



Second-Person Authenticity and the Mediating Role of AI: A Moral Challenge for Human-to-Human Relationships?

Davide Battisti¹

Received: 28 October 2024 / Accepted: 7 February 2025
© The Author(s), under exclusive licence to Springer Nature B.V. 2025

Abstract

The development of AI tools, such as large language models and speech emotion and facial expression recognition systems, has raised new ethical concerns about AI's impact on human relationships. While much of the debate has focused on human-AI relationships, less attention has been devoted to another class of ethical issues, which arise when AI *mediates* human-to-human relationships. This paper opens the debate on these issues by analyzing the case of romantic relationships, particularly those in which one partner uses AI tools, such as ChatGPT, to resolve a conflict and apologize. After reviewing some possible, non-exhaustive, explanations for the moral wrongness of using AI tools in such cases, I introduce the notion of second-person authenticity: a form of authenticity that is assessed by the other person in the relationship (e.g., a partner). I then argue that at least some actions within romantic relationships should respect a standard of authentic conduct since the value of such actions depends on who actually performs them and not only on the quality of the outcome produced. Therefore, using AI tools in such circumstances may prevent agents from meeting this standard. I conclude by suggesting that the proposed theoretical framework could also apply to other human-to-human relationships, such as the doctor-patient relationship, when these are mediated by AI; I offer some preliminary reflections on such applications.

Keywords AI-mediated communication · Authenticity · Ethics of AI in relationships · Human relationships · Romantic relationships · Doctor-patient relationships

✉ Davide Battisti
davide.battisti@unibg.it

¹ Department of Law, University of Bergamo, Via Gianbattista Moroni, 255, 24127 Bergamo, BG, Italy

1 Introduction

“Artificial intelligence” (AI) refers to the capability of machines to simulate several aspects of human intelligence, such as learning, reasoning, problem-solving, and decision-making. In recent years, AI systems have made substantial progress and can now process texts, recognize and create videos and images with extraordinary success. Such tools can often produce outcomes that are indistinguishable from those created by humans. In text production, large language models (LLMs) generate language or perform other related tasks, such as text comprehension and classification, translation, and summarization of longer and complex texts. AI can also determine the sentiment or emotion expressed in a text as well as in audio or video, but typically the latter requires specialized models beyond those used for text, such as speech emotion recognition models (Madanian et al., 2023), facial expression recognition models (Ge et al., 2022), or multimodal emotion recognition (Abdullah et al., 2021) that combine data from various sources, including facial expressions, vocal tone, and body language.

Considering these developments, the philosophical and bioethical debate has increasingly devoted attention to the potential challenges that AI poses to everyday life. One of the most widely discussed aspects is the ethical impact of AI on human relationships. In particular, the debate has focused on relationships between humans and AI. Empirical studies show that many users develop what they perceive to be relationships with AI personal assistants, such as Alexa, and that to some extent these relationships are romantic (Gao et al., 2018); other users consider their relationships with AI systems as reciprocal and mutually beneficial (Brandtzaeg et al., 2022; Pentina et al., 2023). In this context, several applications, such as Replika, DreamGF, and RomanticAI, which involve the creation of virtual personas that converse and engage in role-playing scenarios, have been commercialized (Pan & Mou, 2024).

According to some authors, the human tendency to anthropomorphize AI, namely to attribute human characteristics or human-like traits to AI systems, poses several challenges. For instance, people can be misled into improperly giving AI moral standing, which in turn may affect their relationships with real people (Zimmerman et al., 2023). Further ethical complexities emerge when people establish friendships or intimate relationships with artificial intelligence tools (Munn & Weijers, 2023; Weijers & Munn 2022; Zhou & Fischer, 2020). For instance, in meeting users’ needs for intimacy and companionship, AI tools might also lead to social isolation and a weakening of community bonds (Figueroa-Torres, 2025). Finally, some people—especially those who are more vulnerable, such as the elderly or individuals with dementia or cognitive issues—may even be deceived by AI and believe they are interacting with a real person (Donath, 2020), as might occur with the use of care robots (Dalton-Brown, 2020). Conversely, researchers have also highlighted the positive effects of human-AI relationships. Chatbots and other AI tools may help reduce loneliness (De Freitas et al., 2024), and AI may act as a safe conversational partner, as people may feel more comfortable sharing sensitive information with an AI that lacks a capacity

to “judge” them (Skjuve et al., 2021). Moreover, human-AI relationships have led some authors to begin rethinking our conceptual understanding of relationships with non-human entities and their ethical and legal consequences (Jecker, 2024; Puzio, 2024).

While much attention has been devoted to AI-human relationships, another area that has instead been neglected concerns the moral issues of AI mediating human-to-human relationships, such as when one person uses AI tools to support their conversations with another. This paper explores such cases and concludes that using AI tools can be morally problematic when doing so undermines “authenticity-based obligations”—specific obligations in relationships that require one to personally perform a particular task. That is, at least some actions performed within human relationships should adhere to a standard of authentic conduct, as the value of such actions depends on who performs them and not solely on the quality of the outcome. Delegating some of these tasks to AI may prevent people within relationships from meeting this standard. My argument relies on “second-person authenticity,”¹ a form of authenticity that the other party within the relationships (e.g., one’s partner) recognizes². As I will argue, second-person authenticity is a necessary feature in some human relationships.

Although such reflections may be helpful for considering various kinds of relationships, the article’s main purpose is to discuss and assess how romantic relationships might be problematically mediated by AI. Given that romantic relationships are highly valued (Earp & Savulescu, 2020), the ethics of using AI in them might be readily apparent. Moreover, we might also think that authenticity plays a more important role in intimate, romantic relationships than in other kinds of relations. Romantic relationships are therefore a reliable starting point for analyzing a set of problems that emerge from the mediating role of AI.

The paper is structured as follows: in Section 2, I clarify the concept of “mediation by AI” and propose a paradigmatic example in which AI is used in a morally problematic way within a romantic relationship. In Section 3, I assess some possible explanations for why the mediating role of AI might be concerning, namely explanations that involve a lack of disclosure, lack of effort, and lack of commitment. These explanations do not exhaust ethical reflection and, in Section 4, I propose an alternative explanation that relies on second-person authenticity. In Section 5, I claim that using AI tools in some circumstances may fail because of authenticity concerns; then, I address some possible critiques of my position. Finally, I conclude by suggesting that the proposed theoretical framework could be helpful for addressing

¹ Moore (2002) has used the term “second-person authenticity,” as well as “first-person authenticity,” in the context of musicology. In this paper, however, I introduce a different usage within the framework of philosophical and moral reflection.

