



Introduction to the Special Issue on “Artificial Speakers - Philosophical Questions and Implications”

Hendrik Kempt¹ · Jacqueline Bellon² · Sebastian Nähr-Wagener³

Published online: 2 December 2021
© The Author(s) 2021

1 On the Philosophical Relevance of Artificial Speakers

The often investigated future of human–machine relationships, ranging from cooperation partners at work, to social bots in care homes and personal assistants at home, is based on an often implicit technological requirement of robotics and artificial intelligence (AI): the ability of those machines to communicate with us in a form familiar and comfortable to us.

Thus, those machines will have to learn how to communicate, either through intuitively understandable signs, text, or audio. This issue deals mostly with the latter. We may assume that machines ‘speak’, or rather, have to become speakers. However, this simple statement is laden with philosophical notions both about the conditions of what it means to ‘speak’, and thus become a speaker, and whether machines will ever be able to achieve these conditions. One may argue that the key challenge is to teach machines to use language according to its rules. However, what distinguishes ‘natural’ speakers from artificial ones? There is more to human language use than mere linguistic rule following: What do humans do additionally to follow the rules of language—and the requirements of communication in general—that machines are not currently and may not be able to? Could machines perform speech acts? If yes, which ones can be performed without the presence of underlying conditions such as human intentionality? Should machines count as agents, or should we reconstruct their actions as “quasi-action”? What could be functional equivalents to speech, and where exactly do they

✉ Hendrik Kempt
hendrik.kempt@humtec.rwth-aachen.de

Jacqueline Bellon
j.bellon@posteo.de

Sebastian Nähr-Wagener
sebastian.naehr-wagener@kit.edu

¹ Institute for Applied Ethics, RWTH Aachen, Aachen, Germany

² FoKoS, University of Siegen, Siegen, Germany

³ Institute of Technology Futures (ITZ), Karlsruher Institute of Technology (KIT), Karlsruhe, Germany

differ from human speech? Can we trust a machine's testimony? Do we lie to a machine as we lie to human beings?

These questions are of fundamental philosophical interest, as they cover fields of social philosophy, philosophy of language, philosophy of mind, philosophy of technology, and ethics. The motivation to concentrate scholarly efforts on the issue of artificial speakers sparkles from technological progress and an otherwise conceptually underdeveloped field. Philosophical inquiry in this field is, to a degree, dependent on technological development, and some technological aims are informed by the conceptualizations of language provided by philosophy. Without knowing the latest trends of creating AI using human language, philosophical arguments about the possibility of machines to enter moral discourses are without footing. Yet, without understanding properly what language is, how our relationship to language shapes our understanding of the world, and how we created rules to use language in certain ways, the development of artificial speakers may lead to unwanted, and undesirable, effects.

The technological field contributing to this progress is usually referred to as natural language processing (NLP). This field, originally developing from teaching machines to do translations from one language to another, has recently experienced a huge leap forward with the establishment of surprisingly broad-functioning language models potentially the size of every word written on the internet (OpenAI's GPT-3, Google's BERT and others).

It has become possible to train models to use language with an immense spectrum of abilities. These abilities can be understood as a step towards creating genuinely new philosophical issues.

The determination of whether something is a genuinely new problem, or an old problem in a new light, is not an easy task. Especially in the field of AI ethics, some previously known philosophical conundrums problems are merely reproduced on a bigger technological scale. However, we hold that the issue of artificial speakers, i.e., machines with the ability to hold a phenomenologically rich conversation with human beings, presents a convincing example of technology that brings about new problems.

Even if one remains skeptic that the current particular technology used to program speaking machines will satisfy philosophical requirements, at least our ability to refine the conditions of what may count as an agent will improve. Thus, philosophical progress is possible in concerning ourselves with the phenomenon of speaking machines.

And while many different areas of philosophy can benefit from questioning our assumptions about artificial language use, we would like to focus on three aspects of the nexus between ethics and the challenges of NLP: the phenomenology of language use, the ethics of imitating human beings, and the communicative foundation of human relationships.

