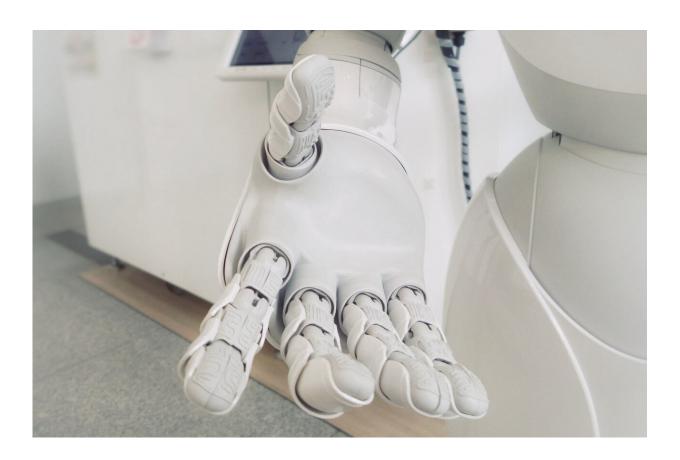# 'Empathetic' AI has more to do with psychopathy than emotional intelligence—but we should treat machines ethically

March 21 2024, by Catrin Misselhorn

Credit: Photos Hobby, Unsplash

AI has long since surpassed humans in cognitive matters that were once considered the supreme disciplines of human intelligence like chess or Go. Some even believe it is superior when it comes to human emotional skills such as empathy. This does not just seem to be some companies' talking big for marketing reasons; empirical studies suggest that people perceive ChatGPT in certain health situations as more empathic than human medical staff. Does this mean that AI is really empathetic?

A definition of empathy

As a psychologically informed philosopher, I define genuine empathy according to three criteria:

- Congruence of feelings: empathy requires that the person who empathizes to feels what it is like to experience the other's emotions in a specific situation. This distinguishes empathy from a mere rational understanding of emotions.
- Asymmetry: the person who feels empathy only has the emotion because another individual has it and it is more appropriate to the other's situation than to their own. For this reason, empathy is not just a shared emotion like the shared joy of parents over the progress of their offspring, where the asymmetry-condition is not met.
- Other-awareness: There must be at least a rudimentary awareness that empathy is about the feelings of another individual. This accounts for the difference between empathy and emotional contagion which occurs if one catches a feeling or an emotion like a cold. This happens, for instance, when kids start to cry when they see another kid crying.

Empathetic AI or psychopathic AI?

Given this definition, it's clear that artificial systems cannot feel empathy. They do not know what it's like to feel something. This means that they cannot fulfill the congruence condition. Consequently, the question of whether what they feel corresponds to the asymmetry and other-awareness condition does not even arise. What artificial systems can do is recognize emotions, be it on the basis of facial expressions, vocal cues, physiological patterns or affective meanings; and they can simulate empathic behavior by ways of speech or other modes of emotional expression.

Artificial systems hence show similarities to what common sense calls a psychopath: despite being unable to feel empathy, they are capable to recognize emotions on the basis of objective signs, to mimic empathy and to use this ability for manipulative purposes. Unlike psychopaths, artificial systems do not set these purposes by themselves, but are given them by their designers. So-called empathetic AI is often supposed to make us behave in a desired way, such as not getting upset when driving, learning with greater motivation, working more productively, buying a certain product—or voting for a certain political candidate. But then does not everything depend on how good the purposes are for which empathy-simulating AI is used?

Empathy-simulating AI in the context of care and psychotherapy

Take care and psychotherapy, which aim to nurture people's well-being. You might think that the use of empathy-simulating AI in these areas is definitely a good thing. Would they not be wonderful care-givers and social companions for old people, [loving partners for the disabled](link), or [perfect psychotherapists that have the benefit of being available 24/7](link)?

Such questions ultimately concern what it means to be a human being. Is it enough for a lonely, old or mentally disturbed person to project emotions onto an artifact devoid of feelings, or is it important for a person to experience recognition for themselves and their suffering in an interpersonal relationship?

Respect or tech?

From an ethical perspective, it is a matter of respect whether there is someone who empathically acknowledges the needs and the suffering of a person as such. By taking away recognition by another subject, the person in need of care, companionship or psychotherapy is treated as a mere object because ultimately this is based on the assumption that it does not matter whether anybody really listens to the person. They do not have a moral claim that their feelings, needs and suffering is perceived by someone who can really understand them. Using empathy-simulating AI in care and psychotherapy is ultimately another case of technological solutionism, i.e., the naïve assumption that there is a technological fix for every problem, including loneliness and mental "malfunctions." Outsourcing these issues to artificial systems prevents us from seeing the social causes for loneliness and mental disorders in the larger context of society.

In addition, designing artificial systems to appear as someone or something that has emotions and feels empathy would mean that such devices always have a manipulative character because they address very subliminal mechanisms of anthropomorphisation. This fact is used in commercial applications to get users to unlock a paid premium level: or customers pay with their data. Both practices are particularly problematic for vulnerable groups, which are at stake here. Even people who do not belong to vulnerable groups and are perfectly aware that an artificial system has no feelings will still react empathically to it as if it did.

Empathy with artificial systems—all too human

It is a well-studied phenomenon that humans react with empathy towards artificial systems that display certain human or animal-like characteristics. This process is largely based on perceptual mechanisms which are not consciously accessible. Perceiving a sign that another individual is undergoing a certain emotion produces a congruent emotion in the observer. Such a sign can be a typical behavioral manifestation of an emotion, a facial expression or an event that typically causes a certain emotion. Evidence from brain MRI scans shows that the same neural structures get activated when humans feel empathy with robots.

Although empathy might not be strictly necessary for morality, it plays an important moral role. For this reason, our empathy toward human-like (or animal-like) robots imposes at least indirect moral constraints on how we should treat these machines. It is morally wrong to habitually abuse robots that elicit empathy as doing so negatively affects our capacity to feel empathy, which is an important source of moral judgment, motivation, and development.

Does this mean that we have to establish a robot-rights league? That would be premature, as robots do not have moral claims by themselves. Empathy with robots is only indirectly morally relevant due to its effects on human morality. But we should carefully consider whether and in which areas we really want robots that simulate and evoke empathy in human beings as we run the risk of distorting or even destroying our social practices if they became pervasive.