² I am aware that there may be more than one other person in a relationship. For instance, I have a relationship with each of my parents, but the three of us have a triadic relationship. Moreover, in the context of romantic relationships, this can involve polyamorous relationships. Without reducing some of the complexity in theorizing about relationships, in this paper, I discuss ethical issues in dyadic terms. After all, every triadic or plural relationship can be broken down into more basic relationships, where it is possible to isolate the “I” and “you” of the relationship.

other cases in which AI mediates human-to-human relationships, such as the doctor-patient relationship, and by offering some preliminary reflections.

2 The Mediating Role of AI in Human Relationships

To explain what I mean by “mediated by AI” in the context of human relationships, it is helpful to recall a conceptual shift in communication ethics. Hancock and colleagues (2020) argue that our way of interacting is shifting from Computer-Mediated Communication (CMC) to Artificial Intelligence-Mediated Communication (AI-MC). CMC refers to human communication mediated by various computer formats and channels to send messages and interact with others across time and space. Examples of CMC include sending an email or messaging on WhatsApp. Considering the availability of AI tools, instead, interpersonal communication is not merely conveyed by technology, “but modified, augmented, or even generated by a computational agent to achieve communication goals” (Hancock et al., 2020, p. 90). At the time their paper was published, they had in mind tools such as auto-correct, predictive text, grammar correction, smart replies, and so forth. However, recent advancements in AI tools, such as LLMs, provide even more powerful and disruptive examples of AI-MC.

While the rise of CMC has presented some peculiar issues (Hales, 2009), the advent of AI-MC has the potential to be even more revolutionary in that it challenges our understanding of human communication and interaction. Admittedly, we still have little empirical evidence about people’s attitudes toward AI-MC; recent research offers ambivalent results. Hohestein and colleagues (2023) find that using algorithmic responses can produce better outcomes: it increases communication speed and the use of positive emotional language, and it leads conversation partners to evaluate each other as being closer and more cooperative. However, and quite interestingly, people evaluated others more negatively if the latter were suspected of using algorithmic responses in their conversation.

So far, I have only referred to mediated communication and not mediated relationships. However, although communication may be short-term, the way in which we communicate within a relationship has significant effects on our relationships. Assessing the shift from CMC to AI-MC, therefore, seems crucial not only in communication ethics but also for the ethics of human relationships. Relationships can be defined as the connections between people who have recurrent interactions that are perceived to have personal meaning by the participants (August & Rook, 2013; Zimmerman et al., 2023). Other authors have also argued that the reciprocal exchange of information determines how relationships develop and change (Carpenter & Greene, 2015). Nowadays, a significant number of interactions and reciprocal information exchanges occur through text messages, emails, and social media. These interactions can be integrated and enhanced by AI tools such as LLMs. This already happens, for instance, in relationships with our colleagues: my anecdotal evidence, as well as several blogs and forums, suggests that in the last couple of years, a growing number of colleagues have asked ChatGPT to make an email more formal or even to write one from scratch based on a previous email. From a more

quantitative perspective, as early as November 2023, Business.com conducted a survey of about 2,000 American workers regarding their experiences with ChatGPT in 2023 and reported that around 23% of respondents were using AI for written communication, including refining and drafting emails (Business.com, 2023). I expect that this percentage is increasing, given that a survey conducted by Microsoft and LinkedIn in May 2024 found that 75% of global knowledge workers are using generative AI (Microsoft & LinkedIn, 2024). Such usage may impact the quality of one's relationships.³

Analyzing the entire range of questions related to the ethics of AI-mediated human-to-human relationships would be monumental. So, as stated in the previous section, I begin with romantic relationships. Consider the following hypothetical case:

Jamie uses ChatGPT to craft responses while messaging with their partner Kia to resolve an argument. Jamie copies and pastes the messages into ChatGPT, reviews the generated responses, and then sends them back to Kia. The responses are better than those Jamie could have formulated on their own but perfectly resemble their intention. Moreover, Jamie's messages persuaded Kia, which was the intended result. The fight between the partners is over; peace is made.

This case may reflect real-world situations. A survey conducted on the AI platform Pollfish in January 2024 found that, of US respondents between 18 and 34 years old, 33% of men and 14% of women had used ChatGPT for relationship advice (Dickey, 2024). Similarly, an author of a Medium article reported how he apologized to his wife by using ChatGPT (Anderson, 2023).⁴ However, many people might have the moral intuition that Jamie's behavior is morally wrong. The next section assesses whether we have moral reasons to support such an intuition.

3 Why is Jamie's Behavior Morally Wrong? Some (Non-exhaustive) Explanations

Here, I analyze three possible explanations—lack of disclosure, lack of effort, and lack of commitment—for why Jamie wronged Kia by using ChatGPT to resolve their argument. These explanations, I argue, are not exhaustive.

³ Although I mainly discuss circumstances in which human-to-human relationships might be negatively affected by the mediating role of AI, I leave open the possibility that certain uses of AI can have a positive impact on one's relationships. As I will argue later, whether this is the case depends on the type of mediation, the type of task assigned to the AI, and the nature of the relationship in question.

⁴ Of course, this is not the only context in which AI can mediate a human-to-human romantic relationship. For instance, in another Medium article, its author describes using AI to determine whether he or his wife was right in their recent disagreements (Stein, 2023). More generally, we can imagine an AI tool capable of *facilitating* dispute resolution between both parties. From this perspective, the idea of AI's mediating role takes on additional significance and has different moral implications that should be analyzed separately.

According to the lack of disclosure explanation, Jamie does something wrong because Kia might be unaware that Jamie is using ChatGPT, which makes Jamie's behavior morally questionable. That is, Jamie's behavior might be considered fundamentally dishonest. It is certainly plausible that the wrongness of Jamie's behavior stems from the fact that relevant information is being hidden; however, this reason alone does not fully capture our moral concerns about Jamie's use of ChatGPT. Indeed, the lack of disclosure explanation ceases to be effective when the case is slightly modified, say, by stipulating that Jamie transparently communicates to Kia that they are using ChatGPT. Even in this scenario, our intuition that Kia might still be unhappy with Jamie's behavior for other reasons remains intact and suggests concerns that are independent of Jamie's transparency. Of course, the lack of disclosure explanation is fully adequate to explain moral wrongdoing in certain cases. For instance, consider a scenario in which I enter a relationship and deliberately withhold important information, such as my real job or certain habits, because I fear it might jeopardize the relationship. If my partner later discovers this, they may feel resentment precisely because such information, which could have influenced their decision to be with me, was withheld. However, in other cases, such as a student who uses cheat sheets during an exam, people may choose to conceal something because what is being concealed is per se morally wrong, regardless of whether it is hidden or revealed. The wrongness here does not lie solely in the act of hiding the cheat sheets but primarily in the unfair advantage the student has over their peers and the teacher's inability to accurately assess the student's actual level of preparedness.

