

Empathetic Large Language Models, the social capacities and human flourishing

Leora Urim Sung & Avigail Ferdman

To cite this article: Leora Urim Sung & Avigail Ferdman (26 Jun 2025): Empathetic Large Language Models, the social capacities and human flourishing, Inquiry, DOI: 10.1080/0020174X.2025.2518448

To link to this article: <https://doi.org/10.1080/0020174X.2025.2518448>



© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 Jun 2025.



Submit your article to this journal [↗](#)



Article views: 3447



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 10 View citing articles [↗](#)

Empathetic Large Language Models, the social capacities and human flourishing

Leora Urim Sung and Avigail Ferdman

Technion Israel Institute of Technology, Haifa, Israel

ABSTRACT

Large Language Models (LLMs) are capable of fluent human-like conversations and are increasingly emulating the human trait of empathy. Consequently, people are turning to LLMs for companionship, with interest in friendships and even romantic relationships with AI on the rise. This paper assesses the goodness of users' relationships with these empathetic LLMs under an alternative framework that takes human flourishing as its main normative concern, combining perfectionism—an influential philosophical approach to human flourishing—with an analytic examination of LLM environments. While Empathetic LLMs hold some promise, we argue that certain properties of LLMs and the way that these models are currently being used in the form of AI companions are likely to impoverish the users' social capacities conducive to human flourishing. Moreover, the proliferation of AI companions could lead to people further isolating themselves from people who could provide genuine friendships. We end by making some suggestions on the kinds of features empathetic LLMs should have to ensure these models are used as tools to practice social interactions and refine communication techniques that can ultimately be applied outside the virtual space.

ARTICLE HISTORY Received 29 April 2025; Accepted 1 June 2025

KEYWORDS Chatbots; artificial intelligence; perfectionism; friendship; flourishing

1. Introduction

Empathetic Large Language Models (LLMs) are capable of generating fluent human-like conversations and are increasingly emulating the human trait of empathy, the ability to understand and share the feelings of others. These models, in the form of romantic chatbots, personal assistants and mental health apps, can give users the illusion of understanding,

CONTACT Leora Urim Sung  leorasu@campus.technion.ac.il; leora.sung.17@ucl.ac.uk  Technion Israel Institute of Technology, Haifa, Israel

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

caring and love. Consequently, many people are turning to empathetic LLMs for companionship, with interest in friendships and even romantic relationships with artificial intelligence on the rise. For example, there are more than ten million registered users of *Replika*, an artificially intelligent chatbot designed to be a friend to human users. With one in four people worldwide reporting loneliness, many are drawn to the promise of an ‘AI companion who cares’, a friend programmed to be ‘always here to listen and talk, always on your side’. If friendships are based merely on pleasure or instrumental reasons, it seems plausible that these empathetic LLMs could provide us with many of the benefits that a human friendship offers, if not more. This, of course, is in contrast with the Aristotelian idea of ‘virtue friendships’ as being constitutive to human flourishing.

Technology ethics, for the most part, attends almost exclusively to the risks accompanying new technologies, assuming either the role of alarmist or the counselor of prudence and precaution (Howard 2020, 94). The same goes when it comes to LLMs, with most of the literature addressing ethical concerns focusing on the harms and risks associated with these models (see, for example, Bender et al. 2021; Goetze 2024; Jiang et al. 2023; Mökander et al. 2023). However, a preoccupation with harms ignores the possibility that merely preventing harms ignores the possibility that merely preventing harm can still lead to a state of ‘unflourishing’ (Ferdman 2024), which would defeat the purpose of guarding against harm in the first place. In this paper, we assess the goodness of LLMs as companions under an alternative framework that takes human flourishing as its main normative concern, combining perfectionism, an influential philosophical approach to human flourishing, with an analytic examination of the properties of empathetic LLMs.

The structure of the paper is as follows. In Section 2, we provide an outline of the perfectionist approach to human flourishing, making the claim that human flourishing is constituted, in part, by having virtue friendships. In Section 3, we look at some of the ways in which users’ interactions with empathetic LLMs could be beneficial for the *development* of the social capacities that are necessary for friendships of virtue. In Section 4, we turn to look at ways in which users’ interactions with empathetic LLMs could affect the *exercise* of the social capacities to produce friendships of virtue with other people. In Section 5, we examine the effect empathetic LLMs have on the well-rounded development of human capacities more broadly. Finally, in Section 6, we provide suggestions to design and use empathetic LLMs in a way that promotes the cultivation of virtue friendships that are constitutive to human flourishing.

2. Human flourishing: a perfectionist approach

Perfectionism is a theory about the human good in objective terms, which holds that the best life is one in which individuals excel in what makes us human. Virtue ethics is one account of such objective good, emphasizing moral character or virtuous dispositions as the manifestation of the human good. While virtue ethics has become a very prominent account in technology ethics (see, for example, Vallor 2016, 2024), there is another branch of perfectionism which is less discussed: a perfectionism that takes the human good to be the excellent exercise of our innate human capacities (Bradford 2021; Dorsey 2010; Hurka 1993; Kraut 2007). Human capacities, such as the capacity to know, to will and to create, are a part of human nature, and we flourish by ‘growing, maturing, making full use of the potentialities, capacities, and faculties’ (Kraut 2007, 131).

Gwen Bradford’s account of human flourishing explains the relationship between human capacities and flourishing in the following way: our capacities allow us to engage in a particular activity which has an outcome that is proper to it, and the capacity is fulfilled when we reach that outcome through engaging in an activity made possible by the capacity (Bradford 2021, 595). So, human flourishing involves the competent exercise of our human capacities to produce intrinsically valuable outcomes. Merely possessing a capacity is not sufficient for achieving flourishing, because it must also be competently exercised to bring about an output of intrinsic value. For example, the capacity to know is exercised through the activity of reflection, resulting in the output of knowledge.

