



Do you feel like (A)I feel?

OPEN ACCESS

EDITED BY

Florin Dolcos,
University of Illinois at Urbana-Champaign,
United States

REVIEWED BY

Annisa Ristya Rahmanti,
Gadjah Mada University, Indonesia

*CORRESPONDENCE

Alexander Tagesson
✉ alexander.tagesson@lucs.lu.se

RECEIVED 04 December 2023

ACCEPTED 15 May 2024

PUBLISHED 30 May 2024

CITATION

Tagesson A and Stenseke J (2024) Do you feel like (A)I feel? *Front. Psychol.* 15:1347890. doi: 10.3389/fpsyg.2024.1347890

COPYRIGHT

© 2024 Tagesson and Stenseke. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Alexander Tagesson* and Jakob Stenseke

Department of Philosophy, Lund University, Lund, Sweden

KEYWORDS

empathy, emotion, artificial intelligence (AI), perspective-taking, experience sharing

Most of us are familiar with the uncomfortable feeling that results from skepticism about others' capacity to see the world as we do. If we dig too deep into this solipsistic worry, we might feel alone in the universe—*no one can feel what it is like to be me*. One might say that it is a question of attitude. Regardless of whether the other actually *can* empathize with us, our attitude prevents us from believing it. In a correspondence article in *Nature Human Behaviour*, Perry (2023) recently made her attitude toward the prospect of empathic AI clear: they will never know how it feels to be human! This sentiment is part of a broader aversion toward the prospect of artificial empathy (AE) (e.g., Montemayor et al., 2022; Zaki, 2023). While we agree that these dystopic concerns should be taken seriously, we also believe that the debate would benefit from additional nuance. More precisely, we argue that the AI systems of today—exemplified by AE skeptics such as Perry—are not the appropriate metric to evaluate the potential for AE and should not be used as support for why people might dismiss AE as non-genuine empathy.

At the core of Perry's critique is the observation that AE is well-received until recipients realize it was generated by an AI. Perry provides two explanations for this "artificial-empathy paradox". Firstly, "AI can learn to say the right words—but knowing that AI generated them demolishes any potential for sensing that one's pain or joy is genuinely being shared". Secondly, human empathy is valued *because* it is demanding and finite, and since "AI entails no emotional or time cost", it fails to indicate that "the recipient holds any unique importance". However, we argue that neither explanation succeeds in discrediting the prospect of artificial empathy.

Empathy is a notoriously convoluted concept (see Cuff et al., 2016 for a review) and researchers often highlight cognitive-, affective-, and motivational components of empathy (Zaki, 2014; Perry, 2023). Cognitive empathy, sometimes called perspective-taking or mentalizing, is the intellectual ability to understand how the other perceives and experiences their situation (Decety and Cowell, 2014; Zaki and Ochsner, 2016; Marsh, 2018). Cognitive empathy is to a degree already achievable for AI, which can detect and identify human emotions (Montemayor et al., 2022; Perry, 2023). Affective empathy, or experience sharing, refers to how one vicariously feels and experiences the other's emotional states (Decety and Cowell, 2014; Zaki and Ochsner, 2016; Marsh, 2018). This kind of experience-sharing is potentially not obtainable for AI. Lacking lived subjective experience (Montemayor et al., 2022; Perry, 2023), trying to share the experience of an emotional human may not resonate adequately as the AI reasonably does not feel anything (Turkle, 2007). The motivational component, also called empathic concern, can be understood as a motivation to support others' wellbeing or help them alleviate suffering (Decety and Cowell, 2014; Zaki and Ochsner, 2016; Marsh, 2018). However, while it is reasonable to contest the extent to which AI can manifest affective empathy and empathic concern, we disagree with the argument that it comes down to the human recipients' attitude toward AE (Montemayor et al., 2022; Perry, 2023).

