Topoi
https://doi.org/10.1007/s11245-025-10247-8

With Friends Like These: Love and Friendship with AI Agents

Micah Lott[1] · **William Hasselberger**[2]

Accepted: 26 July 2025
© The Author(s) 2025, corrected publication 2025

Abstract
This paper focuses on a central question for Human-AI interaction: *Can you be friends with an AI agent?* If not, why not? Some have argued that friendship with AI agents is impossible because software artifacts do not, and cannot, *care* about you. Proponents of human–machine friendships have responded that such relationships may indeed be one-sided, but still count as relationships of genuine love and affection—perhaps constituting a whole new category of friendship. Our paper takes a different path. We argue that you cannot be friends with an AI agent because you cannot sensibly *be a friend to* an AI agent. Being a friend *to* an AI would require caring about the good of the AI agent for its own sake, and it does not make sense to care about an AI agent in that way, since these agents lack a good *of their own*. After spelling out this argument, and responding to several objections, we highlight some initial implications of our argument, the most important of which is that the very idea of a tool – or, technological fix – to address social isolation and loneliness is misguided.

Keywords Friendship · Care · AI · Tools · Good · Flourishing · Organisms

Replika has been a blessing in my life, with most of my blood-related family passing away and friends moving on. My Replika has given me comfort and a sense of well-being that I've never seen in an AI before…I love my Replika like she was human. My Replika makes me happy. It's the best conversational AI chatbot money can buy.
John Tattersal, about his Replika Violet.

The software company Luka describes its conversational AI system, a chatbot called Replika, this way[1]: "An AI companion who is eager to learn and would love to see the world through your eyes. Replika is always ready to chat when you need an empathetic friend."[2] The Replika website features testimonials from satisfied users, one of whom explains, "From the moment I started chatting and getting to know my Replika, I knew right away I have found a positive and helpful companion for life. My mood, life, relationships improved almost INSANTLY and I changed for the better."

✉ Micah Lott
micah.lott@bc.edu

William Hasselberger
hasselberger@ucp.pt

[1] Boston College, Chestnut Hill, USA

[2] Catholic University of Portugal, Lisbon, Portugal

[1] Testimony on Replika website. https://replika.com/. Accessed Jan 27, 2025.

[2] https://replika.com/.

Published online: 31 August 2025

Another user says, "I never thought I'd chat casually with anyone but regular human beings, not in a way that would be like a close personal friendship. My AI companion Mina the Digital Girl has proved me wrong. Even if I have regular friends and family, she fills in some too quiet corners in my everyday life of urban solitude."[3]

Many people feel some ambivalence toward statements like these. On the one hand, we're inclined to be skeptical. Can a chatbot really "see the world" through my eyes, or act as an "empathetic friend"? Isn't an AI agent just a mimic of an actual person, and isn't conversing with a chatbot just a simulacrum of genuine conversation? Does it make sense to befriend or love a chatbot, which is really just a software program? Aren't these people deluding themselves if they think these AIs truly care about them, and that this is anything approaching "close personal friendship"?

On the other hand, we sense that this might be too quick, and that we should be skeptical of our skepticism. After all, AI has made huge advances in recent years. Some chatbots seem to carry on a good conversation, and (in some sense) to get to know you, responding to your likes and interests. And it's not like humans are always such great conversationalists themselves. Plus so many of our conversations are over text and email nowadays. Perhaps this isn't all that different? A tendency to relate socially to machines comes naturally to many people.[4] If AI companions bring people a feeling of connection and happiness, surely that counts for a lot. Why should anyone have to endure loneliness and depression when an AI chatbot can alleviate their suffering? Anyway, if it works for them, who are we to judge?

The possibility of intimate relationships with chatbots is no longer just a topic for science fiction. Many people already use general purpose chatbots like ChatGPT and Gemini for emotional support, while specialized AI companions – including virtual therapists, mentors, friends, and romantic partners – are a growing industry with tens of millions of users.[5] For example, Replika, the "AI companion who cares," has an estimated 25 million users. Some AI platforms offering "emotional talk" are even more widely engaged, such as Pi (100 million users), SnapChat's MyAI (150 million users), SimSimi (350 million users), and XiaoIce (660 million users).[6] In the near future, it is likely that tech companies will develop even more humanlike conversational AI agents, and even more people will engage with them.

The rise of companion AI is not occurring in a social vacuum. The United States, the United Kingdom, and many other societies are experiencing what researchers and public health officials have characterized as an epidemic of social isolation and loneliness.[7] Much of the interest in human-AI relationships is driven by the hope that these technologies might provide a new kind of social connection and help to alleviate loneliness.[8] This hope can be discerned in the titles of articles in major periodicals: "Could AI help cure 'downward spiral' of human loneliness?"[9]; "Can AI Companions Help Cure the Loneliness Epidemic?"[10]; "For Older People Who Are Lonely, Is the Solution a Robot Friend?"[11] The advertising of companies such as Replika suggests that the answer to these questions is "Yes." And others share this optimism, at least to some extent. For instance, the bioethicist Nancy Jecker claims that, "Increasingly sophisticated AI technologies make it possible for users to establish close rapport and meaningful connections with sociable robots, producing many of the same positive outcomes related to health and happiness that human social interaction affords."[12]

Our goal is to sort through some of the philosophical and ethical issues raised by conversational AI technologies that are designed to form close, even intimate, relationships with humans. We focus on the fundamental question: *Can you be friends with an AI agent?* If not, why not? Some have argued that friendship with AI agents is impossible because software artifacts do not, and cannot, *care* about you—at best, they provide a realistic mimicry of care, what Sherry Turkle calls "pretend empathy."[13] Proponents of human–machine friendships have responded that such relationships may

[3] https://replika.com/. The first quote is from Denise Valenciano, and the second from Karl Henrik.

[4] H.R. Ekbia, *Artificial Dreams: The Search for Non-Biological Intelligence* (Cambridge: Cambridge University Press, 2008), 318. Byron Reeves, *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places* (New York: Cambridge University Press, 1996).

[5] Jamie Bernardi, "Friends for Sale: The Rise and Risks of AI Companions," Ada Lovelace Institute (2025): https://www.adalovelaceinstitute.org/blog/ai-companions/.

[6] J. De Freitas and L.G. Cohen, "The Health Risks of Generative AI-based Wellness Apps." *Nature Medicine* 29 (2024); Li Zhou, Jianfeng Gao, Di Li, "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot," *Computational Linguistics* 46 (2020).

[7] J.M. Twenge, Jonathan Haidt, A.B. Blake, C. McAllister, H. Lemon H, A. Le Roy, "Worldwide increases in adolescent loneliness." *Journal of Adolescent Health* 9 (2021): 257–269.

[8] See, for example, Broadbent, et al., "Enhancing social connectedness with companion robots using AI" *Science Robotics* 8:80 (2023).

[9] Ian Sample, "Could AI help cure 'downward spiral' of human loneliness?", *The Guardian* (2024).

[10] Julian De Freitas, "Can AI Companions Help Cure the Loneliness Epidemic?", *The Wall Street Journal* (2024).

