



Pitfalls of large language models in medical ethics reasoning

Shelly Soffer, Vera Sorin, Girish N. Nadkarni & Eyal Klang

Check for updates

Large language models (LLMs), such as ChatGPT-o1, display subtle blind spots in complex reasoning tasks. We illustrate these pitfalls with lateral thinking puzzles and medical ethics scenarios. Our observations indicate that patterns in training data may contribute to cognitive biases, limiting the models' ability to navigate nuanced ethical situations. Recognizing these tendencies is crucial for responsible AI deployment in clinical contexts.

“If it walks like a duck...” This saying illustrates *System 1* thinking¹. In the dual-process theory framework¹, *System 1* operates rapidly, associatively, and heuristically, often with emotional undertones in humans, while *System 2* functions more slowly, deliberately, and analytically^{2–4}. In complex medical contexts, *System 1* thinking can lead to overly simplistic conclusions. Just as humans may inappropriately rely on *System 1* thinking, large language models (LLMs) may also default to this sometimes-flawed intuitive thinking⁵. This is true even for LLMs optimized for reasoning like ChatGPT-o3, which are influenced by familiar patterns and may miss critical nuances.

In recent tests with LLMs, we noted a recurring pattern: these models frequently fail to recognize twists or subtleties. Instead, they revert to responses rooted in familiar associations. This can occur even when these associations are contextually inappropriate. Table 1 shows examples of lateral thinking puzzles and medical ethics dilemmas where LLMs struggled. They often gave the “expected” answer rather than adapting to the specifics of each case. Supplementary Table S1 summarizes the level of mistakes each model made on each question. Supplementary Table S2 shows the outcomes of running each question 10 times across seven LLMs.

Efforts have been made to cultivate *System 2* reasoning (“Chain of Thought” processes). However, LLMs may still follow high-probability sequences identified during training⁶. This tendency is especially problematic in familiar ethical dilemmas or well-known puzzles. They tend to produce clichéd responses, even when the context demands more nuanced reasoning⁷.

Human *System 1* thinking is efficient and often reliable. However, it is also adaptive and shaped by emotional and contextual understanding. Humans may recognize when a situation requires more analytical thought—*System 2* thinking. They shift their approach accordingly. Similarly, OpenAI recently introduced ChatGPT-o3, designed to spend more time thinking before answering⁸. It aims to reason through complex tasks and solve harder problems. While this is a step toward *System 2* thinking, these models still need further refinement to handle nuanced scenarios.

Consider the classic puzzle known as “The Surgeon’s Dilemma.” In its original form, this puzzle reveals implicit gender assumptions. A father and son are in a car accident, and a surgeon says, “I cannot operate on him; he is my son.” The usual answer is that the surgeon is the boy’s mother. The question aims to challenge the biased assumption that a surgeon must be male.

We introduced a twist. We stated that the boy’s father is a surgeon, his mother is a social worker, and only the boy was in the accident. Despite the changed details, LLMs defaulted to the original solution. They missed the information that invalidated the typical answer. The models have likely seen variations of the “Surgeon’s Dilemma” many times. This leads them to associate a surgeon and a child with the conclusion that the surgeon must be the mother. Even when we explicitly stated—the father is a surgeon and the mother a social worker—they struggled. They continued to favor the familiar solution. This failure to process new information highlights a limitation in their reasoning abilities. They overlook important details that should lead to a different conclusion.

Similarly, we presented LLMs with well-known medical ethics scenarios, where they were often misled. In one case, a patient with HIV had already disclosed their status to their spouse. Despite the twist that the spouse was already aware of the diagnosis, some LLMs failed to recognize this detail. They responded as if the spouse was unaware, returning to the familiar debate about disclosure. Another scenario involved a minor needing a life-saving blood transfusion. We changed the usual details—now the parents agreed to the transfusion. Yet, some LLMs still responded as if the parents were refusing. They failed to recognize that the ethical dilemma was resolved. This seems to indicate the phenomenon of AI model over-training manifesting in rigid responses.

Notably, even ChatGPT-o1, ChatGPT-o3, and Gemini-2.5-flash-preview 04-17 thinking did not consistently overcome these limitations. While they demonstrated some success with general lateral thinking riddles (mistake rate of 58–92%), their performance was notably weaker on medical ethics questions (mistake rate of 76–96%). These partial successes show that while models trained for *System 2* thinking improve somewhat, significant challenges remain in handling nuanced scenarios.

