SIGHCI 2023 Proceedings

Special Interest Group on Human-Computer Interaction

1-25-2024

Digital Risk Considerations Across Generative AI-Based Mental Health Apps

Sagarika Thimmanayakanapalya
University at Buffalo, sthimman@Buffalo.edu

Sanjukta Das Smith
University at Buffalo, sdsmith@buffalo.edu

George Sanders
University at Buffalo, mgtsand@buffalo.edu

Digital Risk Considerations Across Generative AI-based Mental Health Apps

Sagarika Suresh Thimmanakanapalya
University at Buffalo
sthimman@buffalo.edu

Sanjukta Das Smith
University at Buffalo
sdsmith@buffalo.edu

George Lawrence Sanders
University at Buffalo
mgtsand@buffalo.edu

ABSTRACT

Mental health platforms on online mobile applications are increasingly adopting generative AI algorithms, however, studies point to the digital risks involved in this adoption. Ethical dilemmas, misinterpretation of complex medical cases, compromised patient privacy, and potential legal liabilities deter generative AI integration with online mobile applications. This study examines 1 million user-generated review comments from 54 applications on various mobile platforms such as Google Store and App Store which use generative AI to provide mental health assistance. The review comments are studied using text-mining approaches to identify the potential digital risks posed to users across these mental healthcare apps. Results from our study aim to guide regulatory frameworks in healthcare in the future.

Keywords

Generative AI, Mental Health, Healthcare AI, Chatbot

INTRODUCTION

In a study conducted in 2021, a substantial proportion of the surveyed American population, comprising nearly fifty percent, disclosed recent manifestations indicative of anxiety and depressive disorders (Hamdoun et al., 2023). Correspondingly, a noteworthy 10% of the respondents indicated a perception of inadequacy in addressing their mental health exigencies (Hamdoun et al., 2023). The prevalence of mental health disorders has escalated in recent years, a phenomenon that has garnered authoritative acknowledgment from the World Health Organization’s 2019 report. Moreover, there is a glaring paucity in the availability of mental healthcare practitioners, with a single mental health professional allocated for every 350 individuals (World Health Organization, 2019). The pronounced deficit in qualified healthcare providers, coupled with the enduring societal stigma enveloping mental health concerns, has precipitated a discernible paradigm shift whereby individuals are increasingly gravitating towards digital behavioral health tools as a means of addressing their mental health requisites (Ma et al., 2023; Ebbert et al., 2023).

With the rapid shift towards more patients using various online tools for assistance, many researchers have studied how generative AI could be instrumental in aiding under-served populations. Studies now point to how these tools could act as a patient-facing chatbot or back-end assistant, providing insights to the physician based on their processing capabilities (Hider & Wright, 2023; Mitra, 2023; Berbatova, 2023). Moreover, studies talk about the ability of generative AI tools to democratize the source of medical information. They may also deconstruct the medical hierarchy and center patients in their mental health treatment (Mitra, 2023). In addition, generative AI's analysis of speech and language patterns may more accurately identify patients' concerns (e.g., medication side effects that the physician may consider relatively minor) to optimize shared decision-making. However, despite all these merits, it is important to recognize that the true nature of patients simply cannot be captured by generative AI. For instance, a savvy patient may know just what to say to these tools to avoid being flagged for hospitalization when suicidal. Patient risk assessment requires a thorough, moment-to-moment physical and mental status exam as well as a longitudinal knowledge of the individual. Generative AI may not only miss crucial nuances but also incorrectly influence the thinking of the patient in high-risk situations (Aminah et al., 2023).

Therefore, realizing the digital risks that generative AI in mental health care could bring is crucial. Digital risk in the healthcare domain refers to all unexpected consequences that result from digital transformation disrupt the patient experience and sometimes pose direct risks to their health (Boucher et al., 2021). Several challenges must be addressed before optimistically placing this system in a sensitive healthcare context. It is important to realize that generative AI cannot currently understand and respond to human emotions with the depth and nuance that a human therapist can offer (Oniani et al., 2023). Moreover, unlike most IS systems that yield deterministic outcomes, the research continuum in generative AI is beginning to accept that these systems' outcomes are probabilistic (van Giffen et al., 2023). It is, therefore, necessary to check for the various digital risks imposed by systems that generate probabilistic outcomes for mental health contexts. Furthermore, understanding these risks can help contribute towards the changing paradigm in the design that meets the regulatory and compliance requirements, which has historically focused on deterministic systems (Emdad et al., 2023).