Whether empathy is valuable is *not* (solely) a question of the recipient's attitude toward the empathizer. The value of the empathy a parent directs toward her child cannot easily be discarded based on whether the child takes the parent to be a genuine empathizer or not. Similarly, the fact that some foster negative attitudes toward psychologists, which prevents them from seeking therapy that would be beneficial for them, does not discredit the value of therapy. If we view empathy as a question of attitude, then we are exposed to the solipsistic worry: what prevents us from having the "wrong" attitude toward genuine empathizers? Perry merely assumes the anthropocentric standard view: only human activities are valuable for humans (Singer, 1975). This view is also prevalent in our attitudes toward animals. Studies show that people's tendency to not attribute mental capabilities (e.g., empathy) to animals fails to live up to their own self-reported normative standards (Leach et al., 2023). This effect was attenuated when ascribing a mind and mental capabilities to humans, displaying another example of anthropocentrism. Fortunately, attitudes can change, and if it is our attitudes toward animals and AIs that are decisive for granting them genuine empathy, couldn't they also change? In another recent study, researchers manipulated participants' attitudes toward AI, making them believe that the algorithm either had a manipulative motive, a caring motive, or no particular motive. This considerably changed how participants perceived and interacted with the AI, where participants in the care motive-, or empathic concern-, group perceived it as more empathetic than participants in the two other groups. Notably, the effect was stronger for more sophisticated AI (Pataranutaporn et al., 2023). These results suggest that our attitudes and the technological sophistication of the algorithm had an impact on participants' perception of AI as being empathetic. Consequently, we believe that our malleable attitudes toward AI, combined with increased technological sophistication, will impact our perception of AE, making it more likely to be perceived as genuine empathy. Furthermore, we are also not convinced by Perry's claim that because human empathy is demanding and limited, choosing to empathize communicates the importance of the recipient to the empathizer. A recent study showed that when people learned that empathy is an unlimited resource, they became more empathic, with real-life consequences, e.g., more likely to hug an out-group member (Hasson et al., 2022). These recipients of empathy and hugs would likely not deem the empathy they received as less genuine if they learned that the empathizer believed empathy to be unlimited and not finite, which led them to feel and behave more empathically.

AI already displays cognitive empathic abilities and is capable of generating generic empathic responses. Due to vast advancements in computational approaches to emotional inference, AI has demonstrated the ability to identify human emotions (Ong et al., 2019), manifest facial emotions (Mishra et al., 2023), and to successfully facilitate empathetic interactions (Ayers et al., 2023; Sharma et al., 2023). For instance, a recent example showed how human-AI collaboration produced more empathic conversations in peer-to-peer mental health support in comparison to human-only responses (Sharma et al., 2023). In another case, healthcare professionals found AI's responses to medical questions to be almost ten times as empathetic as the responses from human

physicians (Ayers et al., 2023). These generic applications of empathic displays are impressive, but when recipients learn that the empathy they received is AI-generated, they will likely not perceive it as genuine empathy. As Perry points out, "AI empathy fails to convey authentic care or to indicate that the recipient holds any unique importance". This might not be surprising as these examples merely attest to AI's capability of emotion detection and empathic signaling and are designed to respond empathetically to anyone repeatedly. In contrast, the motivated nature of human empathy often leads people to empathize in discriminating ways (Zaki, 2014; Bloom, 2016) and with few people (Cameron and Payne, 2011; Cameron et al., 2022). Thus, generic AI applications not only deprive recipients of the feeling of uniqueness; their empathic displays are also dissimilar to how humans empathize, making them seem non-genuine. The real test of artificial empathy—by Perry's own standards—would be through personalized algorithms tuned to their human. We can envision algorithms that respond with enthusiasm and authentic care to their particular person, making *that* person more important than any other. As the relationships deepen, the personalized algorithm will be able to display more empathy toward its person, paralleling how empathy is extended in humans (Depow et al., 2021). Recipients of this kind of AE will likely perceive the care as authentic and the concern as genuine, as there are already examples of people developing strong feelings for their AI companions (Pentina et al., 2023).