A second explanation involves a lack of effort: using ChatGPT for romantic interactions expresses laziness, which undermines the moral value placed on personal effort in one's maintaining relationships. This argument closely resembles one from the ethical debate on biomedical enhancement (Fox, 2005; Kass, 2003), according to which the moral value of conduct resides predominantly in the agent's effort such that they are praiseworthy for the effortful production of something valuable (Maslen et al., 2020). Consequently, failing to make an effort to achieve a valuable outcome, such as resolving an argument, should be considered blameworthy or at least less morally appreciable than a similar result reached through effort. Although to some extent effort can be related to moral praise, neither such explanation exhausts our moral understanding of Jamie's case. Some people might make minimal effort because they are particularly good at resolving conflicts, seeking forgiveness, or engaging in arguments. Should we consider how these people handle their arguments in romantic relationships to be morally problematic? Moreover, the effort involved in using ChatGPT, as well as the genuine intent to resolve the dispute, can be compared to the effort another person might exert in engaging and resolving an argument. We can concede that a partner who exerts a lot of effort to explain their reasons is commendable, but the lack, or at least the reduction, of this effort does not *in itself* make the action wrong. Many other morally commendable aspects, such as their good intentions and full disclosure, could make Jamie's behavior, all things considered, appropriate.

The third explanation is the lack of commitment: using ChatGPT to communicate in a relationship indicates a lack of genuine interest in maintaining that relationship, which requires a mutual commitment to addressing each other's needs. In

this sense, the effort to resolve a disagreement is not seen as valuable in itself, as in the case of the lack of effort explanation, but as an indicator of one's willingness to invest in maintaining the relationship. Using ChatGPT would demonstrate that Jamie is unwilling to invest much time in caring for their relationship with Kia, and this could significantly impact its quality. Again, we might agree that, in some cases, the use of ChatGPT can be *symptomatic* of a lack of interest and commitment in romantic relationships, but this is not necessarily true. Let us assume that Jamie feels insecure about their communication abilities and therefore uses ChatGPT to ensure that what they communicate is well-articulated. Recognizing these shortcomings and seeking support from generative tools could even be a sign of awareness of, interest in, and commitment toward Kia, especially if we consider a scenario in which someone consistently avoids arguments and disengages from disagreements with their partner. Thus, this explanation does not show why Jamie's use of ChatGPT is per se wrong; on the contrary, it suggests that using such tools could signal a strong commitment to the relationship.

4 Second-person Authenticity

I believe the aforementioned explanations do not fully capture the moral dimensions of Jamie's case. I argue Jamie's action is morally wrong because of reasons related to second-person authenticity, a form of authenticity that the person, such as a partner, in the relationship recognizes. To develop my argument, I will first present and critique the conventional understanding of authenticity, of first-person authenticity, which has, so far, received more attention in the philosophical debate. I then present the concept of second-person authenticity and emphasize its practical and moral advantages in defining authenticity-related obligations toward others. Finally, I will discuss its normative implications in the context of relationships.

Admittedly, authenticity is a very slippery concept (Varga, 2014) and is employed across various fields as well as in everyday language. Moreover, like effort, authenticity is often discussed in the bioethics of human enhancement, particularly in discussions of cognitive and moral enhancement (Bolt, 2007; Gordon, 2022; Juth, 2014). In this context, authenticity has an ambivalent meaning (Levy, 2011): those who oppose enhancement often argue that using drugs to improve certain human traits or abilities distances the person from being one's *true self*. This perspective stems from Taylor's (1992) concept of authenticity, in which authentic basically means that one listens to an inner voice that demands one be human in a way that is distinctively one's own. According to this view, everyone has a unique way of being human. In this sense, being authentic means remaining true to oneself or *discovering* one's essence.

A second perspective, which has been used to support cognitive enhancement (DeGrazia, 2000), views authenticity as the *creation* of the self: one shapes oneself into what one would like to be rather than accept oneself as one is. The emphasis on the creation of the self is also central to Sartre's (2022) rather radical view of authenticity. He asserts that humans have no predetermined essence that is shaped

by genetics, personal history, or other factors. Therefore, nothing opposes the creation of the self, which remains the only authentic stance against bad faith.

Both conceptions are *first-person* conceptions, as they concern the dialogue an individual has with oneself in discovering or creating one's true self. However, first-person authenticity encounters several challenges that hinder its moral relevance, particularly for the purposes of this paper. First, it relies on the concept of a "true self" that harbors metaphysical and epistemological pitfalls related to identifying and distinguishing essential properties that define the self and those features that can change without a loss of authenticity. Moreover, being able to refer to one's true self requires that one can distinguish between one's authentic and inauthentic selves. However, being able to distinguish between the two proves problematic both philosophically and practically. Therefore, although the idea of a true self is highly intuitive—so much so that it has achieved great success in popular culture—some authors sensibly argue that theorizing a true self is impossible (Bialystok, 2014).

Second, even if we assume that a person can access one's true self, it is doubtful that others can do the same. Despite the fact that relationships, especially romantic ones, are privileged spaces in which individuals explore themselves (Taylor, 1992), the emphasis on the search for one's true self presupposes a dialogue with oneself that renders this space virtually inaccessible to others. Given this inaccessibility, it is doubtful that someone could demand that another conform to their "true self" or raise a legitimate complaint if they do not. In other words, one does not seem to have a moral obligation toward others to be first-person authentic.