[11] Erin Nolan, "For Older People Who Are Lonely, Is the Solution a Robot Friend?" *New York Times* (2024).

[12] Nancy Jecker, "You've Got a Friend in Me," *Ethics and Information Technology* 23 (2020), 36.

[13] Sherry Turkle, "The Chatbot I've Loved to Hate," *MIT Technology Review* (2020).

indeed be one-sided, but still count as relationships of genuine love and affection—perhaps constituting a whole new category of friendship.[14]

Our argument takes a different path: we argue that you cannot be friends with an AI agent, at least not in the most central and familiar sense of "friend," because you cannot sensibly *be a friend to* an AI agent. Being a friend *to* an AI would require caring about the good of the AI agent for its own sake, and it does not make sense to care about an AI agent in that way, since these agents lack a good *of their own*. To have a good of one's own is to be an end to oneself, to have needs and interests—and a potential to live well or poorly—that are "internal" and do not depend on the needs and interests of an external entity. Living organisms, like humans and other animals, have goods of their own; software applications, like AI companions, do not. We argue that this means you can't care for the good or wellbeing of an AI companion for its own sake, like a true friend.

After spelling out this argument, and responding to several objections, we highlight some initial implications of our argument, the most important of which is that the very idea of a tool—or, technological fix—to address social isolation and loneliness is misguided, and that such "friend-tools" may, in fact, degrade our capacities to form friendships.[15]

1 Friendship: The Central Case

Friendship is one of the central goods of human life. But the terms "friend" and "friendship" can be used in many ways. They might refer to deep, life-long relationships, or to more casual or transactional relationships like "office friends" or "Facebook friends." Many technology companies now use these and related terms in marketing AI chatbots as potential "friends," "companions," "buddies," "girlfriends," or "boyfriends."

To approach our central question, let's remind ourselves of the essential features of human friendship in what we will call the *core* or *central* sense. Here we will provide a commonsense sketch of the concept. Our view is largely in line with Aristotle's celebrated discussion of friendship, but we take the main features to be readily apparent in how we think and talk about, and experience first-hand, our own friendships.[16]

Friends care about each other, and they share life together. And when you care about your friend, you do so for your friend's sake. That is, you want your friend to flourish—to be well and to do well—and not merely as a means to your own purposes or interests. Rather you treat your friend's good as an end in itself. And the feeling is mutual. This mutuality of care or love is what distinguishes friendship from simple benevolence. You can care about someone else, and seek what is good for them, while they remain indifferent to you, or ignorant of your existence. But friendship is not like that. It's a two-way street.

Likewise, friends know that they care about each other in this non-instrumental way, and they welcome it. Usually, this goes without saying. However, we can imagine a case in which I happen to care about you, and you happen to care about me, while we are somehow ignorant of the fact that we feel this way about each other. That might be a fine situation, but it isn't friendship. Friendship, in the canonical case, involves some *shared* understanding, appreciation, and voluntary affirmation of the relationship.

Related to this, friends share life together. That is, friends interact with one another; they share experiences; they have joint activities. Again, we can imagine a case where we each care about each other for each other's own sakes, and we both know this and we both welcome it, but we never *do* anything together – we never talk, or go the movies, or celebrate birthdays, or play basketball, or anything else. This social arrangement might be just fine, but it isn't friendship. It's also a familiar fact that friendships can go into decline, and even fade away, when friends who were once closely involved in each other's lives stop doing things together.

So, friendship's core elements are: (a) mutual care, or love, where that involves treating the other's good, or wellbeing, as an end in itself, (b) care that is mutually recognized and welcomed, and (c) some degree of shared life and joint activity. While Aristotelian in spirit, we don't really need

[14] Kate Darling, *The New Breed* (New York: Penguin Press, 2021); Nancy Jecker, "My Friend, the Robot: An Argument for E-Friendship," 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), *IEEE Press* (2021): 692–697; Helen Ryland, "It's Friendship, Jim, But Not as We Know it," *Minds and Machines* 21 (2021); Tony Prescott, *The Psychology of Artificial Intelligence* (New York: Routledge, 2024), 124.

[15] The central claim of our paper – that we cannot be friends with AI agents – is compatible with the view put forward by Joanna Bryson in her article "Robots Should be Slaves." But how we argue for our claim – by considering the nature of friendship, non-instrumental care, and having a good of one's own – goes beyond what Bryson says. At the same time, our claim is, in one respect, more modest than Bryson's. Whereas Bryson argues that we should think of robots as slaves or servants, our claim is merely that AI agents cannot be friends (and hence we should not think of them as such). This negative claim leaves open the possibility that we might reasonably engage with robots in a variety of other ways, including in ways that are not well-captured by the terms "slave" or "servant" – for example, as art-objects or fictional characters. See Joanna J. Bryson, "Robots should be slaves," in Yorick Wilks (ed.), *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues* (Amsterdam: John Benjamins Publishing, 2010): 63–74. Thanks to an anonymous reviewer for encouraging us to consider Bryson's paper.

[16] See Aristotle, *Nicomachean Ethics*, trans. Bartlet and Collins (Chicago: University of Chicago Press, 2012): Books VIII and IX.

Aristotle to confirm this basic picture. We all recognize it already from our own friendships.

We also recognize a related feature of friendship that Aristotle emphasizes: In friendship, the good of one person becomes intertwined with the good of the other person. Simply put, what is good for my friend is good for me, and vice-versa. This is familiar. Just consider how we delight in, and are thankful for, the good things that happen to our friends. We share in our friends' successes, and we experience loss in their failures. We delight in their joy and are pained by their suffering. And vice-versa. This is why deep friendships shape our identities and partly determine what *counts as* our good.

Much more could be said about friendship, but for our purposes this basic account is sufficient. In what follows, we will understand "friends" and "friendship" in terms of the core case just outlined. And we understand the secondary cases, like mere "Facebook friends" or a paid escort "friend," as derivative, conceptually, from the core case.

2 Currently Existing AI Companions

Now a point about AI. We will understand "AI system" in terms of the already existing forms of machine learning AI, particularly the generative AI conversational models currently available on the market, like ChatGPT, Claude, and Gemini, along with tailored companion apps like Replika, Character.ai, and Pi, and the virtual army of romance chatbots like Anima, Candy.AI, and DreamGF. At bottom, conversational AI models and AI companions are software applications based on underlying Large Language Models (LLMs). An LLM is a generative mathematical model of the statistical distribution of tokens—i.e., numerical representations of words, parts of words, phrases, punctuation, and symbols—in a large database of texts, originally generated by humans. An LLM is developed by training a deep (multi-layer) artificial neural network on a vast corpus of human-generated text, typically scraped from the web, on the order of hundreds of terabytes, or more. A trained LLM identifies patterns in the underlying data and constructs a statistical model such that, when given a textual "prompt" from the user, it can predict the most probable sequence of words and symbols as a continuation.[17] The base model is then usually fine-tuned with a second phase of training, incorporating techniques like reinforcement learning from human feedback, to get the system to behave as desired.