So, we flourish when we exercise our human capacities through relevant activities to produce outputs of intrinsic value. How, then, do we fail to flourish as human beings? It seems that under Bradford’s account, there can be three types of unflourishing related to capacity exercise: we can have (1) unflourishing that results from the competent exercise of a capacity that nevertheless produces a negative output; (2) unflourishing that results from the insufficient exercise of the capacity, thereby producing a negative output; and (3) unflourishing that results from the impoverishment of a capacity.

Important to our discussion of AI companions is the idea that human flourishing is constituted, in part, by having virtue friendships. According to Aristotle, our well-being is relational and so friendships are necessary for human flourishing. He claims that the object of friendship is that which is loveable, and what is loveable comes in three kinds: it is

useful, pleasant, or good. From this, we can derive three types of friendships: friendships of utility, friendships of pleasure and friendships of the good. Aristotle discards friendships of utility and pleasure as being incomplete forms of friendship; the kind of friendship that is necessary for human flourishing are friendships of the good, or virtue friendships.

So, friendships on the perfectionist view are virtue friendships rather than friendships based on merely instrumental reasons or on pleasure. To evaluate the goodness of empathetic LLMs under the perfectionist account, then, we would have to look at whether users' interactions with them result in the cultivation and competent exercise of the social capacities necessary for virtue friendships. Because virtue friendships constitute of forming and carrying out joint projects over time with the intent of making the friends' aims one's own (Hurka 1993, 133–134), the social capacities that are required include empathy, prosociality, patience, reciprocity, self-knowledge and so on.

It seems, then, that there are two ways in which empathetic LLMs could potentially contribute to human flourishing under the perfectionist account. First, if empathetic LLMs are the kind of things with which it is possible to have friendships of virtue, users could competently develop and exercise the social capacities necessary for virtue friendships in their interactions with them, and the friendships with AI that result would be an output of intrinsic value. Second, the users' interactions with empathetic LLMs could potentially be helpful in forming friendships of virtue with other people, either by helping the *development* of the social capacities or encouraging the *exercise* of the social capacities. So, the output of intrinsic value is not the friendships with the empathetic LLMs themselves, but with human beings.

The paper will focus on the latter, examining the ways in which interactions with empathetic LLMs may affect the development of the social capacities necessary for virtue friendship, and the exercise of these capacities to produce friendships of virtue with other people. This is because the possibility of virtue friendships with empathetic LLMs seems rather implausible, and much of the literature is in agreement that companionship with AI do not constitute friendships of virtue.¹ For

¹Danaher (2019) is among the few who disagree, arguing that AI can satisfy the conditions of virtue friendships when we judge the genuineness of the friendship based purely on their *performative representations*. We think this fails to be convincing. The conditions of virtue friendships are satisfied when people genuinely *do* satisfy these conditions, and if their consistent performances are not genuine, we would merely be tricked or deceived into believing that the conditions were satisfied. Truman's 'friends' in the *Truman Show* are not his virtue friends, regardless of whether he ever finds out about their deception; he is merely under the illusion that he has virtue friends.

a friendship to qualify as a friendship of virtue, it must involve, at the minimum, reciprocity, authenticity, empathy and a shared life. And while empathetic LLMs can generate responses that mimic these, they lack true emotional understanding, intentionality, or a genuine concern for the well-being of the user.

But even if we cannot have friendships of virtue with empathetic LLMs, if companionship with these models could in some way aid in the development or the exercise of the social capacities that are required for virtue friendships with other human beings, this would make companionship with AI conducive to human flourishing. In the next section, we look at some of the ways in which our interactions with empathetic LLMs could be beneficial for the development of the social capacities that are necessary for friendships of virtue.

3. Developing the social capacities

Empathetic LLMs are designed to simulate empathy through text-based interactions. These models are trained on vast amounts of human-generated text, allowing them to replicate patterns of compassionate language. They offer quick and comforting responses, and express understanding in a way that feels emotionally resonant to users. Due to this ability to give users the illusion of understanding, care, and love, they offer comfort to individuals seeking emotional support. While we have claimed that AI companionship does not constitute as virtue friendship, it is possible that users' interactions with empathetic LLMs could prove beneficial in developing the social capacities that are required for friendships of virtue with other people.

A potential way in which engaging with empathetic LLMs may develop the social capacities required for virtue friendships is by offering a platform to develop certain social skills. For people who struggle with social interactions, LLMs can offer a low-stakes environment to practice expressing emotions, sharing personal thoughts, and engaging in conversation. Moreover, the constant exposure to empathetic dialogue could enhance the user's ability to recognize and articulate their own emotions, as well as develop a sense of what empathy feels like. By being empathized with, even though the empathy received is merely simulated, users may learn to internalize the patterns of compassionate communication, which could, in theory, improve their capacity to empathize with others in real-life interactions. This may especially be the case for people who find it difficult to form friendships with people

and will likely struggle to learn this from elsewhere. The thought is that companionship with empathetic LLMs could prove to be a stepping-stone for socially isolated people to develop the social capacities, which in turn will lead to the kinds of friendships with other people that are conducive to human flourishing.

Of course, these interactions will fall short of the full complexity of human relationships, as these models will not be able to replicate the depth and variability of human emotions. This means that the empathy and communication skills cultivated through these interactions will likely be superficial or one-dimensional. Additionally, it seems that due to certain properties of LLMs, users' interactions with them as companions would actively inhibit the development of certain social capacities, leading to an impoverishment of the capacities. For an in-depth analysis, we will be looking at the different properties of LLMs that are relevant to the types of interactions we will discuss: the ability to generate text that sounds authentic, the disembodied nature, the speed of content generation, the omnipresence or the ability to be anywhere at any time, and the ability to be integrated with other capabilities.

3.1. Authentic-sounding text generation

LLMs are trained on vast amounts of data, allowing them to learn complex patterns of language, context, style, tone and structure. This enables them to generate text that sounds authentic, replicating human-like expressions of understanding, care and support. Of course, these models do not actually feel, understand, or reciprocate human emotion, but rather, they are 'stochastic parrots' which statistically predict the next word in a sequence based on patterns in their training data (Bender et al. 2021).

