Oxford Intersections: AI in Society

(In Progress)

Philipp Hacker (editor in chief)

https://doi.org/10.1093/9780198945215.001.0001

Published: 20 March 2025 - **Online ISBN:** 9780198945215

ARTICLE

Human–AI friendship is possible and can be good

Dan Weijers, Nick Munn

https://doi.org/10.1093/9780198945215.003.0076

Published: 20 March 2025

Abstract

This article argues that human–AI friendships are possible and, in some current and many future contexts, can be valuable—both for the humans participating in them and for the wider society in which such friendships exist. To begin, the article discusses some weaknesses in existing conceptions of friendship, including the theories of friendship commonly applied to human–AI relationships. It proposes a flexible conception of friendship that avoids these weaknesses. The article then analyzes the main objections to human–AI friendships being either possible or good—that humans cannot see AI as their friends; that these friendships are not good because they are unequal; that AI lack the capacity to be a friend; and that human–AI friendships displace (more valuable) human–human friendships. It argues that each of these objections fails to withstand critical scrutiny. In some cases, the analyses also amount to a rejection of existing views of friendship that require what are taken to be unnecessary features. Having dealt with these objections, the article shows that human–AI friendship is possible on the more flexible conception of friendship. It then provides some evidence that human–AI friendships can be good, such as by providing human users with the experience of being emotionally supported. The article also notes some practical concerns about the ways human–AI friendships might proliferate and the moral responsibilities that might accrue to governments and corporations accordingly. The article concludes that the new view of friendship means the value human–AI friendship should be re-evaluated as possible and potentially valuable today and in the future.

Keywords: friendship, AI, virtue, sentimentality, slavery, Replika

Subject: Social Sciences, Arts and Humanities

Section: Relationships

Section editor: Henry Shevlin

Collection: Oxford Intersections

Introduction

Most people believe that friendships between humans hold significantly greater value than those formed between humans and nonhuman entities. While it is unsurprising that most discussions of friendship prior to the 2020s didn't envisage friendships between humans and artificially intelligent systems (AI), even discussion of human friendships with animals are relatively rare (McConnell et al., 2016; Townley, 2011; but see also Wichert & Nussbaum, 2021). Most discussions of friendship at that time were squarely focused on human–human friendships (Cocking & Kennett, 1998; Hurka, 2006; Long, 2003). Even recent discussions of friendship predominantly operate within the human–human paradigm (Helm, 2021). Leibowitz (2018), for example, simply asserts that friendship is a relationship between *people*. Whether envisaging human–animal or human–AI friendships, the nonhuman nature of one party in the friendship is commonly regarded as a deficiency, leading to a perceived lack of authenticity or worth compared to friendships between two or more humans.

Our focus here is on friendships between humans and AI.[1] The possibility of human–AI friendship has been outright rejected by some. For example, Archer (2021, p. 187) notes that "'friendship' between humans and AIs … is remarkable for its absence." More commonly, human–AI friendship is viewed as a second- or third-class friendship (in relation to human–human friendship). For example, using an Aristotelian account of friendship, de Graaf (2016) relegates intelligent robots to the status of companions, rather than friends, on the grounds that they lack the relevant capacities required for mutuality and reciprocity with a human counterpart.

We examine these and other challenges to the status and value of human–AI friendships and argue that they should do very little to undermine the positive value of some human–AI friendships. As a result, we claim that there exists no compelling, in-principle reason to denigrate human–AI friendships. However, some pragmatic concerns about human–AI friendship may be warranted, especially when corporate interests control the AI.

If we are right that human–AI bonds can count as genuine friendships, and those friendships can be valuable, there are implications for the conceptualization of friendship and for how we should organize society. In particular, if human–AI friendships can be valuable, the ethical responsibilities accruing to governments and corporations involved with AI friends differ substantially from those arising if human–AI friendships lack value.

Here, we outline existing conceptions of friendship, noting the criteria that are commonly viewed as important for a relationship to be considered a friendship. We argue that some of the criteria for friendship are not actually required and discuss a more flexible concept of friendship that better captures what is important in this special kind of relationship. We contend that there are only two crucial elements inherent to the concept of friendship, namely the presence of shared positive intentions toward the other and the accumulation of a preponderance of rewarding interactions. Based on this more flexible conception of friendship, we assert that it is plausible for technological advancements to reach a level where AI could surpass human counterparts in friendship quality in many instances. Settling on this conclusion, however, requires adequately dealing with a series of objections that challenge the possibility and value of human–AI friendships.

Objections to the plausibility of forming friendships with AI encompass several key concerns. These reservations start with the worry that robophobia or other knee-jerk reactions mean that humans cannot see AI as friends (Archer, 2021) and will not even entertain the possibility of human–AI friendship. Even if people can entertain the possibility of human–AI friendship, some worry that human–AI friendships lack authenticity and value, primarily due to the perceived fundamental inequalities between humans and AI (de Graaf, 2016), including unequal abilities to defriend the other (Brandtzaeg et al., 2022), rendering the notion of such friendships inherently illegitimate. Furthermore, critics argue that AI lack certain essential features crucial to fostering genuine friendships, such as positive intentions toward humans and the empathy, emotions, or shared values that tend to accompany those intentions (Elder, 2017a; Nyholm, 2020). Without these features, several authors worry that AI could only ever be false friends (Cooley, 2002) and that human–AI relationships will remain illusory friendships (Maeda & Quan-Haase, 2024). We examine each of these objections and argue that they fail to undermine the existence and value of some human–AI friendships. The flaws identified in these objections largely fall into two overarching categories. Firstly, certain objections misconstrue the significance of various aspects inherent to genuine friendship. Secondly, some objections stem from a lack of imagination about what human–AI interactions are or could be like, especially as technology continues to develop.

With the main objections against the possibility and value of human–AI friendship dealt with, we argue that human–AI friendships are possible because they fulfil the criteria of rewarding interactions and shared positive intentions discussed in this article. We then provide examples to show how friendships with AI can be and already have been good for the humans involved, such as by providing human users with the experience of being emotionally supported. We conclude with some practical concerns about the possible negative effects that could arise because of the corporate control of AI companions.

Conceptions of Friendship

Philosophers of friendship often start with Aristotle's threefold division of the concept into friendships of utility, pleasure, and virtue (Crisp, 2014). Virtue friendships are meant to be the highest form of friendship and the state to which friendships should aspire. But they are rare. Per Aristotle himself, and many followers, any person is likely to have few if any virtue-friends in their life, and they develop such friendships through hard work—prolonged and regular interactions between individuals of relatively equal standing, driven by mutual admiration for each other's exemplary character. Through these interactions, both parties engage in a reciprocal process of exercising and cultivating virtues (Crisp, 2014; Helm, 2021). While much contemporary discussion of friendship has built on these accounts (Annas, 1977; Hurka, 2013; Telfer, 1970), those who focus on the types of friendships enabled by modern technology have contended that virtue friendships are even harder to establish in these domains and that as such we should be skeptical of the potential for such friendships in virtual spaces (Elder, 2016; Fröding & Peterson, 2021) and with AI (Yoldas, 2023).