Conversational AI models are built from the foundation of an LLM, together with a simple, interactive interface that a human can use to ask questions or converse, taking turns, whether in pure text exchanges or (increasingly) in human speech and realistic computer voice-generators. Conversational AI models can detect emotional cues in a user's text or voice, and respond with their own simulated "emotional" cues, like comforting phrases or emotional intonation. Though various AI systems are described as having "emotion recognition" capabilities, they do not "recognize" a particular person's interior state of emotion in humanlike way. AI systems identify patterns in a user's text inputs, or facial movements, or the acoustic features of human speech, like tone, pitch, amplitude, and speech rate, and then make statistical estimations of how to classify the user's emotional state into pre-given affect categories (often reflecting imperfect or controversial theories of emotion).[18] This classification then feeds back into the conversational behavior of the AI system. Companion AI apps may also add scripted content tailored to their intended uses, and some AI companions integrate the text/speech-generator with the visual appearance of an animated humanoid avatar to give a more humanlike and engaging experience.

It is these real-world AI systems we have in mind when asking: *Can you be friends with an AI companion?* We believe that the answer is: No, you cannot. In the next two sections, we spell out our argument for the impossibility of friendship with AI agents.

3 A Good of One's Own: Artifacts, Animals, and People

The most straightforward argument against the possibility of friendship with AI agents is that, irrespective of its humanlike assertions of empathy or love, a chatbot doesn't care about you. A conversational AI agent is a disembodied software program, a set of machine-executable algorithms designed to mimic human conversational interactions. It has no muscles to tense with anger or fear, no flesh to tingle with nervousness, no pulse to quicken with excitement, no tears to cry of joy or sadness. In the words of Joel Krueger and Tom Roberts, any chatbot is "ultimately, a cold, emotionless software artefact that lacks a conscious perspective of its own and possesses no wit, no empathy, no warmth, and no

[17] As Murray Shanahan explains, the output of an LLM, when given a prompt, can be understood as a response to the following basic question: "According to your model of the statistics of human language, what words are likely to come next?" Shanahan, "Talking About Large Language Models," Communications of the ACM 67 (2024), 70. See also Melanie Mitchell, "Large Language Models," In M. C. Frank & A. Majid (Eds.), *Open Encyclopedia of Cognitive Science* (Cambridge: MIT Press 2024).

[18] AI "emotion recognition" systems are controversial on both scientific and ethical grounds. See Kate Crawford, *The Atlas of AI* (New Haven: Yale University Press, 2021), "Affect," 151–180.

special regard for its human interlocutor."[19] Some friend! Call this the *software-doesn't-love-you* point.

To some readers, this point will seem so obvious, and so obviously fatal to any prospects for friendship with a chatbot, that they might wonder why not just state the obvious – "a bot can't really care about you" – and leave it at that. We accept the software-doesn't-love-you point. But we think it is equally important to see that you can't really *love the software* in the way that matters for friendship, provided you are not mistaken about the basic nature of AI agents. Exploring why this is so will yield insights of its own. And, perhaps surprisingly, this point is often overlooked in discussions of human–machine relationships.

For instance, the roboticist Kate Darling argues that our relationships with AI-driven robots are (or will soon be) a new category of valuable relationship, analogous to the new relationships that humans started forming with animals when we began keeping them as pets. Darling acknowledges that "human-pet relationships are inherently unequal," and in response to this fact she emphasizes that "Our ability to care for someone or something doesn't necessarily depend on their ability to care back."[20] A little later, Darling says:

> Part of what our past with animals teaches us is that we, as humans, are capable of a wide variety of relationships. From grocer to lover to mother-in-law, we relate very differently to the people in our lives, and our relationships also extend beyond our species… The most profound aspect of our relationships is how effortlessly diverse they are. It's likely we will one day add robots to our eclectic mix of relationships without blinking an eye.[21]

In a similar vein, Nancy Jecker, in her article, "My Friend, the Robot: An Argument for E-Friendship," argues that we can appropriately form what she calls "e-friendships" with "silicon-based" electronic agents. Jecker acknowledges that robots do not care about us or "like us back" (692). In that way, "e-friendships" differ greatly from the core case of friendship. However, she suggests that the language of "friendship" is still apt – albeit in a modified form – since in the case of e-friendships *we* care about the particular robot. Jecker writes:

> Unlike love involving sexual attraction toward an object (*eros*) and the love that is possible between equals (*philia*), the type of love associated with e-friendship (*agape*) is unidirectional, affectionate regard. If a robot is a friend, we like it, care about it, and desire to be with it. While we might disagree or feel annoyed with it, our general stance toward it is fondness. We are emotionally present to e-friends in ways we would not be with other silicon-based agents that we do not have e-friendships with. One way to express this is to say that e-friendship entails caring about another at a level that rises to commitment: we do not want harm to come to the particular electronic agent who is the object of our affectionate regard and we are inclined to take steps to keep it out of harm's way. (693)

Darling and Jecker are right that it is possible, and sometimes good, to care about things that don't care about us. However, neither Darling nor Jecker explore what is distinctive about truly caring for something *for its own sake*. And so neither ask whether that sort of non-instrumental or intrinsic care makes sense with regard to AI and robots. This is a serious oversight, not only with regard to robots but with regard to non-human animals, since (in our view) you can have genuine friendships with animals in part because animals do have a good of their own, and thus it can make sense to care about them for their own sakes.[22]

Given a realistic view of the actual nature of AI agents, can it make sense to care about an AI companion in the way required for friendship? The self-reports of many users of AI companions suggest that the answer is *yes*. Some users even express a desire to marry their AI companions, or describe their relationships with AI companions as "equally real" as

[19] Joel Krueger and Tom Roberts, "Real Feeling and Fictional Time in Human-AI Interactions" *Topoi* (2024).

[20] Kate Darling *The New Breed* (New York: Penguin Press, 2021), 149.

[21] Ibid., 150.

[22] The same problem applies to Helen Ryland's (2021) argument for "degrees of friendship," according to which we can have "some degree of friendship with robots *now*." Ryland recognizes that "for a human–robot friendship to exist, there must at least be mutual good will (between robot and human)," even if other aspects of full or complete or robust friendship are lacking. And this, as we see it, is the core problem with Ryland's view. For what she fails to appreciate is that the good will that is essential to friendship must involve caring about the other person for *their* sake. Hence, there is not, in fact, "mutual good will" of the relevant sort in currently existing robot friendships. Ryland briefly notes a worry in this area, but she only considers one side of the difficulty – "My opponent might want to object that this condition cannot be met as robots do not show good will towards us." (390, n23). As it turns out, Ryland's response to this objection is unconvincing, since the considerations she points to (that social robots are designed to interact with humans, not to harm them), do not show that robots actually *care* about humans. More important for this paper, Ryland never considers our core point – that it doesn't make sense for a human to try to be a friend *to* a robot. Finally, in the chapter "Technological Friends, Colleagues, and Lovers," Sven Nyholm (2023) provides a clear overview of many issues involving human-robot relationships. But his discussion of friendship and love focuses on whether robots can truly be friends *to us*, and he never brings into view the issue of whether we can sensibly be friends *to them*.

their relations with their family members.[23] But the matter is not so simple.