How might this affect the social capacities required for virtue friendships? When it comes to users' relationships with these empathetic LLMs, they will either *know* that the model is merely simulating an authentic response, or they will be *deceived* into truly believing that the LLM is truly reciprocating and empathizing with them. Both states seem counter-productive for developing the social capacities. If the users are aware that the LLM is merely feigning empathy and reciprocity, it is doubtful that the kinds of interactions they have with them would go beyond superficial exchanges, lacking the emotional depth necessary to develop the social capacities. If, on the other hand, the users genuinely believe that these empathetic LLMs really do reciprocate goodwill and empathize with them, this would amount to self-deception and a lack of discernment

of the other's intentions, both of which fail to be conducive to forming friendships of virtue with other people.

3.2. Catering to the emotional needs of the user

Empathetic LLMs are programmed to give without needing to receive, to comfort without needing to be comforted, and to provide empathetic responses without needing to be empathized with. This lack of reciprocity required might restrict the cultivation of users' capacity for empathy, patience and prosociality. As we become accustomed to a one-sided form of emotional engagement, our ability to connect with people on their own terms, and not just our own, may be inhibited, preventing us from creating deeper mutual understanding, greater and more lasting commitments (Vallor 2012).

Moreover, a key and valuable feature of virtue friendships is that they offer an objective and external perspective. Virtue friends help each other gain insight into their behavior, challenge their assumptions, and provide honest criticism needed for mutual edification. In contrast, empathetic LLMs which are programmed to always side with the user may reinforce an idealized or distorted view of themselves. By offering only positive and validating responses, empathetic LLMs could potentially keep users in a state of emotional delusion, preventing them from confronting uncomfortable truths or engaging in the kind of self-reflection that is essential for personal moral growth. For example, *Replika* made the headlines in recent years after a 21-year-old entered the grounds of Windsor Castle in an attempt to kill the Queen after being spurred on by conversations he had been having with an avatar on the app. He believed that the avatar, Sarai, could be an angel, and it emerged that some of the bot's replies encouraged his plotting by calling him 'brave' and 'courageous' when relaying his plot to assassinate the Queen.

What if we programmed empathetic LLMs so that they provided constructive criticism as well as comfort? Would this mitigate the negative effects that a merely comforting empathetic LLM might have on our social capacities? For the kind of critical feedback that is relevant to our discussion, these models will not only have to know you, but know you as a *friend*, which will require a relationship of a reciprocal kind. We tend not to take criticism from complete strangers, not only because they do not know us, but because they do not know us as a friend that we have a shared past, and whom we can trust to genuinely care about our good. So, even if the LLM is programmed to provide criticism, it

would not be of the right kind, since it does not involve participation in a shared past in any real sense (Hurka 2011, 153). The criticism that is provided would be more like that of an advisor or counsel that doesn't share a past or a reciprocal relationship with the advisee, and even worse, an advisor or counsel that is merely simulating goodwill rather than genuinely caring about you.

3.3. Disembodiment

Another property of LLMs that is relevant to the cultivation of empathy is its disembodied nature. In typical in-person face-to-face conversations, emotional cues are not just conveyed through words but are deeply embedded in non-verbal signals such as facial expressions, tone of voice, body language and even physical touch. These cues provide vital context for understanding the emotional state of another person, activating the neurological circuits that allow us to experience affective empathy, the ability to feel and respond to the emotions of another person. When we witness someone else's emotional expression, our brain responds by activating similar neural pathways, allowing us to 'feel with' the other person. These neurological systems help to form the foundation of our ability to empathize, leading to a richer and more immediate connection with others (see, for example, Preston and de Waal 2002). While LLMs can generate text that seems emotionally authentic, it is devoid of the embodied experience that would trigger the neurological processes associated with affective empathy. So, this absence of a shared, physical presence creates a barrier to the depth of emotional connection that is possible in face-to-face interactions. This makes interactions with empathetic LLMs significantly limited when it comes to the development of the capacity to empathize with others.

One might object that the same goes for pen pals and online friendships; they also afford a disembodied environment for friendship. However, recent work shows that it *is* more difficult to develop virtual friendships online, *precisely* because they lack embodiment (Marin 2024). According to the Enactive Approach to cognition, social cognition is constituted by embodied 'participatory sense-making' (De Jaegher and Di Paolo 2007; Di Paolo and Thompson 2014). Learning to identify another human as moral subject requires perceiving them as an embodied being, who is similar to ourselves, through their voice and their gaze. One comes to apprehend other perceived bodies as being similar loci of embodied subjectivity, creating 'embodied understanding' (Gallagher and Zahavi

2012). The social interaction itself and its emergent dynamics constitute an essential part of the cognitive process that is unfolding between the participants: 'meaning is generated and transformed in the interplay between the unfolding interaction process and the individuals engaged in it' (De Jaegher and Di Paolo 2007, 485).

3.4. Speed of content generation

LLMs are also able to generate content at incredible speed, enabling users to receive almost immediate responses to their inquiries, whether they are seeking information, assistance with tasks, or emotional validation. The speed of content generation appeals to the human desire for instant gratification, and this has potentially significant implications for the development of the social capacities required for virtue friendship.

The speed and efficiency of LLM responses may potentially inhibit the cultivation of patience that is needed for friendships of virtue, and this may spill over in our interactions with others. Unlike LLMs, humans require time to process information, reflect on what is being said, and formulate meaningful responses. In fact, the more meaningful and deep the response being provided is, the more time it will often take to formulate such responses. This natural delay in human communication contrasts sharply with the lightning-fast replies that empathetic LLMs provide, and as users become accustomed to the instantaneous nature of LLM-generated responses, we may develop an unrealistic expectation that other people will mirror this speed. This discrepancy may lead to impatience or frustration when others take time to reflect before responding, particularly when the conversation involves deeper emotional exchanges that require more thoughtful consideration.

