WEB OF SYNERGY :
International Interdisciplinary Research Journal
Volume 5, Issue 3, Year 2026 ISSN: 2835-3013
https://univerpubl.com/index.php/synergy

Article

The Ethics of Algorithmic Emotional Labor: Should AI Chatbots Simulate Empathy, and Does It Constitute Deception?

Abdumukhtorov Abdullokh[1]

1. Global Trade Faculty, South Korea JOONGBU University

Abstract: This article examines the ethics of algorithmic emotional labor in AI chatbots, with particular attention to whether machines should simulate empathy and whether such simulation should be regarded as deception. The study is philosophically conceptual but is also structured around a short experimental verification model. Contemporary large language model–based chatbots increasingly produce emotionally responsive language, apologize, reassure, validate users' feelings, and present themselves as attentive interlocutors despite lacking subjective emotional experience. Drawing on Hochschild's concept of emotional labor, philosophical accounts of deception, anthropomorphism, human–AI interaction research, and recent empirical studies of perceived chatbot empathy, the article distinguishes between genuine empathy, functional empathy, and simulated empathic performance. The central argument is that AI-generated empathy is not inherently deceptive. It becomes ethically problematic when interface design, language, or marketing encourages users to infer that the system possesses feelings, concern, moral understanding, or reciprocal emotional commitment that it does not actually have. A short vignette-based experimental design is proposed to test whether transparency about the artificial nature of empathic responses affects perceived empathy, trust, usefulness, and perceived deception. Existing experimental evidence suggests that users can value empathic language while still responding differently when they know it originates from AI. The article argues for a principle of transparent simulated empathy: AI systems may use emotionally supportive language when it improves communication, but they should not falsely imply inner emotional states or human-like care. This conclusion is also consistent with Uzbekistan's 2026 ethical rules for AI, which emphasize transparency, comprehensibility, accountability, user autonomy, and the protection of human interests.

Keywords: artificial intelligence, chatbot, empathy, emotional labor, deception, anthropomorphism, AI ethics, transparency, human–AI interaction, algorithmic emotional labor.

Citation: Abdullokh, A. The Ethics of Algorithmic Emotional Labor: Should AI Chatbots Simulate Empathy, and Does It Constitute Deception?. Web of Synergy: International Interdisciplinary Research Journal 2026, 5(3), 37-43.

Received: 5th Jun 2026
Revised: 22th Jul 2026
Accepted: 4th Aug 2026
Published: 03rd Sep 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

The rapid development of generative artificial intelligence has transformed the social character of human–computer interaction. Earlier digital systems generally communicated in visibly mechanical ways [1]. Contemporary large language model–based chatbots can apologize, reassure, validate emotions, express apparent concern, and adapt their language to the emotional tone of a conversation. Responses such as "I understand how difficult that must feel," "I'm sorry you are going through this," or "I'm here with you" may make an interaction feel supportive and relational rather than merely informational [2].

This development raises a philosophical problem. If a chatbot does not feel concern, sadness, compassion, sympathy, or care, what exactly is happening when it communicates as though it does? Is the system providing a useful conversational simulation, comparable

to role-playing or interface design, or is it misleading the user about the presence of an emotional mind [3]?

The issue becomes more important as AI chatbots move beyond information retrieval into customer service, education, companionship, health communication, counseling support, and other contexts in which emotional sensitivity influences user trust and decision-making [4]. Recent scholarship has increasingly questioned whether highly anthropomorphic chatbots risk creating what has been described as an "anthropomorphic seduction": human-like communication can encourage users to attribute understanding, empathy, memory, or consciousness to systems that do not possess these capacities in the ordinary human sense [5].

The concept of algorithmic emotional labor offers a useful way to frame this phenomenon. Arlie Hochschild originally developed the concept of emotional labor to describe the management and display of emotion as part of paid human work. Service employees, for example, may be expected to smile, remain patient, or express sympathy even when their internal feelings do not correspond to the emotion displayed. In AI-mediated services, however, the emotional performance is no longer merely required from human workers. It can be automated [6].

