IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Making Emotional AI Ethical and Effective in Mental Healthcare: Addressing Bias, Privacy, and Accuracy

Akriti Shetty

Santa Clara High School, 3000 Benton St, Santa Clara, California, 95051, USA; akritishetty3@gmail.com
Mentors: Dr. Siddharth Krishnan, Samuel Lefcourt

ABSTRACT: Emotional AI is trusted with feelings it cannot feel. AI is increasingly being used in mental healthcare, chatbots are offering therapy, and algorithms are detecting depression through voice. But when it comes to understanding human emotions, even the smartest algorithms can misread, misjudge, or mistreat those they aim to help. Researchers in this field have used both qualitative and quantitative methods to explore options to mitigate biases and misreadings, including surveys on public perception, evaluations of chatbot and LLM performance, and systematic reviews of ethics in AI healthcare. Some studies compare AI responses to professionals in the medical field, while others critique current guardrails using ethical theories. These methods show both the promises and drawbacks of AI in mental health. However, surveys are not always the most accurate way to gain information on public opinion because new research shows it only adds more bias through design flaws. New evaluation metrics need to be used to help creators understand what the public thinks about using AI in this field, and to know what changes need to be made. We argue that emotional AI can be made helpful in mental healthcare, but only if its design and implementation directly address three main problems: privacy/security risks, bias against marginalized groups, and clinical accuracy. We challenge the idea that emotional AI should completely replace human care because of the risks that will arise without proper human oversight. We suggest building AI-driven mental health monitoring systems that move beyond today's mainly screening-oriented tools by integrating multiple signals over time, offering early, proactive check-ins, tracking emotional trends across text, voice, and behavior, and triaging urgent cases directly for human intervention. Because suicide claims more than 700,000 lives each year worldwide, these improvements are not just technical but vital to prevention. By rethinking emotional AI's role through this lens, our research seeks to guide the creation of safer, fairer, and more effective systems for real-world mental healthcare.

KEYWORDS: Robotics and Intelligent Machines, Machine Learning, Emotional Artificial Intelligence, Mental Health, Bias.

Introduction

According to recent global estimates, approximately 703,000 people die by suicide annually, that is, nearly one death every 43 seconds.[1] It is the leading cause of death for countless teens, veterans, and other vulnerable groups. Many people who commit suicide do not want to die; they just want the feeling or situation they are experiencing to end.[2] This distinction is incredibly significant because it means they can be talked through other alternatives. If someone is going through a difficult time but lacks access to mental healthcare, faces stereotypes or biases, or is unable to express their emotions in a way others can understand, their mental health may worsen. Poor mental health leads to suicidal thoughts, which is why helping someone before they reach these thoughts is so vital.[3] Because many mental health organizations struggle with limited resources and accessibility challenges, Artificial Intelligence is now being integrated to help address this ongoing problem.

One major challenge for emotional AI is that people express emotional distress in highly variable ways. Signals of distress can differ depending on personality, culture, or current mental state, making it difficult for AI to detect consistent patterns. Because of this variability, researchers are exploring emotional AI through controlled clinical studies, training it to chat with patients and detect mental illnesses through facial expressions. These methods could potentially help us mitigate this ongoing problem, even if not completely. Despite this, a large number of people believe it will cause more harm than good.[4] A review of recent research reveals a consistent set of obstacles that AI faces in clinical and mental-health settings. Privacy and security, biases and unfairness towards marginalized groups, and accuracy. AI can potentially help us, but if these three problems persist, then it will become part of the problem rather than the solution. It is urgent that we make emotional AI helpful for mental health without being unfair or causing harm. According to a 2017 study by Franklin *et al.*, the field of mental health has made little to no progress in efficiently predicting and preventing suicide.[5] We might have increased awareness, but the reality is that suicide rates are not declining. While AI usage in mental health support is increasing, the overall number of people experiencing suicidal thoughts has also risen, so statistics may not yet show a clear decline. In a recent survey, the final sample consisted of 107 community members (CMs) and 86 mental health professionals (MHPs). General attitudes toward AI varied, with CMs reporting neutral and MHPs more positive. AI usage was reported by 28% of CMs (primarily for quick support [60%] and as a personal therapist [47%]) and 43% of MHPs (mostly for research [65%] and report writing [54%]). While most found AI generally beneficial (77% of CMs; 92% of MHPs), a significant portion reported specific harms or concerns (47% of CMs; 51% of MHPs), and open

DOI: 10.36838/IJHSR89.47

feedback showed a mix of positive and negative sentiment toward the future of AI in mental health care.[6] These statistics show that, although AI adoption is growing and some users find it helpful, current methods are inconsistent and can cause harm, making it evident that better solutions are needed. This paper will argue that emotional AI (EAI) could save lives by helping catch signs of suicide that many people often dismiss, starting with addressing three main problems to try to save the countless lives on the line.