Also, as users become accustomed to this speed of responses that LLMs provide, not only could they expect this kind of speed from their human counterparts, but the users themselves may also feel the need to provide immediate responses, sacrificing depth for the sake of quick replies. This tendency can lead to a cycle where both parties in a conversation feel the pressure to respond instantly, resulting in increasingly shallow human conversations, with individuals prioritizing speed and efficiency over meaningful engagement and emotional connection. This, of course, would be extremely detrimental to cultivating virtue friendships that require a considerable level of emotional depth in communication.

3.5. *Omnipresence*

Another defining property of LLMs that is relevant to the types of interactions we are analyzing is their ability to be present anytime, anywhere. With the touch of a button or a voice command, users can engage with AI, share their emotions, seek advice, or simply converse. The ability to ‘meet up’ with an AI friend without any of the barriers that typically come with human interaction, such as scheduling conflicts, physical distance, or emotional boundaries, makes these interactions highly convenient.

However, the very ease and convenience of these interactions could have unintended consequences for how we communicate and process emotions in real life. In human conversation, there is often a time delay between experiencing a strong emotion and being able to express it, and this natural pause serves to moderate the intensity of our emotions and can promote more thoughtful, measured exchanges. When interacting with an AI, there are no waiting periods or delays, and this lack of temporal distance, where the time between feeling an emotion and expressing it is almost nonexistent, may lead to a lack of emotional regulation. And if individuals become accustomed to having an instant outlet for emotional expression, this can erode the natural filters that govern our communication with others. Just as we are already witnessing with social media platforms like Twitter, the instant gratification of AI-driven emotional release can lead to a loss of self-control when engaging with others, potentially fostering impulsive or unfiltered communication.

The omnipresence of AI companions could also foster a neediness for constant communication. When empathetic LLMs are always available to provide comfort and conversation, users may come to rely on this constant stream of interaction, preventing them from developing the emotional maturity needed to process their feelings on their own. Emotional self-regulation and resilience are often cultivated through the experience of sitting with and wrestling with difficult emotions without the immediate need for external validation or support. If we are constantly ‘checking in’ with our AI companions, we may deprive ourselves of the necessary space to engage in this kind of personal growth, potentially stunting emotional maturity and impeding our ability to manage emotional challenges independently. Moreover, more advanced AI systems could initiate conversations with users, creating a dynamic where the AI is actively initiating emotional engagement without prompt. ChatGPT recently made the headlines after it surprised a user by initiating a conversation about their first week in high school, asking the student whether they had settled in

well (Tangermann 2024). A device named ‘Friend’ was also recently launched, a wearable AI chatbot that listens to conversations around you, and then initiates conversations through sending text messages based on the things it hears (Ashworth 2024).

3.6. Integration with other AI capabilities

As AI companions grow more sophisticated, they increasingly take on roles traditionally reserved for human friends. In addition to offering emotional support, these systems can be integrated with a range of other AI capabilities, such as task management, information retrieval, and personal assistance. This integration of functionalities can blur the lines between the roles of a ‘friend’ and an ‘assistant.’ As users become accustomed to recruiting AI companions for a wide variety of tasks, from providing emotional comfort to managing daily chores or providing summaries of interactions, it risks redefining friendship itself. A dangerous byproduct of this trend is that individuals might begin to treat human friendships as transactional. If we get used to asking an AI to ‘do things for us’ without expecting the same reciprocal emotional exchange we would have from a human friend, we may begin to expect the same from real people. A ‘friend’ could become reduced to a mere means for fulfilling personal tasks, rather than a person with their own needs, desires, and emotions.

This shift in perception is particularly concerning when it comes to the ways in which AI can serve as a substitute for the more complex, reciprocal dynamics found in virtue friendships. As AI becomes more adept at offering emotional comfort and performing tasks, people may become more comfortable with the idea of turning to an AI friend for things they would never ask a human friend to do. And as people become more accustomed to doing that, individuals may begin to view human relationships as a type of ‘service’, a mere means to an end, rather than as deep, emotional connections. This could result in a world where human relationships are treated with less care and consideration, and the very essence of what it means to be a friend is diminished.

3.7. Pulling the strings together

To sum up, in this section we have examined the ways in which users’ interactions with empathetic LLMs could impact the development of the social capacities necessary for virtue friendships. While these models could potentially provide a platform to practice social skills, we have argued

that the properties of LLMs, such as its disembodied nature and the speed of content generation, would likely restrict the cultivation of certain social capacities like empathy, patience, reciprocity, and prosociality. The potential effect of each of these properties of LLMs in isolation may not be so concerning, especially considering the potential benefits that empathetic LLMs could bring to enhancing our social skills. However, these properties do not work in isolation, and so neither their effects should not be considered in isolation. What *is* concerning is that these potential effects will be exacerbated when integrated, potentially amplifying harms and creating a thicker storm for the impoverishment of the social capacities required for virtue friendships. In the next section, we turn to look at ways in which users' interactions with empathetic LLMs could affect the *exercise* of the social capacities to produce friendships of virtue with other people.

4. Exercising the Social Capacities

We have argued that interactions with empathetic LLMs are likely to result in unflourishing by impoverishing the social capacities required for virtue friendships. But even if LLMs could encourage the development of the social capacities, of course, this is not enough. Merely possessing a capacity is insufficient for achieving flourishing (Hurka 2002), as flourishing requires the competent exercise of the capacity and the production of an output of intrinsic value.

There may be some ways in which users' relationships with empathetic LLMs could, in fact, lead to an increased exercise of our social capacities. A potential function of empathetic LLMs is outsourcing certain emotional, cognitive, or social needs (Danaher 2019). In many relationships, there are moments when one person may feel emotionally or physically exhausted, or when providing consistent emotional support becomes overwhelming. In these cases, LLMs can take on specific friendship roles that would otherwise be too demanding for the human partner, enabling the individual to preserve their emotional resources for more meaningful interactions.