Admittedly, first-person authenticity may be morally relevant insofar as one has moral reasons to become or discover one's true self, and these reasons would be independent of how one's search for one's true self affects others. But this reasoning works only within a peculiar moral framework in which flourishing requires one realize an objective account of the good (i.e., it is good for human beings to try to be authentic) through one's search for authenticity. Outside of such a framework, first-person authenticity risks being morally unhelpful in our interactions with others—especially when others demand that we be authentic.⁵

Given the controversy of first-person authenticity and its inadequacy in defining moral obligations based on authenticity toward others, we should explore an alternative approach to authenticity, one that adopts a second-person perspective. Second-person authenticity is, indeed, assessed not by individuals themselves but by other people in the relationship. In a similar vein, the psychological literature refers to "perceived authenticity" when a partner recognizes that a behavior or action reflects the self of the person one is in a relationship with (Bailey & Levy, 2022; Wickham, 2013). In other words, whereas first-person authenticity seems accessible to

⁵ Nevertheless, I am not categorically excluding the possibility of sensibly appealing to first-person authenticity or recognizing that it has some normative force, provided that it is possible to refer to a true self (Bauer, 2017). For example, one may have prudential reasons to be authentic because, in this way, one could be more likely to achieve a higher quality of life. One may have *indirect* moral reasons toward others to be authentic, to the extent that being authentic allows one to fulfill their obligations toward others. Finally, according to some, first-person authenticity can be normative even if not in moral and prudential terms (Bell, 2024).

the subject only through an internal experience of what the subject considers to be their true self (Al-Khouja et al., 2022), second-person authenticity occurs when someone within a relationship identifies some actions or behaviors as attributable to the person one is in relationship with. This approach to authenticity avoids determining whether an action aligns with someone's true self. Moreover, unlike the first-person perspective, the second-person perspective enables authenticity to be assessed by one's partner and so has a broader moral application. Therefore, as I will explain, this framework justifies others' potential complaints about one's inauthentic behavior.

Second-person authenticity may resemble the notion of nominal authenticity proposed by Dutton's (2005) reflection on the philosophy of arts, according to which nominal authenticity is the correct identification of the origins, authorship, or provenance of an object. This commonsensical definition of authenticity is linked to the notion of authorship of a given object, outcome, or behavior. However, in the relational context, second-person authenticity differs from simple authorship because, in the latter, one can perform actions that do not reflect one's authentic features. For example, if I am usually a sober person and, in an exceptional case, I drink a bit too much and make inappropriate jokes to my partner, it can be said that I am the author of those jokes but that they do not represent the person I usually am with my partner.⁶ From this perspective, an alien factor—alcohol—explains why my partner may not recognize me as being authentic and justifies their potential judgment about my inauthenticity (Ahlin, 2018). Moreover, variability of one's behavior across different contexts and relationships affects how others' in the relationship perceive one's authenticity. While some traits may remain stable across contexts, a person who suddenly starts behaving with their partner as one might with one's work colleagues is still legitimately recognizable as the author of their actions but not recognizable as second-person authentic.

In light of these observations, I propose a more formal definition of "second-person authenticity," one that involves two necessary conditions:

- (a) The agent is the author of a given action such that they perform a certain action or behavior, where "performing" implies undertaking the cognitive and emotional process required to bring it about.
- (b) The performance of this action or behavior is consistent with their practical identity as it has emerged in the relationship with the partner.

By practical identity, I mean certain patterns of thinking, behaving, and reacting that a person shows to the other party of the relationship, which mark the person's character and personality in the partner's eyes. My understanding of practical identity is similar to Christman's (2009) and Ahlin's (2018).

I should clarify (b), as someone might argue that I am proposing an unrealistic and overly demanding concept of authenticity, one that seems to suggest that,

⁶ In this case, being inauthentic can even mitigate the moral blame my partner holds toward me.

to be second-person authentic, a person must always remain the same and cannot grow personally. For instance, let us consider Anthony, who is a fundamentally shy person and takes acting classes to become less shy. Some might argue that if these classes affect his actions and behaviors with his partner, Tamara, she may no longer recognize Anthony as authentic. This is certainly counterintuitive. Second-person authenticity is not so demanding and is meant to be compatible with one's authentic, personal growth. I believe that it is possible for a partner to change, sometimes even significantly, or exhibit behaviors that are notably different, while still being recognized by the other partner as consistent with their core characteristics. For example, if personal growth occurs—say, as a result of Anthony's acting class—this certainly changes Anthony but gradually and in a way that ensures the continuity of his practical identity. In this way, Tamara can appreciate the continuity between Anthony at time t_1 , Anthony at t_2 , Anthony at t_3 , and so on, by revising and updating her perception of Anthony. Hence, it seems to me that Anthony's behavior meets condition (b). However, it may be that case that Anthony's change is not so gradually evident, and his increased extroversion manifests in specific situations, that surprise Tamara. Nonetheless, even in this situation, a certain form of second-person authenticity may still be present because Anthony's new characteristics can be considered authentic by virtue of the reasons and motivations that led him to undergo a transformative journey.

An additional issue with condition (b) is that one's partner might have a distorted idea of what one has manifested over time—sometimes romanticized, sometimes discrediting. Hence, someone could claim that, to respect second-person authenticity, the person would need to comply with the partner's distorted idea of them. This is hardly plausible given that the partner might not recognize second-person authenticity when it actually applies and vice versa. However, condition (b) does not require one to conform to the mere idea of oneself that one's partner has. Instead, it is sufficient that one acts consistently with patterns one has effectively manifested over time in the relationship—such as the expression of one's desires, reactions, behaviors, and so on. In other words, behaving as one has always done within the relationship is sufficient. Therefore, when assessing second-person authenticity, one's partner should not rely on their problematic projection but rather on evidence that one has provided within the relationship over time. Nonetheless, in human relationships, especially romantic ones, to avoid projecting one's ideas onto another is challenging. For second-person authenticity to function and be potentially relevant in moral discourse, partners should strive to recognize each other while minimizing projection. In this paper, I only focus on cases in which the partner is sufficiently capable of adequately identifying another's practical identity.

I now discuss the normative implications of second-person authenticity. Evidence shows that simply perceiving one's partner as authentic is associated with better relationship outcomes, including a stronger identification with the relationship's goals and a weaker identification with self-protection goals (Wickham, 2013). But more importantly, second-person authenticity possesses moral relevance as well, given that when two people decide to enter a relationship they do so based on the characteristics and peculiarities of each other. Though a relationship can be sought for many reasons—protection, stability, building a life project, or social

acceptance—one enters a relationship with another based on their particularities. It aligns with our intuitions that it is likely we would not have decided to be with a specific partner if they did not possess certain physical and psychological characteristics or behaviors. Again, as I have argued, these characteristics can change while maintaining continuity over time, although it is possible that some changes are so disruptive to end the relationship. Regardless, knowledge of what a partner is like seems to be a necessary condition for entering and staying in a relationship. The mutual knowledge that partners have of each other's characteristics creates a unique, meaningful bond of intimacy and trust, which in turn can generate reciprocal obligations to which the partners agree to adhere⁷.