Discussion

This paper argues that emotional AI has the potential to improve mental healthcare, but only if three main challenges are addressed. These include privacy/security, biases especially toward marginalized groups, and clinical accuracy. Throughout the paper, we have highlighted how sharing personal data that will be stored in agents can cause barriers in trust between patients and healthcare providers. There are also security risks to this, and if the data is wrongly used, it can lead to a loss of trust completely. This would cause AI's role as a tool to be set back instead of moving forward through innovations. This paper also covers how biases can be introduced through unfair training data or poor training in general. This can worsen stereotypes about marginalized groups, causing more harm than good. Additionally, clinical accuracy is difficult to accomplish due to the complicated and subjective nature of many mental illnesses and disorders. Unlike other uses of these agents, like in customer service, this is quite literally a life-or-death situation. If a misdiagnosis occurs or no diagnosis at all, then someone who is deeply suffering could make a life-altering decision. These issues are deeply interconnected with each other and need ethical frameworks, enforceable policies, and enhanced evaluation metrics to assure AI is used to support rather than destroy.

What Emotional AI Is and Why it Matters:

Emotional AI, also known as affective computing, refers to algorithmic systems designed to automatically detect and interpret human emotions and moods.[7] AI systems can learn from their environments, adapt to new situations, and formulate their claims as humans can. A subset of Artificial Intelligence is Emotional AI, and how it is being used specifically in mental healthcare. EAI is designed to understand and respond to human emotions by using inputs like voice, facial expressions, and online behavior.[8] Although these agents use complex technologies to simulate emotions, most used in healthcare today are basic and do not deeply understand or connect with patients.[8] While AI can analyze heartbeats and scan medical images to diagnose cancer, it cannot reliably detect all signs of emotional distress.[9] This is partly because AI relies on data, and mental health data is often limited, noisy, and highly subjective. Unlike medical images, emotional distress signals vary across individuals, contexts, and cultures, and are dynamic over time, which increases the risk of AI errors. Therefore, if a person is experiencing something and it is the first-ever recorded case, then AI could potentially be completely unhelpful.

While AI shows promise as a tool in mental health, the main concern is that it can unintentionally cause harm, making its unregulated use risky. "CDC" says that suicide rates in our youth have increased by over 50% since 2007.[10] According to "APA"(American Psychological Association), racial and ethnic minority groups or those part of LGBTQ+ have been most affected.[11] "AACAP"(American Academy of Child and Adolescent Psychiatry) says most children who commit suicide have a serious mental disorder, usually depression.[12] Suicide is a leading cause of death, and the bigger concern is that it is in our youth, too. This problem has made a recurring appearance in our youth, highlighting a major health concern.[10] Another factor that makes it difficult to solve this problem is that predictive models that predict and prevent suicide have been no better than chance for the past 50 years.[5] This proves we are failing in the ways we are testing and using AI in mental healthcare, and we need to find new methods. Fortunately, Emotional AI has the potential to help look for mental health issues, but its integration in mental healthcare is still being debated. More research is needed to understand how these functions work and how to help them become better. Since this is a clinical use of AI, without proper oversight, serious problems could occur, including misdiagnosis, inappropriate treatment recommendations, and loss of patient trust. All in all, the same tools that promise to support should be inspected, as their uses bring new concerns, especially for privacy and data security.

Three Core Problems: Privacy, Bias, and Accuracy:"

Although EAI seems promising, there are serious ethical and technical risks. These include how personal data is collected and stored, how AI could be unfair to certain groups, "especially marginalized ones", and how efficient and clinically correct it is. Because unstable mental health can lead someone to take their final breath, dismissing these critical challenges Artificial intelligence is facing could mean a life or death situation.