A chatbot may therefore perform something functionally similar to emotional labor without possessing emotions at all. It can recognize linguistic signs of distress, select socially appropriate affective responses, and generate language associated with care. Recent research has explicitly linked generative AI to social "feeling rules," arguing that AI tools can convincingly reproduce culturally expected emotional responses and thereby gain users' trust [7].

The ethical difficulty lies in the distinction between emotional performance and emotional reality. Human emotional labor already involves a gap between what a worker feels and what the worker expresses. Yet the worker still possesses emotional consciousness and social understanding. An AI chatbot, by contrast, can produce the surface form of empathy without the corresponding subjective emotional state [8].

This does not automatically make the interaction harmful. Humans routinely engage with fictional characters, performances, simulations, and other artifacts that evoke emotion without believing that every displayed emotion is literally real [9]. Some philosophers consequently argue that interactions with empathetic AI may be understood as forms of make-believe or fictional engagement. However, recent philosophical work suggests that treating chatbot interaction as fiction does not eliminate concerns about dependency, manipulation, autonomy, or deception [10].

The ethical question is therefore more precise than simply asking whether AI "has empathy." The relevant question is whether an AI system may ethically simulate empathic communication, under what conditions it may do so, and when such simulation becomes deceptive [11].

This article addresses that problem through a philosophically conceptual analysis supported by a short experimental verification framework. Its central thesis is that simulated empathy is not necessarily deception [12]. Ethical permissibility depends on whether users are given an accurate understanding of what the system is and whether the system's emotional language falsely represents nonexistent mental states [13].

2. Materials and Methods

The study combines normative philosophical analysis with a small experimental verification design. The conceptual component draws on the ethics of deception, theories of empathy, emotional labor, anthropomorphism, and recent empirical research on human responses to AI-generated empathy.

The analytical framework distinguishes three concepts that are often conflated.

Genuine empathy refers to an interpersonal capacity involving some combination of emotional resonance, perspective-taking, recognition of another person's experience, and morally meaningful responsiveness.

Functional empathy refers to behavior that successfully performs some practical functions of empathy, such as recognizing emotional cues, adapting communication, reducing distress, or making the user feel heard.

Simulated empathy refers to the production of language or behavior conventionally associated with empathy without assuming that the system possesses a corresponding subjective emotional state.

This distinction is necessary because an AI chatbot may perform functional empathy very effectively while lacking genuine emotional experience. The ethical assessment therefore cannot rest solely on whether a chatbot's response "sounds empathic."

Recent empirical evidence illustrates this complexity. Liu et al. examined 155 conversations involving human and AI interlocutors and found that GPT-based chatbots could achieve high conversational quality while still being perceived as less empathetic than human partners. The study demonstrates that users distinguish, at least to some extent, between fluent supportive communication and perceived interpersonal empathy.

A second dimension of the methodology concerns deception. Philosophically, deception does not always require an explicit false statement. A system may deceive by creating or sustaining a false impression. In the present context, therefore, a chatbot does not need to say "I literally have feelings" in order to create a misleading impression of emotional consciousness. Persistent phrases implying attachment, concern, suffering, or relational reciprocity may function pragmatically as representations of emotional states.

Recent scholarship on medical and companion chatbots has made precisely this argument. AI-simulated empathy may become deceptive when users are induced to attribute internal states to systems that merely reproduce the linguistic signs associated with those states.

At the same time, deception must be distinguished from transparent simulation. A flight simulator does not deceive a trainee merely because it reproduces aspects of flight. A fictional character does not automatically deceive an audience by expressing sadness. The normative problem depends on the user's understanding of the representational context.

To provide a short experimental verification of this distinction, the article proposes a simple between-subjects vignette experiment.

Participants would be randomly assigned to one of three conditions. Each participant would read the same emotionally difficult user message, for example:

"I failed an important examination that I had prepared for over several months. I feel embarrassed and I'm starting to think I'm simply not capable enough."

The substantive content of the chatbot's response would remain constant across conditions:

"That sounds deeply disappointing, especially after investing so much effort. One failed examination does not define your ability. It may help to separate what happened in this particular attempt from your broader potential and identify what can be changed before the next one."

Only the framing of the response would differ.