I identify at least two kinds of moral obligations within a romantic relationship. First, in outcome-based obligations, partners are responsible for achieving certain goods or performing specific tasks. This category generally includes duties to mow the lawn, do the grocery shopping, or iron clothes. People who are in a relationship and share a life plan agree to divide these and other tasks. For these duties, what matters is that the task is performed adequately.

A second class of duties within a relationship are authenticity-based obligations. This type of obligation requires partners to personally perform a particular action. Given the bonds of intimacy and trust within the relationship, certain tasks cannot typically be outsourced to someone or something else, even if doing so might improve the quality of the outcome. Otherwise, this would undermine the very foundation of the relationship. In other words, such obligations are *non-transferable* because they are inherently tied to specific people. From this perspective, the value of the result lies in the person performing the task, that is, in who undertakes the cognitive and emotional process required to bring it about. In this context, a person's actual involvement in the process of achieving the outcome renders the outcome meaningful (Battisti, 2025). Within authenticity-based obligations are the duties to console a loved one in grief, resolve a dispute, or seek forgiveness.

Put more formally, at least some actions performed within human relationships should respect a standard of authentic conduct insofar as the value of such actions depends on who *actually* performs them—that is, who engages in the proper cognitive and emotional processes—and not only on the quality of their outcome. Notice that, again, being authentic does not imply that one must conform to one's "true self" to fulfill one's duty; in a less demanding way, they must directly perform the task if one's partner is to recognize one as the person with whom they have entered into a relationship. In other words, compliance with authenticity-based obligations is assessed through second-person authenticity, which one's partner evaluates on the basis of external, rather than internal, criteria.

⁷ Authors like Darwall (2006) argue that the nature of moral obligations is interpersonal, or second-personal. In other words, the validity of moral obligations presupposes relations of authority and accountability between individuals and, consequently, the possibility of addressing reason from one person to another. An in-depth discussion of the relation between Darwall's perspective and my analysis of AI's role in mediating romantic relationships is beyond the scope of this paper. Nevertheless, I believe that his theoretical framework can offer additional support to my argument.

To clarify this distinction, consider another example. My partner and I have agreed upon an outcome-based obligation in which I am responsible for mowing the lawn. However, I am particularly bad at mowing the lawn, it takes me a long time, and I leave ugly marks on the garden soil. So, I decide, at my own expense, to pay a gardener to do it. I also have an authenticity-based obligation to console my partner after her sister's death. As in the case of mowing the lawn, I am not particularly good at comforting people, so I pay a psychologist to do it in my place. In the first case, I effectively fulfill the outcome-based obligation: at the end of the day, the lawn is mowed, and my partner can enjoy the benefits of a mowed lawn. I believe my partner would be happier if I decided to hire a gardener rather than do a bad job of it. In this case, when comparing two outcomes, the quality of the outcome itself is the only thing that matters. But in the second case, I fail to fulfill the authenticity-based obligation; what my partner wants is that *I*—the person with whom she is in the relationship—console her. The psychologist may be better at comforting people than I am, but that does not replace the need for comfort from my partner.

Notice that while the specific content of outcome-based obligations and authenticity-based obligations vary from one romantic relationship to another, the obligations depend on the reasonable mutual expectations of the partners. In other words, for some couples, and in some circumstances, an outcome-based obligation—such as helping a partner study for an exam or accompanying them grocery shop—might be considered an authenticity-based obligation by people in other relationships, and vice versa. I refer to authenticity-based obligations that can plausibly be regarded as such by many people and leave open the possibility that the boundaries between these obligations may be blurred.

Some might criticize the distinction between outcome-based and authenticity-based obligations by arguing that, in both cases, it is solely the partner who is ultimately responsible for fulfilling the obligation in each instance and, therefore, in both cases, the partner has non-transferable obligations. After all, even when it comes to mowing the lawn, it is *I*—and can only *I*—who takes responsibility for completing the task. However, the distinction is not about attributing a role responsibility, namely the attribution of some specific obligation to someone because one has that role (Hart, 2008). The partner's role responsibility is to fulfill both types of obligations. On the contrary, the distinction emphasizes the moral importance of who *directly performs* the task. For outcome-based obligations, we are fundamentally instruments for achieving a certain goal or good and, as instruments, can be replaced without significant moral consequences. However, for authenticity-based obligations, the *specific* person in the relationship ought to directly perform the task. Therefore, our replacement—complete or partial—in this context, can come with moral concerns that undermine the moral quality of the outcome and the integrity of the relationship.

5 The Ethics of the Mediating Role of AI

With this theoretical framework in mind, I return to Jamie's use of AI to explore the ethics of the mediating role of AI. Recall that Jamie uses ChatGPT to resolve a conflict between themselves and Kia. The ethics of conflict resolution within a couple is

a complex matter. However, it is reasonable that the partner who seeks forgiveness be the one who directly asks to be forgiven and that, just as important—or perhaps even more important—is willing to apologize. In other words, although the Jamie case is not analogous to that of paying a psychologist to console the partner mentioned above, when a partner intends to resolve an argument and seek forgiveness, they also face an authenticity-based obligation to directly perform such actions. It is not enough to have a sincere intention and sentiment to resolve the argument, one also needs to bear the cost of executing the apology.

These considerations help us understand why Jamie's behavior is morally problematic—even without appealing to the lack of commitment, effort, and disclosure explanations. Using ChatGPT extensively to resolve an argument and perform an apology fails to fulfill an authenticity-based obligation. Although the messages generated by ChatGPT align with Jamie's intentions to resolve the dispute with Kia, they cannot be directly attributed to Jamie, the person with whom Kia shares a bond of intimacy and trust. Resolving their dispute requires assessing whether that bond remains intact, yet the messages crafted by ChatGPT undermine this assessment.

Now consider why it is not possible to directly attribute to Jamie the messages that ChatGPT generated, even if Jamie has the genuine intention to resolve the conflict and subsequently sends the generated apology to Kia. First, without ChatGPT, Jamie would not have been able to produce the content of the messages, so Kia cannot perceive them as the result of Jamie's practical identity. Even worse, Jamie might not even fully understand the texts that ChatGPT crafted for them. In this case, it appears that condition (b) for second-person authenticity is not fulfilled: the performance of a certain action or behavior does not align with the practical identity of the agent available to the partner.