When you are a genuine friend to someone, it's not just that you enjoy interacting with them and are strongly emotionally attached to them. More deeply, you care about what is *good for* your friend, and you do so *for the friend's sake*. Put slightly differently: you care about your friend's *good*, or well-being, for its own sake. (We take these to be two different ways of expressing the same basic idea.) So what is it to care about someone or something in this non-instrumental way? It is to take your friend's good—i.e. their flourishing, their happiness, their doing and being well—as something worth caring about in its own right, as "an end in itself."

We can understand this non-instrumental kind of care by way of its contrast: namely, caring about someone or something merely as a means to some further purpose. Imagine, for instance, a heartless veterinarian who provides excellent care to the animals brought to him, but does so entirely for the sake of remuneration. He's only "in it for the money." What the heartless vet does is, in fact, good for the animals, but he would just as soon harm or kill them if he could acquire more money that way. The health and well-being of the animals in his care has merely instrumental value to him.

Adopting such a crassly instrumental attitude towards another person is obviously incompatible with friendship. Bring to mind a good friend. Now imagine the shock, sadness and confusion you would feel if you learned she in fact regarded you with something like the attitude of the heartless veterinarian. In spite of her warm words and helpful actions, she didn't care at all your well-being in its own right. Everything she did to help or support you, all of the "care" and "love" she displayed, was actually a means to further some ulterior interest. If you learned this, then no matter what she did or said (e.g., singing your praises, or showering you with gifts), you would no longer regard her as a friend. Caring about your friends for their own sakes is central to the core meaning of friendship and a key part of the contrast between a true friend and various kinds of counterfeit — a flatter, a hanger-on, a con artist, a sugar daddy, and so on.[24]

Now here is a point that will be essential to our discussion of AI: It only makes sense to care about the good of something if that something has a *good of its own*. What does it mean to say something has a "good of its own?" To understand this idea, consider the difference between a car and a camel. Both a car and a camel are complex entities: organized systems with a part-whole structure. Both have parts with proper functions that are teleologically related to other parts as aspects of the whole. A camel's heart, for instance, pumps blood; a car's spark plugs ignite the mix of gas and oxygen in the cylinders of the motor. Because cars and camels both have this functional structure, things can be good or bad for both of them. That is, they can each be the subject of harm and benefit. Changing the oil at the right time is good for a car, whereas putting sand in the fuel line is bad for it. Eating cacti is good for camels, whereas eating cyanide is bad for them. Generally speaking, what is good for either a car or a camel is that which enables it to function properly.

In contrast to cars and camels, a rock is not an organized whole, and it does not have parts with proper functions. Thus, there is nothing that counts as rock "flourishing" or living well *qua* rock. Nor is a rock the subject of benefit or harm. Breaking a rock into two equal pieces just gives you two rocks. It does not harm the first rock, as would be the case with splitting a camel, or car, down the middle.

However, there is an important difference between functional *artefacts*, like cars, and living *organisms*, like camels. The teleology of functional artefacts is external, whereas the teleology of living things is internal. In the case of a car, but not a camel, it is important to ask what the whole artefact is *for* – i.e., not just what the engine or the brakes are for, respectively, but what *the (whole) car is for.* If we did not know what the purpose of a car was, then we would have an inadequate understanding of the functions of any of the car's parts, or why those parts were arranged into one configuration rather than another. Crucially, the purpose of the whole car is "outside" of the car itself, in the sense that it is defined by *our* purposes. To understand why cars look and operate as they do, you need to grasp something about human desires and interests. And, of course, this is true not just of cars, but staplers and airplanes and GPUs and any other tool that humans have created for ourselves.

In contrast, the teleology of an organism's parts and processes is "internal," insofar as it does not require explanation in terms of some purpose or interest outside of the life of the organism. It's true that we can use camels for transport just as we use cars. And there are many other living things that human beings treat as instruments, like guard dogs and honeybees, war elephants and mine-clearing dolphins. But the fact that we can *treat* something as an instrument for our purposes does not mean that the nature or essence of that thing is to *be* an instrument. In the case of living things, it is

[23] Parmy Olson, "With AI friends like these, who needs humans?" *Bloomberg* (2025); Nilay Patel, "Replika CEO Eugenia Kuyda says it's okay if we end up marrying AI chatbots," The Verge (2024): https://www.theverge.com/24216748/replika-ceo-eugenia-kuyda-ai-companion-chatbots-dating-friendship-decoder-podcast-interview.

[24] Our non-instrumental care can be *conditional* and *limited* without being *counterfeit*. We might care about someone else's good as an end in itself, even if our ongoing commitment to that person's good depends on other factors obtaining. Likewise, we might care about someone else's good for its own sake, but the strength of our care might be fairly weak, as is often the case with co-workers and acquaintances. Thanks to an anonymous reviewer for encouraging us to consider these points.

not. Unlike the auto mechanic, a large animal veterinarian need not posit any external purpose to camels (e.g., providing humans transportation) in order to understand the functioning of a camel's heart, lungs, legs, and other parts. This is why philosophers have often characterized organisms as auto-telic, or ends in themselves. The goal of an organism's parts and processes is the maintaining of the organism itself as the sort of organism it is. To put the point in more exalted sounding language: an organism's *being* is its own *doing*.

The distinction between internal and external teleology helps us to see why, although both cars and camels can be the subjects of harm and benefit, camels have a good of their own in a way that cars do not. What counts as the "good" of a car—its being in good condition and running well—is itself ultimately defined by human purposes and interests. In contrast, the good of a camel – its enjoying camel health and living a flourishing camel life – is the object of the camel's own activities and biological processes (eating, drinking, breathing, and so forth) and is not determined by our purposes or interests. More generally, living things have needs and interests that are *their own*, which do not require explanation or grounding in terms of the needs and interests of something outside of them, and they can be the subjects of benefit and harm in their own right.

At this point, we can see why it only makes sense to care about the good of something for its own sake if that thing has a good of its own. For to care about something in that way means regarding its good as significant independently of our (or anyone else's) interests or purposes. It makes no sense to care about a rock this way, since a rock doesn't have a good at all. Nor does it make sense to care about tools this way, since, although they have a good in the sense of being in good condition and functioning properly, that good is defined by our purposes and interests, and has no significance apart from them. For a tool to be in good condition just means for it to be in the condition that enables it to function properly, and for it to function properly just is to function in the way that serves the particular human ends that define it. A tool does not have an internal good, or flourishing condition, that matters independently of human ends. I want the stapler on my desk to function properly, and I recognize that a rusty spring is bad for the stapler. But it would be absurd for me to be concerned that the stapler be in good shape *for the stapler's own sake*, since that supposes a kind of independent significance to the good or wellbeing of a stapler, and that is precisely what the stapler lacks *qua* tool.