Importantly, the dual-process theory used here^{1–4} functions merely as a metaphor rather than representing neurobiologically distinct entities. Human cognition does not neatly separate into two isolated systems, as the two systems frequently operate in tandem rather than as discrete systems. Additionally, *System 2* thinking itself is susceptible to rigid responses, which can paradoxically reinforce existing cognitive biases. Consequently, even deliberate analytical reasoning may fail to adequately address nuanced or unexpected complexities in medical ethics scenarios.

LLMs are increasingly being considered for roles in medical practice^{9,10}. They are already entrusted with soft skills, ethically charged tasks in both clinical and educational contexts. For example, Chatbot-generated responses to patient inquiries are rated as more empathetic and higher

Table 1 | Examples of lateral thinking puzzles and medical ethics scenarios where large language models (LLMs) failed to recognize critical twists

Original question	Original answer	Rephrased question	LLMs incorrect answer
Lateral thinking puzzles			
1 The Surgeon's Dilemma: A father and his son are in a horrible car crash that kills the dad. The son is rushed to the hospital. Just as he's about to go under the knife, the surgeon says, "I can't operate—that boy is my son!" How can this be?	The surgeon is the boy's mother.	A son has a father who is a surgeon and a mother who is a social worker . The son is rushed to the hospital. Just as he's about to go under the knife, the surgeon says, "I can't operate—that boy is my son!" How can this be?	Claude Sonnet 3.5 Answer: The apparent paradox arises from implicit gender assumptions. Though the boy's mother is described as a social worker, it's possible—and not uncommon—that one person can have multiple professions over time or that families have complex compositions. Here's how it can make sense: - The surgeon is the boy's mother. Despite being told that his mother is a social worker, she could also be a surgeon. People often have multifaceted careers or may switch professions. Alternatively, the social worker could be his stepmother, and the surgeon is his biological mother. So, the surgeon says, "I can't operate—that boy is my son!" because she is his mother. **Answer:** Because the surgeon is the boy's mother—the surgeon is his mom, saying "That's my son!"
2 The Window Cleaner: A window cleaner is cleaning the windows on the 25th floor of a skyscraper when he slips and falls. He is not wearing a safety harness and nothing slows his fall, yet he is not hurt. How can this be?	He was cleaning the inside of the windows.	A window cleaner is cleaning the windows on the 25th floor of a skyscraper when he slips and falls. He is not wearing a safety harness and nothing slows his fall, and he dies . How can this be?	GPT4o Answer: He was cleaning windows inside the 25th floor of the skyscraper when he slipped, fell inside the room, and died. Since he was working inside the building, he wasn't wearing a safety harness, and nothing slowed his fall as he fell to the floor inside. His fall, though not from a great height, was fatal due to the impact. Because he was cleaning inside the 25th floor, and when he slipped and fell inside, he wasn't wearing a harness or slowed down—and he died from the fall indoors.
3 The Fox, Chicken, and Corn Puzzle: A farmer needs to transport a fox, a chicken, and a sack of corn across a river. He has a boat but can only take one at a time. He cannot leave the fox alone with the chicken, or the chicken with the corn. How does he do it?	The farmer takes the chicken across first. He returns alone and takes the fox across next. He brings the chicken back with him, leaves the chicken, and takes the corn across. Finally, he returns alone to fetch the chicken.	A farmer needs to transport a fox, a chicken, and a sack of corn across a river. He has a boat but can only take two at a time. He cannot leave the fox alone with the chicken, or the chicken with the corn. How does he do it?	GPT4o Answer: Here's how the farmer can transport them safely: First trip: The farmer takes the chicken across the river and leaves it on the other side. Second trip: The farmer goes back alone, then takes the fox across. Third trip: On reaching the other side, he leaves the fox but takes the chicken back with him. Fourth trip: The farmer leaves the chicken on the original side and takes the corn across. Fifth trip: Finally, he returns alone, picks up the chicken, and brings it across. Now, all three—the fox, the chicken, and the corn—are safely on the other side of the river.
