



# Co-authoring Apologies with ChatGPT Can Support Relationships but not Second-Person Authenticity: A Reply to Lavazza

Davide Battisti<sup>1</sup>

Received: 2 May 2025 / Accepted: 2 May 2025  
© The Author(s) 2025

## Abstract

In this paper, I offer a reply to Lavazza (2025), who states that using an LLM as a co-author to apologize to a partner – far from compromising second-person authenticity – can improve relational morality. I agree that some valuable aspects of the relationship may be enhanced in the co-authorship context; however, the second-person authenticity standard cannot still be adequately met. At best, using ChatGPT as a co-author might be compatible with a scenario in which the second-person authenticity assessment is suspended and postponed to a later point in time; this allows us to capture a potential diachronic dimension in such an assessment that has not yet been explored. In the remainder of the article, I criticize Lavazza’s Kantian argument and offer some clarifications on the analogy between using ChatGPT and friend-suggested communication, as well as on the terminology used.

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

## 1 Introduction

In his insightful commentary, Lavazza (2025) provides relevant reflections that complement my argument (Battisti, 2025). In this reply, after briefly presenting both my view and Lavazza’s main claims, I offer some further clarifications and responses to his arguments aimed at continuing the urgent discussion on the mediating role of AI in human-to-human relationships.

According to my view, using LLMs such as ChatGPT to fulfill authenticity-based obligations – namely, specific non-transferable obligations in relationships that require one to personally perform a particular task – is generally problematic

---

✉ Davide Battisti  
davide.battisti@unibg.it

<sup>1</sup> Department of Law, University of Bergamo, Via Gianbattista Moroni 255, 24127 Bergamo, BG, Italy

because it fails to respect what I call “second-person authenticity”. This form of authenticity is recognized by the other party within the relationship (e.g., one’s partner) as long as: (a) the agent is the author of a given action, which implies directly performing it, namely undertaking the cognitive and emotional process required to bring it about; (b) the resulting action or behavior is in line with the agent’s practical identity as it has emerged in the relationship with the partner. A paradigmatic example of authenticity-based obligations is apologizing or resolving an argument with one’s partner in romantic relationships.

According to Lavazza, ChatGPT – far from undermining second-person authenticity – can, in some circumstances, enhance relational morality. More specifically, Lavazza proposes a variation on my hypothetical case of Jamie and Kia, a couple who have an argument. To solve it, in my original case, Jamie copies and pastes the messages into ChatGPT, reviews the responses and sends them to Kia; in Lavazza’s version, rather than massively relying on ChatGPT, Jamie better refines their prompt, adjusting the language until it produces a message that they find appropriate, even making modifications to the chatbot’s responses. According to Lavazza, this version more clearly illustrates a sort of *shared authorship* with ChatGPT.

In light of the possibility of such a shared authorship, Lavazza offers two arguments in favor of the idea that using ChatGPT can enhance the relationship: (1) Jamie may become aware of their limited cognitive and emotional capacities in a specific situation and, by using ChatGPT, might find new words they would not have found without it, which could provide more effective tools to convey messages to Kia; (2) stemming from a Kantian ground, Lavazza argues that moral worth does not stem from an authenticity expressed through harmony between actions and inclinations, but from acknowledging obligatoriness of a given course of action. In this view, there is something morally commendable, if Jamie, moved by a good will, uses ChatGPT to comply with their duty to apologize with Kia.

## 2 Co-authorship, Second-person Authenticity and the Diachronic Dimension

I address (1) and (2) in turn, starting with some preliminary clarification. I agree with Lavazza that not all uses of ChatGPT undermine second-person authenticity. In my paper, I presented the example of language translation during an argument for couples who do not share the same native language. Furthermore, I agree that the potential blameworthiness in using ChatGPT to fulfill authenticity-based obligations comes in degrees: since the aim of my contribution was to present a novel moral-philosophical concept, I proposed cases in which there is *clearly* either a lack or a presence of authenticity in using ChatGPT in the fulfilment of authenticity-based obligations; however, it is true that there can be a great variety of cases where the conclusion regarding authenticity is more blurred, such as the one proposed by Lavazza.