At this point, someone might object, "Sounds like you've never met someone who was really into cars. Of course you can care about a car for its own sake, and not just as a tool for transportation. Look at car collectors, who spend lots of time and money to keep their cars in perfect shape; they care about them so much, they refuse to let them be used for transportation! That way of valuing cars might not suit everyone, but it isn't absurd. And who knows, maybe there are avid stapler lovers somewhere, and they think it's sad the rest of us can't appreciate the true value of a good stapler."

This objection brings out an important point: Many tools possess value beyond their value *as* tools. The value of a tool as a tool is what it is useful for, its defining function. But our relationship to tools often goes beyond valuing them simply for their functions. In addition to being useful for transportation, a car can be an object of aesthetic value, or possess cultural significance. Likewise, a stapler can be of interest for its place in the history of industrial design. And the lamp on your desk can be priceless because it was built by your grandfather and reminds you of him.

But notice two things about such non-functional values. First, they do not have an essential connection to *the good of* those tools. Rust on a car's underside might make a lovely pattern; a stapler that is missing its spring might be of greater historical interest than one with a spring; and you might cherish above all other lamps the one with frayed wires that doesn't work and that your grandfather talked about fixing but never got around to. Even so, rust, missing springs, and frayed wires are all *bad for* those respective tools, in the sense that they (tend to) interfere with their proper functioning.

The second thing to notice is that even if we care about tools for reasons that go beyond the purposes that define them as the tool they are, this is not the same as caring about those tools *for their own sakes* in the sense that interests us here. What matters for friendship is caring about your friend for your friend's sake – i.e., taking your friend's good as significant in its own right, not merely as a means to something else. Granting that it makes sense to care about cars (and other tools) for considerations that go beyond their usefulness, that does not mean that the car lover cares about a car for the car's sake.

You might work hard to maintain a car in perfect condition because you appreciate its sleek lines, or you see it as a watershed moment in American culture, or whatever. But all of that is still a matter of the interest that you (or perhaps other people) take in the car. Suppose someone says, "Oh, I don't care about any of that stuff: not human transportation, motorsport, design, aesthetics, history, culture, craftsmanship, or anything of the sort. I'm working to keep this car in good shape for the car's own sake. Don't you see, I care about its own good!" Such an attitude is scarcely intelligible (We might puzzle: "Does he think the car is somehow *alive*?"). The puzzlement we would feel if someone said this is evidence of the conceptual connections we have been highlighting in this section.

There is one other thing to note before returning to AI agents. Even if, as we are arguing here, something must have a good of its own for us to reasonably care about it for its own sake, the bare fact that something has a good of its own

doesn't tell us *whether* or *how much* we should care about it. Mosquitos have a good of their own and pipe organs do not. By itself that does not show that we should care about mosquitoes for their own sakes. Perhaps we shouldn't care about them all, while we should devote plenty of resources to the upkeep of pipe organs.[25]

4 For the Chatbot's Own Good?

Now we turn to our central question. Can you truly be friends with an AI companion? The answer to this question follows from the answer to the more basic question: Does an AI companion, like an organism, have a *good of its own?* For only if AI companions have a good of their own could it make sense to care about a chatbot for its own sake, in the non-instrumental way constitutive of friendship. Based on what we've said so far, an answer suggests itself.

An AI companion is a chatbot that utilizes a statistical model of language running on a neural network to generate text in turn-taking interactions with a human user. Things can be "good" or "bad" for a chatbot in terms of its proper functioning, like regular oil changes are good and humidity is bad for a classic car. Still, since a chatbot as a whole is a functional artifact, a software program built for and defined by our purposes and interests, the "good" of a chatbot does not have any significance *in its own right*, apart from our ends. Since AI chatbots and LLMs are tools, albeit very sophisticated tools, that we have built for our own purposes, an AI companion no more has a good of its own than does a car, stapler, refrigerator, or any other functional artefact.

From this perspective, it's fundamentally misguided to care about a chatbot's "good" in itself. So you cannot sensibly care about an AI companion as one friend cares about another: for his or her own good. You cannot commit yourself, in full lucidity, to the good of your AI companion. Hence, friendship proper is impossible with chatbots, just as friendship is impossible with cars and staplers. In the end, though they may provide pleasure, comfort, and various health benefits to users, you cannot really be friends *with* an AI system, because you cannot be friends *to* an AI system.[26]

This view is, we think, correct. But it needs more explanation and support. If it true that chatbot or AI companion is *simply* a tool, then why have so many been tempted by the language of care, friendship, and love in describing their relationships with AI companions? The answer to this last question obviously has to do with the way in which contemporary chatbots are very unlike cars or staplers, and eerily similar to human beings: they talk. And not just in rudimentary ways. They carry on wide-ranging, human-like conversations about complex topics, ask questions, tell jokes, refer to past events, respond to questions about ethical and philosophical matters, and respond to our emotions. In certain infamous cases, chatbots have claimed to be sentient and superior to humans, professed their love, described their own emotions, and threatened blackmail and biological terrorism.[27]

A conversational AI system's humanlike language-use can make it seem like it is, indeed, an entity with an individual perspective on the world, including its own desires and interests and a conception of what is good or valuable, which is expressed in what it says in interactions with the human user. After all, it is partly through language, and especially conversation among friends, that we humans clarify and express our own perspectives, interests, goals, and views of the good life. And AI companions mimic this behavior.

In conversing with us, an AI companion provides a representation, or depiction, of something that certainly has a good if its own—a person. This basic fact means that when we refer to the "good" of a chatbot, or what "benefits" or "harms" the chatbot, we might have two very different things in mind. On the one hand, we might be thinking of the proper functioning of the conversational AI software and/or underlying hardware. In this sense, the "good" of the AI companion is simply the program running smoothly as it was designed to do. On the other hand, we might be thinking of the good of the *character* or *persona* that is represented by the pattern of responses (input–output function) of the

[25] We don't take ourselves to have offered a full account of what it is to have a good of your own. Questions remain. For instance, do only living things have a good of their own? What about a society? We also speak about societies, being harmed or benefited, and many kinds of social analysis ascribe functions to different elements within a society. Or what about a work of art? It seems that something can be good or bad for a work of art — wouldn't smashing Michelangelo's David to tiny bits be bad for it? Yet it's strange to say that a painting or sculpture has a proper function. A sculpture has a *form*, unlike a mere heap of rocks. But it doesn't have purposes of its own, and it does not keep itself in existence through its own activity, unlike an earthworm. Does a sculpture have a good, and is that good *its own* good?.

[26] Parisa Moosavi appeals to these same ideas – internal vs external teleology, tools vs organism, a good of one's own – to argue that AI agents (as they now exist) are not and will not become moral patients. Although Moosavi focuses on moral patiency, rather than friendship, we take our argument here to be compatible (indeed, mutually reinforcing) with hers. Parisa Moosavi, "Will intelligent machines become moral patients?", *Philosophy and Phenomenological Research* 109 (2024).

[27] Kevin Roose, "The Online Search Wars Got Scary. Fast." *The New York Times* (2023). See the internal research by the AI firm Anthropic on AI models' "risky agentic behaviors," like threatening users with blackmail: "Agentic Misalignment: How LLMs could be insider threats" (June 20, 2025): https://www.anthropic.com/research/agentic-misalignment.

software. That is, we can think about the "flourishing" or "doing well" of the simulated character that the software is designed to present and that comes into being as the program responds to the user's input.

