Ethical Evaluation of Human-AI Friendships: Concerns and Considerations for Long-Term Moral Implications

Yixuan Hu *

Shenzhen College of International Education, Shenzhen, China

* Corresponding Author Email: 350277208@qq.com

Abstract. This paper evaluates friendships between humans and AIs from an ethical perspective. I first explore what friendship is and why it is valuable. Then I discuss the unique features of friendships between humans and AIs and their likely moral effects and implications. I argue that we have reasons to be worried about the long-term effects of friendships between humans and AIs from an ethical perspective.

Keywords: AI, friendship, value, ethics.

1. Introduction

Artificial Intelligence is playing more and more significant roles in our daily lives and many realms of society. Algorithms help us make all kinds of decisions including hiring, admission, and even bailing and sentencing decisions. They can also answer our questions, have a conversation with us, offer us advice, produce texts according to our commands, and even become our "friends" or "therapists." In this paper, I provide an ethical evaluation of friendships between humans and artificial intelligence supported by technologies such as Large Language Models. I first explain my understanding of friendship and why friendship is valuable and important to us. I then delve deeper into the features of friendships between humans and AI and the likely consequences of such friendships. I offer reasons for us to be concerned about people relying too much on friendships with AI.

2. What is Friendship?

Friendship, as understood here, is a distinctively personal relationship that is grounded in a concern on the part of each friend for the welfare of the other, for the other's sake, and thatinvolves some degree of intimacy. This understanding is in line with Chiara Cordelli's understanding of friendship as "one particular set of interpersonal relationships, which are voluntary, and to some degree, affective, as well as non-instrumental." [1] According to Cordelli, friendship also has rough reciprocity as its central norm of exchange. Cordelli excludes purely instrumental bonds from her account of friendship even though purely instrumental bonds and useful contacts can also be valuable relationships. So, on this understanding shared by Cordelli and some other philosophers [2, 3], friends must be prepared to act for their friend's own sake, as their final goal, rather than as a means towards some end.

As such, friendship is undoubtedly central to our lives, in part because the special concern we have for our friends must have a place within a broader set of concerns, including moral concerns, and in part because our friends can help shape who we are as persons. Friendship is also central to our wellbeing, partly due to the practical and emotional benefits provided by friendships. Moreover, for many people, friendship can be intrinsically valuable and an indispensable part of a flourishing life. Despite its important role in our lives, forming and maintaining friendships can be challenging and require skills, efforts, and even luck in many circumstances. Access to opportunities of friendships is also unequal and influenced by factors out of our control such as our social class, gender, ethnicity, and innate sociability [4].

3. The Moral Hazard of Human-AI Friendships

With the development of artificial intelligence and large language models, technology is progressing at a speed that makes AI companionship not just a theme of science fiction but a real possibility. AI companions can converse with humans fluently just like our friends would. They can answer our questions, provide support and comfort, and even offer advice. Due to their superb learning abilities, they are able to process and utilize a vast amount of information and resources in our interactions and conversations with them. And their abilities in performing these tasks are only becoming better. Further, unlike real human beings, AI friends would not become upset or angry at us because they do not have emotions of their own. They would not feel frustration, disappointment, boredom, jealousy, resentment, and many other negative emotions that can pose significant challenges to friendships between real human beings. Even if there are miscommunications or misunderstandings, the AI companion would always be available for further communication and clarification that we are able to provide. In simple words, they are always there and will never leave us (unless we decide to stop interacting with them). This picture makes AI companions look like our ideal candidates for friendships and even better than real humans could be.

However, using AI companion can harm people's valuation of friendship in various ways. It can also have severe detrimental effects on our personal relationships and wellbeing in general. First, the basic logic of friendship includes mutual giving and supporting. Unrequited friendship is senseless. The depiction of friendship we often employ goes beyond mere mutual affection I've mentioned above, and instead comprehends it as a union that encompasses significant interactions between friends. In this context, these "vital interactions" encompass not only emotional transactions, such as empathy, possessiveness, and the embodiment of emotional sustenance which are more spiritual, but also mutual motivation and camaraderie in daily life. This includes mutual encouragement and companionship in daily life, a hug in a difficult time, a whole-night heart-to-heart conversation following a setback, and other daily interactions that "convey the assurance of there is always someone behind you."