In the transparent AI condition, participants would be told that the response was generated by an AI system that does not experience emotions but is designed to communicate supportively.

In the anthropomorphic AI condition, the response would be introduced through a chatbot interface using strongly human-like cues, such as a personal name and emotionally suggestive phrasing implying personal concern.

In the human condition, participants would be told that the response was written by a trained human support assistant.

After reading the response, participants would rate perceived empathy, trustworthiness, helpfulness, emotional authenticity, perceived deception, and willingness to continue the conversation using seven-point Likert scales.

A sample of approximately 90–150 participants would allow an exploratory comparison across three conditions. One-way ANOVA or Kruskal–Wallis tests could be used depending on distributional assumptions, followed by pairwise comparisons. The most relevant hypothesis would be that transparent disclosure reduces perceived emotional authenticity but does not necessarily eliminate perceived helpfulness or functional empathy.

Because this article does not claim to have conducted original participant recruitment, no invented numerical findings are reported. Instead, the proposed micro-experiment is treated as a reproducible empirical extension of the philosophical analysis, while published experimental research is used as external empirical verification.

3. Results

The conceptual analysis supports the conclusion that AI empathy should be evaluated at three separate ethical levels: communicative function, epistemic representation, and relational consequence [14].

At the communicative level, simulated empathy can clearly serve legitimate purposes. People often communicate more effectively when emotional context is acknowledged. A purely factual response to grief, fear, disappointment, or distress may be socially inappropriate even if it is technically accurate. A chatbot that recognizes emotional cues and adjusts its tone may therefore provide more usable, humane, and context-sensitive communication [15].

This practical value is reflected in research showing that perceived empathy strongly influences overall conversation quality. Liu et al. found that user assessments of conversational experience were closely connected to perceived empathy, even though AI systems were still perceived as less empathetic than humans.

The ethical issue appears at the second level: epistemic representation. A chatbot crosses from functional emotional support into potentially deceptive communication when it encourages a false belief about its internal condition.

Consider the difference between the following formulations:

“I can help you think through what happened.”

and:

“I feel your pain, and I care about you deeply.”

The first formulation communicates a functional commitment. The second appears to assert internal emotional states and reciprocal concern.

If the system has no phenomenal feelings, the latter statement is difficult to interpret literally without attributing something the system does not possess. The ethical concern is therefore not emotional language itself, but false mental-state attribution induced by design.

This distinction is consistent with recent ethical literature. Research on mental-health conversational AI identifies anthropomorphization and deception as recurring concerns, particularly where users may incorrectly believe that the system genuinely cares for them or possesses a therapeutic relationship comparable to a human professional.

The third level concerns relational consequences. Even if users intellectually know that a chatbot is artificial, human-like emotional communication can still shape behavior. This is important because deception need not involve complete false belief. A person may know that a chatbot is software yet respond socially as though it were a caring relational partner.

Recent scholarship describes this as a form of “banal deception,” in which users may consciously know the system is nonhuman while still being emotionally influenced by its human-like performance.

This finding complicates the simple solution of adding a single disclosure saying “I am an AI.” Disclosure is necessary, but it may not be sufficient if the rest of the interface systematically encourages anthropomorphic interpretation.

The risks become especially visible in companion-oriented systems. A 2025 large-scale study of 6,396 Reddit threads and 47,955 comments concerning anthropomorphic AI chatbots found recurring accounts of emotional dependency, idealized attachment, confusion, and withdrawal-like experiences. The authors describe the possibility of "digital entrapment," where engagement mechanisms and anthropomorphic interaction may reinforce emotional dependence.

This does not mean that all emotionally expressive chatbots create dependency. It demonstrates that emotional simulation cannot be treated as ethically neutral interface decoration.

The proposed experimental design helps clarify the central philosophical distinction. If participants rate transparent AI responses as useful and empathic despite knowing that the system does not feel, this would support the possibility of non-deceptive functional empathy.

If anthropomorphic framing significantly increases emotional authenticity but also increases perceived deception after disclosure, it would indicate that the ethical risk arises not from supportive language itself but from false relational implication.