While we take issue with Perry's view, we also acknowledge that much more needs to be said about the potential value and risks of AE. For instance, several ethical issues need to be seriously addressed for the responsible development and deployment of AE, e.g., regarding privacy (Lutz et al., 2019), deception (Park et al., 2023), and negative impacts on human-human relationships (Turkle, 2010). A lot more can also be said about the potential benefits of AE; e.g., how it may alleviate many of the problems associated with human empathy, such as biases, e.g., ingroup empathy bias (Cikara et al., 2014) or compassion fade (Västfjäll et al., 2014). It could potentially also lessen some of the taxing costs associated with empathy (Cameron et al., 2019) and reduce compassion fatigue (Cocker and Joss, 2016) on the part of human empathizers. To this end, researchers have called for strategic regulations of different components of empathy, i.e., affective-, cognitive- and motivational-, to attain critical outcomes (Weisz and Cikara, 2021). Similarly, we could strategically apply different components of AE to enhance human flourishing. While all these considerations must be weighed against each other to determine the overall risk-benefit of AE, contrary to what Perry claims, human attitudes toward AI do not seem to present an insurmountable obstacle.

In sum, Perry's criticism of the value of artificial empathy seems unjustified. It might turn out to be the case that full-fledged AE is unobtainable due to principled obstacles for AI to simulate human empathy. Those are other, more legitimate, reasons for not granting AI empathy than due to recipients' changeable attitudes about AE. Before making any conclusive remarks about AE being perceived as genuine empathy or not, we need to create and evaluate AI applications that better emulate human empathy. It is possible that these applications may change our attitudes

so that we start perceiving artificial empathy, or parts of it, as genuine empathy.

Author contributions

AT: Writing – original draft, Writing – review & editing. JS: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