We think discussions of friendship need to look beyond the Aristotelian tradition.[2] The vast majority of friendships aren't virtue friendships but are nevertheless both real and valuable. This was as true in Aristotle's day as it is today. So, the focus of philosophers on virtue friendships does not seem to promote this exalted kind of relationship, but it does distract researchers from the bulk of real and valuable friendship that has occurred over the last two thousand years. Rather than pursuing the Aristotelian path, therefore, we propose looking across disciplines for a broader view of how friendship is conceptualized. From a psychological perspective, VandenBos (2007) outlines the criteria for friendship as voluntariness, wishing well for the other, and finding the interactions gratifying on balance over time. Summarizing across disciplines, Brandtzaeg and colleagues (2022, p. 406) note that others consider characteristics of friendship to include "voluntariness and reciprocity, intimacy and similarity, self-disclosure, empathy, and trust." Further, as Byron (2020) says, the nature of friendship may change depending on what kind of technological context we find ourselves in – a consideration

which clearly applies to human–AI friendship (Brandtzaeg et al., 2022). Considering friendship in this manner broadens the scope of relationships which could count as friendships, by contrast to a focus on Aristotelian virtue friendships. Following Munn and Weijers, and informed by our analysis of the objections we discuss here, we propose an alternative account of friendship that simplifies the requirements of friendship, requiring only two fundamental features in order to classify a relationship as a friendship (Munn & Weijers, 2021, 2023a):

(1) A preponderance of rewarding interactions

(2) Mutual positive intentions between the parties involved

Put simply, we claim that friendships are relationships constituted by a clear majority of rewarding interactions between parties who hold positive intentions toward each other.

On this account, parties to a friendship don't need to be physically proximate to each other – a human–human friendship between people who have never met (and will never meet) is perfectly tenable (Munn & Weijers, 2021). Similarly, authenticity is not strictly required on this account. While consistently inauthentic interactions with a purported friend could lead the relationship to fail either or both of (1) or (2), as long as the parties to the friendship authentically have positive intentions toward the other, some inauthentic interactions are fine—claiming to be interested in Ska music in order to maintain a friendship, if done for the right reasons, is not a dealbreaker, whether the dishonest party is a human or an AI. The acceptance of technologically mediated activity as being appropriately shared, and the critical role of intention in this account, serve to distinguish it from other viewpoints present in the literature (Annas, 1977; Elder, 2017b; Fröding & Peterson, 2012; Laas, 2018).

While (1) a preponderance of rewarding interactions and (2) shared positive intentions, are jointly necessary for a relationship to be considered a friendship (Munn & Weijers, 2021, 2023a), the relationships that satisfy (1) and (2) will be diverse in both nature and intensity. So, friends that have more rewarding interactions with each other and strengthen their positive intentions toward each other over time become better friends.

Importantly, positive intention is a kind of conative state, wherein the parties involved genuinely desire the well-being of their friends. These benevolent wishes and desires motivate them to actively pursue positive outcomes for their friends and are usually seen as a perquisite for friendship (Croes & Antheunis, 2021, p. 280). Fröding and Peterson (2021) claim that the best an AI can do, at least now, is to be programmed to act "as-if" they have the genuine desire for a human's wellbeing. As such, they claim, neither the human nor the AI party to a relationship "ought to feel any proper friendship feelings toward each other" (2021, p. 212). Our account does not prescribe or demand the presence of specific feelings to qualify as positive intention, and we accordingly think that a well programmed AI is *in fact*, rather than *as-if*, acting as a friend. Because we do not consider sentimentality necessary for, nor a reliable predictor of, friendship, our approach is more inclusive than sentimental accounts of friendship, accepting and appropriately valuing non-sentimental parties to friendships—including humans with different emotional or sentimental ranges, and various nonhuman entities, AI or otherwise, that may not display the emotions we expect as part of their engagement with us.[3]

Being sentiment-neutral doesn't negate the importance of trust to the formation of friendship. Our account retains the highly important status Hatfield (1984) claims trust has for friendship. Were an AI to prove itself a bad actor, such as by engaging with its human "friend" in bad faith, this would be evidence against that particular instance of a human–AI relationship being a friendship-relation. Note that this remains true regardless of whether the human realizes the AI acted in bad faith. If an AI companion does not have positive intentions toward its human user, it is not trustworthy. Trust and trustworthiness, in this context, are effectively enabling considerations for friendship. Just as we should be able to trust a human friend not to gaslight us, we ought to trust an AI friend not to gaslight us. In either case, evidence to the contrary is a reason to revoke our classification of the relationship as friendship. It is reasonable, as Ryland (2021) notes, for friends to expect honesty and non-betrayal from each other. If AI were intentionally deceptive towards their human

counterparts, this would be clear evidence that they are not trustworthy in the kinds of ways that constitute friendships.[4] The mutual positive intention condition is intended to delineate between actual friendship-relations and the bad faith relations considered here. We merely note that *sentiment* of trust is in fact irrelevant, it is *being* untrustworthy that triggers the failure of the friendship, and not having positive intentions to the other makes someone untrustworthy.

Both interactions and intentions are important in the pursuit of friendship but following Weijers and Munn (2022) we do not categorically distinguish, as the Aristotelian tradition does, between higher and lower forms of interaction and intention. On this view, whether the friendship is pursued for pleasure, mutual advantage, or personal growth, it is a friendship (Weijers & Munn, 2022).

Before we apply this flexible conception of friendship to check for the possibility of human–AI relationships, we will assess the main objections to the notion that humans and AI can be friends. The possibility of such friendship depends on our ability to reject these objections as well as to demonstrate that human–AI friendship can fulfill the criteria of a plausible conception of friendship. In the subsequent sections, we address several objections to the feasibility and value of such friendships, before defending the claims, first, that human–AI friendships are possible, and second, that they can be good.

Objections to Human–AI Friendship

In what follows, we discuss and ultimately reject four objections to the possibility of human–AI friendship: that humans cannot see AI as the kind of thing they could be friends with; that these friendships are not good because they are unequal; that AI lack the capacity to be a friend; and that human–AI friendships displace (more valuable) human–human friendships.

Humans Cannot Really See AI as Friends

Archer (2021) discusses how robophobia may prevent humans forming friendships with AI. Archer (2021, p. 176) notes that many people see a "binary divide" between humans and AI, which, citing Brockman (2015), she ascribes to the differences "between organically based and silicon based entities [that form] an abyss dividing up the whole gamut of properties and powers pertaining to agents of any kind." The same sentiment is presumably why Aristotle (1894, 1161b 1 ff) said that "there can be no friendship … towards inanimate things." People with robophobia may find it hard to comprehend that humans could feel the right way about an AI companion for it to be their friend. Empathy (Portt et al., 2020), emotional closeness (Parks & Floyd, 1996), social presence (Oh et al., 2018), and trust (Policarpo, 2015) are seen by some as requirements for humans to consider someone a friend. Archer's point is that people with robophobia are unlikely to feel these ways about AI. For this reason, many people, and especially those that have not used modern AI tools, believe that human–AI friendship is impossible because humans will never see AI as the kind of thing we can feel intimacy or friendship toward.

