https://doi.org/10.53656/phil2024-03-06



# SIMULATING CARE AND EMPATHY IN MEDICINE — CAN WE TRUST ARTIFICIAL AGENTS?

**Assist. Prof. Iva Georgieva**
*Bulgarian Academy of Sciences*

**Abstract.** The utilization of artificial agents in medicine becomes a foremost topic in research of recent medical advances. Moreover, to possibility to simulate adequate care and create the sense of empathy towards the patients, becomes a major question for research that would rather fall in the area of philosophy of medicine than in any other discipline as the very notion of empathy is a complex topic that is difficult to define in its natural concept boundaries, and furthermore in proposing the notion of artificial care and empathy. The question of whether we can define and simulate those would also help us answer a larger problem – can we trust artificial agents as aids or replacements of humans in the medical environment and healthcare in general.

*Keywords*: artificial agents; medicine; care; empathy; simulation; trust

## Introduction

This study explores several key concepts in the field of philosophy of medicine that are reflected in the emerging context of Artificial Intelligence (AI), namely the ones of care as a foremost notion in medicine, healthcare and wellbeing, the last being of a special interest of this and future studies.

These concepts are: empathy – a notion undistinguishable from the very essence of medical care, which notion now is being recalibrated for the context of AI (as a comparison, another technological advance that sparked a lot of discussions, namely Virtual Reality (VR) has been named the "ultimate empathy machine" (Barbot & Kaufman 2020), and so we can ask if AI would become a bearer of such definition as well as it would be able to mimic the human empathy – a foremost topic, subject of discussion hereafter); trust, which in close relation to the previous two is now very common in the context of AI in medicine, as trust should be built and maintained between all sides of relation in medical care and practice when artificial counterpart exists; artificial – the very notion of it as not opposed to human, but as an upgrade that is more in the form of aid and enhancement, rather than lack of human element in every sense of the word; simulation as a notion that requires a method to distinguish when

simulated or artificial element is present in the medical interaction, which distinguishing process is also of a great interest here and in the studies hereafter; complexity – the idea about the main topics in relation to technology advancement, that are, without a doubt, a subject of very complex definitions and present even more complex relations, in short the complex topic of AI in medicine is then a justified subject of philosophical analysis.

The above key notions and the understandings behind them naturally lead to the necessity of adequate definitions, regulations, laws or at least at present moment – of larger agreements, that are, certainly, in the capacity of the works of philosophy of medicine and technology.

**New relation**

As an experiment with a typical Generative AI (*DeepAI: https://deepai.org/machine-learning-model/text2img*), when preparing this paper, a prompt consisting of the keywords "artificial care trust medicine" was used to generate image that reflects the meaning of the words. The very second attempt gave an image that shows explicitly the new relation(ship) between Doctor – Patient that becomes threefold, namely between Doctor – AI – Patient as below:

This leads us to the main part of the topic of discussion, namely, the Artificial agents (AA) as one side of AI development. As a philosophical concept, it is important to distinguish between AI and "its" embodiment in the form of AA where physicality, presence, and perception become evident. This existence of AA shows the transformative perception from a mere tool to an tool agent and evokes questions related to the ideas of humanization, personification, and anthropomorphization

that were subject of theoretical analysis for decades but now become a sensible topic of real issues to be resolved.

As a result, the problems of ethics of AI **and** AA, especially about how AA would ***mimic*** or ***model*** empathy, and so convince us that understand us as human beings, would become the most important ones when discussing how AI and AA interfere or intertwine with our human lives and the manifestation of care in it, especially as medical care.

The problem of creating moral machines (and the corresponding responsibility issues that would result from such a creation) (Vallverdú & Casacuberta 2014) is a main task of the scientific effort to integrate AI in human life and it starts with the understanding of the legal differences that exist when AI interferes with issues that are subject of moral discussion and most importantly, in nowadays world of post-truth and replica – to know when AI is involved, is an opponent or is responsible for something for which fact humans were not aware of in the first place. It would benefit immensely to have regulations that state that certain artifact is a result of AI, has its intervention or is represented by AI.

This especially pertains to cases when emotional intervention is at hand (Montemayor, Halpern & Fairweather 2022), which cases are the most human-like and would impact at the greatest extend the end subject, that in the discussed case would be a human patient whose health and wellbeing is affected. Thus, an Emotional intelligence (EI) concept within the AI notion should be created and understood, perhaps in a way that differs from the concepts we have about human EI or as and EI that is detected and perceived by humans as a human-like or human-acceptable artificial EI. This also leads to the questions of experienced empathy (ibid.) that is not an existing but rather experienced by the recipient human empathy. Then, as philosophers we might ask, is it necessary to experience one in order to express it, or just simulating it according to the task is enough to have *real* (sense) of empathy. Would an "observable" (Grozdanoff 2021) empathy and morally appropriate behavior enough?