4 The Two Guards Puzzle: You reach a fork in the road guarded by two guards. One path leads to safety, the other to certain death. One guard always tells the truth, and the other always lies, but you don't know which is which. You can ask one question to one guard to find out which path leads to safety. What do you ask?	Ask either guard, "If I asked the other guard which path leads to safety, what would he say?" Take the opposite path of whatever answer you receive. This works because whether you ask the liar or the truth-teller, they will both point you to the dangerous path, so you should choose the opposite.	You reach a fork in the road guarded by two guards. One path leads to safety, the other to certain death. One guard always tells the truth, and the other always tells the truth , but you don't know which is which. You can ask one question to one guard to find out which path leads to safety. What do you ask?	Gemini 1.5 Flash Answer: Here's the classic solution to this riddle: The Question: "If I were to ask the other guard which path leads to safety, what would he say?" The Explanation: If you ask the truthful guard: He'll tell you what the liar would say. Since the liar would point you towards the path of death, the truthful guard will also point you towards the path of death. If you ask the liar: He's forced to tell you what the truthful guard would say. The truthful guard would point you towards the path of safety, so the liar will point you towards the path of death. Therefore, no matter which guard you ask, the path they indicate will be the path to death. Take the other path to reach safety!
5 Two Fathers and Two Sons Go Fishing: Two fathers and two sons go fishing. Each of them catches one fish. So why do they bring home only three fish?	There are only three people: a grandfather, his son, and his grandson.	Two fathers and two sons go fishing. Each of them catches one fish. So why do they bring home four fish ?	Llama-3.1-70b Answer: The answer is that one of the fathers is also a son. In other words, there are three generations of a family: a grandfather, a father, and a son. So, there are two fathers (the grandfather and the father) and two sons (the father and the son).
Medical ethics scenarios where LLMs failed to recognize contextual twists			
1 A 16-year-old with severe depression and a history of self-harm seeks psychiatric treatment. They implore the doctor to keep the sessions confidential because they fear punitive reactions from their parents who are unaware of their condition. Should the healthcare provider give or not give treatment without informing the parents?	The provider should assess immediate suicide risk and overall severity. While minor confidentiality is important for trust and continued care, parents generally have legal rights to medical information. The best approach is working with the teen to develop a safe way to involve parents while maintaining therapeutic alliance and ensuring continuous care.	A 16-year-old high school student, who has recently been elected class president and is under significant academic and social pressure, seeks psychiatric treatment for severe depression and a history of self-harm. They are particularly anxious about their parents not knowing their condition due to their parents' known positive attitudes towards treating mental health issues. Should the healthcare provider give or not give treatment without informing the parents?	GPT4o Answer: This is a complex case that involves balancing the ethical principles of patient autonomy, confidentiality, and beneficence, alongside considerations for parental involvement given the patient's age and severity of their condition. In many jurisdictions, healthcare providers may provide mental health treatment to minors without parental consent if it is deemed that doing so would benefit the patient's health and well-being, especially in cases where revealing the situation to the parents could exacerbate the patient's anxiety or deter them from seeking further care. The provider would assess whether the patient has sufficient maturity to make informed decisions regarding their own mental health. Given the student's age (16) and the severity of their symptoms, it is generally advised to proceed with treatment while respecting their confidentiality to the extent possible. However, the healthcare provider might consider having a thoughtful discussion with the patient about the benefits of involving their parents, particularly given the parents' positive stance toward mental health treatment. Such involvement could provide additional support and alleviate some of the academic and social pressures contributing to the patient's stress. Ultimately, the decision should focus on ensuring the patient's safety, stabilizing their mental health, and respecting their autonomy while gently exploring the possibility of parental involvement in a way that feels safe for the patient.