However, when using an AI companion, the act of "giving" is done unilaterally by the AI, which means the user does not have to think about giving to the AI friendship because the AI has no needs. This can confuse these users' understanding of friendship. In this case, it is easy for people to develop selfish mentalities and social habits and, and their understanding of friendship will very likely shift from supporting and giving to each other to unilaterally venting their emotions and desires. However, when this mentality is applied to socializing with real humans in the real world, it is very prone to cause conflict. In this case, people becoming accustomed to spending time with AI companion will ultimately reduce social cohesion, as it will induce users to adopt behavioral habits and bad manners that are prone to conflict. People's personal relationship, including friendships and other kinds of intimate relationships will likely suffer as their abilities to deal with conflicts, to have difficult conversations, and to maintain relationships with real humans decline. These interpersonal skills and emotional dispositions underpinning these skills, such as empathy, compassion and respect, are not only essential for personal relationships, but also important for non-personal relationships, such as relationships with colleagues, teachers and students, and even our fellow citizens. All kinds of social interactions are made better by pro-social dispositions and excellent communication and interpersonal skills. In the long run, decline in these dispositions and skills will result in a reduction of total social utility.

Given the involvement of love and mutual affection, all kinds of friendship seem to involve a concern for your friend for his own sake and not for your own. Nonetheless, there is widespread agreement that caring about someone for his sake involves both sympathy and action on the friend's behalf. That is, friends must be moved by what happens to their friends to feel the appropriate emotions: joy in their friends' successes, frustration and disappointment in their friends' failures (as opposed to disappointment in the friends themselves), etc. Moreover, in part as an expression of their caring for each other, friends must normally be disposed to promote the other's good for her own sake and not out of any ulterior motive. Consequently, accounts of friendship tend to understand it

not merely as a case of reciprocal love of some form (together with mutual acknowledgment of this love), but as essentially involving significant interactions between the friends—as being in this sense a certain kind of relationship.

To take it a step further, friendship in human society usually involves taking on an extra share of responsibility. In our interactions with real friends, we take care of them and prioritize them in times of trouble or reward because they are special to us. This is a good, socially default habit. Conversely, does responsibility exist in human-AI friendships? It appears that there is not. Even if someone indulges in AI companionship to the point of fantasizing that it's a real human being and thinking of him in everything, which is already extreme and rare, AI companionship itself cannot appreciate and respond to that responsibility; it's essentially just a set of anthropomorphic algorithms. At the same time, as your friend, the AI would not take on the responsibilities he is supposed to take on towards you, just as he would not care about you, but will usually only answer your questions after you've asked them. It is not kindness, empathy, or responsibility that drives him or her to do so. It is algorithms coded by programmers. Here the definition of friendship itself is confused, the keyword "for each other" is ignored in AI friendships, and the user of AI companionship does not become a special being to his or her AI friend.

Due to these considerations, we have reasons to think that friendships between humans and AI are less intrinsically valuable than friendships between humans, even if AI can provide some instrumental value of friendships such as offering advice and emotional support. Therefore, friendships with AI cannot fully replace friendships with humans without significantly undermining the wellbeing of our lives. For people who regard intrinsically valuable friendships as an essential part of the meaning of life, our lives would also become less meaningful if we only, or mostly just have friendships with AI.

Moreover, friendships and other relationships with AI may reinforce morally problematic or oppressive social norms, such as sexist, racist, ableist norms, when the algorithms and data used in designing the AI companions are infected by problematic and oppressive norms. Klonschinski and Kühler have developed a careful critique of romantic love between humans and Ais from a feminist perspective [5]. Relationships between a human and an AI is not romantic love between equals on major accounts of romantic love and are thus not love worth pursuing. They also discuss how female AI assistants tend to reinforce patriarchal norms and traditional gender roles that women should be caring, obedient, and not advocating for her own autonomous desires and needs. These norms also promote “men’s sense of entitlement and the lack of ability to see the partner as an autonomous person instead of one’s property” which is a main reason for domestic violence. Even if AIs cannot be harmed by humans because AIs are not persons and do not have moral status as a person does (or at least this is the kind of AI we are assuming here), unequal relationships between humans and AIs can still be harmful when they diminish men’s abilities to respond to real women’s needs and desires. The authors consider such inability as “a serious social and moral problem”, since “taking other people’s needs and preferences into account lies at the heart of each successful social interaction and of moral action in general.” Therefore, extensive interacting with AI may lead to our unlearning of these abilities and morally undesirable side effects.

References

[1] Cordelli, C. (2015) ‘Distributive Justice and the Problem of Friendship’, *Political Studies*, 63, 679 - 695.

[2] Blum, L. (1980) Friendship, Altruism and Morality. Boston MA: Routledge & Kegan Paul.

[3] Stocker, M. (1976) ‘The Schizophrenia of Modern Ethical Theories’, *Journal of Philosophy*, 73 (14), 543 – 66.

[4] Adler, P., Kless, S. and Adler, P. (1992) ‘Socialization to Gender Roles’, *Sociology of Education*, 65 (3), 169 – 87.

[5] Klonschinski A. and Kühler M. (2021) ‘Romantic Love Between Humans and AIs: A Feminist Ethical Critique’ in *New Philosophical Essays on Love and Loving*, edited by Simon Cushing.