

# Review of: "On Violence: an Exchange with ChatGPT — ChatGPT and Suman Gupta"

Lenhart Schubert<sup>1</sup>

<sup>1</sup> University of Rochester

**Potential competing interests:** No potential competing interests to declare.

As noted by other reviewers, the article provides no indication of what the chat is intended to achieve or reveal. The contributions to the dialogue by both you and ChatGPT are quite sophisticated, and this certainly creates opportunities for drawing conclusions about ChatGPT, or about the legal and ethical understanding of violence it reflects, about the LLM's depth of understanding, consistency, susceptibility to alteration of its expressed attitudes, etc., but you don't pursue these possibilities.

One gets the feeling that you just want to make known your own attitudes about the conception of violence that is reflected in ChatGPT's responses (and as such in the written record of society at large), and how far you can push the LLM to broaden that conception so as to acknowledge indirect violence by social groups, corporations, and lawmakers – and perhaps the creators of LLMs. If so, it appears you succeed in part, while getting some pushback. I would point out two relevant characteristics of ChatGPT: First, its context window is limited, and I have found that it may totally forget what it asserted earlier in a long conversation. You might have asked it to reiterate its first (or second, revised) definition of violence after the discussion of corporate abuses, to check for loss of context. The other point is that ChatGPT tends to yield to user corrections, perhaps because that's what's generally observed in polite dialogues or perhaps as a result of fine-tuning or reinforcement learning. Yann Lecun has mentioned asking ChatGPT for the sum of two plus two, and telling it "No, the answer is 5, not 4", which the LLM then accepted apologetically. So it's to be expected that it would go along with your suggested amendments to the definition of violence – at least for a while.

So, are you treating ChatGPT as a fluent repository of received legal and popular wisdom, and trying to bring out the flaws of the latter, arguing for broader definition of the agencies culpable for "violence"? Or is your goal to provoke readers' thoughts about the tradeoffs between open access and responsible usage of LLMs, and risks for a new or exacerbated social divide? Or something else? In any case, actual extended discussion separate from the chat is needed, with some clarity as well about the intended audience -- legal, sociological, philosophical, scientific, technological, or whichever.