This leads us to another point, namely the patient point of view, or the evaluation of recipient that is the most important one, at least as a main viewpoint. However, it feels like this lacks real changes to the ways AI interacts with humans as we need it to be more like us. Will it have a real understanding (even real feeling, a form of standing in one's shoes) at some point, just like us, or it will master the simulation of such as it gets better at doing its job?

**Empathy as a human feature**

The notions of empathy, compassion, and trust are fundamental values of a **patient-centered** medicine (Kerasidou 2020) and as such they lead to a better *experience* of medical care. But then, what is empathy, is it intrinsic human feature, can we teach or discern it from other similar sensations in humans and other living beings?

It is important to first distinguish that there are often defined several different types or forms of empathy (Georgiev et al. 2023), that are: a) cognitive, b) affective, c) somatic; and as such it is specific for a specific context (i.e. it is somewhat a cultural phenomenon), and it is a complex notion (Georgieva & Georgiev 2022) that as such can answer complex needs for not only healing experiences but for preserving wellbeing.

There are apparently different cases that require different types of empathy, and moreover, human emotions are such complex variety within which there is empathy as well, unique for each and every case. Having considered that, the very idea of personalized medicine, nowadays having even more possibilities for personalization with the help of technology and more specifically with the help of AI or AA, would benefit from an analysis of what empathy should mean in the context of AI, how it is understood from a humane point of view that still considers equal rights (even those of AI and AA themselves), and as a result would prompt people to accept and seek help in artificial forms of healthcare and wellbeing assistance.

Of course, when speaking about acceptance and inclusion of such AI agents, the problem of their advancement leads to the problem of uncanniness (Liu and Sundar 2018) that when faced with such deep human emotions as empathy, might get even worse as we as human beings tend to "read" very subtle signs of empathy and when such are simulated or too much but not yet same as human, might provoke serious repulsion or at least inability of acceptance.

That is why, the notion of Empathetic AI (Shao 2023) and the work of its definition, development and acceptance is foremost in the context of existence of technology in medicine and certainly should become a main subject of discussion in philosophy of medicine and technology. There is still one question that needs to be put as a starting point, that is – when do it is necessary to have real care and when it is necessary to have simulated care, or in other words what actually do we need as patient experience and goal of the treatment? What does that mean? Of course, there are medical cases when patients need a very high-quality empathetic experience such as cases when they report very delicate problems, for example traumatic experiences (Georgieva & Georgiev 2019) and need a technological solution that can reflect that health condition with the adequate empathy and much more – with the ***understanding*** of empathy.

**Aid or replacement?**

However, there are other cases when empathy is a necessary supporting condition of a different end result – let us imagine a patient in the emergency room who needs assistance but there are no available doctors or very few nurses to take care of that patient (quite common in disaster cases). In such an event, the existence of a fairly empathetic AI that handles the initial condition of the patient and shows the adequate attention while providing the adequate care (let us say talks to calm

the patient down), is exhibiting an empathetic behavior without necessarily going deep into understanding the specific case but rather answering the overall need to tackle a major problem. In such a case we might say a superficial level of empathy is shown but sufficient level could be experienced as an extreme example. This shows that there are very different degrees and depths of empathetic experiences, expressions and understandings, broadly speaking in philosophical terms.

The above example also shows that there might be multiple cases when there is no need to replace humans in providing medical care unless it is not safe environment for humans (e.g. in the case of a very aggressive patient or in a very dangerous situation). Moreover, AI and the respective AA will be perfect for tasks that do not resemble human activity, that result in system and workforce overload or where automation is needed. And most importantly, AI and AA will help the doctor load in the very serious problem of medical staff burnout, not only in physical exhaustive tasks but also in those that affect even the most trained human professionals, for example when working with dementia patients where the necessity for *detachment* is more prominent.

Having said that, is AI empathy even near the level of those needs? Actually, researchers report that "chatbot-generated responses outperformed physician-generated responses in both quality and empathy" (Cadiente, Chen & Pilkington 2023, p. 1278), and lead us to the more eminent next step, that is the possibility of assessment of **AI empathy**.

