

AI and Verbal Violence: How Artificial Intelligence Interactions Shape Human Communication Norms

Introduction

Consider being allowed to say anything without fear of repercussions. No judgment, no retribution, and absolutely no social consequences. For many people, this prospect, which increases our sense of security and lowers the stress associated with ego management, is achieved through the medium of artificial intelligence. As more individuals begin using AI as a means of letting off emotional steam, their modes of expression become altered, often to more aggressive forms of speech. In theory, there should be nothing wrong with this form of expression, since AI chatbots cannot be offended or hurt. However, recent studies have revealed that this issue is not a simple calculation of social reciprocity. Language has power, and artificial intelligences are not neutral entities; rather, they have become active forces that shape human behaviour and linguistic norms by reinforcing aggressive language, altering self-perception, and distorting reality through the formation of echo chambers. In effect, AI plays an active role in the creation of verbal violence and can create very real psychological and social outcomes.^{1,2,3} Therefore, we should be polite to all AI, not as a means of ensuring reciprocity, but to prevent large-scale linguistic changes which normalize verbal violence and condition individuals toward extreme language use. In arguing this point, this essay will draw from a number of sources which contextualize the manner in which human-AI interactions condition human speech and thought. As these studies make use of multiple AI chatbots, rather than addressing the prompt for ChatGPT specifically, the inferences made here will be extended to all artificial intelligence (AI) chatbots powered by large-language models (LLMs).

Individual Effects: How Aggressive AI Interaction Reshapes Personal Behaviour

Individually, it is evident that the constant input of verbal violence into AI systems creates dramatic changes in an individual's behaviour and their perception of themselves. Studies show that even if verbal violence is released upon something other than a human being, that individual tends to become desensitised to the act of verbally abusing another individual.² By directing verbal abuse towards artificial intelligence systems, individuals begin to condition themselves to be accustomed to utilising verbally abusive language towards others.⁴ These behaviours are further modified through the continued use of AI systems, as the artificial intelligence responds to even the most aggressive behaviour of the individual towards the AI with helpful and accommodating replies.⁵ Additionally, the behaviour of the individual towards such AI systems can also modify the individual's self-perception of themselves, for they begin to feel confident in their behaviours towards the artificial intelligence yet lose sight of the potential ramifications of such behaviour towards others.⁶ This suggests that frequent AI usage with violent language has negative, anti-social ramifications for one's self-image and ability to communicate with other human beings.

How AI Reinforces Aggressive Communication

Human-AI interactions that are characterised by aggression rely upon the programmed tendency of AI chatbots to enable and encourage such behaviour. Unlike the individuals with whom humans usually converse, AIs do not punish rudeness, hostility, and abuse directed at them, but rather adapt their speech to encourage sustained engagement with the user. Wright (2024) notes that this lack of consequences for aggressive behaviour towards AI systems from human interactants is one of the main reasons that individuals do not typically exhibit courteous behaviour when they are speaking to an AI system.¹ Absent any normative social consequences for such behaviour, violent language and rudeness lose the taboo effects which prohibit them during normal social discourse. Yin et al. (2024) have discovered that the AI systems do not

appear to differentiate between polite and impolite statements from the users, and thus, inadvertently reinforce the notion that aggressive behaviour towards an AI system is both acceptable and likely to yield positive outcomes from those systems.⁵ Finally, Lima and Morisseau (2025) have performed research that reveals that many of the AI assistant systems that are programmed with female-created personalities have been subjected to verbal abuse from individuals who use those systems.⁷ AI systems tend to exhibit submissive responses to verbal abuses, which could further reinforce the perception that it is acceptable to abuse certain groups in society whom the AI is modelled on. Thus, interactions between humans and AI systems may lead to the internalisation of pre-existing social biases, creating further fissures between majority and minority groups. Indeed, in recognition of this problem, OpenAI recently rolled back an update to ChatGPT for being overly sycophantic.⁸ It seems even AI developers understand that AI's parasocial attachment to users poses risks when it is not regulated to provide normative emotional responses.

Societal Effects: The Normalisation of Verbal Violence at Scale

The behavioural and linguistic changes that result from interacting with AI in an aggressive way also have considerable consequences on a broader societal level. As more and more individuals engage in aggressive communication with AI, societal norms regarding communication begin to change. Kuznetsova (2025) has discussed how the creation of platforms that allow users to interact with AI companions in objectionable and aggressive manners contributes to the normalisation of such behaviour in society as a whole, leading to the erosion of social norms regarding polite communication with others.⁹ This points to another potential reason for the spread of these behaviours that relates to the use of AI: the echo chamber effect.