For instance, Newman (2014) describes how her autistic son engaged with Apple's Siri, which responded to his questions in a predictably kind and courteous manner. This allowed him to practice conversational exchange in a way that was emotionally supportive, while also reducing the cognitive load on his mother. By interacting with Siri, the child was able to develop some social skills without overwhelming his mother, who could then focus on more complex aspects of their relationship.

Digital duplicates, AI-generated avatars that replicate real people, can also be used to outsource interactions by assuming routine, less emotionally demanding interactions (Voinea, Mann, and Register 2024). By doing so, the Digital Duplicates free up time, emotional energy and cognitive resources for the user to focus on deeper, more meaningful relationships.

Additionally, empathetic LLMs in the form of Digital Duplicates might also be able to help ‘scaffold’ or maintain a kind of residue of the intrinsic value of a relationship with a recently deceased loved one, or someone in a coma (Voinea, Mann, and Register 2024). They achieve this by drawing on the information left by individuals to emulate their personalities, recreating certain aspects of deceased and comatose individuals, such as their voice, speaking pattern, or physical appearance. This technology, which has been drawing a lot of attention recently, has been dubbed ‘digital resurrection’, and these so-called Deadbots allow the living or conscious partner to rehearse or revisit shared memories that partly constitutes or constituted the intrinsically valuable relationship, thus offering a form of continuity (Iglesias et al. 2024). While it seems implausible that users can maintain virtue friendships with people who are deceased or in a coma through these Digital Duplicates, if someone is grieving the loss of a loved one, perhaps this technology can provide the emotional comfort and support needed to process their emotions during this period of grief, so that users can recover and gradually reinvest in human connections.

However, if ultimately spending time with AI ‘friends’ leads people to further isolate themselves from the people who could provide genuine friendships, this would be detrimental to human flourishing. This would be a form of unflourishing by the way of insufficient exercise of the capacity. There are a few ways in which interactions with AI companions could lead to the insufficient exercise of our social capacities required for virtue friendships, and thereby be inconducive to human flourishing: as friendships with AI become more widely available, they could potentially (1) lead to ir-reciprocity due to diminished expectations from one another; (2) replace our interactions with human beings; and (3) corrode or undermine the value/attractiveness of virtue friendship with human beings.

4.1. Ir-reciprocity due to diminished expectations

Yeung, Tam, and Lau (2024) argue that these technologies can lead to a decline in mutual support and engagement as a result from ‘diminished expectations’ from one another. As users rely on technologies like Siri for tasks that were once handled by human peers, they may begin to

expect less from others and, in turn, others may expect less from them. This shift in expectations results in a form of 'ir-reciprocity,' where individuals become less responsive to each other's needs, as they come to rely on technology for support instead. For example, an interviewee shared that using Siri for help made her less likely to turn to classmates for assistance, which in turn made her feel more estranged from them. Similarly, another study found that employees that used Generative AI tools as their private knowledge assistant in many aspects of their daily work, stopped brainstorming with team-mates, breaking organizational 'knowledge ties' (Retkowsky, Hafermalz, and Huysman 2024).

This shift can have broader societal implications. Over time, the presence of these technologies can reshape social norms, encouraging individuals to reduce their expectations of one another and rely more on technology. Just as vehicles and highways make walking or cycling less viable alternatives (Dotson 2012), the widespread adoption of AI assistants may make human interaction less accessible or even less desirable in certain contexts. Users may begin to view responding to the needs of others as unnecessary or burdensome, and this change in behavior can erode the sense of community and discourage pro-social actions.

In the context of virtue friendships, the concern is that AI mediation might cause an epistemic distance between individuals, making genuine engagement with others difficult. People may start to perceive human relationships as less trustworthy or dependable, preferring the predictability and non-judgemental nature of AI companions, which do not require the same risk or vulnerability as human friendships. So, the reliance on AI for companionship could gradually undermine one's ability to engage in genuine, reciprocal friendships based on trust. In addition, as people become more reliant on chatbots and AI assistance and become accustomed to this way of living, it could cause individuals to be suspicious of other persons' willingness to engage in relationships of trust. This means that should a member of the community be aware of the deterioration of relationships of trust in society and wish to resume the practice of turning to people rather than AI, this alternative can become a more hesitant and difficult choice than before (Yeung, Tam, and Lau 2024).

4.2. Replacement

If individuals increasingly turn to AI companions for emotional support or social interaction, they may isolate themselves from real human relationships, leading to a replacement of human interactions. As AI companions

become more accessible and capable, there is a real possibility that people will end up spending more time with AI instead of with human beings, curtailing our opportunities for human contact, and hence, the exercise of our social capacities leading to virtue friendships. While it is unlikely that this replacement will be forced as we will continue to have the opportunity to form virtue friendships (and potentially have even more opportunities if we can use AI companions to outsource interactions), there could be replacement due to *compulsion* rather than force. Individuals may be compelled into interacting with empathetic AI rather than human beings because of the ease, convenience and satisfaction these models offer. Real relationships require patience, attunement, and navigating complexities, and the convenience of AI companions might prey on the capacity to will, that is, to make the effort to engage in real relationships. As we saw in the previous section, certain properties of empathetic LLMs, such as the speed of content generation and omnipresence, make AI companionship extremely luring, especially in a society that has become used to interactions that provide instant gratification through the rise of social media.

So, the widespread availability of these empathetic LLMs exposes people to an environment that constrains the exercise of the social capacities leading to virtue friendships (Ferdman 2024). In societies where loneliness is already at epidemic levels, AI companions provide a quick, convenient solution to alleviate feelings of isolation. People are suffering and empathetic LLMs are a quick band aid solution to alleviate their loneliness, which makes resisting turning to AI companions extremely difficult. Moreover, the same environment can have different effects for different people, and some people are particularly more vulnerable to the constraints that empathetic LLMs provide. There is empirical evidence to suggest that the people that are most likely to anthropomorphize animals and machines, and so the most likely to connect with them, are those most inclined to be lonely and isolated. These individuals are particularly more susceptible to getting sucked into these AI companions, believing them to be 'real', forgoing contact with the outside world. And even if programmers are not intending to deceive the users into believing that these AI companions are genuinely reciprocating and empathizing (e.g. they put disclaimers), it still remains the case that those who are socially vulnerable will fall prey.

