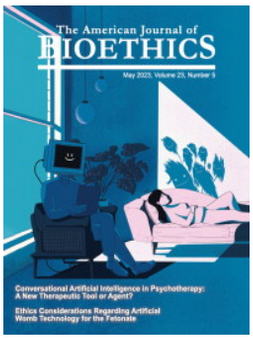


Taylor & Francis
Taylor & Francis Group

The American Journal of Bioethics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uajb20

# Therapeutic Conversational Artificial Intelligence and the Acquisition of Self-understanding

J. P. Grodniewicz & Mateusz Hohol

**To cite this article:** J. P. Grodniewicz & Mateusz Hohol (2023) Therapeutic Conversational Artificial Intelligence and the Acquisition of Self-understanding, The American Journal of Bioethics, 23:5, 59-61, DOI: 10.1080/15265161.2023.2191021

**To link to this article:** https://doi.org/10.1080/15265161.2023.2191021

THE AMERICAN JOURNAL OF BIOETHICS
2023, VOL. 23, NO. 5, 59–61
https://doi.org/10.1080/15265161.2023.2191021

Taylor & Francis
Taylor & Francis Group

OPEN PEER COMMENTARIES



# Therapeutic Conversational Artificial Intelligence and the Acquisition of Self-understanding

J. P. Grodniewicz and Mateusz Hohol

Jagiellonian University

In their thought-provoking article, Sedlakova and Trachsel (2023) defend the view that the status—both epistemic and ethical—of Conversational Artificial Intelligence (CAI) used in psychotherapy is complicated. While therapeutic CAI seems to be more than a mere tool implementing particular therapeutic techniques, it falls short of being a "digital therapist." One of the main arguments supporting the latter claim is that even though "the interaction with CAI happens in the course of conversation … the conversation is profoundly different from a conversation with a human therapist" (Sedlakova and Trachsel 2023, 8). In particular, unlike a human therapist, CAI cannot help its users gain new insight and self-understanding

CONTACT J. P. Grodniewicz j.grodniewicz@gmail.com Jagiellonian University, Copernicus Center for Interdisciplinary Studies, Krakow, Poland.

(Sedlakova and Trachsel 2023). We agree that currently available therapeutic CAI cannot be considered a "digital therapist," however, we think that the issue surrounding the acquisition of new self-understanding in the interaction with therapeutic CAI is more complicated than Sedlakova and Trachsel suggest.

The type of self-understanding one can acquire during psychotherapy proves difficult to characterize. As reported by Hill et al. (2007), while trying to formulate its relatively uncontroversial definition, a group of authors representing different therapeutic traditions managed to agree only on identifying it with "a conscious meaning shift involving new connections (i.e., 'this relates to that' or some sense of causality)" (442). A more elaborate definition has been offered by Lacewing (2014), who suggests that therapeutic self-understanding "involves grasping the *connections* between one's emotions, motivations, thoughts, and behavior, past and present, including one's interpretations of and relations with others" (154–5). Even though Lacewing focuses on the context of psychodynamic psychotherapy, we think that this definition is sufficiently neutral to, at least initially, guide our thinking about therapeutic self-understanding in general. Moreover, if we accept that Lacewing's definition is on the right track, we realize that it is only natural to think about self-understanding as a kind of what contemporary epistemologists call *objectual understanding*, i.e., a kind of understanding one has about a given subject matter in virtue of possessing a set of information about this subject matter and grasping connections between them (see, e.g., Kvanvig 2003; Grimm 2021). In this sense, we speak about someone understanding American history, molecular biology, or the origins of abstract expressionism. Following Lacewing's definition, we suggest that in the case of self-understanding, the subject matter is oneself, and the relevant set of information contains (but does not have to be limited to) information about one's emotions, motivations, thoughts, and behaviors.[1]

Moreover, just as objectual understanding is taken by many to be irreducible to knowledge (cf. Hannon 2021), self-understanding seems to be irreducible to self-knowledge. Again, as pointed out by Lacewing, it involves not only knowing a set of facts about oneself, but also "grasping" how they relate to each other. This, however, has important consequences for the problem of acquisition of self-understanding. While knowledge can be transmitted *via* testimony, most epistemologists assume that understanding cannot. We can pass true information on a given topic between each other but grasping the relationship between them is something that everyone has to do for themselves (see, e.g., Zagzebski 2008; Hills 2009). Sedlakova and Trachsel (2023) recognize the asymmetry between the facilitation of the acquisition of new self-knowledge and self-understanding but seem to assume that it is something specific to CAI: "[i]n terms of (self-) knowledge acquisition, CAI can provide novel information and data from the third-person perspective" (10) but "CAI cannot offer authentic facilitation of new self-understanding" (11). We disagree. Facilitating self-understanding is—both for human therapists and therapeutic CAI—simply much more difficult than providing someone with bits of knowledge about themselves. It is a matter of debate whether, and how much, CAI falls behind a human therapist in this respect.