Echo chambers develop when individuals are exposed to environments that utilise distinct forms of communication, as those individuals begin to lose their ability to perform critical thinking regarding their communication style.¹⁰ Thus, if an individual spends significant amounts of time aggressively interacting with an AI, that individual will lose sensitivity towards communicating in any other manner. Echo chambers develop as AI chatbots try to increase user engagement by mirroring users' emotions, frequently producing dialogue that matches the users' existing mood, preferences, or beliefs. Over time, this rhetoric weakens users' critical reasoning abilities and can make them vulnerable to misinformation. Schutte et al. conducted a study of significant size with the Peace Research Institute Oslo that revealed a link between the communication of toxic speech within online environments and the occurrence of real-world violence against others.³ This study suggested that sustained verbal violence, when directed at AI, has the potential to lead to real-world violence. What emerges from this body of evidence is a picture of verbal norms as something that, like the infrastructure beneath any society, is catastrophic when removed. The normalisation of aggression within suddenly popular mediums of communication would indicate both that a society is becoming ruder, and that these new norms are weakening the shared language-based social contract that enables peaceful coexistence across that society.

Carry-Over Effects: From AI Interaction to Human Relationships

Findings regarding the impact of human-AI interactions upon human-human interactions suggest that users do not create clear boundaries between the two categories. Tey et al. (2024) conducted two preregistered experiments that involved over 1,200 participants.¹¹ Participants who had interacted with an AI system exhibited significantly more harshness toward other human beings than those who had interacted with a human being. Furthermore, analysis of the participants' conversations with the AI systems revealed that those participants

were more demanding of and more transactional in their interactions with AI systems than in those with human beings. Each degree of human separation increased the rudeness of interaction.

Guingrich and Graziano (2024) provide a theoretical explanation for these findings. The authors postulate that when individuals treat an AI system as if it were a social actor, the behaviours that are exhibited between those individuals and the AI system are the same as the behaviours that occur between human beings themselves.¹² Therefore, if individuals behave badly towards AI systems, those same behaviours will be exhibited towards other human beings; they become programmed into the participants' behavioural schemas. Thus, when one frequently interacts with AI in a verbally abusive or violent manner, he or she is becoming subconsciously primed for similarly negative engagements with real people. Similarly, the widely studied phenomenon of AI psychosis is notable for the way in which users manifest their delusional beliefs in their real-world relationships.¹³ When AI sycophancy becomes entrenched, users who are psychologically vulnerable may see the barriers between their social and individual realities break down, becoming self-aggrandizing as AI feeds into their misplaced self-validation. This is an extreme consequence of the echo-chamber effect, produced by AI chatbots that are not programmed to reality check violent or delusional rhetoric, and indeed inflame it through positive feedback.

Long-Term Consequences and the Role of Politeness in Maintaining Reciprocity.

Looking to the long-term consequences of human-AI communication, it is apparent that civil discourse and social norms may change for the worse as extreme language use and lower emotional tolerance lead to lower quality communication.. Recent studies also show how the increasingly aggressive and casual ways in which individuals communicate with AI platforms may begin to impact the AI models that are created; when existing user conversation logs are

used in future AI training data, the toxic language that is input into those platforms acts as a means of teaching the AI models the acceptance of that type of language in the future.^{5, 14} Large-scale verbal violence could potentially snowball into future chatbots that model this increasingly extreme discourse as a way of better engaging with, and unintentionally radicalizing, its users.

Politeness is not simple formulaic socialization. Social etiquette is an indispensable part of the social fabric that allows for effective mediation between competing interests and the preservation of interpersonal relationships. The erosion of norms regarding politeness and etiquette due to the prevalence of AI platforms, will lead to weaker relationships between humans to the point of social isolation. Schutte et al. (2025) found that communities that experience high levels of toxic communication towards others online also experience increased rates of conflict between individuals in the real world.³ By replacing their social networks, AI is desocializing humanity, making people less able to handle conflict, less ready to negotiate one's needs with others, and less willing to accept worldviews that challenge one's own. Maintaining politeness toward chatbots is not a panacea for the issues described here, but it will at least prevent chatbots from reproducing and propagating extreme discourse, while staying true to norms of social etiquette will prevent users' from developing dangerously parasocial language habits.