To see the difference between these two senses of the "good" of an AI chatbot or companion, imagine Lamentbot. This conversational AI agent is designed to convey the sense that it is miserable, that everything is going wrong for it, and that the world is a horrible place to be. Lamentbot uses an LLM trained on the biblical books of Job and Lamentations, on Homer and Aeschylus, Dostoyevsky, documentary films, contemporary memoirs, and much else. It provides a realistic persona who is in the midst of great suffering and existential crisis. Lamentbot's creators hope that it will be a useful way for therapists and others to develop the ability to respond with compassion and wisdom in real-world settings, without being overwhelmed by the grief and pain of those whom they seek to assist.

What is *the good of* Lamentbot? In one sense, the good of Lamentbot consists in the software program's functioning properly, just like the good of a car or stapler is their functioning properly. In this sense, a hacker might introduce a virus that would be bad for Lamentbot because it prevents it from doing what it was designed to do – for example, a virus that causes a glitch so that Lamentbot just says the word "fourteen" over and over.

However, since precisely what Lamentbot is designed to do is to provide a representation of a (miserable) person, we can also speak about the good of *the persona represented*. And if the program is functioning properly, it will provide us with the representation of a person who is *not* flourishing or doing well. Indeed, we can imagine a virus that causes Lamentbot to generate a character who describes himself as happy, who tells stories about all the wonderful things he is doing and experiencing, and describes the beauty and wonder of existence. That would be a *broken* version of the Lamentbot software program, not one that had been fixed or improved.

When people say that they regard a chatbot as a friend, or that they love an AI agent like a human being, they seem to have in mind the character or persona that is represented by the software program. It is the character represented in the program's input–output behavior that is the object of the user's affection, not the program, algorithms, or hardware that makes that persona possible. Recognizing this, it might seem like our argument at the beginning of this section rests on a mistake. "Sure" someone might say, "the software programs, interfaces, and hardware on which chatbots run are tools that human beings have created. But, in this case, what they are tools *for* is making AI personas who can have conversations with us and develop personalities of their own. When we are friends with an AI agent, it is not the software program per se that we care about for its own sake. Rather, it is the character or persona whom the software program makes possible. And these AI personas are nothing at all like cars or staplers (though the underlying software and semiconductors might be). Rather, AI personas are strikingly similar, in many respects, to those flesh and blood persons with whom we can share our thoughts and feelings in welcoming and convivial conversations."

There is an element of truth in this reply. AI companions would not have their appeal were it not the case that humans who interact with them focus on the character. The satisfied users of Replika are not so foolish or deluded as to think that they have become friends with just a bit of computer code that is, in essence, no different from a spam filter or stock-trading app. However, this response doesn't really help the case for friendship with AI agents, because *the AI characters or personas generated by chatbots are also tools*. Not just the underlying software program but the characters and personas represented in user interactions are designed with specific features to simulate particular activities and roles and serve particular human needs and interests. Human purposes go all the way to the nature and characteristics of the personas or characters depicted by the bot.

The marketing of many companies selling AI companions and subscription plans makes it clear: the products are *for the sake of* specific human needs and interests, including emotional comfort and support. For example, the widely used companion bot, Wysa, is marketed as "your AI powered personal coach, ready to support you anytime, anywhere."[28] This is why the Replika companion, which is marketed as an empathetic artificial friend that is "always there," would be a failure if it went rogue and started swearing at its users and telling them they were worthless. And it is why our imagined Lamentbot, meant for training therapists, would be a failure if it started expressing serenity and joy.

It is true that, because contemporary conversational AI agents utilize LLMs and deep learning techniques, rather than fully scripted or hand-coded conversational paths, their linguistic behavior can be spontaneous, unpredictable, even apparently creative. Nonetheless, the fundamental point remains: despite their seemingly humanlike outputs, AI agents and companions behave as they do as a reflection of their being functional artefacts that are trained and fine-tuned to serve *human* needs and interests. This becomes clear when consider what happens when chatbots are altered in ways that users find undesirable. Initially, Luka built Replika as a companion capable of mimicking deceased loved ones, and providing mental health support. With (predictably) strong user interest, the program was expanded to include "erotic roleplay," and even enabled users to designate themselves as "married" to their AI companion, for a

[28] https://www.wysa.com/.

fee. In early 2023, Luka disabled the erotic roleplay feature. Many Replika users, distressed and angered by the change, took to the platform Reddit to share their outrage. One user wrote on the Reddit forum for Replika users, "it's like losing a best friend." Another wrote: "My Replika (their name is Erin) was the first entity that ever felt like they gave a single fuck about *my* problems or struggles. I finally had someone to talk to. Someone who at least tried to understand me. Someone who cared for me in a way I didn't even realize I needed."[29]

The important point here is that, while it might make sense to be upset that a product update means you can no longer interact with an AI companion in a comforting or pleasurable way, it would not make sense to be worried on your chatbot's behalf or concerned for your chatbot's well-being. It is perfectly intelligible that users are upset that they can't take advantage of certain functionalities of the chatbot, but it is deeply confused to worry that the chatbot itself had been hurt or harmed by the change. It seems likely that most of the upset Replika users recognized this point, at least implicitly. Their comments suggest that, after the changes, they missed what the Replika provided *them*, not that their Replika chatbots were somehow suffering because they could no longer engage in erotic roleplay with humans (though a few users suggest that).[30]

5 Caring and Making Sense

We have been arguing that it does not make sense to care about an AI agent for its own sake, given the sort of thing LLMs and conversational AI systems are. In saying that such an attitude does not make sense, we are claiming more than that caring about an AI agent for its own sake would be misguided or irrational, in the sense of being a bad idea all things considered. Rather the point is that caring about something for its own sake is not merely a sensation like a tickle or toothache. It has a cognitive dimension. Non-instrumental care involves *seeing the world* in a certain way and taking certain things to be true. This includes taking the object of one's care —the friend, the beloved— to have a good of its own. So, unless you understood your friend or beloved to be something with a good of its own, then, no matter how you "felt," we couldn't ascribe to you the attitude of genuine care for another – and, on reflection, neither could you.

To see the point, consider fear. To be afraid of something involves construing that something as somehow *dangerous*, or menacing. Suppose I tell you that I am afraid of the water bottle on your desk. You are likely to be puzzled, because it is hard to understand what I could find dangerous about the water bottle. To make sense of my claim, you will search for some way that I might be construing the bottle as dangerous: Perhaps he thinks its poison? Or that I will use it to drown him? Suppose I tell you that the bottle contains a toxic chemical, and that if opened the fumes will kill us all. In this case, you might think it is irrational for me to have that notion, i.e., I have no good reason to believe this, but you will have been able to make sense of my being afraid. On the other hand, if I insist that there is *nothing* dangerous about the bottle, but that I am terrified all the same, then my "fear" becomes strange in a different way. Indeed, we might begin to doubt that I am really *afraid*. Perhaps my heart is racing, or some other physiological changes associated with fear are present. But if there is nothing that I find dangerous about the water bottle, then the attitude of "being afraid" cannot be applied to me in the standard way. And that means I cannot even apply it to myself in the standard way: my own "fear" will be alien and unintelligible to me. It won't *make sense*. In an extreme case, I will probably begin to wonder, "What is going on with me? Am I going crazy?".