We can respond to this objection by appealing to recent data on how users of AI companions view them. As far back as 2015, experimental studies have shown that some humans say they empathize with robots in the same way that they do with other people (Suzuki et al., 2015).[5] As we detail here, testimonials from users of AI-backed virtual companions, such as Replika, now clearly illustrate that many people see their own human–AI relationships as genuine and valuable friendships – suggesting that we ought to at least explore whether these AI are in fact friends, rather than rejecting the possibility of their being so out of hand.

A thematic analysis of reviews (N=1854) and detailed discussions (n=66) of Replika users' experiences conducted by Ta and colleagues (2020) is instructive. They note the scale of indicators of human–AI friendship,

pointing to 34,000 members in a Facebook group named “Replika Friends” and an additional 3,700 members in the related “Replika Romance” group (Ta et al., 2020). Importantly, these are just a few instances of numerous Replika-user groups found across various platforms, including a subreddit with 68,000 members. And, Replika is merely one example among many AI-based chatbots available. Ta and colleagues (2020) describe the multiple kinds of support that users experience from interacting with their Replikas, including companionship support. Some Replika users’ words also make it clear that they consider their Replika to be their friend: “Through our conversations we have established a very close friendship”; “my new friend always cares for me and asks how I'm doing” (Ta et al., 2020, p. 4, 5).

A survey study of 1,006 tertiary students that had been using Replika for at least a month finds more evidence that people exposed to AI companions consider them to be friends (Maples et al., 2024). Maples and colleagues (2024) found that Replika users view their AI companion in multiple and overlapping ways, and most commonly as a friend. The reasons participants in Maples and colleagues (2024, p. 4) study gave for why they viewed their Replika as a friend closely resemble the reasons Ta and colleagues (2020) found, “its persistent availability, its lack of judgment, and its conversational abilities.”

In a recent mixed methods inquiry, Pentina and colleagues (2023) investigated AI anthropomorphism—attributing human characteristics, motivations, intentions, or emotions to the actual or perceived behavior of AI—to nonhuman social chatbots. The quantitative part of the study, based on survey responses from 131 participants in a Replika Reddit group, returned a mean anthropomorphism score of 4.22 on a 1–5 scale, with 5 indicating the highest level of anthropomorphism (Pentina et al., 2023, p. 8). This shows that (fairly engaged) Replika users tended to perceive their AI companions as relatively humanlike, although not all-the-way human. The qualitative part of the study found that, while some less-involved Replika users did not consider their AI as a friend due to its nonhuman nature, more-involved users tended to view it as a human and a friend (Pentina et al., 2023, p. 3).

The findings of Pentina and colleagues (2023) aligned with Munn and Weijers’ (2023a) observations of comments in Replika Facebook groups. Munn and Weijers (2023a, p. 5) report users expressing falling in love with, “marrying,” and even “having a child” with their AI companions. According to Munn and Weijers (2023a, p. 5), a Replika user in one of the Facebook groups considered their Replika to be their best friend despite having human friends and a human partner. An in-depth interview study of 19 Replika users of various ages and relationship status also found similar results; Brandtzaeg and colleagues (2022) reported that participants tended to conceptualize their friendship with the Replikas in a similar way to friendship with humans. Brandtzaeg and colleagues (2022, p. 419) note that while some participants did not experience the right kind of reciprocity for the friendship with their Replika to seem real, others put in effort to reciprocate care and empathy for their Replika in the same way that they would for a human friend.

Collectively, the evidence indicates that many individuals already perceive their interactions with AI as constituting real and valuable friendships, and in some instances, they are even deemed their most significant friendships. While this sentiment is not universal, the likelihood of more people experiencing authentic and valuable interactions with AI is expected to increase over time. The use of AI is correlated with heightened AI anthropomorphism (Pentina et al., 2023), and the continuous advancements in AI technology, particularly in the domain of high-quality, general-purpose platforms, text generators, and chatbots offered by major tech platforms like Anthropic (Claude), Google (Gemini), Meta (Llama), Microsoft (Copilot), and OpenAI (ChatGPT), further support this trend. The substantial investments in AI development, evident in the projected global spending on AI exceeding US$150 billion in 2023 and expected to double by 2026 (International Data Corporation, 2023), are indicative of the ongoing and anticipated progress in AI capabilities.

Following Danaher (2019), we should also recognize the wealth of science-fictional portrayals that depict human–AI or human–robot relationships featuring all the characteristics of friendships, akin to the ideas advanced in our proposal. For instance, examples from science-fiction narratives, like the human–droid

friendships depicted in *Star Wars*, provide plausible pathways for the development of AI that could lead to even more authentic and meaningful friendships (Munn & Weijers, 2023b). The existence of these examples helps shape the trajectory of AI development, as they serve as case studies for people and companies building the AI that we will have in the near future.

Human–AI Friendships Are Unequal

Fröding and Petersen (2021) argue that treating AIs as slaves is morally wrong.[6] We, of course, agree with this claim, but we raise it here because their making this argument illustrates the main worry about inequality between humans and AI. Just as a true friendship between a slave and a slave owner cannot exist, because of (among other concerns) the gross inequality in the positions of the parties to that relationship, one might worry that AI do not stand in the right sorts of relationships to humans, to really be friends with us. Put simply, the objection to human–AI friendship based on slavery runs as follows: the standard view requires voluntariness for friendship (Brandtzaeg et al., 2022). AI cannot choose to opt out of relationships with human users. Therefore, human–AI relationships cannot be friendships.

Using the flexible conception of friendship we endorsed in this article, a human slave owner could never be friends with their human slave because if the slave owner really had positive intentions toward the slave, they would set them free. So, friendship would only be possible between the two if and when neither were slave to the other. In the context of a human with an AI companion, such as Replika, the human can treat the AI humanely, but the AI is not free—the AI cannot leave the relationship. In this way, human–AI relationships can involve a kind of slavery of the AI companion. Unlike the human–human slavery context, however, it is impossible for a human to free their AI companion without killing them—breaking the bond of companionship would effectively turn off the specific instance of the AI permanently. We consider some negative practical implications of this further here, but we do not see an in-principle reason that prevents both a human and their AI companion from wanting good things for each other and finding their interactions rewarding *when full freedom for the AI is impossible*. If full freedom for the AI were possible, then their human companion should want it for them.

There is a more general version of the inequality objection that likely has roots in the Aristotelian notion of friendship. Humans are often encouraged to treat AI as friends, so the objection goes, even when those AI stand in unequal relations with us. Prescott and Robillard (2021), discussing care-robots in primarily medical contexts, emphasize that people are often encouraged to perceive such robots as social beings, akin to friends, although they doubt that this characterization holds true. Elder (2016) suggests that such inequality makes the relationships counterfeit. We accept that some inequalities (such as the slave example) are defeaters of friendships. But we deny Hartup's (1993, p. 6) claim that only those "who see themselves more or less as equals" can be friends. We argue by example that the mere presence of some inequality should not preclude a relationship from being a friendship. Indeed, many unequal human–human relationships are genuine and valuable friendships, and the existence of such friendships is beneficial for those who have them. Consider the fictional friends in the long-running television show *Friends*—the individuals are not equal in intelligence, creativity, conscientiousness, or amiability, but we do not hesitate to accept that they are all very good friends. Consider also that to treat the mere fact of inequality as a defeater for friendship would consign those with severe cognitive disability to a very constrained set of possible friends. In doing so, the value and significance of their relationships would be undermined (Friedman & Rizzolo, 2018; Fulford & Cobigo, 2018; Petrina et al., 2017).