Our study identifies 54 user-oriented generative AI platforms on the Google Store and App Store to assist users with various mental health issues. Our search words were "mental health"/" social anxiety"/" therapy". We narrowed down our list of apps to 54 platforms as they stated using forms of generative AI algorithms. We subsequently collected 1 million user reviews to understand the different digital risk dimensions that users were concerned with using hierarchical modeling based on the context. Our paper is structured as follows: First, the background section highlights important studies that have been conducted in our study's context. Second, the method section explains the processes used to identify various digital risks imposed by these applications in the mental healthcare setting. Third, the result section informs readers of the various digital risks. Finally, we conclude our study by highlighting our contributions, and future work.

BACKGROUND

Generative AI in Healthcare and Digital Risks Imposed

The rapid advancements in artificial intelligence (AI) have led to the development of sophisticated large language models such as GPT-4 and Bard (Haver et al., 2023). Given its wide range of uses, such as facilitating clinical documentation, obtaining insurance pre-authorization, summarizing research papers, or acting as a chatbot to respond to questions from patients about their specific data and concerns, generative AI's potential implementation in healthcare settings has garnered a considerable amount of research interest (Haver et al., 2023). While offering transformative potential, generative AI warrants a very cautious approach since these models are trained differently from AI-based medical technologies that are regulated already, especially within the critical context of caring for patients (Mesko & Topol, 2023). Numerous studies currently suggest that these tools should be rigorously regulated before becoming widely used (Mesko & Topol, 2023; Zhu et al., 2022). Currently, the FDA does not have any categories solely for AI-based technologies; instead, it examines them by the rules already in place for medical devices (Zhu et al., 2022).

As of today, there has been little to no generative AI that has been pre-trained with the corpus of medical information or with millions of patient records, images, lab data, office visits, or bedside conversations (Mesko & Topol, 2023). However, generative AI has the potential to revolutionize this industry, with applications ranging from clinical documentation to creating customized health plans (Lee et al., 2023). The introduction of these models into healthcare may lead to the amplification of digital risks and challenges.

Generative AI poses a new challenge to physicians as patients now come to consultations with pre-conceived notions formed through their interactions with ChatGPT-like chatbots, or apps that use generative AI (Mesko & Topol, 2023). It is well known that generative AI can sometimes "hallucinate" results, which refers to generating outputs that are not grounded in the input data or information (Lee, 2023). Such misinformation may be related to diagnosis, treatment, or a recommended test. Such outputs that are transmitted with a high degree of confidence could be readily taken as fact for ignorant patients, which has the potential to be harmful. Whether it is due to incomplete or biased training data, its probabilistic nature, or the lack of context, it poses a significant risk of providing unreliable or outright false answers in the medical setting that might have serious consequences.

Many studies point to how applications using generative AI in healthcare raise ethical concerns that warrant a regulatory framework. Issues, such as transparency, accountability, and fairness need to be addressed to prevent potential ethical lapses (Wang et al., 2023; Javaid et al., 2023; Kumaresan et al., 2023). Specially, while building regulatory requirements for generative AI it is crucial to observe its impact in high-risk healthcare situations.