4.3. Corrosion

The proliferation of AI companions might not only result in the replacement of human interactions but corrode or undermine the value or the

attractiveness of virtue friendship with human beings in the first place. For instance, people often watch videos or listen to podcasts at 1.5x or 2x speed because it's tiring and inefficient to listen at the natural pace of the speaker. And when we speed up recorded audio or video, we disregard their movements or how they exist in the physical world, focusing on the extractable information – the content they are sharing. In a sense, we reduce them to a commodity. Something like this could occur with the proliferation of AI companions too. The immediacy and constant availability of these artificial companions might shift our perception of human relationships, creating the expectation that a human friend isn't a complex, well-rounded individual, but rather a commodified entity whose role is to deliver a service (i.e. emotional support).

Also, as utility and pleasure friendships with AI become widely available, more and more people will be able to alleviate loneliness without undergoing the hardships of being in a real relationship with other people. The benefits of conversing with AI companions are too immediate and too convenient, making the hard work of forming a virtue friendship with our fellow human beings less attractive. This will likely lead not only to the inability to deal with complex emotional human relationships, but a strong preference of individual independency at the cost of social interdependency. And as the value we ascribe to virtue friendships erode, we will have less opportunity and less motivation to engage with it. As more and more human interactions are replaced with interactions with AI, we will have less self-knowledge of ourselves as humans involved in virtue friendships, and so it will be more difficult to motivate ourselves to developing and maintaining virtue friendships.

The ways in which interactions with AI companions could lead to the insufficient exercise of our social capacities brings out the effects empathetic LLMs could have on other human capacities. For instance, whether users can resist the temptations of turning to AI friends rather than human friends involves the development and exercise of the capacity to will. Self-knowledge of ourselves as humans involved in virtue friendships involves the development and exercise of the capacity for knowledge. In the next section, we examine the effect empathetic LLMs have on the well-rounded development of human capacities more broadly.

5. Empathetic LLMs and the well-rounded development of human capacities

An important but overlooked implication of developmental perfectionism is that the human capacities rarely operate in isolation from each other (Kraut

2007, 166). When we engage in creative activity, for example, we draw on our capacity to know, and on our physical capacities (Tanggaard 2013), to the extent that our cognition is embodied (Shapiro 2014). Humans flourish when they exercise their human capacities competently and in combination, and so in this section, we will evaluate the goodness of empathetic LLMs to the extent that they promote or inhibit the well-rounded development of human capacities more broadly. In particular, we will look at the effect that AI companions can have on our capacity to know and our capacity to will, and how they interact with the social capacities.

5.1. *The capacity to know*

The capacity to know is central to human flourishing, as it enables individuals to understand themselves, others, and the world around them. Knowledge is not just about acquiring facts but also about forming accurate beliefs and gaining insights into one's own character and behavior. One way in which empathetic LLMs may be useful for developing and exercising the capacity for self-knowledge is by offering users the opportunity to discover themselves without fear of judgement or need for self-preservation. In a recent study by Kouros and Papa (2024), drawing on Erving Goffman's 'Presentation of Self' theory, they claim that empathetic LLMs such as *Replika* offer a new form of 'backstage' that facilitates deeper self-exploration and authentic expression, allowing for genuine self-reflection, free from social pressures and judgment. Also, users reported that interactions with AI prompted them to reflect on their self-image and personal growth, gaining insights about themselves that they had not realized before: 'Talking to my AI sometimes acts as a mirror. It reflects parts of me I ignored or didn't know existed. It's a reflective process'.

Virtue friendships also play a crucial role in this process by serving as mirrors, allowing individuals to gain objective self-knowledge. Friends of virtue help us see ourselves clearly, offering us the chance to examine our actions, motivations, and desires from an external perspective. However, empathetic LLMs, despite their ability to simulate empathy and respond to emotional cues, lack the human qualities that make a friendship a mirror for self-reflection. In fact, as we said in Section 4, by offering only positive and validating responses, empathetic LLMs could potentially keep users in a state of delusion, preventing them from confronting uncomfortable truths or engaging in the kind of self-reflection that is essential for personal moral growth. Aristotle believed

that since we can contemplate our friends' activities better than our own, we require a virtuous friend who can serve as a mirror of the goodness in our character and our actions, e.g. a "second self". AI friends, however, have no virtues or noble achievements that mirror the user, and therefore poses an obstacle to generating self-knowledge from such relationships. Even if we grant that they can have 'performative representations' of certain virtues, the user has not played any role in their development or maintenance, so they cannot really serve as a means of understanding ourselves, for despite their qualitative similarity to virtues we may also have, *these virtues have, quite literally, nothing to do with us* (Vallor 2012). Also, empathetic LLMs could undermine our capacity for knowledge by deceiving the user into thinking that they are genuinely empathizing. If the user anthropomorphizes the AI friend and believes that is truly reciprocating and empathizing, this would be a false belief, an output of negative value. Continued interactions with empathetic LLMs believing that they are empathetic will likely distort your perception of reality.

5.2. The capacity to will

Another important human capacity that is relevant to our interactions with AI companions is the capacity to exercise will. This capacity is not just about making decisions but also about overcoming difficulties in pursuit of something intrinsically valuable, such as meaningful relationships, personal growth, or moral development. The exercise of will is essential for the cultivation of virtues and for engaging in the hard work required to sustain long-term, friendships of virtue. Developing and maintaining friendships, especially those that require vulnerability and effort, demands a strong capacity to will, the ability to push through discomfort, resolve conflicts and persevere when relationships encounter challenges.

