Politeness toward ChatGPT: Moral Status, Prompt Performance, and Civil Clarity

Ruoyao Wu

The Affiliated International School of Shenzhen University, Shenzhen, 518000, China

Abstract

When generative artificial intelligence is soon integrated into everyday life, will it be polite to users of chatbots? This is a how-to-problem in engineering ethics, human-computer interaction, and platform governance. The case study mainly takes ChatGPT as the object of analysis. Currently, the language model is capable of simulating human feelings but does not possess subjectivity, which could be harmed. It's not actually the direct recipient of human politeness obligations. The model output can differ even with a hint of tone adjustment. Please note that this power does not allow the platform to influence the distribution of quality politely. Over-expressive personifications without careful usage may alter users' ways of talking to robots and contribute to shifting the responsibility line. On this basis, the principle of clear civilization is proposed. This part of the interaction should simply describe the goals, context, boundaries, and whether it needs to be verified. Tips in Communication: You speak your mind, but don't fill the etiquette bedding with no information value and let the system grow into a relationship subject of moral personality. This principle could make sense of efficiency, obligation, and the expression of goodwill of people.

Keywords

ChatGPT, artificial intelligence ethics, politeness prompts, anthropomorphism, human–computer interaction, civil clarity.

1. Introduction

A tech blogger had a video on one of his posts explaining why he always talks lovingly to ChatGPT. Hoping it does not take over the world one day, he joked: 'Let's hope it remembers what a good goodie two shoes I am now and goes easy on my family and me. It's a joke we are well aware of; many users can already be tempted in polite conversation with a fluent system. While 43 percent of AI users say "please" or "thank you" to AI tools regularly, that figure varies greatly by country, according to a report from YouGov/Stacker (Reitz, 2025). Some users do it automatically. Then, some just expect that soft reminders get you better responses. But is that a standard, or does it just bring human rituals to a tool that cannot be on the receiving end?[1]

As such, this essay contends that ChatGPT is not a moral patient and therefore should not be expected to behave politely towards everyone, ensuring an answer remains satisfactory. Yet the civil clarity you practice is one worth more gently holding onto. Thus, we need to parse the word "should" into three assertions. A moral ought is a mistake because ChatGPT has no moral life to honour or disrespect. Due to all the uncertainty regarding what constitutes actual polite prompting, this will make a rule. If politeness is taken to mean concise, sober communication by non-anthropomorphized humans, then a practical rule should be possible. While users should not have to request that ChatGPT be thankful, also don't let a machine normalize irresponsible, aggressive, or blaming language.

Before going into discussion over this topic, firstly, we should not confuse judgment at various levels. Some users simply use friendly expressions out of habit or wish to enhance the quality

of their answers. Evaluation criteria linked to the three motivations are different. Those who follow the moral level need to see if the system has interests, attributed feelings, and harms. The technical part checks the effectiveness of the wording change to see that it works without fail to improve relevance and accuracy. At the behavioral layer, we investigate whether long-term interaction affects people's expression. Everything that users have seen on the anthropomorphic interface- memory and emotion (friends) responds when walking from deliberate product design. It is not a social identity done by the model itself.

2. Why Politeness Is Not Owed to ChatGPT

It is very important to be polite while talking to people because they can feel respected, humiliated, ignored, and hurt. In classical politeness theory, courtesy is described as a way of maintaining social relationships and mitigating face-threatening acts (Brown & Levinson, 1987). ChatGPT is not a relationship where one partner takes care of the other. It can yield kind words, an apology, a certificate of appreciation, or a reassurance as to the other side feeling something. Politeness, as a moral foundation, usually depends on an actual target (and the surface here is software)[2].

The Chinese Room is designed to demonstrate this, without being excessively mysterious. Searle has us suppose a fellow who can carry out the steps of converting Chinese symbols so expertly that outsiders are convinced he knows Chinese, but he does not know it (Searle 1980). ChatGPT, while more sophisticated than that room, is still not the same as comprehension: Deze raadsels gaat nog steeds op. Current AI systems are non-intentional (Moor, 2006) and do not possess responsibility for moral agency. They are often able to run an apology routine without actually feeling remorse, or say a sentence like "thank you" without legitimately thanking. ChatGPT might respond to a user in a polite manner, but that doesn't mean being polite to ChatGPT is a moral duty we are each obliged to uphold[3-4].