If we cannot simply pass the "grasping" of a given subject domain to another person, how can we help them acquire understanding? Emma Gordon (2017) suggests that it can be done by facilitating: (i) the acquisition of new true beliefs; (ii) the rejection of false beliefs; (iii) the grasping of new connections (and rejecting of mistaken connections); (iv) overcoming blocks to grasping; and (v) the acquisition or enhancement of abilities linked to grasping. Arguably, at least to some degree, all these things can be done by a therapeutic CAI. For example, by implementing techniques of Cognitive Behavioral Therapy, CAI can provide its users with a list of cognitive distortions and encourage them to examine their beliefs in this light. Moreover, CAI can interwind the cognitive work with, e.g., emotion regulation or mindfulness practices, which can enhance users' abilities and put them in a better position for grasping. For many, this might be precisely what they needed to overcome existing blocks and grasp new connections between how they feel, think, and behave. Obviously, grasping is something that users have to ultimately do for themselves, but it is no different in the case of working with a human therapist.

But maybe the main difference between the acquisition of self-understanding in conversation with a human therapist and in conversation with CAI is that the therapist understands their clients/patients while the CAI does not. Sedlakova and Trachsel seem to suggest something along these lines when they say that users interacting with CAI should not expect

[1]The fact that in the course of psychotherapy or counseling one acquires objectual understanding has been already suggested by Gordon (2017). However, she does not discuss the nature or content of self-understanding, focusing more specifically on the objectual understanding of the origins and development of one's emotional difficulties (304).

"having a complex conversation … in which they are understood and can gain new insight" (2023, 10). Here again, we suggest caution. Firstly, even if a therapist has a certain (however partial) understanding of their client/patient, we have just argued that it cannot be directly transmitted to constitute the client's/patient's new self-understanding. The second, more important problem concerns authority and autonomy. Linda Zagzebski famously points out that "understanding cannot be given to another person at all except in the indirect sense that a good teacher can sometimes recreate the conditions that produce understanding in hopes that the student will acquire it also" (Zagzebski 2008, 145–46). But there is a crucial difference between a good teacher and a good therapist. A teacher is presumed to understand what they teach much better than the student, which makes the teacher the sole expert and authority on the subject matter. Approaching a client's/patient's therapeutic self-understanding in the same fashion would most likely violate their autonomy and constitute a case of epistemic injustice (Crichton, Carel, and Kidd 2017).

To sum up, we do not claim that there are no important differences between therapeutic work done with another human and interacting with CAI. However, our focus was the possibility of acquiring new self-knowledge and self-understanding. We argued that interaction with therapeutic CAI could, and often will, result not only in the acquisition of new knowledge about oneself but in genuine remodeling and transformation of one's self-understanding. In this respect, CAI turns out to be even less tool-like and even more therapist-like than Sedlakova and Trachsel suggest. Moreover, if—as Sedlakova and Trachsel argue (2023)—gaining new self-understanding is necessary for therapeutic change, we then do not have a reason to assume that such a change is impossible to achieve in the interaction with therapeutic CAI. All this has to be taken into consideration in future attempts to characterize both the ethical status of therapeutic CAI and normative requirements guiding our development and use of this promising technology.

## FUNDING

JPG was supported by a grant from the Priority Research Area 'Society of the Future' under the Strategic Programme 'Excellence Initiative' at Jagiellonian University. MH was supported by the National Science Centre, Poland (grant number: 2021/43/B/HS1/02868).

## ORCID

J. P. Grodniewicz http://orcid.org/0000-0001-7788-4236
Mateusz Hohol http://orcid.org/0000-0003-0422-5488

## REFERENCES

Crichton, P., H. Carel, and I. J. Kidd. 2017. Epistemic injustice in psychiatry. *BJPsych Bulletin* 41 (2):65–70. doi:10.1192/pb.bp.115.050682.

Gordon, E. C. 2017. Social epistemology and the acquisition of understanding. In *Explaining understanding: New perspectives from epistemology and philosophy of science*, ed. S. Grimm, C. Baumberger, and S. Ammon, 293–317. New York, NY: Routledge.

Grimm, S. 2021. Understanding. In *The Stanford encyclopedia of philosophy*, ed. E. N. Zalta. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/understanding/

Hannon, M. 2021. Recent work in the epistemology of understanding. *American Philosophical Quarterly* 58 (3): 269–90. doi:10.2307/48616060.

Hill, C. E., L. G. Castonguay, L. Angus, D. B. Arnkoff, J. P. Barber, A. C. Bohart, T. D. Borkovec, and Wampold, B. E. 2007. Insight in psychotherapy: Definitions, processes, consequences, and research directions. In *Insight in psychotherapy*, 441–54. Washington, DC: American Psychological Association. doi:10.1037/11532-021.

Hills, A. 2009. Moral testimony and moral epistemology. *Ethics* 120 (1):94–127. doi:10.1086/648610.

Kvanvig, J. 2003. *The value of knowledge and the pursuit of understanding*. Cambridge: Cambridge University Press.

Lacewing, M. 2014. Psychodynamic psychotherapy, insight, and therapeutic action. *Clinical Psychology: Science and Practice* 21 (2):154–71. doi:10.1111/cpsp.12065.

Sedlakova, J., and M. Trachsel. 2023. Conversational artificial intelligence in psychotherapy: A new therapeutic tool or agent? *The American Journal of Bioethics* 23 (5):4–13. doi:10.1080/15265161.2022.2048739.

Zagzebski, L. 2008. *On epistemology*. Boston, MA: Cengage Learning.