Privacy and Security:

Emotional AI systems raise many red flags on how personal data is being used and stored, aside from assisting patients in mental healthcare. When finding out why someone commits or wants to commit suicide, lots of personal, family, and medical data are needed.[13] This is because previous records help doctors pinpoint problems linked to DNA or previous illnesses the patient faced. Because of this, a patient's willingness to share this data could vary depending on how AI will use it. If the data is not shared, it makes pinpointing the cause particularly difficult. Finding the cause of someone's instability allows doctors to stop these thoughts and feelings from progressing before it is too late.

Another issue EAI is facing is that different countries, cities, states, etc., have distinct databases. Certain data that is available in one country does not guarantee it will be available in another. According to the World Health Organization(WHO), half the world does not have access to healthcare services, and 100 million people are being pushed into poverty due to these

expenses.[14] This means that all these people might not have any pre-existing medical or personal records in the electronic health record (EHR) system, which limits AI's ability to accurately identify risk factors or provide personalized support. A WHO report shows that 75% of people with mental disorders in low-income countries do not receive any treatment at all.[14] That is over half of the population of those with these mental illnesses/disorders in low-income countries, meaning a vast number of civilians in need of help may not be able to access AI agents that are only available through formal healthcare systems. That is because if these people do not have any data in EHR, then the AI systems cannot pull from any previously known information, so everything they will hear from patients will be new information. This leaves room for a huge amount of error to occur because if the patient shares incorrect info or the AI assumes something from the info given, then misdiagnosis could occur.[4] Alongside privacy and security, AI must also be examined for how it might amplify social biases, especially against marginalized groups.

Bias and Marginalization:

Biases implemented into emotional AI agents worsen stereotypes already made against marginalized groups. Shanley Corvite *et al.* claim that they addressed the insufficient amount of empirical knowledge by analyzing 365 U.S. adults with an open-ended survey; however, it is important to understand that surveys are a limited tool for understanding complex and varying perspectives.[4] Surveys do not allow for thorough learning, and this can miss small details that matter, leading to bias or even harm.[15] This potentially means that using surveys to gain more empirical knowledge is just creating more biases. The method to gain information could have been improved, but despite this, the results prove why marginalization is a leading challenge AI faces in mental healthcare. 32% of participants, while 71.1% of them were identified with a marginalized background, claimed EAI in the work field was not helpful to them at all.[4]

Those who claimed EAI was beneficial also stated various concerns about how it could harm workers. Although less than 50% of workers expressed major concern, over half of those who were worried were from a marginalized identity. This shows how much these vulnerable groups are being affected by Emotional AI. Even those who were not directly affected were concerned, showing how much harm these biases can truly do. If biases are this strong, it can make it difficult for people of these groups to talk to someone about their mental disorders, illnesses, or just thoughts and feelings in general. A prime example of this is in veterans. According to the 2023 National Veteran Suicide Prevention Report, there were 6,392 veteran deaths by suicide in 2021 in the U.S.[16] Since veterans are often told to keep a brave face even while facing both mental and physical obstacles, opening up to a bot instead of an actual person could make them feel less judged.[13]

However, this could leave room for more problems trailing back to our first point about privacy and security. Veterans might not want to open up if the agent can be used on a remote site out of fear that their comrades can access or trace the information given to the bot. Although AI bots offer a promising solution for veterans, many may be hesitant to share sensitive information due to concerns that their data could be misused or accessed by others. Even if the AI is designed to flag suicide risk, the information would eventually need to be reviewed by a human, raising additional privacy concerns. Moreover, even with strong privacy protections and unbiased algorithms, AI still faces a critical challenge: ensuring clinically accurate information.

Accuracy and Zero-Shot Failure:

A major barrier to emotional AI is its inaccuracy and issues performing when encountering context-specific emotional problems.[18] As we have stated before, a study published in *Psychological Bulletin* shows that traditional predictive models for suicide have not performed better than chance for 50 years.[5] These older models often relied on limited datasets, simplistic statistical methods, and narrowly defined risk factors, which reduced their accuracy. This highlights the need for new methods to implement predictive models or utilize more effective training techniques. More recent AI-driven approaches have begun to address these limitations: newer models, including neural networks and gradient-boosted algorithms, have achieved AUC values between 0.60 and 0.95 in specific contexts.[13] Unlike customer service bots, which can be retrained if they make mistakes, EAI in mental healthcare cannot afford errors. Different people show emotional distress in various ways, and these can vary by person depending on personality, culture, or mental state. This is why it is harder for AI to detect patterns since each person is so different. Many "… mental disorders such as depression are highly subjective, with complex symptoms, individual differences, and strong socio-cultural ties, meaning that their diagnosis requires comprehensive consideration".[9] This proves how complicated symptoms of a great deal of mental disorders/illnesses can be, making it challenging for even qualified professionals to notice.