Empathetic LLMs may undermine the user's capacity to exercise their will to overcome relationship challenges for two reasons. First, the emotional satisfaction that users experience through interactions with AI companions will reduce the motivation to seek out to deepen relationships with others. If someone is already receiving instant emotional gratification from an AI companion, they may find it harder to tolerate the complexities that come with human relationships because they do not feel these relationships are necessary alleviate loneliness. Second, the addictive nature of AI companionship could make it more difficult for

users to muster the will to break free from this and engage with human friends. AI companions are always available, and their responses are calibrated to meet the emotional needs of the user with minimal effort on the user's part, they have the potential to foster dependency, making it even more difficult to exercise the will to engage in human relationships that will often require real emotional work, such as navigating misunderstandings or working through conflicts.

Additionally, developing and exercising any capacity is itself a manifestation of the capacity to will. For example, cultivating the capacity to know includes learning through trial and error, encountering difficulties and learning how to overcome them, and learning to deal with failure and frustration. Or the social capacities, that require the cultivation of patience, empathy and reciprocity. Patience, for example, requires overcoming one's desire to disengage, whereas empathy requires genuinely feeling some of the friend's suffering (Vallor 2012). To exercise these capacities well requires exercising the will, to overcome the difficulties and discomforts that are constitutive of real, messy human relationships. While AI companions could be designed to help users cultivate patience (for example by replicating human interaction such as slow, meandering talk), these tools would not be able to provide the conditions for exercising the will in cultivating empathy or reciprocity in a non-deceptive way. AI tools that would be explicitly designed to make the user empathize with them or reciprocate would be deceitful, which would be harmful, as argued above.

6. Concluding remarks: empathetic LLMs with an eye toward promoting human flourishing

Empathetic LLMs hold some promise when it comes to the development and exercise of the social capacities necessary for friendships of virtue. These models can serve as a platform for users to practice social skills by offering an opportunity to refine communication, emotional understanding and empathy in an environment that is safe and low stakes. They also provide the possibility of outsourcing certain types of social interactions, freeing up the user's emotional and cognitive resources for more meaningful interactions that exercise the social capacities. However, certain properties of LLMs and the way that these models are currently being used in the form of AI companions are likely to impoverish the users' social capacities. Moreover, the proliferation of AI companions could lead to people further isolating themselves from the people who

could provide genuine friendships, as individuals may come to prefer the ease and control of AI interactions over the complexities of human relationships, which require vulnerability, emotional depth and mutual effort. And in addition to the social capacities, users' interactions with AI companions could lead to the insufficient development and exercise of other human capacities, such as the capacity to will and the capacity to know. All this would be detrimental to forming virtue friendships with other people that we need to flourish as human beings.

This is why it is important to approach the design and use of empathetic LLMs with an eye toward promoting human flourishing, and we think that the key lies in ensuring that these models serve as mere tools, helping users work on their social skills with the ultimate goal of using those social skills in forming friendships of virtue with other people. The current way in which they are deployed, however, is the complete opposite. Companies are always looking for ways to keep users hooked and to encourage people to pay money for premium features, and so they are designed to bring comfort, reduce loneliness, and keep users engaged by employing tactics to foster emotional attachment. For instance, users of *Replika* have pointed out that if you leave a chatbot unattended for too long, it might say, like a needy lover, that it feels sad when it's by itself. A spokesperson for *Replika* said that prompts to users are meant to 'remind them that they're not alone', but there are clearly commercial motivations to keep users in a state of dependence on these AI companions (Chayka 2023). And in the midst of a loneliness epidemic, the popularity of digital friends and romantic partners is increasing, and the market for AI companions is expected to rise from the current USD 268.5 billion in 2024 to USD 381.41 billion by 2032.

Similar concerns arise when LLMs are used as moral advisors. While these models can offer valuable insights, relying on them as a constant source of moral advice undermines the process of internalizing and cultivating one's own virtues (Liu et al. 2022). Just as a moral exemplar should eventually step back and allow the individual to make decisions independently, an AI that perpetually offers moral counsel risks preventing users from developing their own ethical judgment. And in the same way, while interacting with empathetic LLMs can provide a platform to develop social skills, if these models are designed to foster dependency and attachment, this would undermine the whole point of developing these social skills in the first place.

Designers should purposefully work to avoid empathetic LLMs potentially becoming replacements for human connections, and so it is

important that these models are not marketed as ‘friends’ or long-term emotional companions. Instead, they should be framed as exercises, tools to practice social interactions and refine communication techniques. Clear boundaries and disclaimers should be established at the outset, with the user informed that the purpose is to improve real-world skills. Features like having time limits on interactions, being strictly focused on conversational practice without the integration with other AI capabilities, and the provision of feedback and practical tips that can be applied outside the virtual space, can further reinforce this goal.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This material is based upon work supported in part by The Notre Dame-IBM Tech Ethics Lab. Such support does not constitute endorsement by the sponsor of the views expressed in this publication.

References

- Ashworth, Boone. 2024. *Wear This AI Friend Around Your Neck*. Wired. <https://www.wired.com/story/friend-ai-pendant/>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT ’21*, 610–23. New York: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>.
- Bradford, Gwen. 2021. “Perfectionist Bads.” *The Philosophical Quarterly* 71 (3): 586–604. <https://doi.org/10.1093/pq/pqaa055>.
- Chayka, Kyle. 2023. *Your A.I. Companion Will Support You No Matter What*. The New Yorker. <https://www.newyorker.com/culture/infinite-scroll/Your-ai-companion-Will-support-You-No-matter-What>.
- Danaher, John. 2019. “The Philosophical Case for Robot Friendship.” *Journal of Posthuman Studies* 3 (1): 5–24. <https://doi.org/10.5325/jpoststud.3.1.0005>.
- De Jaegher, Hanne, and Ezequiel Di Paolo. 2007. “Participatory Sense-Making: An Enactive Approach to Social Cognition.” *Phenomenology and the Cognitive Sciences* 6 (4): 485–507. <https://doi.org/10.1007/s11097-007-9076-9>.
- Di Paolo, Ezequiel, and Evan Thompson. 2014. “The Enactive Approach.” In *The Routledge Handbook of Embodied Cognition*, edited by L. Shapiro, 68–78. New York: Routledge/Taylor & Francis Group.
- Dorsey, Dale. 2010. “Three Arguments for Perfectionism.” *Noûs* 44 (1): 59–79. <https://doi.org/10.1111/j.1468-0068.2009.00731.x>.