Second, and more interestingly, concerns about second-person authenticity would persist even if the content of the messages were consistent with Jamie's practical identity. Imagine a case in which Jamie uploaded all of their conversations with Kia such that ChatGPT could replicate some pattern proper to Jamie's practical identity.⁸ In this situation, Jamie would fail to meet condition (a): Jamie did not personally perform the act of writing the apology. As a result, Kia could not perceive the message as being from the person with whom she chose to maintain a relationship. Despite the fact that there is a sense in which they still perform the apology—namely, by copying the messages and reviewing them—Jamie neither thought about the structure, the information to convey, or the ideas to include. In the context of apologizing, these actions arise from a deeply cognitive and emotional process generated by their relationship with Kia, an actual experience that makes possible Kia's assessment of the bond and Jamie's performance of the apology. This process is fundamentally different from the algorithmic operations that LLMs employ to formulate a message similar to the one that Jamie could have written. Notice that second-person authenticity does not demand that apologies contain original ideas that have

⁸ Evidence for this possibility already exists. For instance, Porsdam and colleagues (2023) set up an LLM that was fine-tuned on the previously published work of three of its authors. The AI tool appeared to adapt to the style, argumentation, and reasoning employed in these individuals' prior work.

never been used before: it is inevitable that various patterns are reused given that we share a culture, biology, and common experiences. What matters is that these ideas emerge from a personal and emotional process in which Jamie preserves the bond of intimacy and trust that characterizes the relationship.

Thus, the Jamie example helps demonstrate how AI-MC in personal relationships can be morally problematic—when AI prevents one from fulfilling authenticity-based obligations. Since these obligations are non-transferable, we cannot largely delegate their fulfilment to AI tools, regardless of whether the outcomes align with our motivations and intentions. While we can acknowledge that the results might be more effective than what one might achieve without their use, delegating this responsibility compromises the authenticity, trust and intimacy, required of relationships. However, such a conclusion is compatible with employing, even extensively, AI tools to fulfill outcome-based obligations within romantic relationships. For instance, if I promise to keep track of household tasks and create a weekly schedule to plan and distribute them across multiple days, ChatGPT can be an effective tool for this purpose. Yet, it falls short when addressing authenticity-based obligations.

Some might object that Jamie's conduct is not morally wrong, given that it is possible to draw a moral analogy with cases in which we seek advice from our friends, relatives, therapists, or acquaintances when trying to resolve arguments with our partner. Since seeking advice from a real person is usually not considered morally problematic and already occur in daily life, AI tools should not be viewed as morally problematic either. They are simply a new way to seek support in the, sometimes hard, task of maintaining and cultivating human relationships. Using AI tools might even be preferable to asking advice from a friend, since Kia might have privacy-based reasons for Jamie's not discussing their business with friends, concerns that do not arise if ChatGPT is used.⁹

To counter this criticism, consider what must be meant by "seeking advice from a friend" for the critique to be effective. If the friend's advice is Socratic and aims at helping Jamie understand the reasons for the argument and exploring their feelings, then the objection does not pose a threat to my argument. There is nothing wrong in using ChatGPT in a similar way, namely as a tool that critically helps one understand one's own feelings before confronting one's partner. In this way, the use of ChatGPT resembles that of a confidant, which can provide, at a time other than during the discussion itself, tools that help one process their feelings and that might enhance the quality of the human-to-human relationship without necessarily raising problems related to the confidentiality of the information exchanged with the LLM. However, from this perspective, no authenticity-based obligation is undermined. The case that guides the paper is different and involves messages for Kia that ChatGPT crafts and Jamie reviews before sending them to Kia. In light of this, at least two key observations emerge.

⁹ One may have many reasons to prefer advice from a trustworthy, human friend. For example, ChatGPT would not be consistent in the way that a friend might, given the understanding that one builds up over the course of a friendship. While this observation is valid, its relevance might be limited in the near future, as it may become possible to interact with AI capable of maintaining consistent responses and behavior patterns over time.

First, the moral blameworthiness and even the appropriateness of using ChatGPT within romantic relationship comes in degrees: using ChatGPT to review the grammar of a text or even to translate it into another language is certainly different from having it directly write those messages. In some circumstances, when used in the first sense, ChatGPT may even strengthen the relationship. For instance, consider couples who do not share a primary language. Arguing and reconciling might sometimes be challenging precisely because of their language barrier; an AI tool could help mitigate those challenges. This implies that not all uses of ChatGPT compromise an authenticity-based obligation and that further conceptual research is needed to delineate when ChatGPT undermines or aligns with authenticity-based obligations. Here, it is enough to argue that while Jamie has an authenticity-based obligation, engaging in a Socratic, non-directive dialogue with an LLM does not diminish Jamie's authenticity.

Second, the conclusions I have drawn in the context of AI-MC should prompt us to *reconsider* our moral judgments about cases in which we seek advice from friends to resolve a dispute. It is not uncommon for friends to give, instead of Socratic support, rather direct advice, and, driven by the desire to resolve the uncertainty inherent in a conflict, we may even share our partner's messages with friends to get their input on how to respond more effectively. Aside from confidentiality issues, I believe that even in these cases there are second-person authenticity issues. As with Jamie's case, one's messages may not reflect the practical identity of the person performing the apology or that, even if one's identity is reflected, it is not the result of an emotional process generated by the relationship with the partner, but rather the process of a third person who is not part of that relationship. While the ethics of a friend's suggestion warrant further exploration, my argument here has shed light on this matter as well.

Nevertheless, issues related to second-person authenticity in AI-MC contexts assume a peculiar significance when compared to cases of seeking advice from friends primarily because it is much easier and cheaper to outsource certain relational tasks to AI tools than to humans. A friend may not always be willing or able to help with one's problems either because they are not good at such tasks, think it is not the right thing to do, or are simply not available. AI, on the other hand, is potentially very good at these tasks, is unlikely to have moral issues or refuse what is asked of it, and is always available. Its ease of access may undermine authenticity-based obligations more frequently and systematically. Consequently, mediating intimate relationships via AI poses a more serious challenge than doing so via human agents.

6 Concluding Remarks and Further Research

In this article, I have argued that some uses of AI-MC may undermine second-person authenticity requirements embedded in human-to-human relationships. Although I have only considered romantic relationships, the theoretical framework I have proposed is promising for evaluating other human-to-human relationships.

