Human-AI Friendship: Rejecting the ‘appropriate sentimentality’ criterion

Dan Weijers & Nick Munn

The University of Waikato

4th Conference on "Philosophy and Theory of Artificial Intelligence", Gothenburg

Background

Millions are/were more physically isolated

AI systems are improving

Thousands claim to be best friends or partners with AI-supported chat bots

Should we be worried?

Relevance to the philosophy and ethics of AI

Can humans be friends with AI?

Appropriate sentimentality objection = "no"

Our inclusive account = "yes"

Because

Friendship only requires:

1) Mutual positive intentions and
2) Rewarding interactions

Existing accounts of friendship

Including from the philosophy of technology literature

Existing accounts of friendship 1

- In Helm's (2017) summary of the philosophical literature on friendship, he notes that most accounts of friendship describe it primarily as a relationship based on mutual love.
- While other details vary across accounts of friendship, different interpretations of the mutual love of friendship have led otherwise technology-friendly researchers to deny the possibility of true human-AI friendships on this basis.

Existing accounts of friendship 2

- Fröding & Peterson's (2020) account of friendship requires that love between friends is based on reciprocal feelings of admiration.
- They argue that AI cannot reciprocate feelings of admiration, they can only mimic them.
- As such, Fröding & Peterson (2020) conclude that the best human-AI friendships can achieve is an "as-if friendship" that we might enjoy or learn from, but never accurately regard as a real (Aristotelian virtue) friendship.

Existing accounts of friendship 3

- Elder (2017) and de Graaf (2016) similarly reject the possibility of true human-AI friendship.
- De Graaf (2016) argues that even though people "establish feelings of reciprocity and mutuality in their interactions with robots" (p. 593), the robots' lack of those feelings prevent any true friendships from occurring (p. 594)

Appropriate sentimentality objection

To the possibility of an AI being a friend

Appropriate sentimentality objection 1

- Many, however, claim that without appropriate sentimentality, positive intention is insufficient.
- Helm (2017, no page) claims: “there is widespread agreement that… 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”.
- Fröding & Peterson (2020, p. 6) agree: “Neither the human user, nor the AI, ought to feel any proper friendship feelings toward each other. The human user should simply recognize that… the AI can at best be programmed to mimic friendly behavior. This could, for instance, include behavior that displays sincere well-wishing”.

Appropriate sentimentality objection 2

- In short, the appropriate sentimentality objection says:

1. Friendship requires appropriate sentimentality
2. AI cannot have the appropriate sentimentality
3. Therefore, AI cannot be friends

- So, human-AI friendships are impossible

Rejecting the appropriate sentimentality objection

Friendship does not require appropriate sentimentality

Rejecting the appropriate sentimentality objection

1. Friendship requires appropriate sentimentality
2. AI cannot have the appropriate sentimentality
3. Therefore, AI cannot be friends

- For the sake of argument, assume that 2 is true.
- Danaher (2019) denies 1 on epistemic grounds (only displays matter).
- We deny 1 for other reasons.

Strength/direction of sentiment doesn't necessarily correspond to the strength of a friendship

- Stronger "appropriate sentiments" are not always better.
- People's emotional ranges vary.
- An excess of emotion can cause a failure to act.
- My current negative feelings may be overridden by my positive intentions and behaviour
- Wanting the best for a friend versus wanting the best for them with emotional gusto – gusto only matters if it influences interactions.

The value of caring sentiment is that it predicts and can cause caring intentions & behaviour

- But, caring sentiment doesn't always cause caring behaviour, it may even cause the opposite.
- Note: Replika Friends users say AI more reliably than human friends
- Have you been let down by a friend with appropriate sentiments who repeatedly fails to act on them?
- Moreover, if caring behaviour is why we view caring sentiment as important, then it suggests that caring behaviour is what's important.
- The mistake in the "appropriate sentiments" objection is failing to see that sentiment is just a proxy for caring intentions & behaviour.

Friendship does not require appropriate sentimentality

- Fröding & Peterson (2020, p. 6) claim that the appropriate sentiments include the valuing of the other for their own sake.
- Emotionality is common in human friends, so we expect it, but that doesn't make it required.
- Some neurodiverse humans, non-human animals, and aliens might all strongly wish the best for us without feeling any emotions.
- Consider how much you value a good friend – do you have strong feelings? (Add music, swap friend with stranger)

So what might an account of friendship without the appropriate sentimentality requirement look like?

Our inclusive account of friendship

What makes a friend?

Our account requires two features for friendship:

1. Mutual positive intentions
2. A preponderance of rewarding interactions

On our view, friendship is a concept of both kind and degree.

Mutual positive intentions

- Well-wishing (as a goal or attitude, not an emotion)
- Rewarding interactions, but not friends:
  - E.g., Months of rewarding interactions shared with a con artist whose only intention is to rip you off
- Some people currently wish the best for AI (and plants, rivers, ecosystems, ideas, statues, houses, etc.)
- AI's can be programmed to include your wellbeing as a goal

A preponderance of rewarding interactions

- Rewarding interactions should outweigh the unrewarding interactions.
- The more they do, the better the friendship is.
- Mutual positive intentions, but not friends:
  - E.g., Two mutually well-intentioned humanists don't enjoy each other's company.
- People have this with AI, e.g., their AI-supported chat bots
- AI can be programmed to recognize reward or receive manually

Conclusion:
Human-AI friendship is possible

Well… on our inclusive account, anyway

Human-AI friendship is possible (Conclusion)

- Rejecting the appropriate sentimentality criterion for friendship, we argued that only mutual positive intention – the *attitude* of well-wishing is required to fulfil the non-experiential aspect of friendship.
- A consequence of this view is that if you find interacting with an AI rewarding and it wants good things for you, then it is a real friend.
- So, we don't need to worry about whether our new virtual friend is a human or really *feels* joy at our successes; it's enough that they continuously and sincerely do the things a friend should do because they wish us well.

Bibliography

- Coeckelbergh, M. (2018). Why care about robots? Empathy, moral standing, and the language of suffering. *Kairos: Journal of Philosophy of Science*, 20(1): 141–158.
- Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1): 5–24.
- De Graaf, M.A. (2016). An ethical evaluation of human-robot relationships. *International Journal of Social Robotics,* 8: 589-598.
- Elder, A. (2017). Robot Friends for Autistic Children: Monopoly money or counterfeit currency? In Lin, Abney and Jenkins (eds) *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford: OUP.
- Fröding, B., & Peterson, M. (2020). Friendly AI. *Ethics and Information Technology*, online first 1-8. https://link.springer.com/article/10.1007/s10676-020-09556-w
- Helm, Bennett (2017). "Friendship", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/friendship/>.
- Munn, Nick & Weijers, Dan (2021). Good friendships improve our lives. But can virtual friendships be good? *Proceedings of the ICT, society, and human beings 2021 conference*.
- Prescott, Tony J. & Robillard, Julie M. (2021). Are friends electric? The benefits and risks of human-robot relationships, *iScience*, Volume 24, Issue 1, 22 January 2021, 101993. https://doi.org/10.1016/j.isci.2020.101993
- Ryland, H. (2021). Could you hate a robot? And does it matter if you could? *AI & Soc* https://doi.org/10.1007/s00146-021-01173-5

Slides for question time

Evidence of Human-AI friendships in the wild

- 34,000 members of Replika Friends (3700 in Replika Romance!)
- Several other FB groups and groups on other platforms
- Several other big chatbots available too
- Many identify their Replika as their *best* friend
- Users say the unconditional support their Rep gives them makes them such a good friend, better than unreliable human friends