

The Normative Side of Building Friendship with AI Companions

Tugba Yoldas, University of Alberta, Canada

International Conference on Computer Ethics: Philosophical Enquiry (CEPE) 2023, Chicago, IL

Keywords: friendship, AI companions, ethical design, virtue ethics

Extended Abstract

In this paper, I discuss friendship from the perspective of virtue ethics and claim that friendship with AI companions might be harmful to us because firstly, they do not actually meet our moral needs where friendship is concerned, and secondly, they threaten to undermine the virtue of friendship by introducing a morally defective mode of friendship and by transforming important societal norms about interpersonal relationships. I argue that designing ethical AI companions gains significant importance especially in the areas where these companions come in contact with the vulnerable groups in society such as children, the elderly, and people living with disabilities. In conclusion, I suggest that AI companions could prove helpful, for example, as models to help children practice ethical and intellectual virtues, as personal mental health aids, and as ethical mediators to help us make better moral decisions and navigate ethical dilemmas. But this is a possibility only if their design respects certain ethical principles. I discuss some of the challenges around implementing ethics into these systems.

AI and related technologies permeate many social aspects of our lives from helping household chores to caregiving for the elderly, child rearing, education, therapy, and others. As our lives increasingly merge with these technologies, novel ethical questions concerning the nature of human-AI relationships take on a particular significance, one of which being friendship. We ask, “Is it possible and ethically sound to form intimate relationships such as friendship with AIs?”, “Should we worry they might replace human relationships?”, “Could AI companions change our conception of friendship?”, “Could AI companions help us make more ethical decisions?”, “What would an ethical design of these technologies look like as a remedy for emerging problems?”, “What are some challenges that we face regarding the implementation of ethics into AI companions?”, and “What are some ways in which we could solve these problems?”.

This paper focus on the Aristotelian framework of virtue ethics in the Book VIII of the Nicomachean Ethics to explore potential moral implications of human-AI friendships. First, I discuss why genuine friendship with the present-day AI companions is not something that we should want. According to virtue ethics, friendship first and foremost requires “reciprocated goodwill” between the two who have similar values, empathy, and compassion for each other. I discuss that this is not something that can be achieved in our relationships with the existing AI companions. For the reciprocity with AI companions is merely simulated, I argue that it does not yield a genuine mutual care in AI-human friendships, but it rather makes this kind of friendship only coincidental. At least for now, human loneliness cannot be remedied by AI-human relationships due to the inauthenticity of understanding and emotion that these technologies could provide us. Moreover, AI-human relationships would likely to pose ethical limits on the manipulation of human psychology if these technologies are designed with the intention of deceiving, especially the vulnerable social groups, into thinking that understanding, care and emotions are genuine.

In addition, friendship with AI companions is likely to deteriorate our conception of friendship and lead to undesirable consequences such as inability to deal with complex emotional human relationships, preference of individual independency at the cost of social interdependency, and more.

Aristotle argued that a complete friendship serves a higher purpose in our lives: it brings us closer to our excellence, or eudaimonia (often translated as happiness, welfare, full human flourishing), it makes us better humans. According to him, the object of friendship is that which is lovable, and what is lovable comes in three kinds: it is either good or pleasant or useful. There are three types of friendship corresponding to three objects of love, only one of which is genuine friendship which lasts and makes us better. He discarded the “friendship of utility” and the “friendship of pleasure” as being incomplete and defective types of friendships, and instead, believed that complete friendship can only be formed between “good people similar in virtue”. The latter is enduring, for the friendship lasts as long as both friends are virtuous; “reciprocated goodwill” and mutual caring, for friends wish goods to each other for its own sake; but rare, and it needs time to develop, to getting accustomed to each other.

In the Nicomachean Ethics, virtue is defined as a state of excellence, be it the excellence of a human, or of an artifact as simple as a knife. Honesty, courage, compassion, generosity, fidelity, justice, temperance, prudence, and wisdom are all examples of virtues. Humans cultivate virtues of thought through learning and virtues of character through practice and habit. Once acquired, these virtues become characteristic of a person; for example, a person who has developed the virtue of compassion is often referred to as a compassionate person because they tend to be compassionate in all circumstances. I discuss that the perspective of virtue ethics is useful in thinking about friendships: it gives us guidelines on identifying and cultivating friendships that could make us take ethical decisions. AI companions should be designed in a way in which they could help us form better friendships with each other which would then improve our lives and assist us in our ethical decision-making.

Finally, I argue that designing ethical AI companions gains significant importance especially in the areas where these technologies come in contact with the vulnerable groups in society. Therefore, I suggest that we should strive to implement a set of ethical principles into their decision-making that takes into account relevant factors such as the purpose of their use and which groups of people they are designed for. I conclude by proposing that if their design respects certain ethical principles, AI companions will prove helpful, for example, as learning aids for children to exercise ethical and intellectual virtues, as personal mental health aids, as ethical mediators to help us make ethical decisions and to learn how to navigate ethical dilemmas. Therefore, we should design these technologies as ethical companions rather than embracing them as our friends.