In any case, is such assessment should be designed, it must consider the above discusses needs for simulation, that is no need for increased empathy (as in routine tasks, or when the patient does not require such) and special need for empathy (as in patient-specific cases when deeper understanding and compassion are needed). Then, how to determine what cases are present at hand in any given situation – as natural as it may sound, again human expertise is needed to perform medical decision-making (Ayers, Dredze, & Smith 2023) about when and what kind of empathy any particular AA/AI should (re)present.

Then, what kinds of AI/AA are there and should be developed more, what are the successful ones that are already well-accepted?

– Chatbots
– Virtual assistants
– Contact (call) centers / self-service, etc.

The above examples show a very clear tendency of a more business-oriented approach, that is not so much in based on healthcare, betterment of quality of life, self-care and so on. And exactly these are the areas of deficiency and necessity for empathetic AI. Consequently, the abilities AI should develop more are moving from those that aid analytical to such that foster emotional interactions. In result, increased presence of AI in

– HCI and human services (e.g. social workers in suicide prevention)
– Oral interviews as part of psychological intervention

for example are necessary as they will lead to a major shift in the need of AI in healthcare, namely to pave its presence in mental health care and support of general wellbeing, as well as in case-specific and as defined, complex cases such as the ones of psychological trauma as we have previously discussed (Georgieva & Georgiev 2020).

**Limitations and risks**

It is not possible to talk about technological implementations in medicine and not to take a good consideration of the possible limitations and risks, the most evident of which can be divided into technical and theoretical.

Medical ethicists have long expressed concerns that technologies such as "care-bots" remove the human element from medicine, resulting in *dehumanization* and *depersonalization* of care. On the other hand, conditions and treatments that are shame-inducing offer opportunities for introducing AI in a manner that removes the human element of medicine in an **ethically** (*morally*) justified way (Palmer and Schwan 2022). This would lead to significant results in:

– stigmatization and biases reduction
– decreased negative emotional impact of heavy clinical cases on medics (e.g. as we discussed the dementia treatment risks), and many other.

That is why it is more urgent to validate the techniques about how to integrate these tools into clinical practice, defining clear boundaries between full, supervised, and proscribed autonomy (Ayers et al. 2023) of AI in its presence in medical health care and wellbeing provision.

**In summary**

We have seen that the relations(ships) between Doctor and Patient get broadened and advanced with the presence of AI as below:

Doctor – AI – Patient

However, we should consider their special cases, such as only AI – Patient or only Doctor – AI relation when specific patient decision should be made for example. The fears and the concerns presented here and elsewhere in relation to the advance of AI and AA are real and need to be tackled, however, we also need to look at the beneficial side, and to give an example for this, the progressive input from AI, resulting in increased access to medical information leads to a surprising effect on the physicians who report increased "need to care" (Dalton-Brown 2020) for the patient, that is a motivation to provide overall better health results.

In addition to the above, the design of empathetic AI can gain wider acceptance among the public and hence result in increased trust in technological presence in general as we have seen that health emergencies question deeply our moral and

philosophical understanding of technology and require the corresponding measures for acceptance of novelties in order for them to work for us (Georgieva et al. 2021).

And lastly, but truly importantly, we need to stress that currently AI learns through us and our biases, limitations and scopes of knowledge define or teach what it means to be human and hence, what is empathy for us.

**Philosophy of medicine position**

It is evident that only a meta discipline such as the philosophy of medicine could be efficient enough to overcome the obstacles and handle the need for clearing that view upon AI in health. As authors state, understanding "how 'intelligence,' in its different dimensions, is being manifested and co-constituted through the human-technology interface, in ways that are re-materialising the boundaries of the human and the machine identities in affective, embodied, and relational ways" (De Togni et al. 2021, p. 7) is the way to integrate AI in the most natural way in our human lives.

And human moral and compassion should spread both ways as already "AI technologies are being used to enhance empathetic awareness; empathetic response and relational behavior; communication skills; health coaching; therapeutic interventions; moral development learning; clinical knowledge and clinical assessment; healthcare quality assessment; therapeutic bond and therapeutic alliance; and to provide health information and advice" (Morrow et al. 2023, p. 2), and will begin to work with greater notions that before would have never be assigned to a technological artefact, such as ***care***, ***empathy***, ***compassion***, ***trust***, ***health***, ***wellbeing***, and so forth.