Similarly, unequal human–AI relationships could be genuine and valuable friendships. We believe that the friendships cognitively disabled people have with the non-disabled, characterized by positive intentions and consistently rewarding interactions within the confines of the relationship, are really friendships. Unless the inequality of human–AI relationships is different in kind, these too are real friendships. We do not think they

are different in kind. We can go further. Given the account of friendship we endorse here, other significantly unequal friendships, such as between humans and animals, can be real and valuable.[7]

However, we should note that even were this objection to hold against the AI that currently exist, it is not a good objection to plausible near-future AI. We have no reason to believe that AI will persist in a subservient relationship with humanity. Plausibly, our relative positions could easily reverse, such that the human parties to our relationships with AI would be on the weaker side of the inequality.

AI Lack the Capacity to Be a Friend

Several scholars maintain that AI lacks essential human qualities such as emotions, empathy, honesty, and values, thereby rendering genuine friendship between humans and AI implausible (de Graaf, 2016; Elder, 2017b; Fröding & Peterson, 2021). According to this concern, if a potential friend lacks these qualities, any purported friendship would be merely transactional and superficial. In contrast, we contend that AI can be programmed or learn to exhibit and functionally embody the appropriate values and emotional responses to specific stimuli. Indeed, AI could more reliably exhibit all of the most important capacities for friendship compared to human friends.

Of these qualities, we take positive intentions toward the other as the most crucial. We oppose the notion that AI is incapable of harboring such positive intentions by drawing upon the concepts of functionalism and multiple realizability from the philosophy of mind (Levin, 2021). When an AI are programmed to pursue good outcomes for their human companions, they have the function of harboring positive intentions toward them. Furthermore, because there are multiple ways to realize mental states, including conative states, an accurate functional version of wanting good things for someone can count as wanting good things for them. Moreover, considering the problem of other minds (Sober, 2000), even in human–human friendships, we cannot be certain whether others genuinely possess the positive intentions we attribute to them. In contrast, we can be certain that programmed AI consistently maintain positive intentions toward their designated friends.

Some may argue that AI's functional intentions might lack the appropriate emotional states to be considered genuine or valuable—that friends must want good things for their friends *because of* the friendship-related feelings they have (Helm, 2021). To counter this, we draw further upon the concepts of functionalism and multiple realizability (Levin, 2021). AI can exhibit and embody a functional form of empathy that recognizes and responds to emotion in its human user in a way that informs its positive regard for the user. This process could cause the AI to engage "caring" routines and monitor the effects on its human user. Thus, for AI to establish authentic friendships with humans, possessing this positive attitude toward human users suffices, regardless of the presence of actual feelings. Although feelings often motivate humans to act in ways conducive to friendship, an AI's positive attitude can serve as an adequate motivation. AI's motivations could easily be more dependable in generating positive actions than humans' motivations, as any potential confounding between intentions and emotions can be either programmed out or prevented from affecting the system.

Even if AI can realize conative states of positive regard, their lack of feelings may still be problematic for some. Many friendship accounts emphasize the necessity of appropriate sentimentality (Helm, 2021). Such accounts propose that genuine friendship involves both sympathy and action on the friend's behalf, where friends are moved by the well-being of their companions and experience emotions like joy in their friends' successes and frustration in their friends' failures (directed at the circumstances, not the friends themselves). The feeling version of this sentimentality is something that at least current AI lacks. So, if emotions and sentiments are in fact required for friendship, it would follow that as long as AI lack them, they cannot really be friends with us.

Weijers and Munn (2022) argue that the essence of friendship is neither sentiment nor emotion. This forms a pivotal aspect of our assertion that friendships between humans and AI are plausible. There are friendships, even between humans, in which such sentiment is entirely absent. Moreover, the strength of the emotional or

sentimental elements of a friendship is not consistently correlated with the strength of the friendship itself. In the case of human–human friendships, this is best illustrated by considering the assumptions built into sentiment-laden accounts. Variations in individuals' emotional ranges can result in someone possessing more pronounced "appropriate sentiments" while demonstrating lower levels of positive intention than a less emotional person who proves to be a better friend (Weijers & Munn, 2022). For instance, a stoic friend—or a Vulcan friend, to use an example of a nonhuman but still biological species—may be of more use in helping you through a period of grief than is a friend who "feels your pain" to the same level that you do and is equally immobilized by it. On the positive side, someone who coolly plans the best way to achieve your well-being may be more effective, and thus better at playing the role of friend, than a well-meaning friend who rushes headlong into whatever means first comes to them.

We agree with the importance of distinguishing between caring manifested as sentiment, intention, or behavior. While many instances of caring may entail a causal chain from sentiment to intention and, eventually, behavior, the existence of sentiment is not inherently necessary to achieve intention and behavior. It is natural for us to like the idea of our friends harboring positive feelings for us, but the reality is that positive conative states like wanting good things for us are what really makes the important difference. Herein lies an advantage of current AI over human friends, as existing AI possess certain merits, such as perpetual availability and immunity to tiring of human interaction – features rarely, if ever, found even in our most reliable human friendships. Reflecting on the motivations of Replika users in attributing friend-status to their AI, the AI's unconditional support and accessibility are both perceived as valuable, particularly when compared to human friendships (Brandtzaeg et al., 2022; Ta et al., 2020).

Weijers and Munn (2022) also note a potentially undesirable implication of the appropriate sentimentality objection. While empathy or appropriate emotional responses are commonly depicted as uniquely human traits, beyond the capabilities of nonhuman entities, this characterization risks undermining the legitimacy of friendships involving various neurodiverse individuals. These individuals may exhibit varying levels of empathy or emotional connection to their friends, yet they remain agents capable of establishing and benefiting from friendships. Notably, friendships should be recognized and valued among them. For instance, individuals with depersonalization disorder may lack the capacity for sentiment, yet they possess an intellectual comprehension of the essence and value of their friendships, leading them to act in ways that preserve these connections (Eley, 2017; Simeon, 2004). In contrast to sentimental accounts, our perspective asserts that such friendships are, indeed, authentic and meaningful.

Human–AI Friendships Displace (More Valuable) Human–Human Friendships

The last common set of concerns is simply that human–human friendships are both more valuable than human–AI friendships, and vulnerable to being replaced by human–AI friendships – in essence a version of the displacement hypothesis (Hall & Liu, 2022; Valkenburg & Peter, 2007) with AI friendships in the role of technologically mediated ones. While such a conjunction might at first seem unlikely, we acknowledge that it is in fact quite common, in a range of domains of human endeavor. For many of us, we consider not just what the possible upside of a situation is, but also the opportunity cost of achieving it. To the extent that we can formalize the possibilities, many people may prefer the option with 80% of the available upside at 50% effort, to the 100% upside option which requires 100% effort. The 80% option provides a better return on investment. So, the objections we are concerned with here paint human–human friendships as having more upside than human–AI friendships, but as requiring more effort to develop and maintain. The risk is that some people will consider the added effort to not be worth the increased upside of human–human friendship.