Zero-shot models attempt to get around this by applying knowledge learned from one domain to unforeseen circumstances. In mental health, this might mean detecting suicidal intent from a type of speech pattern the model has not seen before. In their 2023 study, Workman *et al.* employed zero-shot learning to identify suicide-related language in Veterans Affairs clinical notes without relying on task-specific labeled data.[19] The results demonstrated that this approach effectively flagged relevant content, highlighting the potential of zero-shot learning in detecting suicidality in unstructured clinical text.[19] Similarly, Gerczuk *et al.* showed that zero-shot personalization of speech models could improve mood monitoring for depression.[20] But these successes come with a warning: zero-shot models involve a tradeoff between sensitivity and specificity. Developers must decide whether to prioritize correctly identifying as many patients with suicidal thoughts as possible (high sensitivity) or minimizing false alarms for patients who are not at risk (high specificity). Failing to strike the right balance could mean missing someone in crisis entirely. We need EAI to be tested not just for how well it works, but also for ethical reasons. Future research must focus on evaluat-

DOI: 10.36838/IJHSR89.47

ing AI, concentrating on fairness, privacy, and accuracy. These intertwined issues make it evident that AI needs both ethical grounding and enforceable standards. Researchers are now arguing whether using AI is just masking the problems it is creating, instead of helping solve them.

Ethics, Policy, and Fixing the Flaws:

These interconnected issues make it clear that AI needs both ethical grounding and enforceable standards. Many researchers are now arguing that AI usage is masking problems by creating its own. Managing these problems is not just a technical challenge, but an ethical and policy-driven one as well. Once issues like privacy, bias towards the marginalized, and accuracy failures are pinpointed, the next step is finding out how to prevent them from harming people. This requires strong ethical guardrails, enforcing policies that hold companies accountable, and creating better evaluation methods that are more than just how clinically accurate an agent is. If these safeguards are not there, then AI in mental healthcare could be rushed into use, which is dangerous. By examining these ethical frameworks, we can see what changes need to be made and how to protect patients while further advancing.

Ethical Frameworks:

Ethical models provide a base for analyzing whether EAI should be used in mental healthcare. Patient autonomy, evolving decision-making ability, and continuous monitoring of AI performance in psychiatric contexts are principles being emphasized by the Integrated Ethical Approach for Computational Psychiatry (IEACP).[21] An approach like this is incredibly significant when dealing with vulnerable groups who may not give full consent or whose conditions change rapidly. Broader AI ethics research also stresses that, because of the emotional vulnerability of most patients, empathy is crucial.[22] We know that AI cannot genuinely show empathy. Empathy is the capacity to understand and share the feelings of another person.[23] AI does not have feelings; it simply mimics human emotion, but without empathy, these agents will fail to truly help patients.[24] Although AI is trying to actively solve problems, we fear that since it is not putting human feelings first, it will ultimately fail. AI often wants to solve the problem and find the solution. When working with these vulnerable patients, dismissing their experiences for the sake of putting a label on them, like depression or anxiety, can be overwhelming. Nowadays, most providers do not care about patients' concerns and put biomedical expertise over lived experiences.[8] This same ideology has been integrated into AI bots. Now, the problems should still be discussed, and solutions should be found, but putting that as the priority instead of connecting with the patients is unacceptable. Solutions cannot be found properly if one cannot even clearly understand their patients' concerns and feelings. If healthcare providers today cannot show genuine human empathy, then how will AI bots be trained to? Most AI bots are trained by using clinically provided notes or research, and if these are the same providers who are not helping their patients correctly, that data could be passed down to bots, harming more people. Beyond theory, practical implementation must be enforced through clear policies.