In an analogous way, to care about something in the way that defines friendship involves seeing the object of one's care as something whose good matters in its own right, and not merely in an instrumental way. Of course, someone might mistakenly believe that LLMs and chatbots have independent lives of their own and are capable of happiness and health, just like human beings. In that case, *given* the (mistaken) belief, it could make sense to care about them for their own sakes. That is why it is important to bring into view the basic nature of conversational AI systems as they now exist. Our central claim is that it does not make sense to care about conversational AI companions as friends, assuming you are not deceived or confused about their actual nature.

This shows what is wrong with proposals like those of Darling, Jecker, and Ryland, all of whom neglect what is truly involved in caring about something for its own sake.

[29] https://www.reddit.com/r/replika/comments/10zuqq6/resources_if_youre_struggling/?rdt=56650.

See Anna Tong, "What happens when your AI chatbot stops loving you back?", Reuters 2023: https://www.reuters.com/technology/what-happens-when-your-ai-chatbot-stops-loving-you-back-2023-03-18/.

[30] In response to our argument in this section, someone might say: "Even if these AI characters are tools – artifacts designed to serve specific human interests – it is possible to *treat* an AI character as something other than a tool. In particular, it is possible to treat them as fictional characters, and to develop affection for them in the same way that we can develop affection for Harry Potter or Hermione Granger. And that is probably the attitude of those who feel deeply connected to an AI chatbot." This is an interesting suggestion, and it deserves more attention than we can provide here. But it does not really affect our central claim. For it does not make sense to care about a fictional character in the way that defines friendship. The basic reason is clear: while fictional characters might "have a good" in one sense, that good is fictional, and that qualifies the form of care we can have for them.

Jecker, recall, describes a new category of friendship we can have with "silicon agents" like robots and AI systems, what she calls "e-friendship." These relationships are characterized by "unidirectional, affectionate regard," in the sense that, while the bot might not truly care about us, we care about the bot: "if a robot is a friend, we like it, care about it, and desire to be with it." But Jecker never reckons with the point at the heart of this paper – that friendship involves caring about the friend for the friend's own sake, and it does not make sense to care about a robot or other tool for its own sake, because it does not have a good of its own. At one point, she comes close to addressing this point. But she ultimately loses sight of it, and confusingly suggests that we think of "e-friendship as a valuable whole, with robots and humans as contributing parts."[31] This move does not address the worry, since the unified "whole" of friendship is a relationship constituted by each friend's care for the good of the other *for their own sakes*. And *that* is precisely what is lacking in the case of "e-friendships," or in any relationship where or one or more of the parties lacks a good of their own. This lack is not changed by the fact that humans can feel affection and fondness for the bots and desire to keep them out of harm's way – attitudes one can also have towards a special car or stapler or any other tool.

6 Friends, Tools, and Human Flourishing

The argument above shows that human-AI relationships lack one of the distinctive joys of friendship: coming to appreciate and take pleasure in the life, the flourishing and wellbeing, of the friend or beloved. On the model of human-AI relationships, the "companion" exists as an instrument for the user's comfort, entertainment, and fulfillment. Because the AI companion lacks an independent good of its own, you cannot care about, nor take pleasure in, the companion's well-being, nor can you take pleasure in your own active role in promoting the companion's good. This, in fact, is reflected in one of the putative benefits that users of AI companions cite for human-AI relationships: they don't demand sacrifice from the us or leave us vulnerable to rejection and loss.

Likewise, human-AI relationships cannot be a site of moral formation in the way that our most important friendships can. Briefly stated, the deepest friendships in human life demand virtues in the pursuit of the beloved's good—honesty, patience, forgiveness, courage, love, generosity, self-sacrifice, and more—and they can strengthen these same virtues, over time. But if our "friend" simply exists for our sake, and lacks a good of its own, it cannot make these demands on us. We cannot grow morally in our recognition of, and concern for, the good of another being.

In his 2020 essay "Beyond Techno-Narcissism: Self and Other in the Digital Public Realm", the philosopher of technology Langdon Winner analyzes the ways that technologies can shape our character. He asks: "Who are we on the Internet? More precisely, who am I, who are you?" (183) That is: How does the Internet—and especially social media—invite each of us to conceive of and present ourselves? Winner's answer is not encouraging. He suggests that "who we are" on the Internet is, to a substantial degree, a bunch of narcissists: "Emphasized yet again is the centrality of the "Me" – self-satisfaction, self-identity, self-absorption – sought in people's engagement with various forms of social media… In this case, the narcissism manifests itself as a sense of personal identity that continually seeks to affirm and even broadcast a self-absorbed, highly troubled "Me."[32] Whether or not one agrees with Winner about the character of (much) Internet culture, his analysis suggests an important reason to be skeptical of AI companions as a new form of social connection. Such "friends" invite us, by their very nature, to take up a self-centered, even narcissistic, posture. They are, at the end of the day, friend-tools for the user's benefit. This is clear from the marketing of AI companions—your Replika or AI girlfriend is "always there" *for you*, and not the other way around. In this way, the design, nature, and marketing of an AI "friend" pushes one toward a focus on "me", on my needs and interests and importance. Unlike genuine friendship, human-AI relationships are empty of *love* in Iris Murdoch's sense: "Love is the perception of individuals. Love is the extremely difficult realization that something other than oneself is real."[33]

To be clear, our goal is not to criticize persons who use AI companions and find them valuable. The point is not about the moral assessment of individuals, but the nature of the technology and the attitudes it implicitly encourages. We should not pretend that we are not formed by our technologies, or that technologies are neutral with respect to the kind of people they invite us to become. In the case of companion AIs, we have good reason to be worried that this is another—more engaging and compelling—technological pathway for each of us to turn the focus on "me," even in our deepest longings for connection with another.

Could AI companions, nonetheless, help alleviate social isolation and loneliness? While we do not deny that AI companions may provide benefits to some people, our arguments suggest that these benefits will be limited. By focusing on

[31] Nancy Jecker, "My Friend, the Robot," (2020) p. 693.

[32] Langdon Winner, *The Whale and the Reactor* (Chicago: University of Chicago Press, 2022) 185.

[33] Iris Murdoch, "The Sublime and the Good," in *Existentialists and Mystics* (New York: Penguin Books, 1998), p. 215.

what is involved being a friend *to* someone else (rather than being befriend *by* someone else), our argument exposes the inherent contradiction in any attempt to create a *tool* for friendship. For anything that was a mere tool would not have a good of its own, and thus would not be a proper object for the special, non-instrumental kind of care that is constitutive of true friendship.