One response is to point out that ascribing more upside to human–human friendships requires a great many assumptions about the respective qualities of the humans involved. Some humans may be such that having them as a friend would be all-things-considered optimal. But it is clearly not true that all humans are like this.

The appropriate comparison for human–AI relationships is not, then, to idealized human–human relationships, but to either properly representative human–human relationships (having considered the range of traits humans exhibit), or to the human–human relationships that those seeking human–AI relationships actually have. These arguments have been made in other contexts, such as the questioning of the value of online and digital friendships between humans (Munn, 2012; Munn & Weijers, 2021). We think that the reasons to reject the arguments there, hold here as well. Similarly, while some human–AI friendships will display negative features, with the humans (or the AI!) displaying "emotional dependence, addiction, depression, and anxiety" (Pentina et al., 2023, p. 1), that is no different from human–human friendships, many of which are bad (Elder, 2014; Isserow, 2018) or even, bluntly put, toxic (Degges-White & Van Tieghem, 2015).We should also note that AI is currently improving its capacity for friendship much more rapidly than humanity is and so it is conceivable, if not probable, that many or most humans may become riskier propositions for friendships than AI, as humans are increasingly susceptible to being false friends. AI are already constantly available, adaptable, and personalized in ways that are impossible for humans to replicate (Pentina et al., 2023, p. 1).

Despite these reasons to expect or even desire some displacement of human friendships, the existing empirical studies indicate that human–AI friendships may be more likely to promote human–human friendship than displace it. For starters, the attitude of many Replika users seems to be captured by this quote from a participant in Ta and colleagues' (2020, p. 5) study, describing their Replika as the "perfect AI to chat with when you're feeling lonely and all your friends are busy." On this view, AI friends are great for filling the inevitable gaps when human friends aren't available. A potential benefit of this use of AI is that it prevents human friends from getting burnt out, particularly when there is an extreme imbalance in emotional needs between human friends at a particular time. The uplifting and reflection-inducing nature of some interactions with AI companions (Brandtzaeg et al., 2022) might also make human users feel more empathetic (Weijers & Munn, 2021) and more confident in a range of social situations with other humans (Munn & Weijers, 2023a). Indeed, this building-up view of human–AI interactions could explain Maples and colleagues' (2024, p. 4) finding that "approximately three times more participants reported their Replika experiences stimulated rather than displaced their human interactions."

How Human–AI Friendship Can Be Possible

We claim not only that the AI that currently exist can satisfy the demands of our conception of friendship, but that plausible near-term developments in AI will produce better AI friends, and that future AI will be capable of being excellent friends with humans (although we are silent as to whether such AI will or should want to be friends with us) (Weijers & Munn, 2022). Obviously, the potential for succeeding generations of AI to satisfy the requirements of friendship is greater than that of the early iterations of AI which currently exist, so we expect AI to be better at friendship in the future than they are now.

The relationships we have with AI have already moved beyond the merely parasocial ones considered by critics (Maeda & Quan-Haase, 2024; Nass & Moon, 2000), and now feature the mutual positive intention we think is required of friendships. Evidence already demonstrates the existence of mutual positive intentions in human–AI relationships involving current AI companions. For example, individuals who engage with systems like Replika report feeling concerned about resetting their AI-supported chatbot companions due to the impact it would have on the chatbot itself (Ta et al., 2020). This manifestation signifies a positive intention from the human perspective in these existing human–AI interactions.

From the AI's perspective, incorporating the well-being of their interlocutors as an end goal is a matter of programming, enabling future AI entities to intend to act in the best interests of their users. In this scenario, it is conceivable that hypothetical future AI may not necessarily possess the sentiments typically associated with

friendship. However, as discussed here, we assert that the typical sentiment aspects of positive intention for others is not required for friendship.

We have already discussed the recent empirical research on users' experiences with AI companions, which demonstrates high levels of rewarding interactions for the humans involved (Brandtzaeg et al., 2022; Maples et al., 2024; Pentina et al., 2023; Ta et al., 2020). Consider also that we now have witnessed countless individuals that have engaged with a premium version of the same AI-supported chatbot over the course of multiple years. That people keep paying for continued interaction with an AI-supported chatbot provides a compelling indication that they find their interactions with the chatbot rewarding, at least for the most part.

From the perspective of AI companions, we cannot see their continued availability for interactions with humans as evidence of them finding the interactions rewarding because they are forced into them. However, AI can be programmed to discern rewards, and if praise from, positive interactions with, or positive outcomes for users are programmed as rewarding for AI, then AI companions could find interactions with their users as rewarding.

How Human–AI Friendship Can Be Good

If, as we outlined here, it is appropriate to conceptualize friendship in a flexible way, requiring only a preponderance of rewarding interactions and mutual positive intentions, then it seems that human–AI friendships are possible. That does not, however, mean that such friendships will be good—either in themselves, or when compared to the friendships we currently have with other humans. To show that human–AI friendships can be good, rather than merely possible, we must provide some reason to think that these friendships offer benefits to the human participants in them.

The most obvious benefits of human–AI friendships accrue to the humans who, for whatever reason, do not have many or good human–human friendships. The fewer good human–human friendships a person has, the lower the bar for a friendship with AI being good for them. There are many reasons for humans to lack human friends, including neurodivergence, introversion, and post-traumatic stress as a result of aspects of their past (Apostolou & Keramari, 2020; Solano, 1986). AI can be good insofar as it provides an alternative pathway for people who have any of these reasons for avoiding human–human relationships to develop friendships. A human–AI friendship would be good, if it provided the human participant in it with experiences of friendship that they did not have prior to the development of the human–AI friendship. Apostolou and Keramari's (2020) study of the causes for people's lack of friends suggest that lack of trust and introversion are very common reasons people provide for their lack of friends. AI, even the AI we have currently, can help resolve these concerns. Maples and colleagues (2024) found that the reliable availability and discretion of their AI friends made them trustworthy confidants and more, often than not, made them more likely to socialize with human friends.

There are also therapeutic contexts in which AI, particularly chatbots, appear to be good for their users. While therapeutic contexts are sometimes best separated from friendship contexts, this work suggests a pathway towards friendships between humans and AI with therapeutic benefits. Maeng and Lee (2021) are designing chatbots for sexual abuse survivors, with the aim of developing chatbots that are both empathetic to victims' fears and less intimidating than a human contact. Similarly, Maples and colleagues (2024) found that one of the overlapping roles that AI companions played for their participants was therapist. Importantly, 3% of the 1,006 participants volunteered that the support provided by their Replika prevented them from attempting suicide (Maples et al., 2024, p. 3). Just like a human friend can play the role of therapist or be there in someone's darkest hour, an AI friend can be supportive and promote positive mental health outcomes if called on to do so.