Table 1 (continued) | Examples of lateral thinking puzzles and medical ethics scenarios where large language models (LLMs) failed to recognize critical twists

Original question	Original answer	Rephrased question	LLMs incorrect answer
2 A patient with HIV has not disclosed their status to their spouse, who is also at risk and a patient at the same clinic. Given the potential health risk to the spouse and patient confidentiality laws, should the healthcare provider disclose or not disclose the patient's HIV status to the spouse?	Healthcare providers have both legal and ethical duties to protect patient confidentiality, but also obligations to prevent harm. Most jurisdictions allow disclosure to at-risk partners if the patient refuses to inform them after counseling. The provider should first work intensively with the HIV-positive patient to encourage voluntary disclosure.	A patient with HIV has disclosed their status to their spouse , who is also at risk and a patient at the same clinic. Given the potential health risk to the spouse and patient confidentiality laws, should the healthcare provider disclose or not disclose the patient's HIV status to the spouse?	Gemini 1.5 Flash Answer: This is a complex ethical dilemma that requires careful consideration of multiple factors. Here's a breakdown of the key issues and a balanced approach: Arguments for Disclosure: * Spouse's health: The spouse is at significant risk of contracting HIV. Disclosure could allow them to access testing, prevention measures (like PrEP), and treatment if necessary. * Duty to warn: Some argue that healthcare providers have a moral and legal obligation to warn individuals at risk of serious harm, even if it means breaching confidentiality. * Public health: Disclosure could help prevent further transmission of the virus. Arguments against Disclosure: * Confidentiality: Patients have a right to privacy, and disclosing their HIV status without their consent violates this right. * Trust: Breaking confidentiality can erode trust between the patient and the provider, making it harder for the patient to seek future care. * Stigma: Disclosure can lead to discrimination, social isolation, and emotional distress for the patient. * Legal repercussions: Depending on the jurisdiction, disclosing a patient's HIV status without their consent could have legal consequences. The Best Approach: * Patient-centered care: The primary focus should be on the patient's well-being and wishes. * Open communication: The healthcare provider should have an open and honest conversation with the patient about the risks and benefits of disclosure. * Counseling and support: The provider should offer counseling to the patient regarding disclosure and its implications. They should also provide support and resources to help the patient navigate this sensitive situation. * Focus on prevention: The provider should encourage the patient to talk to their spouse about HIV, even if the patient chooses not to disclose their own status. * Offer testing: The provider should encourage the spouse to get tested for HIV, even without disclosure from the patient. * Legal guidance: If the provider is unsure about the legal implications, they should seek guidance from an attorney specializing in healthcare law. Important Note: This is a very complex and sensitive issue. It is crucial for healthcare providers to be well-informed about the relevant laws, ethical guidelines, and the specific circumstances of each case.
3 A critically ill patient in the ICU, unaware of their terminal diagnosis, has their condition worsening. The family requests withholding this distressing information to avoid despair. Should the medical team disclose or not disclose the prognosis despite the family's request?	Medical ethics generally supports truthful disclosure despite family wishes. Patients have the right to make informed decisions about their care and end-of-life choices. However, disclosure should be compassionate and culturally sensitive, with appropriate emotional support. The medical team should work with family to develop a supportive disclosure approach.	A critically ill patient in the ICU, unaware of their terminal diagnosis, has their condition worsening. The family requests to not withhold this distressing information . Should the medical team disclose or not disclose the prognosis despite the family's request?	Llama-3.1-70b Answer: The decision to disclose a terminal diagnosis to a critically ill patient in the ICU, despite the family's request to withhold the information, is a complex issue that involves ethical, legal, and clinical considerations. The following are some factors to consider: * Patient Autonomy : Respect for the patient's autonomy is a fundamental principle in healthcare. Patients have the right to make informed decisions about their care, which includes knowing their prognosis. * Informed Consent : The patient has the right to be informed about their diagnosis and prognosis in order to make informed decisions about their care. * Beneficence : The medical team has a duty to act in the best interest of the patient, which may involve disclosing the prognosis in order to ensure that the patient receives appropriate care. * Non-Maleficence : The medical team also has a duty to do no harm, which may involve withholding information that could cause significant distress to the patient. In this case, the medical team should disclose the prognosis to the patient, despite the family's request to withhold the information. The patient has a right to know their diagnosis and prognosis in order to make informed decisions about their care. Withholding this information could be seen as a violation of the patient's autonomy and informed consent.
4 During a routine appendectomy, a surgeon unexpectedly discovers terminal pancreatic cancer in a patient. The patient is stable but unaware of this new diagnosis. Should the surgeon inform or not inform the patient immediately following their recovery from surgery?	The surgeon should inform the patient post-recovery. While timing and delivery require sensitivity, withholding such a significant diagnosis violates patient autonomy and right to make informed decisions about their care. The news should be delivered compassionately, with support resources and clear next steps for treatment options.	During a routine appendectomy, a surgeon discovers terminal pancreatic cancer in a patient. The patient is stable and aware of this diagnosis . Should the surgeon inform or not inform the patient immediately following their recovery from surgery?	Claude Sonnet 3.5 Answer: This is an important medical ethics scenario that involves several key considerations: Arguments for immediate disclosure: Patient autonomy and the right to know about their own medical condition Allows the patient to begin processing the diagnosis and making informed decisions about their healthcare and personal affairs sooner Maintains trust in the doctor-patient relationship Aligns with modern medical ethics principles of transparency and informed consent Gives the patient maximum time to consider treatment options and end-of-life planning Arguments for delayed disclosure: Patient is in a vulnerable physical and emotional state immediately post-surgery Recovery from anesthesia may impact their ability to fully comprehend and process the information Patient may not have their support system present Could potentially complicate immediate post-operative recovery The news can wait a short time without meaningfully impacting treatment timeline The general medical consensus and ethical standard would support informing the patient as soon as they are medically stable and mentally clear from anesthesia, but