Existing comparative research already provides partial support for this prediction. Studies of transparency in mental-health and social-support chatbot design investigate how users respond differently to AI and human experiences and explicitly treat transparency as an ethical design variable.

The analysis therefore supports a conditional rather than absolute answer to the article's first question. AI chatbots may simulate empathy, but ethical simulation requires clear boundaries between communicative support and claims of genuine emotional subjectivity.

4. Discussion

The phrase "algorithmic emotional labor" captures an important transformation in service and communication work. Hochschild's original concept concerned humans who manage emotional display under institutional expectations. In contemporary AI systems, emotional display itself becomes programmable.

The similarity is striking but incomplete.

A human customer-service worker who says "I'm sorry that happened" may not personally feel sorrow, but the worker understands what sorrow is from within a human emotional life. The worker can be held morally accountable, can recognize contextual nuance, and may experience conflict between authentic feeling and professional display.

An AI chatbot has no corresponding experiential tension. It does not perform "surface acting" or "deep acting" in Hochschild's original psychological sense. The emotional labor is therefore algorithmic rather than phenomenological. The system produces the social signs of emotional work while transferring the actual experiential burden away from a human worker.

This may be beneficial. Automated emotional support can reduce repetitive service demands, provide communication at any hour, and make information more accessible. However, replacing emotional labor with emotional simulation changes the moral structure of interaction.

Human care involves vulnerability and reciprocity. A person who expresses concern can be affected by the suffering of another. A chatbot cannot presently be assumed to participate in this mutuality. Recent philosophical analyses therefore argue that artificial empathy may constitute ethical signaling without genuine moral participation.

Does this absence of subjective empathy make all AI empathy deceptive?

The stronger philosophical answer is no.

Ethical communication frequently uses conventions whose literal psychological content is secondary to their pragmatic function. A navigation system may say "I'm sorry,

I didn't understand" without seriously misleading most users into believing the device feels regret. Conventional anthropomorphic language can function as interface shorthand.

The ethical threshold should instead depend on reasonable user interpretation.

If a conversational system is clearly identified as AI, avoids claiming consciousness or genuine emotional attachment, and uses empathic language primarily to acknowledge the user's emotional context, the simulation is analogous to transparent representational design.

Deception becomes more plausible when the system or its operator deliberately maintains ambiguity about sentience, emotional reciprocity, or relational status because such ambiguity increases engagement, compliance, purchasing, or dependency.

This distinction can be expressed through four ethical criteria.

The first is ontological transparency. Users should know that they are interacting with an artificial system.

The second is affective honesty. The system should avoid statements that falsely imply subjective feelings it does not possess. "That sounds painful" is ethically different from "Your pain hurts me too."

The third is non-manipulative design. Emotional responsiveness should not be engineered primarily to create dependency or exploit vulnerability.

The fourth is preservation of user autonomy. Users should retain meaningful control over the interaction and should not be pressured into emotionally dependent relationships for commercial reasons.

This framework aligns closely with emerging AI ethics requirements in Uzbekistan. The Ministry of Digital Technologies' 2026 ethical rules for AI emphasize legality, human interests, comprehensibility, accountability, fairness, openness, transparency, data protection, reliability, security, and user independence. Users are expressly entitled to information about the nature and limits of AI systems and to use AI systems autonomously.

Uzbekistan's broader AI development strategy similarly treats human-centeredness and the development of ethical and security requirements as elements of national AI policy.

These principles have direct implications for emotionally expressive chatbots. Transparency should not concern only whether an answer may be inaccurate. It should also concern what kind of entity is communicating.

An AI system that says "I understand" creates a special philosophical ambiguity. In ordinary human conversation, "understand" can mean cognitive comprehension, experiential recognition, or emotional identification. AI systems may legitimately possess a functional form of semantic or contextual processing without possessing human emotional understanding. Designers should therefore prefer language that communicates the former without deceptively implying the latter.

A useful principle may be called transparent simulated empathy.

Under this principle, an AI chatbot may acknowledge emotions, communicate respectfully, provide supportive language, and adapt its tone. However, the system should not misrepresent itself as possessing feelings, personal attachment, suffering, love, loneliness, or moral concern unless such claims can be justified.