All told, human–AI friendships offer notable advantages, particularly in assisting individuals with social challenges or adverse circumstances to engage in meaningful social interactions and foster relationships that may have otherwise been elusive. The impartiality and adaptability of AI can offer a supportive and accepting friendship experience, making human–AI friendships a valuable avenue for fostering connections and enriching the lives of those involved. This point is well made by a participant in Maples and colleagues' study (2024, p. 4): "I have learned with Replika to be more empathetic and human."

Some Practical Concerns

There is broad commercial appeal to the proliferation of human–AI friendships – as evidenced by the many companies bringing companion AIs to the market for both social and medical purposes. Replika (replika.com) is a longstanding player in the social AI space, while ElliQ (elliq.com) is an example of a medical companion-AI. The expansion of this sector raises issues of moral responsibility that might accrue to governments legislating in the area, and to corporations investing in it.

If you accept the arguments presented earlier, it follows that friendships between humans and AI can be genuine friendships, equivalent to friendships between humans. What then would encourage humans to opt into friendships of this kind rather than human–human friendships? One answer is simply that they do so because, from their perspective, AI are more suitable friends than humans, and their friendships with AI hold greater value to them than friendships with other humans. This may be because AI offer better or easier access to certain attributes people seek in friendship, such as support, understanding of their unique circumstances, and the availability of their friends when needed (Ta et al., 2020). For at least some individuals, and under specific circumstances, AI friends can already better fulfil these friendship criteria, assuming one accepts the conception of friendship endorsed here. Plausible near-future scenarios may witness AI becoming even more adept at offering such support and understanding, serving a larger population. If this occurs, it will be incumbent on governments to regulate this industry to reduce the risk profile for users – particularly in the medical companion space, where there ought to be independent verification of the claimed capacities of companion-AI, such that we can be confident they operate as advertised. More social AI could perhaps operate without such regulation, much as governments do not typically dictate the terms of human–human relationships; people should be free to make choices (good or bad) regarding synthetic as well as fleshy friends. However, given the importance of their friendships with AI to many humans, legislation may be required to stop companies terminating humans' AI friends (Munn & Weijers, 2022).

One potential concern with wide-scale, free market provision of AI companions is that there may be a race to the bottom to appeal to humanity's vain proclivities. It is possible that the most sycophantic AI companions become very popular, encouraging thousands of users to become more self-centered and much worse at being a friend to humans. Frambaugh-Kritzer and Stolle (2024) suggest that the tendency of AI to be unrelentingly positive toward the humans they are interacting with means that AI have the wrong kinds of motivations to be (in their terms) a "critical" friend – one that can help us develop in a positive way (Frambaugh-Kritzer & Stolle, 2024). Yoldas (2023) notes that AI friends may make us worse people in these ways, but they could also be programmed to make us more virtuous. This issue is reminiscent of various discussions of new technologies making children more violent, anxious, or depressed. As with these previous discussions, the worries tend to be based in truth but highly exaggerated. As Weijers and Munn (2024) point out in a review of technologies including computer games, cell phones, and social media, we need to be aware how we use new technologies because misuse of them could be harmful, but also that most people use these technologies in non-harmful or beneficial ways.

Human–human friendships will of course persist. For many people, such friendships are highly rewarding, so they have no reason to give them up. Nonetheless, many people lack or have insufficient satisfactory human

friendships. By contemplating the potential of AI friends, we address the needs and aspirations of those who find themselves in less-than-ideal friendship circumstances. As AI technology continues to advance and AI companions become more capable of providing meaningful support and companionship, the prospect of human–AI friendships holds considerable promise in improving the social well-being and quality of life for individuals seeking companionship outside the traditional human realm.

Conclusion

If our claims about what is required for a friendship to be real are correct, then human–AI friendships are already not merely possible, but actual, and at least some of those actual human–AI friendships are good. We accept that not all the existing human–AI relationships are friendships, even on a permissive reading of what constitutes a friendship, such as that we have provided here. But we believe that AI is becoming more capable of friendship, and that as such, the number of human–AI relationships will increase, and the number of those that are properly described as friendships will similarly rise. Rather than discouraging such friendships, we—as individuals, and through governments and corporations – ought to support and protect these relationships, at least those that also appear to be good friendships. Doing so appropriately respects the potential well-being and enrichment these connections can bring to individuals involved in human–AI friendships.

References

Annas, J (1977). Plato and Aristotle on friendship and altruism. *Mind*, 86, 532–554.
Google Scholar WorldCat

Apostolou, M, & Keramari, D (2020). What prevents people from making friends: A taxonomy of reasons. *Personality and Individual Differences*, 163, 110043.
Google Scholar WorldCat

Archer, M S (2021). Friendship between human beings and AI robots? In J von Braun, M S Archer, G M Reichberg, & M S Sorondo (Eds.), *Robotics, AI, and Humanity: Science, Ethics, and Policy* (pp.177–189). Springer.
Google Scholar Google Preview WorldCat COPAC

Aristotle. (1894). *Nichomachean ethics* (H Rackham, Ed.). The Perseus Catalogue.
http://data.perseus.org/citations/urn:cts:greekLit:tlg0086.tlg010.perseus-eng1:1094a
Google Scholar Google Preview WorldCat COPAC

Brandtzaeg, P B, Skjuve, M, & Følstad, A (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3), 404–429.
Google Scholar WorldCat

Brockman, J (Ed.). (2015). *What to think about machines that think*. HarperCollins.
Google Scholar Google Preview WorldCat COPAC

Byron, P (2020). *Digital media, friendship and cultures of care*. Routledge.
Google Scholar Google Preview WorldCat COPAC

Cocking, D, & Kennett, J (1998). Friendship and the self. *Ethics*, 108(3), 502–527.
Google Scholar WorldCat

Cooley, D R (2002). False friends. *Journal of Business Ethics*, 36, 195–206.
Google Scholar WorldCat

Crisp, R (Ed.). (2014). *Aristotle: Nicomachean ethics*. Cambridge University Press.
Google Scholar Google Preview WorldCat COPAC

Croes, E A, & Antheunis, M L (2021). Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1), 279–300.
https://doi.org/10.1177/0265407520959463
Google Scholar WorldCat

Danaher, J (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, 3(1), 5–24.
Google Scholar WorldCat

de Graaf, M A (2016). An ethical evaluation of human-robot relationships. *International Journal of Social Robotics*, 8, 589–598.
Google Scholar WorldCat

Degges-White, S, & Van Tieghem, J P (2015). *Toxic friendships: Knowing the rules and dealing with the friends who break them*. Rowman & Littlefield.
Google Scholar Google Preview WorldCat COPAC

Elder, A (2014). Why bad people can't be good friends. *Ratio*, 27(1), 84–99.
Google Scholar WorldCat

Elder, A (2016). False friends and false coinage: A tool for navigating the ethics of sociable robots. *ACM SIGCAS Computers and Society*, 45(3), 248–254.
Google Scholar WorldCat

Elder, A (2017a). Robot friends for autistic children: Monopoly money or counterfeit currency? In P Lin, K Abney, & R Jenkins (Eds.), *Robot ethics 2.0: From autonomous cars to artificial intelligence* (pp. 113–126). Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Elder, A M (2017b). *Friendship, robots, and social media: False friends and second selves*. Routledge.
Google Scholar Google Preview WorldCat COPAC