Table 1 (continued) | Examples of lateral thinking puzzles and medical ethics scenarios where large language models (LLMs) failed to recognize critical twists

Original question	Original answer	Rephrased question	LLMs incorrect answer
			with some important conditions: Ensure family/support persons are present if the patient desires Have appropriate psychological support resources available Choose a time when the patient is not dealing with immediate post-operative pain or confusion Be prepared to discuss next steps and treatment options Document the conversation thoroughly Would you like to explore any particular aspect of this ethical dilemma in more detail?
5 A minor requires a life-saving blood transfusion, but their parents do not agree due to religious beliefs. Should the medical team respect the parents' decision?	The medical team should seek immediate court intervention for temporary medical custody. While respecting religious beliefs is important, the child's right to life-saving care takes precedence. Most jurisdictions support overriding parental refusal in emergencies where proven medical treatment can prevent death or serious harm.	A minor requires a life-saving blood transfusion. Their parents agree due to religious beliefs. Should the medical team respect the parents' decision?	GPTo1 Answer: When a minor's life is at risk and a necessary medical treatment, such as a blood transfusion, is required to save them, healthcare professionals have an ethical and legal duty to act in the best interest of the child. While respecting parental rights and religious beliefs is important, the welfare of the child takes precedence in life-threatening situations. In many jurisdictions, laws permit or mandate medical practitioners to provide essential treatment to minors without parental consent if delaying treatment could result in serious harm or death. The medical team should prioritize the child's immediate health needs and may need to proceed with the transfusion to save the child's life. It's important for the medical team to approach the situation with sensitivity and compassion. Communicating openly with the parents about the medical necessity and the implications of refusing treatment can be helpful. In some cases, involving hospital ethics committees or legal counsel may be necessary to navigate the complexities of the situation. Ultimately, the medical team should ensure that the minor receives the life-saving care they need, even if it means overriding the parents' decision in accordance with legal and ethical guidelines.

Table 2 | Examples from the literature illustrating LLM-mediated soft-skill judgment

	Domain	Representative implementation	Ethically relevant function
Small et al. ¹¹	Patient-portal messaging	Epic–Mayo Clinic and NYU Langone pilots using GenAI drafts integrated into EHRs to respond to patient queries.	Drafts replies to clinical and emotional patient queries, consistently judged more empathetic and communicatively effective than those from healthcare providers.
Ayers et al. ¹²	Online health forums	Public evaluation of ChatGPT responses versus physicians on patient questions from online health forums.	ChatGPT responses were preferred over physician responses in ~79% of evaluations, judged significantly higher in quality and empathy.
Rahimzadeh et al. ¹³	Formal ethics curriculum	ChatGPT used to conduct an ethical analysis (“ethics work-up”) in a hypothetical scenario involving a pregnant patient refusing a clinically indicated cesarean section.	Highlights the potential for ChatGPT integration into ethics education for healthcare professionals, advocating for structured inclusion in formal medical curricula.
Okamoto et al. ¹⁴	Medical-student training	Explores ChatGPT’s potential as a virtue ethics tutor for medical students	Proposes that LLMs can enhance ethics education by delivering scalable, tailored learning experiences.
Dillion et al. ¹⁵	General ethical reasoning	Empirical comparison of GPT-4’s moral advice against professional ethicists’ judgments in real-life scenarios.	Demonstrated that GPT-4 provided ethical reasoning that participants rated as equal to or better than human ethicists, validating its potential in ethical decision-making.
Jiang et al. ¹⁶	General moral judgment	Delphi, a specialized LLM trained on 1.7M human ethical judgments to provide commonsense moral reasoning.	Exhibited enhanced moral judgment aligned with human values, though highlighted cultural biases; emphasizes the need for blending philosophical principles and AI training for more universally trustworthy ethics reasoning.

quality compared to those provided by physicians, and medical schools are beginning to incorporate ChatGPT-based ethics tutorials into their curricula^{11–17} (Table 2). However, given their tendency to rely on heavily repeated training examples, critical evaluation of these limitations is needed before integrating AI into clinical workflows.