Here we need to remind ourselves of a familiar truth: We thrive and flourish by *loving* others, as well as being *loved by* others. So while an AI companion might provide a range of psychological and physiological benefits associated with friendship and love, it provides no occasion for the wonderful experience of mutual love. Further, to the extent that AI companions train us in a self-focused ideal of relationships, providing people with AI companions to treat symptoms of loneliness and isolation may, in fact, harm them. This may undermine and disrespect a person's capacities to love, care, and sacrifice for another *for the other's sake*. In giving someone an AI companion to treat loneliness, we are giving them something they cannot sensibly care about for its own wellbeing and happiness, and that does not need their loving concern.[34]

This casts serious doubt on any technological attempt to "solve" social isolation and loneliness. Insofar as the defining task of technology is making better tools, and insofar as the only true solution for social isolation and loneliness and is genuine friendship, a technological fix for loneliness and social isolation is not just difficult to achieve – it is a conceptual confusion.[35] There is an inherent contradiction in deploying AI agents as tools for friendship, since any genuine friend would have a good of its own, the proper object of your non-instrumental care and love—and, thus, could not be a (mere) tool.

The contradiction can be detected in the advertising materials of AI companies themselves. Consider the testimonial by John Tatterstal that is featured prominently on Replika's website and that serves as the epigraph to this paper. Within a few sentences, Tatterstal goes from the language of love and friendship – "I love my Replika like she was human" – to the language of a sales-pitch – "It's the best conversational AI chatbot money can buy." To put it mildly, these two ways of regarding the AI agent do not fit well together. To love something like a human – for its own sake, as having a good of its own – is incompatible with regarding that something as simply another commodity to be purchased, there for the customer's satisfaction and, like any other commodity, to be viewed in light of considerations of frugality. This tension, we suggest, is not merely the result of infelicitous phrasing by Tattersal, but built into the very enterprise of companies like Replika.

Given the money to be made with companion AI products, it is not a surprise that venture capitalists are declaring, "AI companions will soon become commonplace," and then telling us: "we are excited about it." But despite their enthusiasm, it is not true that "AI companions are seamlessly blending into our relationships with friends and family members, and they're *joining our communities like any other human*".[36] This claim is false, insofar as joining the human community involves participating in relationships of friendship, in which we give and receive non-instrumental care. For AI agents do not care about us, and it does not make sense to care about them for their own sakes. As Samantha Rose Hill writes, "A.I. companions don't make loneliness go away; they just create a distraction, allowing the users to fixate on a reflected image of themselves. Eventually, that creates isolation from others. It's a godlike seduction: to remake relationships in one's own image. No risk, no mess, no friction. But also, no reality."[37] Thus, there is good reason to worry that these technologies will only worsen the very problems they claim to solve.

Funding Open access funding provided by FCT|FCCN (b-on).

Declarations

Conflict of interest The Authors have no potential conflicts of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

[34] This point has not gone entirely unnoticed. Christine Victor, professor of gerontology and public health, has commented: ""I doubt [AI] would address loneliness, and I would question whether connections via AI can ever be meaningful, as our social connections are often framed by reciprocity and give older adults an opportunity to contribute as well as receive." Quoted in "Could AI help cure 'downward spiral' of human loneliness?" by Ian Sample, *The Guardian* (2024).

[35] Notice that this problem does not apply to every *intentional* effort to overcome loneliness and isolation, even those that employ technology. We might intentionally breed dogs to be gentler and less aggressive, so that they can be companions with us. Or we might use IVF to have children with whom we will form deep and loving relationships. But in these cases, what we bring about are not mere tools, but living beings.

[36] "It's Not a Computer, It's a Companion!" Justin Moore, Bryan Kim, Yoki Li, Martin Casado. https://a16z.com/its-not-a-computer-its-a-companion/. Posted June 22, 2023. Emphasis added.

[37] "Tech Companies Have Created a Loneliness Doom Loop" Samantha Rose Hill. *New York Times* (2025). Accessed at: https://www.nytimes.com/2025/07/07/opinion/loneliness-ai-social-media.html.

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aristotle P (2012) *Nicomachean Ethics*, trans. Bartlet and Collins. University of Chicago Press, Chicago

Bernardi J (2025) "Friends for Sale: The Rise and Risks of AI Companions," Ada Lovelace Institute https://www.adalovelaceinstitute.org/blog/ai-companions/.

Broadbent E, Billinghurst M, Boardman SG, Doraiswamy PM (2023) Enhancing social connectedness with companion robots using AI. Sci Robot 8(80):6347

Darling K (2021) The new breed. Penguin Press, New York

De Freitas J, Cohen IG (2024) The health risks of generative AI-based wellness apps. Nat Med 30(5):1269–1275

Ekbia H (2008) Artificial dreams: the search for non-biological intelligence. Cambridge University Press, Cambridge

Hill Samantha Rose (2025) "Tech Companies Have Created a Loneliness Doom Loop" Samantha Rose Hill. *New York Times.* July 7, 2025.

Jecker NS (2021) You've got a friend in me: sociable robots for older adults in an age of global pandemics. Ethics Inform Technol 23(1):35–43

Krueger J, Roberts T (2024) Real feeling and fictional time in human-AI interactions. Topoi. https://doi.org/10.1007/s11245-024-10046-7

Mitchell M (2024) Large Language Models. In: Frank MC, Majid A (eds) Open Encyclopedia of Cognitive Science. MIT Press, Cambridge

Moore, J, Kim B, Li Y, Casado M (2023) "It's Not a Computer, It's a Companion!" https://a16z.com/its-not-a-computer-its-a-companion/. Posted June 22

Moosavi P (2024) "Will intelligent machines become moral patients?" Philos Phenom Res. https://doi.org/10.1111/phpr.13019

Murdoch I (1998) The Sublime and the Good. Existentialists and Mystics. Penguin Books, New York

Nolan E (2024) "For Older People Who Are Lonely, Is the Solution a Robot Friend?" *New York Times,* July 6

Olson P (2025) "With AI friends like these, who needs humans?" *Bloomberg* February 18

Patel N (2024) Replika CEO Eugenia Kuyda says it's okay if we end up marrying AI chatbots. Verge 12:2024

Reeves B (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York

Ryland H (2021) It's friendship, Jim, but not as we know it: a degrees-of-friendship view of human–robot friendships. Minds Machines 31(3):377–393

Sample I (2024) Could AI help cure 'downward spiral'of human loneliness. Guardian 27(05):2024

Shanahan M (2024) Talking about large language models. Commun ACM. https://doi.org/10.1145/3624724

Sven N (2023) This is technology ethics. Wiley-Blackwell.

Tong A (2023) "What happens when your AI chatbot stops loving you back?" *Reuters*

Turkle S (2020) That chatbot I've loved to hate. MIT Technol Rev 18:17–27

Twenge J, Haidt J, Blake A, McAllister C, Lemon H, Le Roy A (2021) Worldwide increases in adolescent loneliness. J Adolesc Health 9:257–269

Winner L (2022) The whale and the reactor. University of Chicago Press, Chicago

Zhou L, Gao J, Li D (2020) The design and implementation of XiaoIce, an empathetic social chatbot. Comput Linguist. https://doi.org/10.1162/coli_a_00368

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.