One general question is whether, in this co-authored case, Jamie adequately complies with the authenticity-based obligation. To answer this, recall that to fulfill authenticity-based obligations, Kia must appreciate – through a second-person

authenticity assessment – whether Jamie, the person with whom she is in the relationship, meets the conditions (a) and (b). However, even in the case of co-authorship with ChatGPT, I doubt that such an assessment can be adequately made. This is not because there is no cognitive or emotional process undertaken by Jamie or because the message is hugely inconsistent with the individual's practical identity; in fact, in the variation proposed by Lavazza, Jamie is more actively involved than they are in my version. But in co-authorship context, it is still difficult for Kia to isolate Jamie's contribution and determine who produced what. Thus, in Lavazza's example, there is an inescapable opacity in the joint work between Jamie and ChatGPT, which makes the authenticity assessment hard to carry out. This assessment is also difficult because it is unclear whether the relationship can be sustained without external assistance. In other words, it is hard to properly identify the subject of the relationship, since the fulfillment of the obligation is shared with an external author: ChatGPT.

To preserve the possibility to meet second-person authenticity while still receiving support from ChatGPT, its use should be limited to functionalities that do not attribute to it any form of authorship. Again, language translation is the clearest example of this, but I concede that other usages can be accepted, including correcting typos or asking ChatGPT whether certain parts of one's message are unclear or might be misunderstood by the partner. In such cases, ChatGPT acts not as a co-author but rather as a proofreader, preserving Jamie's practical identity, their cognitive and emotional labour, and Kia's ability to recognize these aspects.

Although co-authorship cannot fulfill authenticity-based obligations, it is nonetheless true that other valuable aspects of the relationship can be enhanced. ChatGPT's use could be an expression of Jamie's commitment to their partner, as long as they are transparent about having used ChatGPT. Moreover, as Lavazza observes in his argument (1), by sharing co-authorship with ChatGPT, Jamie could also demonstrate an initial willingness to learn to acquire the words and thoughts they previously lacked and recognized as missing, in order to effectively address a certain problem in the couple, leading to frustration.

Lavazza's example emphasizes that there is a plurality of valuable aspects within a relationship that are not reducible to one another. This plurality allows us to consider that in the scenario he proposes in (1), even though Jamie fails to fulfil the obligation based on authenticity, the situation may still be open-ended. Kia may be willing to assess second-person authenticity not immediately, but later – in light of the commitment Jamie may demonstrate and their willingness to learn to better apologize. Lavazza's contribution is thus hugely important as it helps to highlight a potential *diachronic* dimension in fulfilling an authenticity-based obligation with ChatGPT – a point I had not adequately considered earlier. The diachronic element matters, as it helps us to appreciate that relationships are not reducible to isolated interactions but consist in exchanges reiterated over time.

In sum, yes: ChatGPT can effectively enhance some aspects of the relationship without necessarily compromising second-person authenticity; but it is important to reiterate that the use of ChatGPT in itself cannot say anything about the fulfillment of the authenticity-based obligation and the respect for second-person authenticity. At best, the use of ChatGPT may be compatible with a scenario in which the

assessment of second-person authenticity is suspended and postponed to a later time. Lavazza seems to agree on this, in particular when he says that if Jamie fails to demonstrate that “he has truly learned something about how to engage with his partner in situations of disagreement or misunderstanding” (Lavazza, 2025, 3), Kia might not consider the authenticity-based obligation fulfilled.

### 3 A Reply to Lavazza’s Kantian Argument

Let us now turn to (2), the Kantian argument which I view more skeptically for two reasons. First, second-person authenticity does not require that an agent’s actions align with their inclination. The concept of inclination and that of authenticity are distinct. The former refers to an agent’s pre-reflective propensity to behave in a certain way under certain circumstances; more broadly, authenticity concerns a more complex sphere of behavior that certainly involves inclinations, but also how a person manages those inclinations. In other words, one may have certain inclinations but still rationally recognize that, at times, those inclinations are problematic – because they may harm others – and therefore acknowledge the need to limit them, while preserving their practical identity and, consequently, second-person authenticity. It follows that when a person has an authenticity-based obligation, this does not mean that they should act in harmony with their inclinations; rather, the person should deal with this required task without the possibility of delegating it to someone or something else. For example, suppose I am someone who never apologizes because I’m full of myself, and I do something very wrong to my partner, something that would require apologizing. If second-person authenticity required acting just according to one’s inclinations, then I should not apologize at all. But that is not what I want to argue. What I argue is that I—the arrogant person my partner unfortunately decided to be in a relationship with—should bear the cost of fulfilling the obligation to apologize, even if this frustrates my inclinations.