Policy Recommendations:

Policy gaps are being filled by organizations working to help shape responsible AI deployment. While ethics provide us with morality, policy determines how these principles are applied in practice. But certain places internationally have specific rules and regulations on how AI can be used. This makes it more challenging to implement these agents worldwide. For example, the EU AI Act prohibits certain high-risk AI applications in healthcare where the positives outweigh the negatives. However, some criticism and concerns have been brought up through this act. One concern is that the act tries to define what counts as "AI", but the definition is still too general. This causes legal confusion and makes it hard for companies to innovate.[25] The EU AI Act also organizes AI into four categories: Unacceptable risk (AI that causes serious harm), High risk (AI used in important areas like safety, critical infrastructure, or big decisions), Limited risk (AI like chatbots answering customer questions), and Minimal risk (AI like spam filters with almost no rules). The problem with this is that if an AI system wrongly gets put into one of these categories, such as High risk, it could be too strict and make it hard for companies to manage. This impacts EAI profoundly because, since these agents are meant to help people who have mental disorders/illnesses, it could be labeled as high risk for affecting people's well-being and rights. Mental health startups or researchers may struggle to meet these requirements and the costs of high amounts of testing, leading to less helpful AI tools. There are many other risks, too, like reduced access to mental health tools.

In North America, organizations like the National Institute of Standards and Technology (NIST) have made guidelines for fairness, transparency, and accountability. The Center for Security and Emerging Technology (CSET) also warns us that failing to test for bias could harm marginalized users. A more specialized technique comes from the Canada Protocol for AI in suicide prevention, which provides a validated ethical checklist going over transparency, privacy, bias, and clinical risk.[26] The main issue is that these policies are voluntary and lack enforcement, meaning companies are free to ignore them. But policy alone is not enough; how we test these systems matters just as much.

Evaluation Metrics:

Table 1 highlights how emotional AI applications in mental healthcare offer meaningful benefits such as early detection, increased access, and personalized support, but also introduce significant ethical, privacy, and bias-related risks. Overall, it emphasizes that while these tools can improve care delivery, their limitations and potential harms must be carefully managed.

Table 1: Double-edged sword scenarios for Emotional AI in Mental Healthcare.

Application	Benefit	Consideration/Risk
AI-powered suicide risk assessment (analyzing speech, text, and facial cues)	Can detect early signs of suicidal thoughts, potentially intervening in a life-altering decision	False positives or false negatives; missing a case will have consequences; over-flagging causes unnecessary interventions; loss of trust
Chatbot therapy for underserved communities	Offers 24/7 low-cost access to mental health support in areas without many professionals	Limited empathy and cultural nuance; could misinterpret user input; give unsafe advice; or fail to flag serious cases to human professionals
Emotion detection in schools (real-time monitoring via cameras/microphones)	Allows early intervention for students showing signs of distress from bullying-related trauma	Raises privacy concerns due to continuous surveillance; potential misuse by administrators or data leaks affecting minors
Emotion-based patient triage in hospitals	Helps prioritize patients in crisis by assessing urgency through vocal tone, body language, and text analysis	Biased training data may lead to misprioritization of marginalized groups, worsening healthcare inequality
AI-driven mood tracking apps	Empowers individuals to monitor mental health trends over time and share with clinicians for personalized care	Sensitive personal data could be hacked, sold, or misused; continuous monitoring may cause anxiety or overreliance on the app

For example, surveys are so commonly used, but research shows they allow a lack of understanding and raise room for bias.[15] A prominent example of the effects of poor evaluation metrics is when a commonly used healthcare prediction algorithm showed significant racial bias. The algorithm predicted healthcare costs, not actual illness, and since Black patients receive less medical spending than White patients due to unequal access to healthcare, the agent underestimated their health needs.[27] This resulted in fewer Black patients receiving extra help. The study shows that simple proxies, such as cost, can cause serious bias in algorithms instead of relying on genuine health measures. Table 1 shows that when agents try to prioritize patients' needs, biased training data can easily misprioritize the needs of marginalized groups, making healthcare inequality worse. In mental healthcare, similar risks will occur if evaluation methods do not change. Agents could focus on things that are easy to measure, such as appointment attendance or prescription fill rates, instead of actual well-being. This is a problem because AI could miss the bigger picture of how someone is doing. This is why it is so urgent for ethical and political frameworks to prioritize patients' well-being, equality, and mitigation of biases. Without these frameworks, these flawed agents will continue to pass through regulatory systems unchecked, leading to thousands of people being harmed. Although solving all of AI's problems is extremely difficult, we argue that solving the first three main problems of privacy/security, marginalization of biases, and accuracy will boost efficiency. Times are changing, and because AI has stepped into the picture, we now know why certain evaluation methods are not working as well as they used to. Suicide rates are currently on an upward trend, and we need to put a stop to it, and fast. The longer we wait, the fewer lives we save.