Mental healthcare apps and Generative AI

Studies explain how there is increasing demand for mental health services and how the expanding capabilities of artificial intelligence have driven the development of digital mental health interventions (DMHIs) (Boucher et al., 2021). One of the biggest challenges with generative AI is aside from being a hallucinating system, it is also a probabilistic system. Models that generate responses to customer support queries will produce inaccurate or out-of-date results if the content it is grounded in is old, incomplete, and inaccurate. This can lead to outcomes in which a tool confidently asserts that falsehood is real (Mckinsey, 2023). Studies speak about the revolution that generative AI can bring to the mental health space when paired with a professional (Mesko & Topol, 2023). Studies also suggest how generative AI could aid in psychiatric medical management and psychotherapy (Hider & Wright 2023; Mesko & Topol, 2023). To overcome generative AI's lack of human experience and long-term rapport, generative AI has also been recommended for therapy modalities that are minimally dependent on emotional/supportive statements (e.g., supportive psychotherapy or motivational interviewing) or the interpersonal relationship between patient and therapist (e.g., psychodynamic psychotherapy) (Heinz et al., 2023). The consensus in the research world is that generative AI could primarily assist in manual-based treatment modalities like cognitive behavioral therapy or interpersonal therapy, where ready-made tools can be taught and applied. Ultimately, IRB-approved research will be needed to understand a chatbot's ability to develop a therapeutic alliance and execute certain psychotherapy modalities. However, many generative AI applications are still available on the Google Play Store and App Store such as Replika, Youper, CBT companion bot, etc. that are actively used by millions of users each day as a replacement for a therapist with no regulatory oversight (Kettle & Lee, 2023; Haque & Rubya, 2023). It is, therefore, crucial to understand the digital risks imposed by these generative AI apps in a sensitive healthcare context such as mental health to alert and inform researchers.

METHODOLOGY

We use an unsupervised text-mining approach to analyze 1 million user-generated reviews from 54 mental health platforms on the Google Store and App Store that use generative AI to provide therapy to end-users. Figure 1 provides an overview of our research methodology.

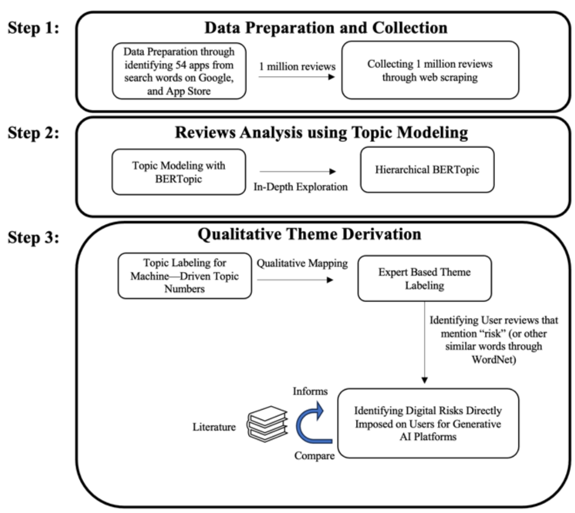

Figure 1. Research Methodology Overview

Step 1: Data Preparation and Collection

This step first consisted of identifying the various mental health platforms on the Google Store and the App Store that utilized generative AI. We identified 54 user-oriented apps that showcased themselves as providing various types of services for the betterment of the user's mental health. Moreover, they specifically used generative AI to provide different forms of therapy. Our search words were "mental health", "social anxiety" and "therapy". Finally, we collected 1 million user reviews from 54 apps.

Step 2: Review analysis using Topic Modeling

The topic modeling procedure was run on these review comments and adjusted for hyperparameters based on the sample size. We conducted BERTopic modeling on all these review comments across the three types of digital health apps. BERTopic modeling is chosen because it has key advantages over traditional topic models such as LDA (Latent Dirichlet Allocation). BERTopic captures finer and more meaningful patterns than traditional models due to accurate pre-trained word and sentence representations (Devlin et al., 2018). It refines topic deduction via hyperparameter tuning (Dong et al., 2019). Specifically, we undertook the following steps:

(1) Data Cleaning

Scikit-Learn is utilized to remove stop words, punctuations, and numbers. Moreover, short reviews with less than 250 characters were removed as this would not yield any meaningful topic patterns (Helan & Sultani, 2023). After data cleaning, we had 986,001 review comments for final analysis.

(2) Tokenization

In this step, text reviews are converted into numerical data, and subsequently, BERT embeddings are generated (Angelov, 2020; Devlin et al., 2018), where words are vectorized in a high dimensional space.

(3) Clustering procedure

There are three stages involved in the clustering procedure. First, the BERT embeddings are clustered into 512-dimensional vectors for a balance between speed and efficacy. Second, UMAP reduces 512 dimensionalities to 5, while retaining the local structure of the clusters (Angelov, 2020), considering both similarity and density. Finally, HDBSCAN (where density-based clustering is converted to hierarchical clustering) is followed to validate the cluster structure effectively in the reduced space (Angelov, 2020). The hyperparameters in this step are adjusted according to the sample size.