2 Phenomenology of Speakers

The development of machines that can communicate with us in a certain way has some immediate phenomenological implications. Having machines create and communicate information causes epistemic uncertainty on the one hand, and philosophical opportunities on the other.

The epistemic insecurity emerges, for example, where we cannot be certain anymore whether we are dealing with a human being or a machine in an interaction. This genuinely new question is not only relevant for social interactions, but also concerning what is called 'fake news', catfishing and other misinformation attempts. Keeping one's own communication machine-free becomes an epistemically and socially costly burden, and will most likely not be afforded by many. Thus, the collection of our communicative interactions will probably eventually be enriched by the ubiquitous presence of artificial speakers. This does not require artificial speakers of extraordinary levels of complexity to be achieved, but merely the increased cost of trying to distinguish them.

Philosophically speaking, then, distinctions between what it is to speak may become less human-centric and will be extended towards artificial speakers. While in some philosophical discourse theories agency has been understood as independent of consciousness, some speech act theories may have different implications, such as the assumption that to converse is generally to speak intentionally (and that intentionality requires consciousness). Yet, it appears more difficult to justify a concept of agency requiring mental processes if artificial speakers can behave to some extent in the exact same way as a human without possessing features necessary for agency. In turn, this offers possibilities to reframe and rethink psychological positions such as behaviorism and has implications for the theory of mind as well.

3 The Ethics of Imitating Human Beings

Creating artificial speakers is sometimes guided by the aim to imitate natural speakers, i.e., human beings. However, the choice of which human features are being imitated is always a normative choice requiring ethical justification (Kempt, 2020). While the specific choice of imitation can be motivated by other reasons, the intended wide-spread use of specific human traits in a generally subservient role, e.g., in customer service contexts, can reproduce and sediment problematic stereotypes. Consider for example how most artificial speakers are by default female-gendered, suggesting the subservient role women ought to play (UNESCO, 2019), a similar issue can be found in the names of most chatbots, who most often carry female-read names. Imitating human features may also lead to the activation of a range of unjustified expectations in human interactants. Sometimes, this is desired by a system's design, sometimes it is unintentional. In both scenarios, it can be beneficial, but also harmful to the human interaction partner.

Moreover, the purpose and contents of what artificial speakers can and will say are to be reckoned with. As the infamous and most relevant paper on statistical parrots has shown and problematized on several key elements (Bender et al., 2021), the way the data is sourced, processed, and reproduced in these large linguistic models is not free of controversy. From omitting key terminology for subcultural groups, concerns of linguistic mainstreaming of artificial speakers become more prevalent: who decides what and in which way artificial speakers are allowed to speak? The negative experiences Microsoft made with their Twitter bot Tay has been an instructive example (Wolf et al., 2017) of the need for limited speaking abilities (to avoid

insults and other offensive speech), while at the same time exposing the core issue the current machine-learned NLP approach faces: a chatbot behaves only as (morally) good as the data it is trained on. However, human communication is an often biased, discriminatory, and otherwise morally dubious source of such data, and available data does not always represent ‘real-life’ analogue world communication. Thus, a main way to mitigate risks of unwanted outcomes, like discrimination, is to curate the sources, contributing to a moralized, potentially human-made bias in the data once more. Though even with such presumably curated data, the lack of common sense of such machines, usually one of the more common arguments against considering artificial speakers as anything but tools, will keep them vulnerable to exploitation and, at worst, turning them into harmful tools.

With the introduction of phenomenologically rich artificial speakers, we may also assume, that the tendency for technosolutionism for social issues is not shrinking. From medical and care-perspectives, for example, this may be perceived or framed as progress. However, while we may provide higher quality care for those in need, we may also contribute to increased distance between humans, disregard their need for empathic communication, and undermining their autonomy. Some of these issues have become center of philosophical and ethical attention, with philosophers actively engaging with the technology and its ethical concerns from creation to implementation. Whether philosophical and ethical considerations will have a lasting impact on how artificial speakers are conceptualized in the future, and what kind of human traits, moral code, and linguistic variety they are allowed to display, remains to be seen. However, philosophy can and should help make informed decision before and while these systems are being designed and integrated.