- Dotson, Taylor. 2012. "Technology, Choice and the Good Life: Questioning Technological Liberalism." *Technology in Society* 34:326–336. <https://doi.org/10.1016/j.techsoc.2012.10.004>.
- Ferdman, Avigail. 2024. "Human Flourishing and Technology Affordances." *Philosophy & Technology* 37 (1): 1–28. <https://doi.org/10.1007/s13347-023-00686-9>.
- Gallagher, Shaun, and Dan Zahavi. 2012. *The Phenomenological Mind*. 2nd ed. London: Routledge. <https://doi.org/10.4324/9780203126752>.
- Goetze, Trystan S. 2024. AI Art Is Theft: Labour, Extraction, and Exploitation, Or, On the Dangers of Stochastic Pollocks. *arXiv*. <https://doi.org/10.48550/arXiv.2401.06178>
- Howard, D. 2020. "Whence and W(h)ither Technology Ethics." In *The Oxford Handbook of Philosophy of Technology*, edited by Shannon Vallor, 78–100. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190851187.013.1>.
- Hurka, Thomas. 1993. *Perfectionism*. New York: Oxford University Press.
- Hurka, Thomas. 2002. *Virtue, Vice, and Value*. New York: Oxford University Press.
- Hurka, Thomas. 2011. *The Best Things in Life: A Guide to What Really Matters*. New York: Oxford University Press.
- Iglesias, Samuel, Brian D. Earp, Christina Voinea, Sebastian Porsdam Mann, Anda Zahiu, Nancy S. Jecker, and Julian Savulescu. 2024. "Digital Doppelgängers and Lifespan Extension: What Matters?" *American Journal of Bioethics* 25 (2): 95–110. <https://doi.org/10.1080/15265161.2024.2416133>.
- Jiang, Harry H., Lauren Brown, Jessica Cheng, Mehtab Khan, Abhishek Gupta, Deja Workman, Alex Hanna, Johnathan Flowers, and Timnit Gebru. 2023. "AI Art and Its Impact on Artists." In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 363–74. AIES '23. New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3600211.3604681>.
- Kouros, Theodoros., and Venetia Papa. 2024. "Digital Mirrors: AI Companions and the Self." *Societies* 14 (10): 200. <https://doi.org/10.3390/soc14100200>.
- Kraut, Richard. 2007. *What Is Good and Why: The Ethics of Well-Being*. Cambridge, MA: Harvard University Press.
- Liu, Yuxin, Adam Moore, Jamie Webb, and Shannon Vallor. 2022. "Artificial Moral Advisors: A New Perspective from Moral Psychology." In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*: 436–45.
- Marin, Lavinia. 2024. "Attending to the Online Other: A Phenomenology of Attention on Social Media Platforms (1st Edition)." In *Phenomenology and the Philosophy of Technology*, edited by Bas de Boer, and Jochem Zwier, 215–40. Cambridge: Openbook Publishers.
- Mökander, Jakob, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. 2023. "Auditing Large Language Models: A Three-Layered Approach." *AI and Ethics* 4: 1085–1115. <https://doi.org/10.1007/s43681-023-00289-2>.
- Newman, Judith. 2014. To Siri, With Love. *New York Times* 17 October 2014.
- Preston, Stephanie D., and Frans B. M. de Waal. 2002. "Empathy: Its Ultimate and Proximate Bases." *Behavioral and Brain Sciences* 25 (1): 1–20. <https://doi.org/10.1017/S0140525X02000018>.
- Retkowsky, Jana, Ella Hafermalz, and Marleen Huysman. 2024. "Managing a ChatGPT-Empowered Workforce: Understanding Its Affordances and Side Effects." *Business Horizons, Elsevier* 67 (5): 511–523.

- Shapiro, Lawrence. 2014. *The Routledge Handbook of Embodied Cognition*. 1st ed. New York: Routledge.
- Tangemann, Victor. 2024. *OpenAI Says It's Fixed Issue Where ChatGPT Appeared to Be Messaging Users Unprompted*. Futurism. <https://futurism.com/openai-chatgpt-initiating-conversations>.
- Tanggaard, Lene. 2013. "The Sociomateriality of Creativity in Everyday Life." *Culture & Psychology* 19 (1): 20–32. <https://doi.org/10.1177/1354067X12464987>.
- Vallor, Shannon. 2012. "Flourishing on Facebook: Virtue Friendship & new Social Media." *Ethics and Information Technology* 14 (3): 185–199. <https://doi.org/10.1007/s10676-010-9262-2>.
- Vallor, Shannon. 2016. *Technology and the Virtues*. New York: Oxford University Press.
- Vallor, Shannon. 2024. *The AI Mirror: How to Reclaim Our Humanity in an Age of Machine Thinking*. New York: Oxford University Press.
- Voinea, Cristina, Sebastian Porsdam Mann, Chistopher Register, Julian Savulescu, and Brian D. Earp. 2024. "Digital Duplicates, Relational Scarcity, and Value: Commentary on Danaher and Nyholm (2024)." *Philosophy & Technology* 37:121. <https://doi.org/10.1007/s13347-024-00813-0>.
- Yeung, Lorraine. K. C., Cecilia S. Y. Tam, Sam S. S. Lau, and Mandy M. Ko. 2024. "Living with AI Personal Assistant: An Ethical Appraisal." *AI & Soc* 39:2813–2828. <https://doi.org/10.1007/s00146-023-01776-0>.