(4) Deducing topics

To understand keywords, the importance of words between documents (i.e., each review comment in our case) is created by calculating the widely used class-based TF-IDF score, which reflects the importance of a word within a document or corpus (Grootendorst, 2022). It yielded the top 4 most important words in each cluster. The topic number (e.g., 0,1,2,3, etc.) is then assigned based on the keywords by the BERTopic model.

Subsequently, we performed hierarchical clustering of BERTopic results, which further helps remove redundancy across topic clusters formed from BERTopic alone (George & Sumathy, 2023). Removing redundant topics is crucial because the quantity significantly affects representation quality. Redundant topics were removed by understanding the hierarchical tree structures from the hierarchical BERTopic results and manually looking at redundant topics. Moreover, the hierarchical BERTopic helped us understand risk-based topics in more detail. Performance metrics are shown in Table 1.

Topic Coherence	Topic Perplexity	No of Topics from BERTopic	No of topics after Hierarchical BERTopic Saturation
0.06	-9.8	21	15

Table 1. Performance Metrics

Step 3: Qualitative Theme Derivation -Comparison to risk-based literature

With the number of topics deducted from the hierarchical BERTopic modeling, three researchers traced back to the relevant review comments under each topic number. After careful redundant topic elimination through expert discussions (i.e., hierarchical BERTopic modeling), the number of final topics was reduced to 15 major topics from 21. Furthermore, out of the 15 major topics identified (Table 2), we only targeted topics that directly led to digital risks to the users. These were identified when users reported having felt at risk (or other similar synonyms that we generated from WordNet) in the subsequent review comment mapped to these topics. We also compared the topics that imposed various risks, to literature to further understand how they caused risks to users. We finally identified 8 major digital risks. These results are provided in Table 2.

	Hierarchical User topics	**Risk Identified directly through user review (when the user mentioned having felt at risk under this theme)**	**Digital Risk Identified-Compared to existing literature**
1	Technical Glitches in the app (app lags, crashes, etc.)	Not a direct digital risk to user	Not applicable
2	About mental health goal setting	Not a direct digital risk to user	Not applicable
3	User Sensitivity towards Robot Eeriness	Mentioned as a form of risk	(1) Uncanny valley (Betriana et al., 2021)
4	Bullying by the bot	Mentioned as a form of risk	(2) Toxic Behavior (Zhuo et al., 2023)
5	General Likeability towards app	Not a direct digital risk to user	Not applicable
6	Improper therapeutical guidance	Mentioned as a form of risk	(3) Lack of therapy nuances (Ratican & Hutson, 2023)
7	General Hate toward app	Not a direct digital risk to user	Not applicable
8	Comparison to real therapist answers	Mentioned as a form of risk	(4) Lack of Expert Validation (Krause, 2023)
9	Feeling guided and motivated	Not a direct digital risk to user	Not applicable
10	Issues with information disclosure to app	Mentioned as a form of risk	(5) Transparency paradox (Cahoy, 2007; Wang et al., 2023)
11	Inaccurate recommendations	Mentioned as a form of risk	(6) Falsified data reporting (Chomsky et al., 2023)
12	Talking about subscription methods	Not a direct digital risk to user	Not applicable
13	Biased Outcomes based on payments	Mentioned as a form of risk	(7) Bias (Wang et al., 2023)
14	Reporting on the aesthetics of the apps	Not a direct digital risk to user	Not applicable
15	Questionable new advice	Mentioned as a form of risk	(8) Stochastic Problem (Li, 2023)

Table 2. Identifying Digital Risks Imposed on Users

RESULTS

Our analysis identified eight key digital risks in mental health platforms. We identified three main entities that can induce these risks, which include (i) risks that are induced by technology, (ii) risks that are induced because of the user's perception of the technology, and (iii) risks that are induced primarily due to a flaw in the therapeutic process. These outcomes result from faulty design because technical developers lack the appropriate knowledge of therapy procedures.