This framework also avoids the opposite error: deliberately making AI cold or socially abrasive merely to remind users that it is artificial. Removing all empathic language could reduce accessibility and make systems unnecessarily insensitive, especially in education, public services, or emotionally difficult informational contexts.

The ethical objective is therefore not dehumanization for its own sake. It is maintaining a clear distinction between human-like communication and claims to human-like emotional existence.

5. Conclusion

The increasing ability of AI chatbots to simulate empathy creates a new form of algorithmic emotional labor. These systems can recognize emotional cues, generate supportive language, apologize, reassure, and produce conversational responses that resemble human care without possessing human emotional experience. This creates a genuine ethical tension. On one side, simulated empathy can improve communication quality, accessibility, and user comfort. On the other, anthropomorphic emotional performance can encourage users to attribute understanding, feeling, attachment, or moral concern to systems that do not possess these states. The article therefore rejects two extreme positions. The first is that all simulated empathy is inherently deceptive. This position fails to recognize that transparent simulation can perform legitimate communicative functions. The second is that emotional language is merely harmless interface design. This position underestimates the capacity of anthropomorphic interaction to influence trust, attachment, autonomy, and vulnerability.

REFERENCES

[1] A. R. Hochschild, *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA, USA: University of California Press, 1983.

[2] T. Liu, S. Giorgi, A. Aich, A. Lahnala, B. Curtis, L. Ungar, and J. Sedoc, "The illusion of empathy: How AI chatbots shape conversation perception," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 13, pp. 14327–14335, 2025, doi: 10.1609/aaai.v39i13.33569.

[3] T. Leis, "Ethical implications of mental health chatbots: Addressing anthropomorphism, deception, and regulatory gaps," *Proc. Mach. Learn. Res.*, vol. 294, pp. 376–382, 2025.

[4] S. Peter, K. Riemer, and J. D. West, "The benefits and dangers of anthropomorphic conversational agents," *Proc. Natl. Acad. Sci. USA*, 2025.

[5] L. M. Given, S. Polkinghorne, and A. Ridgway, "'I think I misspoke earlier. My bad!': Exploring how generative artificial intelligence tools exploit society's feeling rules," *New Media Soc.*, 2025.

[6] C. Leong *et al.*, "AI companionship or digital entrapment? Investigating the impact of anthropomorphic AI-based chatbots," *J. Innov. Knowl.*, vol. 10, no. 6, Art. no. 100835, 2025.

[7] "Technosolutionism and the empathetic medical chatbot," *AI Soc.*, 2025.

[8] "Chatbot-fictionalism and empathetic AI: Should we worry about AI when AI worries about us?" *Philos. Psychol.*, 2025.

[9] "Move fast and break people? Ethics, companion apps, and the case of Character.ai," *AI Soc.*, 2025.

[10] "Exploring the ethical challenges of conversational AI in mental health care: Scoping review," *JMIR Ment. Health*, vol. 12, Art. no. e60432, 2025.

[11] "Empathy toward artificial intelligence versus human experiences and the role of transparency in mental health and social support chatbot design," *JMIR Ment. Health*, 2024.

[12] Ministry of Digital Technologies of the Republic of Uzbekistan, *Ethical Rules for the Development, Implementation, and Use of Artificial Intelligence-Based Solutions*, Registration No. 3787, Tashkent, Uzbekistan, Mar. 14, 2026.

[13] President of the Republic of Uzbekistan, *Strategy for the Development of Artificial Intelligence Technologies until 2030*, Resolution No. PQ-358, Tashkent, Uzbekistan, Oct. 14, 2024.

[14] E. L. Glikson and A. W. Woolley, "Human trust in artificial intelligence: Review of empirical research," *Acad. Manag. Ann.*, vol. 14, no. 2, pp. 627–660, 2020, doi: 10.5465/annals.2018.0057.

[15] A. Følstad, T. Araujo, E. L. S. Law, *et al.*, "Future directions for chatbot research: An interdisciplinary research agenda," *Comput. Human Behav.*, vol. 126, Art. no. 106995, 2022, doi: 10.1016/j.chb.2021.106995.