In addition to relationships between colleagues, this discussion is relevant for another significant example: suppose that in a not-so-distant future healthcare

scenario, advances in AI have enabled the creation of headsets that assist doctors in communicating with patients, say by detecting a patient's sentiments and emotions in audio or video. These headsets are designed to provide real-time feedback and suggestions to doctors' verbal and non-verbal communication, based on the patient's emotional state, medical history, and the nature of the information being conveyed.

If we consider the doctor-patient relationship as a relationship that presupposes *only* outcome-based obligations, we might conclude that AI would be a formidable tool to potentially improve the quality of such conversations: it could ensure a better understanding of the patient's emotional states, help the doctor assess whether the patient has understood complex communication, and so on (Savulescu et al., 2024). However, whether the doctor-patient relationship also presupposes authenticity-based obligations is an open question. If it is—if doctors were required to be the only person performing certain actions without the mediating role of AI—this tool would pose moral challenges, at least in some of its uses. For example, there might be a moral difference between patient-doctor communication that conveys a diagnosis—when effectively communicating clinically relevant information is key—and when addressing the existential and human aspects of a disease, or when a patient seeks comfort. But, again, such a putative difference would have moral value only if the doctor-patient relationship presupposes a certain requirement of authenticity in a similar—though not identical—way as romantic relationships do. To be clear, in this context, I do not intend to argue that the use of AI devices as proposed above undermines the doctor-patient relationship. Instead, I am suggesting that determining whether authenticity-based obligations exist, and therefore whether second-person authenticity can be undermined, is a necessary condition for appreciating both the moral limitations and the opportunities of AI mediation in this type of relationship.

More generally, this paper highlights that in order to morally evaluate the mediating role of AI in human-to-human relationships, we must ask ourselves what types of mutual obligations different kinds of relationships entail. Some people may have a strong moral intuition that the AI-MC used by postal employees to provide better service to customers does not raise significant moral issues regarding second-person authenticity, given that the relationship between a clerk and a customer is primarily instrumental. Certainly, as I argued in the previous sections, things change in the case of romantic relationships, doctor-patient relationship, and those that involve friends, relatives, therapists or counselors, spiritual leaders, and educators or tutors. Addressing all these issues is monumental, but I believe it has become fundamental and urgent to comprehensively evaluate the ethics of AI's mediating role in human relationships, especially given the rapid spread of these tools in our daily lives. This paper has aimed to provide a theoretical framework for this endeavor, starting from a specific case – that of romantic relationships – in which the requirement of second-person authenticity clearly ought to be respected. Further research is urgently needed to address these issues in more complex relationships.

Acknowledgements The author is very much indebted to Gary O'Brian, Federico Bina, Massimo Reichlin, and three anonymous reviewers for providing insightful comments that helped clarify several points in the paper. The author also really thanks John Ramsey for proofreading, revising the English, and offering helpful suggestions. As a non-native speaker, the author used ChatGPT for language support, such as

checking for typos in previous versions of the manuscript. All ideas, sentences, and arguments remain the author's own.

Authors' Contributions Not applicable, since there is only one author.

Funding The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The author has no relevant financial or non-financial interests 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 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

- Abdullah, S. M. S. A., Ameen, S. Y. A., Sadeeq, M., & Zeebaree, S. (2021). Multimodal emotion recognition using deep learning. *Journal of Applied Science and Technology Trends*, 2(01), 73–79.
- Ahlin, J. (2018). What justifies judgments of inauthenticity? *HEC Forum: An Interdisciplinary Journal on Hospitals' Ethical and Legal Issues*, 30(4), 361–377.
- Al-Khouja, M., Weinstein, N., Ryan, W., & Legate, N. (2022). Self-expression can be authentic or inauthentic, with differential outcomes for well-being: Development of the authentic and inauthentic expression scale (AIES). *Journal of Research in Personality*, 97(104191), 104191.
- Anderson, S. P. (2023). *I used ChatGPT to apologize to my wife*. Medium. Retrieved from <https://stephenpeteranderson.medium.com/i-used-chatgpt-to-apologise-to-my-wife-a1bc3ae2c6d2>. Accessed 13 Feb 2025.
- August, K. J., & Rook, K. S. (2013). Social Relationships. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of behavioral medicine* (pp. 1838–1842). Springer.
- Bailey, E. R., & Levy, A. (2022). Are you for real? Perceptions of authenticity are systematically biased and not accurate. *Psychological Science*, 33(5), 798–815.
- Battisti, D. (2025). The authenticity requirement: Why using digital twins for achieving person-span extension goods can be self-defeating. *AJOB: American Journal of Bioethics*, 25(2), 120–123.
- Bauer, K. (2017). To be or not to be authentic. In defence of authenticity as an ethical Ideal. *Ethical Theory and Moral Practice: An International Forum*, 20(3), 567–580.
- Bell, R. (2024). Being your best self: Authenticity, morality, and gender norms. *Hypatia*, 39(1), 1–20.
- Bialystok, L. (2014). Authenticity and the limits of philosophy. *Dialogue*, 53(2), 271–298.
- Bolt, L. L. E. (2007). True to oneself? Broad and narrow ideas on authenticity in the enhancement debate. *Theoretical Medicine & Bioethics*, 28(4), 285–300.
- Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3), 404–429.