Conclusion

These challenges underscore the importance of cautious integration; while EAI is being rapidly deployed in clinical settings and supported by significant investment, its widespread use without comprehensive validation and oversight could lead to unintended consequences. If this happens, then many lives could be in danger. Until better evaluation metrics are used, more acceptable policies are implemented, and ethical frameworks are reevaluated, more research should be done. This will take time, but urgency is also needed because of the continuous upward trend of suicide rates.

Several questions remain: How can evaluation metrics incorporate cultural and individual differences in mental health experiences? What policies should be enforced so companies are held accountable for their agents while still allowing ongoing innovation? How can future AI models improve their ability to understand complex and subjective mental disorders or illnesses? Future research could focus on context-specific scenarios and mitigating bias through fairer training.

Acknowledgments

Thank you to Dr. Siddharth Krishnan and Samuel Lefcourt for all your help and support.

I attest that the ideas, graphics, and writing in this paper are entirely my own.

References

1. *About 740,000 global deaths from suicide occur annually-that's one death every 43 seconds*. Institute for Health Metrics and Evaluation. (n.d.-a). https://www.healthdata.org/news-events/newsroom/news-releases/about-740000-global-deaths-suicide-occur-annually-thats-one
2. Marchionatti, L. E., Amaral, R. R., Barcellos, C., Duarte, S., Campello, A. C., Virtuoso, E., & Magalhães, P. V. (2024). "I don't wanna die, but my brain insists that I should": A big qualitative data approach to the lived experiences of suicidal thoughts. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1420287/
3. Moitra, M., Santomauro, D., Degenhardt, L., Collins, P. Y., Whiteford, H., Vos, T., & Ferrari, A. (2021). Estimating the risk of suicide associated with mental disorders: A systematic review and meta-regression analysis. *Journal of Psychiatric Research, 137*, 242–249. https://doi.org/10.1016/j.jpsychires.2021.02.053
4. Shanley Corvite, Roemmich, K., Rosenberg, T., & Nazanin Andalibi. (2023). Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–38. https://doi.org/10.1145/3579600
5. Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143* (2), 187–232. https://doi.org/10.1037/bul0000084
6. *Cross, S., Bell, I., Nicholas, J., Valentine, L., Mangelsdorf, S., Baker, S., Titov, N., & Alvarez-Jimenez, M. (2024). Use of AI in mental health care: Community and mental health professionals survey. JMIR Mental Health, 11. https://doi.org/10.2196/60589*
7. *Ingber, A. S., Haimson, O. L., & Andalibi, N. (2025). Distinguishing emotion AI: Factors shaping perceptions, including input data, emotion data recipients, and Identity. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, 498–510. https://doi.org/10.1145/3715275.3732034*
8. Roemmich, K., Corvite, S., Pyle, C., Karizat, N., & Andalibi, N. (2024). Emotion AI use in U.S. mental healthcare: *Potentially unjust*