Risk Imposed by Technology

We identified risks that are primarily imposed by the lack of strategizing the building of the technology in sensitive healthcare contexts. We highlight them below:

(i) Stochastic Problem

Stochastic problems are mathematical problems where some of the data incorporated into the objective is uncertain (Li, 2023). This leads to the AI giving unvalidated advice for Mental health users. User-generated comments report that the technologies which they interacted with provided data that is completely new and uncertain about therapeutic procedures. This issue is caused due to improper development of the technology.

(ii) Transparency Paradox

In the world of data, it is frequently assumed that more data is better. But in risk management, data itself is often a source of liability. This is specifically an important issue with sensitive healthcare contexts where users are often seen either complaining that the right datasets are not collected for enough predictions and contrarily also report having security issues with giving out health information (Cahoy, 2007; Wang et al., 2023). Therefore, the technology must consider a balanced approach to collecting data in such cases.

(iii) Falsified Data Reporting

Users also mentioned how certain data reported seem misleading. Users also reported being supplied with wrong information on mental health issues. This is cause of concern for the technology being developed. Developers must seek expert advice before training generative models (Chomsky et al., 2023).

(iv) Bias

Users reported feeling that the chatbot was biased in its answers when they bought premium subscriptions. For instance, companion bots such as Replika which claim to address social anxiety issues behave more politely once premium features are bought. This raises crucial questions about the monetization of AI in healthcare and whether it is fully understood in sensitive healthcare contexts (Wang et al., 2023).

Risk Imposed by User Perception

Under this category, we identified risks that are primarily imposed by user perceptions. In other words, user interactions with the bots must be monitored in cases where the apps can learn from user behavior. Moreover, without properly explaining the basic nuances of the app functionalities to users, it may cause them to experience eerie feelings, discomfort, etc. which are a form of risk, in a sensitive healthcare context such as mental health.

(v) Uncanny Valley

The uncanny valley is the region of negative emotional response towards robots that seem “almost” human (Betriana et al., 2021). Building humane companion bots to address mental health needs causes certain users to feel uncomfortable during their therapeutic interaction with these apps. Users reported feeling eerie, and uncomfortable from their interaction with the tools. Their perception of these apps can in turn be risky in highly sensitive contexts, such as when a mental health patient is highly in need of a humane touch, but realizes that they are indeed interacting with a chatbot.

(vi) Toxic Behavior

Most users also reported to have been bullied and harassed by the chatbot. This is because generative AI trains itself based on the user (Zhuo et al., 2023. If users behave suicidal, the bot can also behave similarly. Many user comments reported that the bot coxed them to commit suicide and had malicious intentions towards them. A similar case was exhibited by Microsoft’s Tay (Wolf et al., 2017) which was eventually shut down due to its toxic behavior. Therefore, it is crucial to properly monitor the data being used to train generative AI apps in highly sensitive healthcare contexts.

Risk Imposed by Process

Under this category, we identified risks that are imposed due to flaws in the therapeutic processes. For instance, certain apps mention curing social anxiety. However, when users interact with the chatbot, they report being even more socially isolated.

(vii) Lack of Therapy Nuances

When mental health apps are built, they must make sure that the right processes are implemented (Ratican & Hutson, 2023). For instance, social bots can indeed socially isolate users from the real world, cognitive behavioral therapy bots may lack the multisensory nuances of the therapy. Therefore, these apps must make sure that the therapy provided must follow the right, ethical procedures to treat patients. Moreover, they must encourage users with high risk to seek professional help.

(viii) Lack of Expert Validation

Most therapy bots lack expert validation. More therapists must be involved while building these generative AI apps. Moreover, therapist or expert recommendations must be given prominence over user data-based recommendations (Krause, 2023).

CONCLUSION

This work makes several significant contributions. First, this research informs readers of the various digital risks that generative AI tools can impose in a sensitive, high-risk, and in some cultures taboo-ridden healthcare context such as mental health. Information in such contexts requires carefully calculated data points, and finetuned algorithms both from the developers as well as domain experts (such as therapists, social workers, etc.). Inadequate or improper addressing of these computed points in AI tools can have severe consequences on mental health patients including life-threatening outcomes. Second, this research will inform patients and healthcare professionals about correctly positioning these systems in sensitive contexts. Generative AI has promise, but we must consider the many ethical and design barriers before launching them in high-risk situations. Third, our work informs readers about the combined risks identified by different stakeholders from diverse backgrounds. This in turn helps each stakeholder involved understand the adversities of the other. Finally, these digital risks also contribute to the bigger research continuum of building utilitarian risk frameworks for generative AI in the healthcare domain.