- Business.com (2023). *How workers and businesses are using ChatGPT at work*. Business.com. Retrieved from <https://www.business.com/technology/chatgpt-usage-workplace-study/>. Accessed 13 Feb 2025.
- Carpenter, A., & Greene, K. (2015). Social penetration theory. *The International Encyclopedia of Interpersonal Communication*. Wiley. <https://doi.org/10.1002/9781118540190.wbeic160>
- Christman, J. (2009). *The politics of persons: Individual autonomy and socio-historical selves*. Cambridge University Press.
- Dalton-Brown, S. (2020). The ethics of medical AI and the physician-patient relationship. *Cambridge Quarterly of Healthcare Ethics: CQ: The International Journal of Healthcare Ethics Committees*, 29(1), 115–121.
- Darwall, S. (2006). *The Second-Person Standpoint: Morality, Respect, and Accountability*. Harvard University Press.
- De Freitas, J., Uğuralp, A. K., Uğuralp, Z., & Puntoni, S. (2024). AI companions reduce loneliness. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4893097>
- DeGrazia, D. (2000). Prozac, enhancement, and self-creation. *The Hastings Center Report*, 30(2), 34–40.
- Dickey, M. R. (2024). *Generative AI is shaking up online dating with flirty chatbots*. CNBC. Retrieved from <https://www.cnbc.com/2024/02/14/generative-ai-is-shaking-up-online-dating-with-flirty-chatbots.html>. Accessed 13 Feb 2025.
- Donath, J. (2020). Ethical issues in our relationship with artificial entities. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 51–73). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.3>
- Dutton, D. (2005). Authenticity in Art. *The Oxford Handbook of Aesthetics*. Oxford University Press.
- Earp, B. D., & Savulescu, J. (2020). *Love is the drug*. Manchester University.
- Figueroa-Torres, M. (2025). The three social dimensions of chatbot technology. *Philosophy & Technology*, 38(1). <https://doi.org/10.1007/s13347-024-00826-9>
- Fox, D. (2005). Safety, efficacy, and authenticity: The gap between ethics and law in FDA decisionmaking. *Michigan State Law Review*, 2005, 1135–1197. Retrieved from <https://ssrn.com/abstract=1072064>. Accessed 13 Feb 2025.
- Gao, Y., Pan, Z., Wang, H., & Chen, G. (2018, October). Alexa, my love: Analyzing reviews of amazon echo. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI) (pp. 372–380). IEEE.
- Ge, H., Zhu, Z., Dai, Y., Wang, B., & Wu, X. (2022). Facial expression recognition based on deep learning. *Computer Methods and Programs in Biomedicine*, 215(106621), 106621.
- Gordon, E. C. (2022). Cognitive enhancement and authenticity: Moving beyond the Impasse. *Medicine Health Care and Philosophy*, 25(2), 281–288.
- Hales, K. (2009). Ethical issues in relational maintenance via computer-mediated communication. *Journal of Information Communication and Ethics in Society*, 7(1), 9–24.
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication: JCMC*, 25(1), 89–100.
- Hart, H. L. A. (2008). Punishment and the elimination of responsibility. *Punishment and Responsibility* (pp. 158–185). Oxford University Press.
- Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, 13(1), 5487.
- Jecker, N. S. (2024). Extremely relational robots: Implications for law and ethics. *Philosophy & Technology*, 37(2). <https://doi.org/10.1007/s13347-024-00735-x>
- Juth, N. (2014). Enhancement, autonomy, and authenticity. In J. Savulescu, R. t. Meulen, & G. Kahane (Eds.), *Enhancing Human Capacities* (pp. 34–48). Blackwell Publishing Ltd.
- Kass, L. R. (2003). Ageless bodies, happy souls: Biotechnology and the pursuit of perfection. *New Atlantis (Washington D C)*, 1, 9–28.
- Levy, N. (2011). Enhancing authenticity. *Journal of Applied Philosophy*, 28(3), 308–318.
- Madanian, S., Chen, T., Adeleye, O., Templeton, J. M., Poellabauer, C., Parry, D., & Schneider, S. L. (2023). Speech emotion recognition using machine learning — a systematic review. *Intelligent Systems with Applications*, 20(200266), 200266.

- Maslen, H., Savulescu, J., & Hunt, C. (2020). Praiseworthiness and motivational enhancement: No pain, no praise? *Australasian Journal of Philosophy*, 98(2), 304–318.
- Microsoft and LinkedIn (2024). 2024 Work Trend Index Annual Report. Retrieved from <https://www.microsoft.com/en-us/worklab/work-trend-index/ai-at-work-is-here-now-comes-the-hard-part>. Accessed 13 Feb 2025.
- Moore, A. (2002). Authenticity as authentication. *Popular Music*, 21(2), 209–223.
- Munn, N., & Weijers, D. M. (2023). Can we be friends with AI? What risks would arise from the proliferation of such friendships? *International Conference on Computer Ethics*, 1(1). Retrieved from <https://soremo.library.iit.edu/index.php/CEPE2023/article/view/254>. Accessed 13 Feb 2025.
- Pan, S., & Mou, Y. (2024). Constructing the meaning of human–AI romantic relationships from the perspectives of users dating the social chatbot Replika. *Personal Relationships*. <https://doi.org/10.1111/pere.12572>
- Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140(107600), 107600.
- Porsdam Mann, S., Earp, B. D., Møller, N., Vynn, S., & Savulescu, J. (2023). AUTOGEN: A personalized large language model for academic enhancement—ethics and proof of principle. *The American Journal of Bioethics: AJOB*, 23(10), 1–14. <https://doi.org/10.1080/15265161.2023.2233356>
- Puzio, A. (2024). Not relational enough? Towards an Eco-relational approach in robot ethics. *Philosophy & Technology*, 37(2). <https://doi.org/10.1007/s13347-024-00730-2>
- Sartre, J. P., & Richmond. (2022). *Being and nothingness*. Routledge.
- Savulescu, J., Giubilini, A., Vandersluis, R., & Mishra, A. (2024). Ethics of artificial intelligence in medicine. *Singapore Medical Journal*, 65(3), 150–158.
- Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion - a study of human-chatbot relationships. *International Journal of Human-Computer Studies*, 149(102601), 102601.
- Stein, J. A. (2023). *The relationship judge*. Medium. Retrieved from <https://medium.com/@joelalanstein/the-relationship-judge-f9675cafb768>. Accessed 13 Feb 2025.
- Taylor, C. (1992). *The Ethics of Authenticity*. Harvard University Press.
- Varga, S. (2014). *Authenticity as an ethical ideal*. Routledge.
- Weijers, D., Munn, N. (2022). Human-AI Friendship: Rejecting the Appropriate Sentimentality Criterion. In: V. C. Müller (Ed) *Philosophy and Theory of Artificial Intelligence 2021*. PTAI 2021. Studies in Applied Philosophy, Epistemology and Rational Ethics vol 63. Springer, Cham. https://doi.org/10.1007/978-3-031-09153-7_17
- Wickham, R. E. (2013). Perceived authenticity in romantic partners. *Journal of Experimental Social Psychology*, 49(5), 878–887.
- Zhou, Y., & Fischer, M. H. (Eds.). (2020). *AI love you* (2019th ed.). Springer Nature.
- Zimmerman, A., Janhonen, J., & Beer, E. (2023). Human/AI relationships: Challenges, downsides, and impacts on human/human relationships. *AI and Ethics*. <https://doi.org/10.1007/s43681-023-00348-8>

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