4 Human Relationships with Machines

Once certain technologies become not only sufficiently ‘social’, but socially accepted, they will create a slew of new ethical issues—by extending not only to our technological, but our social interactions (Bellon et al., 2021). The better machines become at simulating and stimulating social interaction, and the more widespread their use will be, the more it stands to assume that philosophical arguments against the possibility of ‘genuine’ human–machine relationships may sound hollow to those in these kinds of relationships. Issues of human–machine friendships, the demand for moral patiency and moral consideration of those artificial speakers will be supported by the lived experiences of people interacting with machines they grow close with. Whether those perceptions should be philosophically validated, by incorporating certain social machines as proper moral patients, or rejected, by insisting on the fundamental differences between machines and humans, is an open question of research and subject of ongoing controversy.

While a wide range of prominent philosophers have taken up tackling the issues arising from those growing relationships and their problematic social ontology, we ought to remember that they are usually based on the fact that these machines will become, or are, speakers. Using the same speech as humans do appears to be the quickest way to warrant philosophical debate about their moral

status. This, presumably, lies in our construction of relationships partly based on speech, our willingness to suspend disbelief about our communication partners, and the tendency to project onto those speakers. Yet, empirical-psychological and conceptual-philosophical research are catching up on some of the newest developments. For example, the evolving thinking in the concept of friendships, which for the history of ethics had a dimension of physical presence and interaction (e.g., Danaher, 2019), may begin to account for those that do not have a physical presence in a friend's life, including 'unembodied' speaking machines.

5 A Research Agenda

The technological creation, ontological classification, and social integration of speaking machines as artificial speakers is an urgent matter for philosophers to analyze. Insights of this debate can be understood as prerequisite for other debates surrounding our shared future with machines. We can delineate norms for engineering purposes to avoid recreating harmful stereotypes through artificial speakers, and advise lawmakers in their challenges to regulate the use of artificial speakers to shield people from exploitation and other forms of harm. The presence and improvement of artificial speakers to become social agents provides genuinely new philosophical issues and is spreading into other fields of humanity at the same time. Thus, the question of when speaking machines actually become artificial speakers describes more of a research agenda rather than a single research question, and we recommend treating it as such.

Lastly, the developing field of AI ethics stands in a special relationship with the ethics of artificial speakers. As stated above, some debates in the general field of AI ethics are based on the assumption that machines may speak. In turn, the general principles discussed in AI ethics, ranging from explainability and transparency to data collection and privacy, all ought to feature in reflections on artificial speakers as well.

However, it is important to note that artificial speakers, with their rich phenomenology of human speech, their utility in both ethically problematic and unproblematic areas, and their future sociality with humans, will provide their own philosophical challenges. This Special Issue provides an opportunity to delve into this rich and exciting field of research, with many important questions being asked and, as we hope, provided with some first answers.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission

directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bellon, J., Eyssel, F., Gransche, B., Nähr-Wagener, S., & Wullenkord, R. (2021). *Theorie und praxis soziosensitiver und sozioaktiver systeme*. Springer.
- Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT 21). Association for Computing Machinery, pp. 610–623.
- Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1), 5–24.
- Kempt, H. (2020). *Chatbots and the domestication of AI*. Palgrave Macmillan.
- UNESCO (2019) I'd blush if I could: closing gender divides in digital skills through education. <https://unesdoc.unesco.org/ark:/48223/pf0000367416>. Accessed 15 Nov 2021
- Wolf, M. J., Miller, K. W., & Grodzinsky, F. S. (2017). Why we should have seen that coming: Comments on Microsoft's Tay "experiment", and wider implications. *The ORBIT Journal*, 1(2), 1–12.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.