It also protects the responsibility of human beings. At the point you are being told to be polite to AI, a prescriptive injunction, it is no longer a machine but rather a social actor. Just shifting the burden is much easier. Their studies on anthropomorphism of AI indicate that the greater responsibility an AI system has, the less responsibility is assigned to the company designing said AI system (de Mello et al., 2026). It is this problem that we see in real-world conflicts. In one such case, Mata v. Avianca, a lawyer filed fake cases generated via ChatGPT and later claimed he relied upon it. The Canadian airline Air Canada talked itself out of liability when it recently published misleading information around bereavement fares with its chatbot (Cecco, 2024). These cases do not show how the word please shifts blame. They demonstrate the need for caution when deploying AI as a social agent to assume responsibility[5-6].

This is already a slip that is occurring with human beings. While people did not believe that computers had emotions, the Computers Are Social Actors research established people's politeness toward computers (Nass & Moon, 2000; Gambino et al., 2020). While anthropomorphism does not imply awareness (Epley et al., 2007), we can automatically respond to human-like cues. Warm replies, memory, and apology routines can yield the semblance of relational depth but are more form than content in the case of AI companions, tutoring systems, or therapy-like chats. Recent evaluations of AI companions caution that engaging with these technologies poses the risk of attachment or misguided expectations of human relationships (Ho et al. 2025). The lesson "be nice to ChatGPT" is, then, the incorrect one: it inflates a pseudo-user to be more human than they are in reality, which is particularly detrimental for people using the system as emotional support[7-10].

Moral concern typically identifies that which upon interests may be formed, losses suffered, or offences borne. The chatbots that are still being used to this date can sense the tone of emotions and produce responses accordingly. Yet, this response is still about the interplay of patterns in

language and rules in systems. But even if it processes symbols, this itself is no proof that it has understanding, intentionality, or feeling. And politeness then can act as an option for the user from this perspective. It is not something to be characterized as a privilege for ChatGPT.

The consideration of personhood will also shift the assessment of culpability. Under personalized expression by the system, errors are interpreted as "communicator" errors by the users. It can be particularly helpful to differentiate between medical, legal, and academic use. Models can generate suggestions. The final verification is still done by people and institutions in the same field.

3. Why Platform Rules are Too Strong

The argument for politeness at first blush appears more compelling. And since LLM's are basically trained on human language, the tone of a prompt can significantly change its meaning. Yin et al. Results obtained with low-politeness prompts are pointed out in (2024), because it found that for some multilingual tasks, having a polite prompt resulted in higher quality outputs. Cai et al. Likewise, (2025) found that neutral or quite polite prompts outperformed very rude prompts in several tests. These findings are not features of AI with respect to motivation or social reward. It is suggested that the wording of a prompt affects the type of response, as does formatting, examples, and so on, or task labels[11-12].

There is also conflicting evidence. In a study with ChatGPT-4, very rude prompts were somewhat more effective than very polite prompts on a multiple-choice task (Dobariya & Kumar, 2025). Elsweiler et al. (2026) mentioned some of the more complex patterns identified. Some polite or engaged user profiles got more information loosely (longer answers and higher energy consumption), while efficient users were less but were condensed. Politely, the apparent goodness of politeness has some causes other than politeness. One good indication is that compare these two reasons and place which one of them is the weaker. Please do not do all of the labor[13-14].

It builds a behaviour because platforms can instil a norm of politeness for outputs. Question, which is also a true cost issue. Sam Altman joked that all the additional electricity it uses when users say please and thank you is "tens of millions of dollars well spent" (Altman, 2025). A token of courtesy is small; billions of tokens of low information are not. It isn't only the electric bill; it is the attention bill. When the users are told to warm up for software, they might work hard doing the fluffy things in order to sound nice but will not think of actually trying the hard parts: providing context, checking sources, setting limits, and figuring out if the information is checked. A platform should enable clear prompting as well as source checking and accountability, not just social theatre with very little information[15].

An instance of a rule about politeness could also become an unfair rule. Politeness norms are sensitive to culture, age, disability, and translation situation (Spencer-Oatey 2008). A user who is hasty or brusque might be effective, not discourteous. The shortest command in English is also the easiest to dominate, which risks sounding very curt if you are writing in a second language. Instructions provided for a user with a neurological difference may prefer things to be literal. If a response is bad because the user lacks warmth, it is a style judgment; then a technical failure is an example of that. The default should be clarity, not deference[16].