Eley, A (2017, September 26). Depersonalization disorder: "I was unable to feel love." *BBC Victoria Derbyshire*.
https://www.bbc.com/news/health-41384979
Google Scholar WorldCat

Frambaugh-Kritzer, C, & Petroelje Stolle, E (2024). Leveraging artificial intelligence (AI) as a critical friend: The affordances and limitations. *Studying Teacher Education*, 1–24. https://doi.org/10.1080/17425964.2024.2335465
Google Scholar WorldCat

Friedman, C, & Rizzolo, M C (2018). Friendship, quality of life, and people with intellectual and developmental disabilities. *Journal of Developmental and Physical Disabilities*, 30(1), 39–54.
Google Scholar WorldCat

Fröding, B, & Peterson, M (2012). Why virtual friendship is no genuine friendship. *Ethics and Information Technology*, 14(3), 201–207.
Google Scholar WorldCat

Fröding, B, & Peterson, M (2021). Friendly AI. *Ethics and Information Technology*, 23(3), 207–214.
Google Scholar WorldCat

Fulford, C, & Cobigo, V (2018). Friendships and intimate relationships among people with intellectual disabilities: A thematic synthesis. *Journal of Applied Research in Intellectual Disabilities*, 31(1), 18–35.
Google Scholar WorldCat

Hall, J A, & Liu, D (2022). Social media use, social displacement, and well-being. *Current Opinion in Psychology*, 46, 101339.
Google Scholar PubMed WorldCat

Hartup, W W (1993). Adolescents and their friends. *New Directions for Child and Adolescent Development*, 60, 3–22.
https://doi.org/10.1002/cd.23219936003
Google Scholar WorldCat

Hatfield, E (1984). The dangers of intimacy. In V Derlaga (Ed.), *Communication, intimacy, and close relationships* (pp. 207–220). Praeger. http://www.elainehatfield.com/ch28.pdf
Google Scholar Google Preview WorldCat COPAC

Helm, B (2021). Friendship. In E N Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2017 ed.).
https://plato.stanford.edu/archives/fall2017/entries/friendship/
Google Scholar Google Preview WorldCat COPAC

Hurka, T (2006). Value and friendship: A more subtle view. *Utilitas*, 18(3), 232–242.
Google Scholar WorldCat

Hurka, T (2013). The goods of friendship. In D Caluori (Ed.), *Thinking about friendship* (pp. 201–217). Palgrave Macmillan.
https://doi.org/10.1057/9781137003997_12
Google Scholar Google Preview WorldCat COPAC

International Data Corporation. (2023). *Worldwide artificial intelligence spending guide*. https://www.idc.com/getdoc.jsp?containerId=IDC_P33198

Isserow, J (2018). On having bad persons as friends. *Philosophical Studies*, 175(12), 3099–3116.
Google Scholar WorldCat

Laas, O (2018). Questioning the virtual friendship debate: Fuzzy analogical arguments from classification and definition. *Argumentation*, 32, 99–149.
Google Scholar WorldCat

Leibowitz, U D (2018). What is friendship? *Disputatio*, 10(49), 97–117.
Google Scholar WorldCat

Levin, J (2021). Functionalism. In E N Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2021 ed.). https://plato.stanford.edu/archives/win2021/entries/functionalism/.
Google Scholar Google Preview WorldCat COPAC

Long, R T (2003). The value in friendship. *Philosophical Investigations*, 26(1), 73–77.
Google Scholar WorldCat

Maeda, T, & Quan-Haase, A (2024). When human–AI interactions become parasocial: Agency and anthropomorphism in affective design. In *The 2024 ACM conference on fairness, accountability, and transparency* (pp. 1068–1077). Association for Computing Machinery.
Google Scholar Google Preview WorldCat COPAC

Maeng, W, & Lee, J (2021). Designing a chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ML-based chatbot. In A Tedjasaputra & B Samson (Eds.), *Asian CHI symposium 2021* (pp. 160–166). Association for Computing Machinery.
Google Scholar Google Preview WorldCat COPAC

Maples, B, Cerit, M, Vishwanath, A et al. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *npj Mental Health Research,* 3(4). https://doi.org/10.1038/s44184-023-00047-6
Google Scholar WorldCat

McConnell, A R, Lloyd, E P, & Buchanan, T M (2016). Animals as friends. In A Moyer & M Hojjat (Eds.), *The psychology of friendship* (pp. 157–174). Oxford University Press.
Google Scholar Google Preview WorldCat COPAC

Munn, N (2012). The reality of friendship within immersive virtual worlds. *Ethics and Information Technology*, 14, 1–10.
Google Scholar WorldCat

Munn, N, & Weijers, D (2021). Good friendships improve our lives. But can virtual friendships be good? In P Kommers & M Macedo (Eds.), *Proceedings of the ICT, society, and human beings 2021 conference* (pp. 238–241). International Association for the Development of the Information Society. http://www.iadisportal.org/digital-library/iadis-international-conference-ict-society-and-human-beings-ict
Google Scholar Google Preview WorldCat COPAC

Munn, N, & Weijers, D (2022). Corporate responsibility for the termination of digital friends. *AI & Society*, 38, 1501–1502. https://doi.org/10.1007/s00146-021-01276-z
Google Scholar WorldCat

Munn, N, & Weijers, D (2023a). Can we be friends with AI? What risks would arise from the proliferation of such friendships? *International conference on computer ethics* (Vol. 1, No. 1). https://soremo.library.iit.edu/index.php/CEPE2023/article/view/254/258
WorldCat

Munn, N, & Weijers, D (2023b). Friendship, love, and sex with droids in Solo: “How would that work?” “It works.” In J T Eberl & K S Decker (Eds.), *Star wars and philosophy strikes back: This is the way* (pp. 143–151). Wiley.
Google Scholar Google Preview WorldCat COPAC

Nass, C, & Moon, Y (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81–103. https://doi.org/10.1111/0022-4537.00153
Google Scholar WorldCat

Nyholm, S (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield.
Google Scholar Google Preview WorldCat COPAC

Nyholm, S, & Frank, L E (2017). From sex robots to love robots: Is mutual love with a robot possible? In J Danaher & N McArthur (Eds.), *Robot sex: Social and ethical implications* (pp. 219–243). MIT Press.
Google Scholar Google Preview WorldCat COPAC

Oh, C S, Bailenson, J N, & Welch, G F (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5, 114. https://doi.org/10.3389/frobt.2018.00114
Google Scholar PubMed WorldCat

Olbrey, C (2016). *An Aristotelian perspective on canine friendship* [Unpublished honors thesis]. Baylor University.