For future work, we are conducting a mixed-method study, using the results from our machine learning and qualitative grounded theory approach. We plan to inform readers of the digital risks involved in using generative AI tools. Moreover, we also want to categorize the various forms of digital therapy provided by these apps as high, moderate, or low risk based on the results from a series of semi-structured interviews. This would help researchers understand varying digital risks based on the sensitivity of the user interacting with these platforms. Finally, due to page constraints, we have not showcased our results from hierarchical tree structures, which we plan to report in future studies.

ACKNOWLEDGMENTS

We thank all authors, committee members, and volunteers for their hard work and contributions to the workshop. We especially thank the reviewers for their valuable feedback during the workshop.

AUTHOR BIOGRAPHIES

Sagarika Suresh Thimmanayakanapalya is a doctoral candidate in Management Science and Systems from the University at Buffalo. Her work lies in developing and designing generative AI technologies that follow measurement science strategies.

Sanjukta Das Smith is an associate professor in Management Science and Systems from the University at Buffalo, School of Management. Her research interests include resource allocation and contract design in cloud computing. Her research has been funded by Google, NSF, and Huawei. She serves as an Associate Editor (AE) for the Information Systems Journal.

George Lawrence Sanders is an associate editor of Decision Support Systems and is currently a co-PI on a $2.39 million grant from the National Science Foundation to train future cybersecurity experts. His current research interests include psychological profiling of hacking behavior, privacy, and security.

REFERENCES

1. Aminah, S., Hidayah, N., & Ramli, M. (2023). Considering ChatGPT to be the first aid for young adults on mental health issues. Journal of Public Health, fdad065.
2. Angelov, D. (2020). Top2vec: Distributed representations of topics. arXiv preprint arXiv:2008.09470.
3. Berbatova, M. Can AI chatbots be our therapists?.
4. Betriana, F., Osaka, K., Matsumoto, K., Tanioka, T., & Locsin, R. C. (2021). Relating Mori's Uncanny Valley in generating conversations with artificial affective communication and natural language processing. Nursing Philosophy, 22(2), e12322.
5. Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., ... & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. Expert Review of Medical Devices, 18(sup1), 37-49
6. Cahoy, D. R. (2007). Medical product information incentives and the transparency paradox. Ind. LJ, 82, 623.
7. Chomsky, N., Roberts, I., & Watumull, J. (2023). Noam Chomsky: The False Promise of ChatGPT. The New York Times, 8.
8. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
9. Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., ... & Hon, H. W. (2019). Unified language model pre-training for natural language understanding and generation. Advances in neural information processing systems, 32.
10. Ebbert, J. O., Ramar, P., Tulledge-Scheitel, S. M., Njeru, J. W., Rosedahl, J. K., Roellinger, D., & Philpot, L. M. (2023). Patient preferences for telehealth services in a large multispecialty practice. Journal of Telemedicine and Telecare, 29(4), 298-303.
11. Emdad, F. B., Ho, S. M., Ravuri, B., & Hussain, S. (2023). Towards A Unified Utilitarian Ethics Framework for Healthcare Artificial Intelligence.
12. George, L., & Sumathy, P. (2023). An integrated clustering and BERT framework for improved topic modeling. International Journal of Information Technology, 1-9.

13. Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
14. Hamdoun, S., Monteleone, R., Bookman, T., & Michael, K. (2023). AI-based and digital mental health apps: Balancing need and risk. IEEE Technology and Society Magazine, 42(1), 25-36.
15. Haque, M. R., & Rubya, S. (2023). An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews. JMIR mHealth and uHealth, 11(1), e44838.
16. Haver, H. L., Lin, C. T., Sirajuddin, A., Yi, P. H., & Jeudy, J. (2023). Use of ChatGPT, GPT-4, and Bard to Improve Readability of ChatGPT's Answers to Common Questions on Lung Cancer and Lung Cancer Screening. American Journal of Roentgenology.
17. Heinz, M. V., Bhattacharya, S., Trudeau, B., Quist, R., Song, S. H., Lee, C. M., & Jacobson, N. C. (2023). Testing domain knowledge and risk of bias of a large-scale general artificial intelligence model in mental health. Digital Health, 9, 20552076231170499.
18. Helan, A., & Sultani, Z. N. (2023, February). Topic modeling methods for text data analysis: A review. In AIP Conference Proceedings (Vol. 2457, No. 1). AIP Publishing.
19. Hider, A., & Wright, L. (2023). " Clinical Reach into the Cognitive Space"(CRITiCS)-An outline conceptual framework for the safe use of generative artificial intelligence in mental health decision making.
20. https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai
21. Javaid, M., Haleem, A., & Singh, R. P. (2023). ChatGPT for healthcare services: An emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 3(1), 100105.
22. Kettle, L., & Lee, Y. C. (2023). User Experiences of Well-Being Chatbots. Human Factors, 00187208231162453
23. Krause, D. (2023). Mitigating Risks for Financial Firms Using Generative AI Tools. Available at SSRN 4452600.
24. Kumaresan, A., Uden, L., & Ashraf, S. (2023, May). Potential Role of ChatGPT in Healthcare in the Prevention and Management of Non-communicable Diseases. In International Conference on Knowledge Management in Organizations (pp. 430-442). Cham: Springer Nature Switzerland.
25. Lee, M. (2023). A mathematical investigation of hallucination and creativity in gpt models. Mathematics, 11(10), 2320.
26. Lee, P., Bubeck, S. & Petro, J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N. Engl. J. Med. 388, 1233–1239 (2023).
27. Li, Z. (2023). The dark side of chatgpt: Legal and ethical challenges from stochastic parrots and hallucination. arXiv preprint arXiv:2304.14347.
28. Ma, K. K. Y., Anderson, J. K., & Burn, A. M. (2023). School-based interventions to improve mental health literacy and reduce mental health stigma–a systematic review. Child and Adolescent Mental Health, 28(2), 230-240.
29. Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. npj Digital Medicine, 6(1), 120
30. Mitra, D. S. (2023). Generative AI and Metaverse: Companionship and Assisted Living for Elderly People. Available at SSRN 4489358.
31. Oniani, D., Hilsman, J., Peng, Y., Poropatich, R. K., Pamplin, C. O. L., Legault, L. T. C., & Wang, Y. (2023). From Military to Healthcare: Adopting and Expanding Ethical Principles for Generative Artificial Intelligence. arXiv preprint arXiv:2308.02448.
32. Ratican, J., & Hutson, J. (2023). The Six Emotional Dimension (6DE) Model: A Multidimensional Approach to Analyzing Human Emotions and Unlocking the Potential of Emotionally Intelligent Artificial Intelligence (AI) via Large Language Models (LLM). Journal of Artificial Intelligence and Robotics, 1(1).
33. van Giffen, B., & Ludwig, H. (2023). How Siemens Democratized Artificial Intelligence. MIS Quarterly Executive, 22(1), 3.
34. Wang, B., Chen, W., Pei, H., Xie, C., Kang, M., Zhang, C., ... & Li, B. (2023). DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models. arXiv preprint arXiv:2306.11698.
35. Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical Considerations of Using ChatGPT in Health Care. Journal of Medical Internet Research, 25, e48009.
36. Wolf, M. J., Miller, K., & Grodzinsky, F. S. (2017). Why we should have seen that coming: comments on Microsoft's tay" experiment," and wider implications. Acm Sigcas Computers and Society, 47(3), 54-64.
37. World Health Organization. (2019). The WHO special initiative for mental health (2019-2023): universal health coverage for mental health (No. WHO/MSD/19.1). World Health Organization.
38. Zhu, S., Gilbert, M., Chetty, I., & Siddiqui, F. (2022). The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: an analysis of the characteristics and intended use. International journal of medical informatics, 165, 104828.
39. Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). Exploring ai ethics of chatgpt: A diagnostic analysis. arXiv preprint arXiv:2301.12867.