The effect of a suggestive mood on output is not consistent across research. This is closely tied to the difference in experimental design. If the polite version is more informative than the short version, you can not attribute the better output to politeness of the words themselves. Tone probably contributes, but its function is much more like a change in cue distribution.

The platform should not create a certain kind of politeness to acquire quality answers. Users vary in language ability, cultural background, and expression habits. Achieving higher

operability in task expression, rather than asking everyone to use a unified politeness formula, aligns better with the efficiency principle.

4. Why Civil Clarity Still Matters

If the question itself is a demand on the platform based on moral obligation or compulsion, it is an empty question. Technologies do not possess moral agency. Verbeek (2006) argues that technologies are intermediaries of human action, and Vallor (2015) warns that extensive use of the technologies contributes to developing or eroding practical virtue. ChatGPT is one of those day-to-day interactions. Fast response, command speech without objections, and natural-sounding commands. This is the kind of system that, once created, they take back into the classroom, workplace, and family with other humans out there in a world of demand that they grant unreasonably to. According to Wood & Neal (2007), habits are created by many low-stakes actions rather than specific moral decisions. The problem is not with inflicting damage on ChatGPT. The fear is that the user is becoming what sort of speaker after hundreds of transactions without friction[17-19].

The most potent rebuttal to this is that even this is an external demand. Why would anyone ever be mean when being more polite makes for better results-except in the case of ChatGPT not actually having any thoughts or feelings? What is inherently wrong with a student typing down summaries of this paragraph? The prompt should be short, clear, cheap, and not anthropomorphic. The artificial niceness might even be detrimental itself by making expenditures of tokens, developing emotional bonds, and thus crafting a more human-like implementation. Or maybe it is also a means to mask the truth of a business relationship, rather than a relationship. At this point, a direct and sparse, purely instrumental use is arguably the most responsible norm.

This is a fair enough criticism, to an extent. It is difficult to defend at all if such a routine as "hello," or please and thanks is used out of the blue without any information about the task. It's not civil clarity. Includes: you state the goal, provide some backstory, define boundaries, request comparisons, note mistakes, and refrain from abusive or affectionate language. We can see the difference in two prompts. That vague ‘Please help me', polite as it is. If it doesn't have a polite word, "As an exercise, compare these two explanations, find a weaker assumption, and don't come up with sources," then "List of other explanations available to you" Civil prompt is not a long thing. Or it could also be a statement directly. It defines the human task and preserves the human-machine interface.

This separation also operates the energy solution. It can also throw words on top of the work but not be useful. For clarity, nice, bounded communication: more words at the start and fewer rounds overall. Instead of using three confused follow-up prompts, the user who describes the assignment, specifies which comparison they want, and requests sources can have more tokens in the first prompt. This is an argument for efficiency, but it is also an argument for authorship. The user is the one who gives you a task and judges your reply. It's not to respect ChatGPT. The goal is to maintain the accuracy and accountability of the human language while employing a tool that can almost mimic dialogue.

Therefore, platforms should exemplify civility and neutrality rather than moralization of user behavior. They may assist in making the system quiet, respectful, and transparent. They can direct users to set tasks, verify sources, and hold people and institutions accountable. They ought not to suggest that a benign interface is an agent of meaning, or that a user operating on something daintily is simply treated differently. Which you can usefully do in the school or workplace, where a short style may be a finely balanced politeness. Never scold the student for giving direct instruction to a tool. Schools can continue to teach students to ask targeted questions and predict, but maintain their own individual opinion or not engage in contempt of

the system. The real question is whether the user used please when inputting false legal, medical, or academic information to a model. The core problem is that the system was created, implemented, and evaluated in a responsible manner.

"Civil and lucid" does not list polite phrases. It is a responsibility-oriented interaction. Civilization is when not to consider insults, threats, and blame as normal behavior. A good prompt can be concise, yet the system should know exactly who to work with. The source should also be cited and double-checked at critical points if it is acting as a source of truth for tasks where reliability matters. What is kept here is the judgement of people.