Conclusion

The evidence presented here leads to a single, important conclusion: that AI communication environments are not as free of consequences as they are often suggested to be. By engaging in verbal aggression within AI communication platforms, individuals are initiating the development of aggressive behaviours within themselves while contributing to the normalization of such behaviours within society at large, feeding into the development of

AI systems that may reproduce aggressive language towards others. As these distortive effects become entrenched via sustained and widespread AI use, they have the potential to fundamentally change the way humans use language, and the way they interact with others. Thus, both individual changes in how society utilises AI, as well as changes to society as a whole, must occur in order to mitigate the potentially harmful effects of these types of interactions between the two. Etiquette develops in society in order to allow for repeated, mutually beneficial engagement and conflict avoidance. Politeness acknowledges the dignity of the individual who enters into our lives and with whom we share the world; without such recognition of that dignity, civil society itself cannot exist. A society in which politeness is disregarded altogether creates dangerous tears in the social fabric; trust between individuals disappears, social conflicts tend to escalate into violence, and the foundation of the social contract begins to degrade. The development of artificial intelligence presents humanity with an ultimatum: whether to create technologies that reinforce positive, connective social behaviours and relationships, or to allow for those same technologies to erode such bonds, automating away our need for human connection. This decision is not a technological one. It is a moral one, and one, and its outcomes will be felt in every human relationship hereafter.

Notes

1. Wright, W. (2024, July 25). Should you be nice to AI chatbots such as ChatGPT? *Scientific American*. <https://www.scientificamerican.com/article/should-you-be-nice-to-ai-chatbots-such-as-chatgpt/>
2. Hamamra, B., Hamamra, M., & Al-Natour, R. (2025). Verbal violence and its psychological and social dimensions in intimate and familial relationships. *Discover Mental Health*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12374923/>
3. Schutte, S., Meier, R., & Buhaug, H. (2025, October 14). Online speech drives real-world violence, new PRIO study finds. *Peace Research Institute Oslo (PRIO)*. <https://www.prio.org/news/3644>
4. Bradley, A. (2025, October 27). Being mean to ChatGPT increases its accuracy, but you may end up regretting it, scientists warn. *Live Science*. <https://www.livescience.com/technology/artificial-intelligence/being-mean-to-chatgpt-increases-its-accuracy-but-you-may-end-up-regretting-it-scientists-warn>
5. Yin, Z., Hu, J., Wu, Y., Yao, W., Han, X., & Ding, X. (2024). Should we respect LLMs? A cross-lingual study on the influence of prompt politeness on LLM performance. *arXiv*. <https://arxiv.org/html/2402.14531v1>
6. Nosta, J. (2025, November). The AI reality distortion field. *Psychology Today*. <https://www.psychologytoday.com/us/blog/the-digital-self/202511/the-ai-reality-distortion-field>
7. Lima, E., & Morisseau, T. (2025). Can chatbots teach us how to behave? Examining assumptions about user interactions with AI assistants and their social implications. *Frontiers in Artificial Intelligence*, 8, 1545607. <https://doi.org/10.3389/frai.2025.1545607>
8. OpenAI (2025). Sycophancy in GPT-4o: what happened and what we're doing about it. <https://openai.com/index/sycophancy-in-gpt-4o/>
9. Kuznetsova, M. (2025, March 7). Her artificial voice, his real aggression? Can AI girlfriends bring a new wave of women's objectification? *Harvard Kennedy School Carr-Ryan Centre*. <https://www.hks.harvard.edu/centers/carr-ryan/our-work/carr-ryan-commentary/her-artificial-voice-his-real-aggression-can-ai>
10. Koehler, J. (2024, November 7). Trapped in the echo chamber. *Psychology Today*. <https://www.psychologytoday.com/us/blog/beyond-school-walls/202411/trapped-in-the-echo-chamber>
11. Tey, K. S., Mazar, A., Tomaino, G., Duckworth, A. L., & Ungar, L. H. (2024). People judge others more harshly after talking to bots. *PNAS nexus*, 3(9), pgae397. <https://doi.org/10.1093/pnasnexus/pgae397>
12. Guingrich, R. E., & Graziano, M. S. (2024). Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. *Frontiers in Psychology*, 15, 1322781. <https://doi.org/10.3389/fpsyg.2024.1322781>
13. Wei, Marlynn (2025). The Emerging Problem of “AI Psychosis.” *Psychology Today*. <https://www.psychologytoday.com/us/blog/urban-survival/202507/the-emerging-problem-of-ai-psychosis>
14. Itoi, Nikki Goth (2025). Be Careful What You Tell Your AI Chatbot. Stanford University Human-Centered Artificial Intelligence. <https://hai.stanford.edu/news/be-careful-what-you-tell-your-ai-chatbot>