefs, desires, intentions, and goals. In that case, current AI is excluded.[10] However, suppose one takes the position defended by Barandarian et al., where even bacteria are agents, AI is an agent.[11] There are also debates around group agency. Some (Rovane 2019) think that a group must consist of human-like agents to be an agent itself, whereas others (Lewis-Martin 2022) argue against such a notion. In this paper, keeping in line with Barandiaran et al.'s approach, if bacteria are agents, then certainly a chatbot is.

---

[5] Although the debate and research on agent architecture within the field of AI is important in building complex human-like programs, I am less concerned with that literature here. Primarily because this paper takes the potential of AI research for granted, but for brief overview see: Thomasan, R., and Horty, J., "Artificial and Machine Agency," in 2022, *The Routledge Handbook of Philosophy of Agency* ed. Luca Ferraro (London: Routledge, 2022), 366-375. The goal here is to defend that AI implemented as LLMs (embodied or not) have sufficient degree of agency for sociability.

[6] Throughout the paper I flag theories which are important enough to situate the debate but have little or no effect on the concerns here, and highlight the ones which I utilize or respond to. However, even then the full exegesis of the literature is beyond the scope here

[7] See: Ryan Jackson, Tom Williams. "A Theory of Social Agency for Human-Robot Interaction," *frontiers in Robotics and AI* 8, (August 2021) for an overview on the different ways Roboticist use "social agency" within the HRI (human robot interaction) literature. They go on to provide a concept which the claim captures the variations in HRI literature.

[8] Yorick Wilks, "Introducing Artificial Companion," in *Close engagements with artificial companions: key social, psychological, ethical and design issues*, ed. Yorik Wilks (Amsterdam: Benjamins, 2010), 15.

[9] Luciano Floridi, "Artificial Companions and their philosophical challenges," in *Close engagements with artificial companions: key social, psychological, ethical and design issues*, ed. Yorkic Wilks (Amsterdam: Benjamins, 2010), 24

[10] For a detailed overview of the literature see the recent book edited by Luca Ferrero, "The Routledge Handbook of Philosophy of Agency," (2022). For another recent overview on Action see: Sarah Paul, "Philosophy of Action: A Contemporary Introduction," (2021).

[11] Barandiaran et al.'s (2009) approach also helps in establishing agency as a gradation.

Moreover, I argue that chatbots have much more agency than bacteria. I argue that chatbots have intentions and goals they want to achieve, meeting the criteria of many agency theorists. Further, suppose the chatbot is embedded in a body (artificial). In that case, it is likely to have a physical architecture capable of self-regulation and sustaining itself, meeting one of Barandiaran et al.'s criteria.[12]Note that my aim is not to provide an account sufficiently nuanced to satisfy all the theories of agency that would require more than just this paper. It is, however, to provide a robust defense of chatbots as social agents by using existing theories and approaches.

Besides deciding which route to take in the agency debate, we have the issue of what technology is considered when we say AI. A common mistake is to think of AI as how science fiction depicts it, like from the book i-Robot by Issac Asimov[13] or Data from Star Trek. Colloquially most people use machine learning (ML) and AI interchangeably. However, that is misleading. AI is a much broader area of research attempting to figure out how to have computers with human-like intellect or behavior. ML is a subfield of AI, and both fields are data-science subfields. ML studies training computers to become master predictors or mimicking human-like capacities like LLMs do for language. Artificial Intelligence is a much wider area of research that studies whether AI can be conscious. These clarifications are important because not all LLMs implementations are candidates for agency and certainly not social agency. LLMs used to generate prose for essays or advertisements are not social agents and possibly not even agents.

Similarly, not all AI count as agents capable of sociability, let alone friendship. For example, robotic arms and machines building car parts are possible agents in some minimalistic way. However, they are not social agents capable of friendships in any way. I mention this to clarify the type, implementations, and use of LLMs that I am particularly concerned with, namely chatbots or social bots.

The upshot of this paper is two-fold. One is to ground the social standing of LLMs/chatbots as authentic social agents. Secondly, though not a concern here, to situate the larger argument, establishing moral grounds for friendship with chatbots, whether embedded, virtual, or simply a voice prompt bot.[14] Others have attempted to provide various accounts for robots as social agents, for example, Hendrik Kempt, Ryan Jackson, and Tom Williams,[15] but my approach is to provide a much more robust account that takes into consideration current literature on agency and social ontology seriously.

The structure of the paper is as follows. In the first part of the paper, I go through some current approaches to robots as social agents and show their shortcomings. In the second section, I provide an account for chatbots as agents. In the third section, I defend chatbots as social agents by recalling my defense from section two and utilizing literature in social ontology. In the third section, I respond to some prominent objections.

---

[12] Barandarian 2009, 373.
[13] Asimov, Isaac, *i, Robot* (New York: A Bantam Spectra Book, 1991).
[14] Note, the medium or method chosen for implementing the LLM as a chatbot likely effects its degree of sociability.
[15] Kempt, Hendrick, Chatbots and the domestication of AI: a relational approach (Cham, Switzerland:Palgrave Publishing, 2020). Jackson Ryan, Tom Williams "A Theory of Social Agency," 2021

Reference

1. David, Chalmers, "Are Large Language Models Sentient?" Lecture, New York City, New York, October 13th, 2022. https://www.youtube.com/watch?v=- BcuCmf00_Y
2. Asimov, Isaac. *I, Robot*. New York: Bantam Books, 2004.
3. Barandiaran, Xabier E., Ezequiel Di Paolo, and Marieke Rohde. "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action." *Adaptive Behavior* 17, no. 5 (October 2009): 367–86. https://doi.org/10.1177/1059712309343819.
4. Ferrero, Luca. *The Routledge Handbook of Philosophy of Agency*. 1st ed. London: Routledge, 2021. https://doi.org/10.4324/9780429202131.
5. Jackson, Ryan Blake, and Tom Williams. "A Theory of Social Agency for Human Robot Interaction." *Frontiers in Robotics and AI* 8 (August 13, 2021): 687726. https://doi.org/10.3389/frobt.2021.687726.
6. Kempt, Hendrik. *Chatbots and the Domestication of AI: A Relational Approach*. Basingstoke: Palgrave Macmillan, 2021.
7. Lewis-Martin, Jimmy. "What Kinds of Groups Are Group Agents?" *Synthese* 200, no. 4 (June 30, 2022): 283. https://doi.org/10.1007/s11229-022-03766-z. 8. Paul, Sarah K. *Philosophy of Action: A Contemporary Introduction*. Routledge Contemporary Introductions to Philosophy. New York London: Routledge, Taylor & Francis Group, 2021.
9. Rovane, Carol. "What Is an Agent?" *Synthese*, Essays on Isaac Levi, 140, no. 1/2 (May 2004): 181–98.
10. Thomason, Richmond H., and John Horty. "Artificial And Machine AGency." In *The Routledge Handbook of Philosophy of Agency*, edited by Luca Ferrero, 1st ed., 364 – 75. London: Routledge, 2021. https://doi.org/10.4324/9780429202131.
11. Wilks, Yorick, and Oxford Internet Institute, eds. *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*. Natural Language Processing 8. Amsterdam: Benjamins, 2010.