References

- Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., et al. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* 183, 589–596. doi: 10.1001/jamainternmed.2023.1838
- Bloom, P. (2016). *Against Empathy: The Case for Rational Compassion*. New York, NY: Ecco.
- Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., Inzlicht, M., et al. (2019). Empathy is hard work: people choose to avoid empathy because of its cognitive costs. *J. Exp. Psychol. General* 148, 962–976. doi: 10.1037/xge0000595
- Cameron, C. D., and Payne, B. K. (2011). Escaping affect: how motivated emotion regulation creates insensitivity to mass suffering. *J. Pers. Soc. Psychol.* 100, 1–15. doi: 10.1037/a0021643
- Cameron, C. D., Scheffer, J. A., Hadjiandreou, E., and Anderson, S. (2022). Chapter four - motivated empathic choices. *Adv. Exp. Soc. Psychol.* 66, 191–279. doi: 10.1016/bs.aesp.2022.04.005
- Cikara, M., Bruneau, E., Van Bavel, J. J., and Saxe, R. (2014). Their pain gives us pleasure: how intergroup dynamics shape empathic failures and counter-empathic responses. *J. Exp. Soc. Psychol.* 55, 110–125. doi: 10.1016/j.jesp.2014.06.007
- Cocker, F., and Joss, N. (2016). Compassion fatigue among healthcare, emergency and community service workers: a systematic review. *Int. J. Environ. Res. Public Health* 13:618. doi: 10.3390/ijerph13060618
- Cuff, B. M. P., Brown, S. J., Taylor, L., and Howat, D. J. (2016). Empathy: a review of the concept. *Emot. Rev.* 8, 144–153. doi: 10.1177/1754073914558466
- Decety, J., and Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends Cogn. Sci.* 18, 337–339. doi: 10.1016/j.tics.2014.04.008
- Depow, G. J., Francis, Z., and Inzlicht, M. (2021). The experience of empathy in everyday life. *Psychol. Sci.* 32, 1198–1213. doi: 10.1177/0956797621995202
- Hasson, Y., Amir, E., Sobol-Sarag, D., Tamir, M., and Halperin, E. (2022). Using performance art to promote intergroup prosociality by cultivating the belief that empathy is unlimited. *Nat. Commun.* 13:7786. doi: 10.1038/s41467-022-35235-z
- Leach, S., Sutton, R. M., Dhont, K., Douglas, K. M., and Bergström, Z. M. (2023). Changing minds about minds: Evidence that people are too sceptical about animal sentience. *Cognition* 230:105263. doi: 10.1016/j.cognition.2022.105263
- Lutz, C., Schöttler, M., and Hoffmann, C. P. (2019). The privacy implications of social robots: scoping review and expert interviews. *Mobile Media Commun.* 7, 412–434. doi: 10.1177/2050157919843961
- Marsh, A. A. (2018). The neuroscience of empathy. *Curr. Opin. Behav. Sci.* 19, 110–115. doi: 10.1016/j.cobeha.2017.12.016
- Mishra, C., Verdonshot, R., Hagoort, P., and Skantze, G. (2023). Real-time emotion generation in human-robot dialogue using large language models. *Front. Robot. AI* 10:1271610. doi: 10.3389/frobt.2023.1271610
- Montemayor, C., Halpern, J., and Fairweather, A. (2022). In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI Soc.* 37, 1353–1359. doi: 10.1007/s00146-021-01230-z
- Ong, D. C., Zaki, J., and Goodman, N. D. (2019). Computational models of emotion inference in theory of mind: a review and roadmap. *Top. Cogn. Sci.* 11, 338–357. doi: 10.1111/tops.12371
- Park, P. S., Goldstein, S., O'Gara, A., Chen, M., and Hendrycks, D. (2023). AI deception: a survey of examples, risks, and potential solutions. *ArXiv [preprint]. abs/2308.14752*.
- Pataranutaporn, P., Liu, R., Finn, E., and Maes, P. (2023). Influencing human-AI interaction by priming beliefs about AI can increase perceived trustworthiness, empathy and effectiveness. *Nat. Mach. Intell.* 5, 1076–1086. doi: 10.1038/s42256-023-00720-7
- Pentina, I., Hancock, T., and Xie, T. (2023). Exploring relationship development with social chatbots: a mixed-method study of replika. *Comput. Human Behav.* 140:107600. doi: 10.1016/j.chb.2022.107600
- Perry, A. (2023). AI will never convey the essence of human empathy. *Nat Hum Behav.* 7, 1808–9. doi: 10.1038/s41562-023-01675-w
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., and Althoff, T. (2023). Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat. Mach. Intell.* 5, 46–57. doi: 10.1038/s42256-022-00593-2
- Singer, P. (1975). *Animal Liberation: A New Ethics for Our Treatment of Animals*. New York, NY: The New York Review.
- Turkle, S. (2007). Authenticity in the age of digital companions. *Interact. Stud.* 8, 501–517. doi: 10.1075/is.8.3.11tur
- Turkle, S. (2010). "In good company? On the threshold of robotic companions," in *Close Engagements With Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. Y. Wilks (Amsterdam: John Benjamins Publishing), 3–10.
- Västfjäll, D., Slovic, P., Mayorga, M., and Peters, E. (2014). Compassion fade: affect and charity are greatest for a single child in need. *PLoS ONE* 9:e100115. doi: 10.1371/journal.pone.0100115
- Weisz, E., and Cikara, M. (2021). Strategic regulation of empathy. *Trends Cogn. Sci.* 25, 213–227. doi: 10.1016/j.tics.2020.12.002
- Zaki, J. (2014). Empathy: a motivated account. *Psychol. Bull.* 140, 1608–1647. doi: 10.1037/a0037679
- Zaki, J. (2023). *How Artificial Empathy Will Change Work and Life, According to a Stanford Professor*. The Future of Work. Available online at: <https://www.fastcompany.com/90984653/how-artificial-empathy-will-impact-workers-and-customers-according-to-a-stanford-professor> (accessed February 08, 2024).
- Zaki, J., and Ochsner, K. N. (2016). "Empathy," in *Handbook of Emotion*, 4th Edn, eds. L. Feldman-Barrett, M. Lewis, and J. M. Haviland-Jones (New York, NY: Guilford Press), 871–884.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.