

# ON THE MEANING AND USE OF ARTIFICIAL EMPATHY

## IN SHORT

- The chess computer marked the beginning of artificial intelligence. Today, artificial systems are no longer "only" intelligent, but also emotional. And in the future, they will even be empathetic.
- The ability to recognize and respond appropriately to emotions is grounded in many complexities that pose high technical hurdles to the development of empathetic AI.
- Empathy-enabled artificial systems offer opportunities in geriatric care, psychology, and therapeutic areas (such as therapies with autistic individuals) – but philosophy professor Catrin Misselhorn advises against using empathy-enabled robots as friends or love partners.



Watch episode  
[trivadis.com/sparx](https://trivadis.com/sparx)



Do you remember the year 1997? On May 11, IBM's "Deep Blue" computer won a chess match against Garri Kasparov, the reigning world chess champion. From then on, the system behind "Deep Blue" was considered intelligent. Over the years, artificial intelligences have been trained with vast amounts of analytical data from various areas of life, which brings us to today's emotional artificial intelligence. This emotional AI is not only able to analyze past human actions and apply them to new situations. Using various additional methods such as face-based emotion recognition, sentiment analysis, voice analysis or biometric methods, it is also able to understand humans even better and predict their possible actions far beyond the chessboard.

Catrin Misselhorn, Professor of Philosophy at the University of Göttingen, describes in her book "Künstliche Intelligenz und Empathie – Vom Leben mit Emotionserkennung, Sexrobotern & Co" (Artificial Intelligence and Empathy – Living with Emotion Recognition, Sex Robots & Co) that emotional AI had a somewhat bad reputation or was criticized to not be empathetic enough and

manipulate humans too strongly into a desired behavior (p. 41). Because of this criticism, a new research program was created called Artificial Empathy. The goal of this research program is to equip artificial systems with the ability to empathize in order to counteract this tendency to manipulate. For, equipped with empathy, it would be possible for artificial systems to put themselves in the shoes of humans and – in contrast to emotional artificial intelligence – primarily comply with the wishes of humans and not those of the profiteer behind the AI. Artificial empathy is thus intended to avert the danger of AI becoming an instrument of purely manipulative purposes (p. 134 et seqq.).

Before Misselhorn goes into more detail about artificial empathy, she enlarges upon the topic of human empathy and explains, among other things, which conditions are necessary for us to feel empathy at all. Misselhorn names three necessary and sufficient conditions for empathy: congruence, asymmetry and foreign consciousness (p. 44 et seqq.).

- Congruence means that one can in some way understand the feelings of the other person or feel a feeling of the same kind. Congruence is a necessary condition for empathy, but it is not sufficient on its own.
- Asymmetry is a consequence of congruence and means that one has the empathetic feeling only because another person has it and, consequently, that person's feeling is more appropriate than one's own empathetically felt feeling.
- In addition, there is the foreign consciousness: one knows that the sensed feeling is that of another person.

Together, these three points are sufficient conditions for the presence of empathy. However, as soon as one of these three conditions is missing, one cannot speak of empathy, as Misselhorn points out in this example: "For instance, one could take something away from someone in the knowledge that the person will be angry, because people are usually angry when something is taken away from them, without feeling this emotion at that moment. In this case the congruence condition is not fulfilled and consequently, neither is the condition of asymmetry. Foreign consciousness, on the other hand, can be attributed, since it is clearly the feeling of another person" (p. 45).

To further understand what it takes to endow artificial systems with artificial empathy, Misselhorn describes two types of human empathy: perceptual empathy and empathy through taking perspective (reenactive empathy). While perceptual empathy is innate and a reflexive, quick as well as cognitively sparse process (for example, helping in response to a person lying on the ground in pain), reenactive empathy represents the opposite process of consciously putting oneself in another person's shoes. For the question of how to endow artificial intelligence with the capability of empathy, innate perceptual empathy is central (p. 48).

Perceptual empathy thus describes the ability to perceive the feelings of other sentient beings without having to rely on theoretical assumptions or inferences (p. 48 et seqq.). For this ability, a kind of system of mirror neurons in our nervous system or neuronal associations play a role, which is where the replication of the feelings of others takes place: In the brain, the same areas are activated whether we ourselves feel pain or empathetically perceive that another individual is in pain.