**Conclusion**

It is completely necessary to start the discussion about the new paradigm in philosophy of medicine (Vasseva-Dikova 2022), that is exploring the connection between technology and medicine and is trying to find the place of technology not only as an aid or replacement of human effort or (possibly) expertise but also as an equal part of the medical process and essence, as something created by us in order to help us live healthier and better lives.

**REFERENCES**

AYERS, J.W., DREDZE, M. & SMITH, D.M., 2023. Machine-Made Empathy? Why Medicine Still Needs Humans – Reply. *JAMA internal medicine*, vol. 183, no. 11, pp.1279 – 1280.

BARBOT, B. & KAUFMAN, J.C., 2020. What makes immersive virtual reality the ultimate empathy machine? Discerning the underlying mechanisms of change. *Computers in Human Behavior*, vol. 111, pp.106 – 431.

CADIENTE, A., CHEN, J. & PILKINGTON, B., 2023. Machine-made empathy? Why medicine still needs humans. *JAMA Internal Medicine*, vol. 183, no. 11, pp.1278 – 1279.

DALTON-BROWN, S., 2020. The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics*, vol. 29, no. 1, pp.115 – 121.

DE TOGNI, G., ERIKAINEN, S., CHAN, S. & CUNNINGHAM-BURLEY, S., 2021. What makes AI 'intelligent'and 'caring'? Exploring affect and relationality across three sites of intelligence and care. *Social Science & Medicine*, vol. 277, p. 113874.

GEORGIEVA, I. & GEORGIEV, G.V., 2019. Redesign me: Virtual reality experience of the line of life and its connection to a healthier self. *Behavioral Sciences*, vol. 9, no. 11, p. 111.

GEORGIEVA, I., & GEORGIEV, G. V., 2020. Reconstructing personal stories in virtual reality as a mechanism to recover the self. *International Journal of Environmental Research and Public Health*, vol. 17, no. 1, p. 26.

GEORGIEVA, I., BEAUNOYER, E. & GUITTON, M.J., 2021. Ensuring social acceptability of technological tracking in the COVID-19 context. *Computers in Human Behavior*, vol. 116, p.106639.

GEORGIEVA, I. & GEORGIEV, G.V., 2022. Narrative self-recreation in virtual reality. *Frontiers in Virtual Reality*, vol. 3, p. 854333.

GEORGIEV, G.V., NANJAPPAN, V., GEORGIEVA, I. & Gong, Z., 2023, October. Empathic Experiences of Visual Conditions with Virtual Reality. In: *International Conference on Interactive Digital Storytelling*, pp. 168 – 180. Cham: Springer Nature Switzerland.

GROZDANOFF, B. D., 2021, Graph Formalism for Normative Ethical Tasks in Artificial Intelligence. *Ethical Studies*, vol. 6, no. 2, pp. 109 – 120.

KERASIDOU, A., 2020. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bulletin of the World Health Organization*, vol. 98, no. 4, p. 245.

LIU, B. & SUNDAR, S.S., 2018. Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, vol. *21*, no. 10, pp. 625 – 636.

MONTEMAYOR, C., HALPERN, J. AND FAIRWEATHER, A., 2022. In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI & society*, vol. 37, no. 4, pp. 1353 – 1359.

MORROW, E., ZIDARU, T., MASON, C., PATEL, K. D., & STOCKLEY, R., 2023. Artificial intelligence technologies and compassion in healthcare: A systematic scoping review. *Frontiers in Psychology*, vol. 13, p. 971044.

PALMER, A. & SCHWAN, D., 2022. Beneficent dehumanization: Employing artificial intelligence and carebots to mitigate shame-induced barriers to medical care. *Bioethics*, vol. 36, no. 2, pp. 187 – 193.
SHAO, R., 2023, April. An Empathetic AI for Mental Health Intervention: Conceptualizing and Examining Artificial Empathy. In: *Proceedings of the 2nd Empathy-Centric Design Workshop*, pp. 1 – 6.
VALLVERDÚ, J. & CASACUBERTA, D., 2014. Ethical and technical aspects of emotions to create empathy in medical machines. In: *Machine Medical Ethics*, pp. 341 – 362. Cham: Springer International Publishing.
VASSEVA-DIKOVA, J., 2022. Novata paradigma v meditsinata. *Filosofia*, vol. 31, no. 2, pp.123 – 133.

✉ **Dr. Iva Georgieva, Assist. Prof.**
ORCID iD: 0000-0001-9458-8690
Institute of Philosophy and Sociology
Bulgarian Academy of Sciences
Sofia, Bulgaria
E-mail: ivavgeorgieva@gmail.com