DOI: 10.36838/IJHSR89.47

and techno-solutionist. Proceedings of the ACM on Human-Computer Interaction, *8* (CSCW1), 1-46. https://doi.org/10.1145/3637324
9. Yan, W.-J., Ruan, Q.-N., & Jiang, K. (2022). Challenges for Artificial Intelligence in Recognizing Mental Disorders. *Diagnostics, 13*(1). https://doi.org/10.3390/diagnostics13010002
10. *Suicide*. (n.d.). Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/hus/topics/suicide.htm
11. *American Psychological Association. (n.d.). Monitor on psychology. https://www.apa.org/monitor/2023/07/psychologists-preventing-teen-suicide*
12. *Aacap. (n.d.). Suicide in children and teens. https://www.aacap.org/AACAP/Families_and_Youth/Facts_for_Families/FFF-Guide/Teen-Suicide-010.aspx*
13. Lejeune, A., Le Glaz, A., Perron, P.-A., Sebti, J., Baca-Garcia, E., Walter, M., Lemey, C., & Berrouiguet, S. (2022). Artificial intelligence and suicide prevention: a systematic review. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *65*(1), 1–22. https://doi.org/10.1192/j.eurpsy.2022.8
14. World Health Organization. (2017, December 13). *World Bank and WHO: Half the world lacks access to essential health services, and 100 million still pushed into extreme poverty because of health expenses*. World Health Organization; World Health Organization. https://www.who.int/news/item/13-12-2017-world-bank-and-who-half-the-world-lacks-access-to-essential-health-services-100-million-still-pushed-into-extreme-poverty-because-of-health-expenses
15. Tahaei, M., Wilkinson, D., Frik, A., Muller, M., Ruba Abu-Salma, & Wilcox, L. (2024). *Surveys Considered Harmful? Reflecting on the Use of Surveys in AI Research, Development, and Governance. 7*, 1416–1433. https://doi.org/10.1609/aies.v7i1.31734
16. Lauderdale, S. A., Schmitt, R., Wuckovich, B., Dalal, N., Desai, H., & Tomlinson, S. (2025). Effectiveness of generative AI-large language models' recognition of veteran suicide risk: A comparison with human mental health providers using a risk stratification model. *Frontiers in Psychiatry, 16*. https://doi.org/10.3389/fpsyt.2025.1544951
17. Ganz, A., Yamaguchi, C., Parekh, B., Koritzky, G., & Berger, S. (2021). Military culture and its impact on Mental Health and Stigma. *Journal of Community Engagement and Scholarship, 13*(4). https://doi.org/10.54656/zzhp1245
18. *Wu, J. (2024). Social and ethical impact of emotional AI advancement: The rise of pseudo-intimacy relationships and challenges in human interactions. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1410462*
19. Workman, T. E., Goulet, J. L., Brandt, C. A., Warren, A. R., Eleazer, J., Skanderson, M., Lindemann, L., Blosnich, J. R., O'Leary, J., & Zeng-Treitler, Q. (2023). Identifying suicide documentation in clinical notes through zero-shot learning. *Health Science Reports, 6*(9). https://doi.org/10.1002/hsr2.1526
20. Gerczuk, M., Triantafyllopoulos, A., Shahin Amiriparian, Kathan, A., Bauer, J., Matthias Berking, & Schuller, B. W. (2023). Zero-shot personalization of speech foundation models for depressed mood monitoring. *Patterns, 4*(11), 100873–100873. https://doi.org/10.1016/j.patter.2023.100873
21. Putica, A., Khanna, R., Wiliam Bosl, Saraf, S., & Edgcomb, J. (2025). Ethical decision-making for AI in mental health: the Integrated Ethical Approach for Computational Psychiatry (IEACP) framework. *Psychological Medicine, 55*. https://doi.org/10.1017/s0033291725101311
22. Rubeis, G. (2022). iHealth: The ethics of artificial intelligence and big data in mental healthcare. *Internet Interventions, 28*(1), 100518. https://doi.org/10.1016/j.invent.2022.100518
23. *Schramme, T. (2024). Empathy as a means to understand people. Philosophical Explorations, 27(2), 157–170. https://doi.org/10.1080/13869795.2024.2344975*
24. Tavory, T. (2024). Regulating AI in Mental Health - the Ethics of Care Perspective (Preprint). *JMIR Mental Health, 11*(e58493). https://doi.org/10.2196/58493
25. Higgins, T. (2023, April 6). *The EU AI Act: concerns and criticism*. Clifford Chance. https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2023/04/the-eu-ai-act--concerns-and-criticism.html
26. Mörch, C.-M., Gupta, A., & Mishara, B. L. (2020). Canada Protocol: An Ethical Checklist For The Use Of Artificial Intelligence In Suicide Prevention And Mental Health. *Artificial Intelligence in Medicine*, 101934. https://doi.org/10.1016/j.artmed.2020.101934
27. Z, O., B, P., & C, *et al.* V. (2020). Dissecting racial bias in an algorithm used to manage the health of populations. *Yearbook of Paediatric Endocrinology*. https://doi.org/10.1530/ey.17.12.7

Author

Akriti Shetty is a sophomore at Santa Clara High School. She hopes to attend Princeton or UC Berkeley and plans to major in Computer Science, Psychology, or Neuroscience. She is passionate about artificial intelligence and aims to pursue a career integrating AI with real-world problem-solving.