

Letter to the Editor

Empathy Cuts Both Ways in Clinical AI

Vera Sorin¹, MD; Eyal Klang², MD

¹Mayo Clinic, Rochester, MN, United States

²Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:

Eyal Klang, MD
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Boston, MA 02215
United States
Phone: 1 617-754-2519
Email: eyalklang@gmail.com

Related Articles:

Comment in: <https://humanfactors.jmir.org/2026/1/e110739>

Comment on: <https://humanfactors.jmir.org/2026/1/e89005>

JMIR Hum Factors 2026;13:e108778; doi: [10.2196/108778](https://doi.org/10.2196/108778)

Keywords: artificial intelligence; AI; clinical decision support systems; large language models; empathy; explainable AI; trust; human factors; reliance; human-AI interaction

Zhang et al [1] found that participants rated an emotion-adaptive clinical decision support system as clearer, easier to use, more reliable, and more trustworthy than a basic system. Both systems used the same core prediction model and explainability outputs. The empathic system added emotion sensing and a dialogue model that changed how the explanation was delivered.

A clinician who is stressed or confused may understand a difficult explanation better when the system changes how it speaks. But empathy cuts both ways. It can make correct advice easier to accept. It may also make wrong advice harder to reject.

The study measured users' ratings of trust, reliability, and clarity. It did not test whether the empathic system helped them make more accurate decisions or spot errors. The prediction models were tested separately, but that does not show whether the extra trust was deserved. A system can feel more reliable without being more accurate. A systematic review of large language model empathy found that 11 of 12 studies assessed empathy subjectively [2]. That gap matters because the goal is not the highest possible trust. Trust should match what the system can actually do. Users should accept correct AI advice and reject wrong advice [3].

The authors already identify one part of the problem. Participants reported that emotion adaptation sometimes changed the focus or specificity of the answer. Some replies became overly supportive and did not address the clinical

question directly. The authors therefore call for controls that protect clinical precision and decision clarity [1]. But even a direct and precise answer can be wrong. When that answer is also empathic, it may be harder to reject.

This is not only a theoretical concern. In a clinical study, inaccurate advice reduced physicians' diagnostic accuracy [4]. Another study of health chatbots found that empathic responses increased perceived warmth, trust, and intentions to use the system [5]. These findings do not prove that empathy increases acceptance of wrong clinical advice. They show why the question needs to be tested. One objection is that empathy may raise trust because it improves understanding, not because it persuades. That may be true. The test should separate these effects. Better understanding should help users accept correct advice and reject wrong advice. Persuasion alone may increase acceptance of both.

Future studies should compare correct and wrong AI advice in empathic and neutral language. The clinical content should stay the same. Users could make a judgment before seeing the advice and again afterward. The main outcomes should be final decision accuracy, acceptance of correct advice, rejection of wrong advice, and a direct test of what users understood. Trust ratings can remain, but they should not stand alone.

Emotion-adaptive clinical AI should not make every answer more trusted. It should help trust follow correctness.

Acknowledgments

ChatGPT (GPT-5.6 Pro; OpenAI), accessed through the ChatGPT web interface, was used to improve grammar, spelling, and readability. The authors take full responsibility for the final content.

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

VS and EK conceived the argument, drafted and revised the manuscript, and approved the final version.

Conflicts of Interest

None declared.

References

1. Zhang T, Bae SW, Chung T, Dey AK. Emotion-adaptive large language model-driven clinical decision support: user evaluation of the empathic clinical decision support system framework for trust and explainability. *JMIR Hum Factors*. May 22, 2026;13:e89005. [doi: [10.2196/89005](https://doi.org/10.2196/89005)] [Medline: [42172616](https://pubmed.ncbi.nlm.nih.gov/42172616/)]
2. Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review. *J Med Internet Res*. Dec 11, 2024;26:e52597. [doi: [10.2196/52597](https://doi.org/10.2196/52597)] [Medline: [39661968](https://pubmed.ncbi.nlm.nih.gov/39661968/)]
3. Schemmer M, Kuehl N, Benz C, Bartos A, Satzger G. Appropriate reliance on AI advice: conceptualization and the effect of explanations. In: *IUI '23: Proceedings of the 28th International Conference on Intelligent User Interfaces*. Association for Computing Machinery; 2023:410-422. [doi: [10.1145/3581641.3584066](https://doi.org/10.1145/3581641.3584066)]
4. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit Med*. Feb 19, 2021;4(1):31. [doi: [10.1038/s41746-021-00385-9](https://doi.org/10.1038/s41746-021-00385-9)] [Medline: [33608629](https://pubmed.ncbi.nlm.nih.gov/33608629/)]
5. Seitz L. Artificial empathy in healthcare chatbots: does it feel authentic? *Comput Hum Behav Artif Hum*. Jan 2024;2(1):100067. [doi: [10.1016/j.chbah.2024.100067](https://doi.org/10.1016/j.chbah.2024.100067)]

Edited by Andrea Schaffeler; This is a non-peer-reviewed article; submitted 04.Aug.2026; accepted 17.Sep.2026; published 02.Oct.2026

Please cite as:

Sorin V, Klang E

Empathy Cuts Both Ways in Clinical AI

JMIR Hum Factors 2026;13:e108778

URL: <https://humanfactors.jmir.org/2026/1/e108778>

doi: [10.2196/108778](https://doi.org/10.2196/108778)

© Vera Sorin, Eyal Klang. Originally published in *JMIR Human Factors* (<https://humanfactors.jmir.org>), 02.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Human Factors*, is properly cited. The complete bibliographic information, a link to the original publication on <https://humanfactors.jmir.org>, as well as this copyright and license information must be included.