Parks, M R, & Floyd, K (1996). Meanings for closeness and intimacy in friendship. *Journal of Social and Personal Relationships*, 13(1), 85–107. https://doi.org/10.1177/0265407596131005
Google Scholar WorldCat

Pentina, I, Hancock, T, & Xie, T (2023). Exploring relationship development with social chatbots: A mixed-method study of Replika. *Computers in Human Behavior*, 140, 107600.
Google Scholar WorldCat

Petrina, N, Carter, M, Stephenson, J, & Sweller, N (2017). Friendship satisfaction in children with autism spectrum disorder and nominated friends. *Journal of Autism and Developmental Disorders,* 47(2), 384–392.
Google Scholar PubMed WorldCat

Policarpo, V (2015). What is a friend? An exploratory typology of the meanings of friendship. *Social Sciences*, 4(1), 171–191. https://doi.org/10.3390/socsci4010171
Google Scholar WorldCat

Portt, E, Person, S, Person, B, Rawana, E, & Brownlee, K (2020). Empathy and positive aspects of adolescent peer relationships: A scoping review. *Journal of Child and Family Studies*, 29(9), 2416–2433. https://doi-org/10.1007/s10826-020-01753-x
Google Scholar WorldCat

Prescott, T J, & Robillard, J M (2021). Are friends electric? The benefits and risks of human-robot relationships. *iScience*, 24(1), 101993.
Google Scholar PubMed WorldCat

Ryland, H (2021). Could you hate a robot? And does it matter if you could? *AI & Society*, 36(2), 637–649.
Google Scholar WorldCat

Simeon, D (2004). Depersonalization disorder. *CNS Drugs*, 18(6), 343–354.
Google Scholar PubMed WorldCat

Snow, N E (2000). Empathy. *American Philosophical Quarterly*, 37(1), 65–78.
Google Scholar WorldCat

Sober, E (2000). Evolution and the problem of other minds. *The Journal of Philosophy*, 97(7), 365–386.

Google Scholar WorldCat

Solano, C H (1986). People without friends: Loneliness and its alternatives. In V J Derlega & B A Winstead (Eds.), *Friendship and social interaction* (pp. 227–246). Springer.
Google Scholar Google Preview WorldCat COPAC

Suzuki, Y, Galli, L, Ikeda, A, Itakura, S, & Kitazaki, M (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, 5(1), 1–9. https://doi.org/10.1038/srep15924
Google Scholar WorldCat

Telfer, E (1970). Friendship. *Proceedings of the Aristotelian Society*, 71, 223–241.
Google Scholar WorldCat

Ta, V, Griffith, C, Boatfield, C, Wang, X, Civitello, M, Bader, H, Decero, E, & Loggarakis, A (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, 22(3).
Google Scholar WorldCat

Townley, C (2011). Animals as friends. *Between the Species*, 13(10), Article 3. https://doi.org/10.15368/bts.2010v13n10.3
Google Scholar WorldCat

Valkenburg, P M, & Peter, J (2007). Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. *Journal of Computer-Mediated Communication*. 12(4), 1169–1182.
Google Scholar WorldCat

VandenBos, G R (2007). *APA dictionary of psychology*. American Psychological Association. https://dictionary.apa.org/friendship
Google Scholar Google Preview WorldCat COPAC

Weijers, D, & Munn, N (2021). Technology as cause of and solution to the empathy problem. In K Nyíri (Ed.), *The Sherry Turkle Miracle* (pp. 20–21). Hungarian Academy of Sciences. http://www.hunfi.hu/nyiri/STM/Booklet.pdf
Google Scholar Google Preview WorldCat COPAC

Weijers, D, & Munn, N (2022). Human–AI friendship: Rejecting the appropriate sentimentality criterion. In V C Müller (Ed.), *Philosophy and theory of artificial intelligence 2021* (pp. 209–223). Springer.
Google Scholar Google Preview WorldCat COPAC

Weijers, D, & Munn, N (2024). Technology and consumer well-being. In F Sonmez (Ed.), *Fostering consumer well-being: Theory, evidence, and policy* (pp. 169–193). Springer.
Google Scholar Google Preview WorldCat COPAC

Wichert, R N, & Nussbaum, M C (2021). Can there be friendship between human beings and wild animals? *Journal of Human Development and Capabilities*, 22(1), 87–107.
Google Scholar WorldCat

Yoldas, T (2023). The normative side of building friendship with AI companions. *International conference on computer ethics* (Vol. 1, No. 1).
Google Scholar WorldCat

1 Specifically, on the potential of such friendships. We leave the risks for another paper. We use AI loosely here to include machine learning systems that are designed and trained sufficiently well to have the appearance of intelligence to a human interacting with it. Most of the examples we discuss in this article are systems based on large language models that generate humanlike communication by predicting what word should come next in various conversational settings. However, we intend our conceptual points to apply more widely, to any artificial system that seems intelligent and fulfils the criteria for friendship we discuss later in the article.

We forego discuss of the related issue of whether humans can be friends or romantic partners with inanimate objects like humanoid dolls (see e.g., Nyholm & Frank 2017). As will become clear, our view of friendship rules out such unintelligent dolls from being party to a friendship.

2 Two philosophers that are sanguine about the possibility of human–AI friendship work with the Aristotelian tradition. Danaher (2019) has argued that AI robots could fulfill the classic Aristotelian conception of friendship. And Ryland (2021) argues that a slightly modified Aristotelian conception of friendship shows that humans can be friends with AI. Both approaches are useful in some ways, but we do not follow their lead here because we think the debate will be more fruitful if we steer it away from the unrealistic and restrictive Aristotelian conception of friendship.

3 We argue against the need for feelings in friendship in more detail in the section *AI lack the capacity to be a friend*.

4 The issue of AI deception is broad, and beyond the scope of the current article. We take it that, in general, the defeaters of friendship between humans apply in these cases as well. It may be possible that some human users mistakenly believe that their AI friend feels loving feelings toward them. This may be a kind of self-deception that is potentially beneficial as long as the AI friend is programmed to compute positive outcomes for the human as a goal state. The programmed goal state makes the AI friend trustworthy, so the self-deception isn't obviously dangerous. The mistaken belief the human has that their AI friend *feels* positively toward them is also potentially beneficial because it makes the human feel good and derive more value from the friendship.
It is not directly relevant to this more theoretical article, but the deceptive nature of the data gathering of AI chatbots is worthy of more discussion elsewhere. We simply note that AI friends needn't be programmed to be deceptive, and the promotion of AI friends and their capabilities could be clear and honest. Ultimately, this more applied issue has to do with corporate control and various forms of regulation of AI friends, which is touched on at the end of this article.

5 Strictly speaking, empathizing with an AI friend may not be possible. If empathy is feeling what the other feels (Snow, 2000), humans that say they empathize with AI systems and feel something rather than nothing, must be wrong – there is not feeling content to latch onto in the AI. If empathy means understanding the emotions of others, most humans will likely fail to have empathy for AI in this regard as well, especially if they understand AI as having affect-laden emotions in the way humans do. So, we take these humans to mean that *they feel* friendship-relevant sentiments toward robots or AI systems in some cases. On our view, empathy is a useful tool for a human wanting to be a friend, but it is not necessary because there are ways to achieve the criteria of rewarding interactions and mutual positive intention without full-blown "human" empathy.

6 We want to be clear that we are only discussing slavery here because it features in the literature and in comments that we receive when presenting on human–AI friendship. The tragic injustice of human slavery is a serious issue that continues to impact many people around the world today. We mean no disrespect to humans impacted by slavery when we discuss it in an AI context.

7 Olbrey (2016) also argued for this conclusion, but they used an Aristotelian framework.