How strongly we feel this empathetic emotion depends on the intensity with which the emotion is expressed. Whereas similarity and familiarity enhance the feeling, it is weakened, , when several people show different emotions at the same time or when a person's emotion is considered inappropriate (p. 50).

Misselhorn's remarks on human empathy already give an idea of the great technical hurdles that have to be overcome in implementing artificial empathy in artificial systems. However, it is not completely impossible, and Misselhorn describes two methods by which a computational model that can be implemented on a computer can be obtained: by an analytical or an empirical method (p. 58 et seqq.).

While the theory-based, analytic method "relies on psychological and neurophysiological theories of interpersonal empathy based on behavioral observation and physiological measurements," the second basic method is "empirical and makes use of data collected in studies of human empathetic interaction" (ibid.). In the first method, the theories are "computationally modeled and implemented in an artificial system" (ibid.). An example of this method is developmental robotics, in which empathy is supposed to be "produced by artificial systems interacting with the environment and with humans from other cognitive processes or mechanisms" (ibid.).

In the second method, the data collected from the studies are used "to train systems to recognize general patterns of empathetic behavior. These patterns can then be transferred to similar new situations" (p. 59). Misselhorn notes strengths and weaknesses in both methods. While the empirical method is very data-intensive and the correct selection of data raises many questions, the analytical method is already based on abstractions from the ground up (p. 60).

Let's leave these difficulties for a moment and look at what opportunities would arise with artificial empathy. On the one hand, robots and virtual agents can be used in elderly care and in psychology. Experiments already conducted show that patients to some extent open up to a virtual agent in psychological conversations even more than to a human counterpart. This is due to the subjectively perceived anonymity and impartiality of the system as well as the feeling of not being observed or judged (p. 73). Such avatars, which are in a virtual environment but can communicate with humans in real time using speech, facial expressions, and gestures, could not only prevent learning frustration in classrooms, but also, for example, detect early burnout among employees or help increase efficiency and keep motivation high. For example, an avatar could take on research tasks or schedule appointments – always in exchange with the real person, who would then be alerted by their avatar to take a break or drink more water or be cheered up by a joke, depending on their mood. Artificial empathy can also be used in the HR department: during a job interview with a virtual agent, they could assess whether the applicant is suitable for the job in terms of resilience, for example.

Misselhorn states that artificial intelligence in the "Working World 5.0: Cooperation between Humans and Machines" will no longer just take over routine tasks, but increasingly also more demanding social activities – for this, the artificial systems must be able to recognize human emotions and respond to them appropriately (p. 7 et seqq.). Misselhorn says that, for example, care robots equipped with empathetic artificial intelligence would work well in a group setting, where they could

even stimulate social interactions and bring fun to the elderly (YouTube, 2021). However, it would be wrong to assume that such a robot would then be alone at home with the elderly person, overtaking those tasks that would otherwise be done by relatives or friends (ibid.). "Empathy in humans requires that we empathize with the emotions of our counterparts. And that also means to some extent sharing the experience. However, artificial systems cannot have such an experience at all," Misselhorn summarizes (ibid.).

At the end of the book, Misselhorn paints a rather cautious picture of "the project of endowing artificial intelligence with the ability to empathize. These doubts concern both the technical feasibility as well as the ethical desirability" (Misselhorn, 2021, p. 134). Misselhorn notes that there are certainly sensible applications for artificial empathy. For example, in the therapeutic field, empathy-enabled robots could help people with autism learn to recognize emotions and respond to them adequately. In the entertainment or gaming industry, the technology holds the possibility of intensifying and expanding the user experience by, for example, adjusting turns in the game to the user's state of mind. However, robots capable of empathy should not act as friends or love partners. Misselhorn sees the extensive use of empathetic artificial intelligence more as a threat to the foundations of our coexistence (p. 135). Therefore, technology should be used wisely in the sense of making all of our lives and work easier – but here too, humans and the human will always be at the center.

## SOURCES

- Misselhorn, Catrin, 2021, Reclam, Künstliche Intelligenz und Empathie – Vom Leben mit Emotionserkennung, Sexrobotern & Co, ISBN: 978-3-15-011344-8
- YouTube, channel of the University of Göttingen, Künstliche Intelligenz und Empathie: Was sagt die Philosophin? <https://www.youtube.com/watch?v=vpFeHmxwoKs>, last accessed 21.01.2022