

## Should Dr Robot possess moral empathy?

Elisabetta Sirgiovanni (Sapienza University of Rome)

The paper challenges the widely popularized idea that it is imperative to imbue artificial devices in healthcare with empathy. One of the most frequently discussed ethical risks related to clinical AI involves the possibility of dehumanizing the doctor-patient relationship by eradicating a unique human trait such as moral empathy. This is said to negatively affect the fiduciary relationship between clinicians and patients and, consequentially, therapeutic benefit and clinical improvement.

Conversely, it will be argued that natural empathy has negative aspects, including partiality, favoritism, and paternalism, which can harm patients' autonomy and equality, thus limiting their opportunity to receive proper treatment and efficient, high-quality care. These drawbacks, evidenced in humans in the neuro and psychological processing of empathy, are likely to transfer to AI devices and prompt a reconsideration of whether fostering genuinely empathic relationships in AI systems with patients is ideal. As we know from AI research, biases can covertly emerge at any stage of AI information processing, and thus they may presumably emerge in the artificial realization of empathy. Due to the epistemic opacity of AI systems, detecting the origin of such biases can be notably challenging. Moreover, research shows that human-machine interaction tends to reinforce the circle of biases, instead of reducing it.

The aim of the paper is not, however, to completely reject affective AI programs in healthcare. While critiquing empathy's reliability as a moral driver, the implementation of alternative moral emotions like sympathy will be favored. This recently revalued emotion, which comprehends a detached commitment to the welfare of others, may better prevent paternalistic attitudes, possibly also when AIs are employed. The paper states also the importance of some form of direct normativity for clinical AI systems. This work is theoretical, and intertwines different areas of research such as neuroethics, medical ethics and the ethics of technology.

### References

- Bloom, P. (2016). *Against Empathy: The Case for Rational Compassion*. New York: Ecco HarperCollins.
- Bloom, P. (2017). Empathy and Its Discontents. *Trends Cogn Sci*. 21(1), 24-31.
- Liu, B., & Sundar, S.S. (2018). Should Machines Express Sympathy and Empathy? Experiments with a Health Advice Chatbot. *Cyberpsychology, Behavior, and Social Networking*. 21 (10), 625.
- Montemayor, C., Halpern, J., & Fairweather, A. (2021). In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI & Soc.* published online 2021, May 26.
- Paiva, A., Samuel, M., Petisca, S., et al. (2018). Towards more humane machines: creating emotional social robots In: Graca Da Silva S, *New Interdisciplinary Landscapes in Morality and Emotion* (pp. 125-139). New York, NY: Routledge: 125.
- Prinz, J.J. (2011). Against empathy. *The Southern Journal of Philosophy*. 49, 214-233.
- Prinz, J.J. (2011). Is empathy necessary for morality? In: Coplan A, Goldie P (eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 211—229). New York: Oxford University Press.
- Stefanello, E. (2021). Your pain is not mine: A critique of clinical empathy. *Bioethics*, 1– 8.
- Vicente, L., Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*, 13, 15737, 1-13.