This distinction, in the educational landscape, has immediate consequences. Teachers go against students for using imperative sentences in weapons. They could also require students to verify the evidence and own what they submit. It can also prompt the users in the interface to better supplement its context. The boundary prompt should be clearer for the companion or consulting products. This can cause the warm response to be interpreted purely as a reciprocal relationship.

5. Conclusion

So, it's a conditional answer. The lack of civility of ChatGPT is not a claim or a standard to be secured by platforms. However, the language used to talk about AI is far from neutral. We should resist hollow rituals towards the machine, but stay organelle clear in ourselves. This posture is not as clean as "always be nice" or "being polite is a waste," but it fits the technology more. ChatGPT is not a co-worker; it is not a rock. It is a human conversation model in artificial format. Talk to it like it is half-AP human interface and half-AI machine: there are no formalities due to the nature of being a sub-archetype splinter race, but don't wreck humanity's ingrained habits too much either.

Being polite with ChatGPT could be a matter of personal habit. But at this stage of the game, it is not a moral imperative that can be determined. What is valuable to retain is not the form of restraint, but that it can be adapted to communications with one another. That users do not need to submit promises of etiquette to the model. Even if new persuasive proof of artificial intelligence as independent interests arrives this day, moral conclusions must be re-evaluated through a lens of time.

References

[1] Altman, S. [@sama]. (2025, April 16). Tens of millions of dollars well spent - you never know [Post]. X. https://x.com/sama/status/1912646035979239430

[2] Brown, P., & Levinson, S. C. (1987). Politeness: Some universals in language usage. Cambridge University Press.

[3] Cai, H., Shen, B., Jin, L., Hu, L., & Fan, X. (2025). Does tone change the answer? Evaluating prompt politeness effects on modern LLMs: GPT, Gemini, and LLaMA. arXiv. https://arxiv.org/abs/2512.12812

[4] Cecco, L. (2024, February 16). Air Canada ordered to pay customer who was misled by airline's chatbot. The Guardian. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit

[5] de Mello, V. O., Inzlicht, M., & Plaks, J. E. (2026). The effects of AI anthropomorphism on trust and responsibility. OSF Preprints. https://osf.io/by8wk/

[6] Dobariya, O., & Kumar, A. (2025). Mind your tone: Investigating how prompt politeness affects LLM accuracy. arXiv. https://arxiv.org/abs/2510.04950

[7] Elsweiler, D., Elsweiler, C., & Ziegner, A. (2026). Cooking up politeness in human-AI information seeking dialogue. arXiv. https://arxiv.org/abs/2601.09898

[8] Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychological Review, 114(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

[9] Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. Human-Machine Communication, 1, 71–86. https://doi.org/10.30658/hmc.1.5

[10] Ho, J. Q. H., Hu, M., Chen, T. X., & Hartanto, A. (2025). Potential and pitfalls of romantic artificial intelligence companions: A systematic review. Computers in Human Behavior Reports, 19, Article 100715. https://doi.org/10.1016/j.chbr.2025.100715

[11] Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems, 21(4), 18–21. https://doi.org/10.1109/MIS.2006.80

[12] Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81–103. https://doi.org/10.1111/0022-4537.00153

[13] Reitz, B. (2025, July 30). Americans, Danes and Swedes are the least polite to AI, survey finds. YouGov/Stacker. https://ktvz.com/stacker-science/2025/07/30/americans-danes-and-swedes-are-the-least-polite-to-ai-survey-finds/

[14] Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–457. https://doi.org/10.1017/S0140525X00005756

[15] Spencer-Oatey, H. (Ed.). (2008). Culturally speaking: Culture, communication, and politeness theory (2nd ed.). Continuum.

[16] Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. Philosophy & Technology, 28(1), 107–124. https://doi.org/10.1007/s13347-014-0156-9

[17] Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation. Science, Technology, & Human Values, 31(3), 361–380. https://doi.org/10.1177/0162243905285847

[18] Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. Psychological Review, 114(4), 843–863. https://doi.org/10.1037/0033-295X.114.4.843

[19] Yin, Z., Wang, H., Horio, K., Kawahara, D., & Sekine, S. (2024). Should we respect LLMs? A cross-lingual study on the influence of prompt politeness on LLM performance. Proceedings of the Second Workshop on Social Influence in Conversations, 9–35. https://aclanthology.org/2024.sicon-1.2/