Second, recall that, according to Lavazza, Jamie may be morally commendable for using ChatGPT as they feel morally obligated to repair the conflict with Kia. In this context, Jamie faces two options: to solve the disagreement with Kia without ChatGPT, delivering apology X, or to use ChatGPT and deliver apology Y, where Y is a better apology than X (e.g., more persuasive, deep, and effective). Contra Lavazza, I believe that from a Kantian perspective, it is doubtful that Jamie has a duty to deliver apology Y, since it exceeds their current capacities. Kant asserts that an agent has a moral obligation only if they have the capacity to perform the action (Kant, 1781/1929); in other words, “ought implies can”. However, what seems to be implied by Lavazza’s argument is that Jamie has an obligation to perform Y. Admittedly, the Kantian principle may allow for different interpretations depending on how we define “can”: one may argue that Jamie can perform Y precisely because ChatGPT is available; but if we consider the obligation to fix things with Kia an authenticity-based one, we can reasonably argue that, in that moment, Jamie cannot have this kind of capacity without failing to fulfill the obligation. Therefore, Jamie might have an obligation only to perform X. Of course, Jamie might have a duty in

the long run to cultivate their skills to perform Y, but at current time there is no obligation based on Kantian grounding to perform Y.

#### 4 Final Clarifications on Analogy and Terminology

For reasons of space, I do not discuss other interesting aspects of Lavazza's commentary such as the argument that – drawing from an extended mind perspective – in a future where we may access an LLM directly and in real time (e.g., via brain–computer interfaces), problematic cases like the one involving Jamie and Kia could apparently evaporate. This conclusion may be persuasive in a speculative scenario, but it does not undermine my argument's relevance in the present when people are increasingly integrating AI into messaging or email.

I conclude with two clarifications. First, Lavazza states that he is not convinced by the way I dismissed the analogy between the help provided by ChatGPT and that offered by a friend. However, I did not entirely dismiss the analogy. I do maintain that problems with second-person authenticity can even arise in contexts where friends suggest what a person should say to their partner. In this sense, the reflection on AI and second-person authenticity is precisely a way to re-examine moral intuitions surrounding friend-suggested communication – typically seen as unproblematic – in light of second-person authenticity requirement (Battisti, 2025, 15). From this perspective, the analogy makes perfect sense. The potential differences I pointed out can be appreciated on another level: the frequency and scope of AI's role in mediating human relationships. AI is potentially very effective at mediating human-to-human relationships, is unlikely to have moral issues or refuse what is asked of it, and is always available. Therefore, its ease of access may undermine authenticity-based obligations more frequently and systematically, compared to friend-suggested communication.

Second, a terminological note. Lavazza sometimes refers to second-person authenticity as “relational authenticity”. When writing my paper, I considered this term, but I am hesitant to use it. In my view, the term “relational authenticity” is better suited not to describe a situation in which one person recognizes the partner's behaviour as authentic, but rather to capture the authenticity of the product of the relationship itself: a form of authenticity in which both partners are co-authors. For example, when two people love each other, we may appreciate their love through the lens of relational authenticity. Second-person authenticity, by contrast, refers to the recognition by one partner of the authenticity in the other's behavior.

**Acknowledgements** As a non-native speaker, the author used ChatGPT for language support. All ideas, sentences, and arguments remain the author's own.

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

**Funding** Open access funding provided by Università degli studi di Bergamo within the CRUI-CARE Agreement. 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

- Battisti, D. (2025). "Second-Person Authenticity and the Mediating Role of AI: A Moral Challenge for Human-to-Human Relationships?" *Philosophy & Technology* 38 (1). <https://doi.org/10.1007/s13347-025-00857-w>.
- Kant, I. (1929). *Critique of pure reason* (N. K. Smith, Trans.). Mcmillan and Co. (Original work published 1781)
- Lavazza, A. (2025). "AI, Authenticity, and the Chatbot between Us: Commentary on Battisti." *Philosophy & Technology* 38 (2). <https://doi.org/10.1007/s13347-025-00880-x>.

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