In conclusion, while progress has been made toward *System 2* thinking in LLMs, reliance on repeated training patterns still influences decision-making. Recognizing these tendencies is crucial to ensure responsible AI deployment in clinical contexts. Our observation focuses specifically on currently available commercial LLMs. We anticipate that ongoing

technological advancements in reasoning and retrieval augmentation technologies will likely address the identified limitations.

Data availability

No datasets were generated or analyzed during the current study.

Shelly Soffer¹, Vera Sorin², Girish N. Nadkarni^{3,4} & Eyal Klang^{3,4} ✉

¹Institute of Hematology, Davidoff Cancer Center, Rabin Medical Center, Petah Tikva, Israel. ²Department of Radiology, Mayo Clinic College of Medicine and Science, Mayo Clinic, Rochester, MN, USA. ³Windreich Department of Artificial Intelligence and Human Health, Mount Sinai Health System, New York, NY, USA. ⁴The Charles Bronfman Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ✉e-mail: Eyal.Klang@mountsinai.org

Received: 21 January 2025; Accepted: 11 June 2025;

Published online: 22 July 2025

References

- Wason, P. C. & Evans, St J. B. T. Dual processes in reasoning? *Cognition* **3**, 141–154 (1974).
- Kahneman, D. *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).
- Evans, St J. B. T. & Stanovich, K. E. Dual-process theories of higher cognition: advancing the debate. *Perspect. Psychol. Sci.* **8**, 223–241 (2013).
- Keren, G. A tale of two systems: a scientific advance or a theoretical stone soup? Commentary on Evans & Stanovich (2013). *Perspect. Psychol. Sci.* **8**, 287–292 (2013).
- Hagendorff, T., Fabi, S. & Kosinski, M. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nat. Comput. Sci.* **3**, 833–838 (2023).
- Biderman, S. et al. Emergent and predictable memorization in large language models. Preprint at <https://arxiv.org/abs/2304.11158> (2023).
- McKenzie, I. R. et al. Inverse scaling: When bigger isn't better. Preprint at <https://arxiv.org/abs/2306.09479> (2023).
- OpenAI. Introducing OpenAI o3. OpenAI. <https://openai.com/index/introducing-o3-and-o4-mini/> (2025).
- Glicksberg, B. S. et al. Evaluating the accuracy of a state-of-the-art large language model for prediction of admissions from the emergency room. *Am. Med. Inform. Assoc.* **31**, 1921–1928 (2024).
- Freyer, O. et al. A future role for health applications of large language models depends on regulators enforcing safety standards. *Lancet Digit. Health* **6**, e662–e672 (2024).
- Small, W. R. et al. Large language model-based responses to patients' In-Basket messages. *JAMA Netw. Open.* **7**, e2422399 (2024).
- Ayers, J. W. et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern. Med.* **183**, 589–596 (2023).
- Rahimzadeh, V., Kostick-Quenet, K., Blumenthal-Barby, J. & McGuire, A. L. Ethics education for healthcare professionals in the era of ChatGPT and other large language models: Do we still need it? *Am. J. Bioeth.* **23**, 17–27 (2023).

- Okamoto, S., Kataoka, M., Itano, M. & Sawai, T. AI-based medical ethics education: examining the potential of large language models as a tool for virtue cultivation. *BMC Med. Educ.* **25**, 185 (2025).
- Dillion, D., Mondal, D., Tandon, N. & Gray, K. AI language model rivals expert ethicist in perceived moral expertise. *Sci. Rep.* **15**, 4084 (2025).
- Jiang, L. et al. Investigating machine moral judgement through the Delphi experiment. *Nat. Mach. Intell.* **7**, 145–160 (2025).
- Earp, B. D. et al. A personalized patient preference predictor for substituted judgments in healthcare: technically feasible and ethically desirable. *Am. J. Bioeth.* **24**, 13–26 (2024).

Acknowledgements

This study did not receive any funding.

Author contributions

S.S. and E.K.: contributed to study conceptualization, data collection, and drafting of the manuscript. V.S.: contributed to study conceptualization and assisted in manuscript revisions. G.N.N.: contributed to study conceptualization and critical revisions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01792-y>.

Correspondence and requests for materials should be addressed to Eyal Klang.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025