Check for updates

Surviving ChatGPT in healthcare

Zhengliang Liu[1], Lu Zhang[2], Zihao Wu[1], Xiaowei Yu[2], Chao Cao[2], Haixing Dai[1], Ninghao Liu[1], Jun Liu[3], Wei Liu[4], Quanzheng Li[5], Dinggang Shen[6,7,8], Xiang Li[5], Dajiang Zhu[2] and Tianming Liu[1]*

[1]School of Computing, University of Georgia, Athens, GA, United States, [2]Department of Computer Science and Engineering, The University of Texas at Arlington, Arlington, TX, United States, [3]Department of Radiology, Second Xiangya Hospital, Changsha, Hunan, China, [4]Department of Radiation Oncology, Mayo Clinic, Scottsdale, AZ, United States, [5]Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States, [6]School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, [7]Department of Research and Development, Shanhai United Imaging Intelligence Co., Ltd., Shanghai, China, [8]Shanghai Clinical Research and Trial Center, Shanghai, China

At the dawn of of Artificial General Intelligence (AGI), the emergence of large language models such as ChatGPT show promise in revolutionizing healthcare by improving patient care, expanding medical access, and optimizing clinical processes. However, their integration into healthcare systems requires careful consideration of potential risks, such as inaccurate medical advice, patient privacy violations, the creation of falsified documents or images, overreliance on AGI in medical education, and the perpetuation of biases. It is crucial to implement proper oversight and regulation to address these risks, ensuring the safe and effective incorporation of AGI technologies into healthcare systems. By acknowledging and mitigating these challenges, AGI can be harnessed to enhance patient care, medical knowledge, and healthcare processes, ultimately benefiting society as a whole.

1 Introduction

Large language models (LLM) such as ChatGPT and GPT-4 are making significant strides towards the development of Artificial General Intelligence (AGI) (1, 2). Advanced generative AGI holds the potential to improve patient care (3, 4) and streamline healthcare processes (3, 5). However, without proper oversight and regulation, the integration of LLMs into the healthcare system could introduce a range of unintended risks and consequences (4, 6). It is essential to explore these potential risks and address them effectively to ensure that AGI serves as a beneficial aid in the medical field.

One of the foremost concerns regarding AGI in healthcare is the risk of providing inaccurate medical advice (6, 7). Since AI-generated content (AIGC) is based on vast amounts of internet data, there is a possibility that the information provided may be misleading or outright incorrect. For instance, ChatGPT might offer treatment suggestions that are either outdated or not suitable for a specific patient's condition. Such inaccuracies could result in patients receiving inappropriate treatments or even exacerbate their health issues.

In addition to the significant concern of inaccurate advice and conclusions, the potential violation of patient privacy is another area of concern (8). Although AGI systems like ChatGPT are intended to be securely designed, they may not yet fully comply with privacy regulations like HIPAA (9). Consequently, patient data may be compromised, leading to the unauthorized access, harvesting and sharing of sensitive personal information. This risk highlights the importance of ensuring AGI technologies adhere to strict privacy standards before their widespread adoption in healthcare.

Powerful generative AGI could also be exploited to create fake documents or images, resulting in misinformation and negative consequences (10). For example, unscrupulous individuals or companies might use AGI-generated medical images (11) to support false claims or promote unproven treatments, thereby misleading patients and healthcare professionals.

Moreover, the integration of large language models into medical training could have a detrimental effect on the education of future healthcare professionals (12, 13). Students may rely too heavily on AGI-generated content, neglecting to develop the critical skills required for effective medical practice. By using AGI tools as shortcuts, they might not acquire the ability to differentiate between relevant and irrelevant information, which is crucial in the fast-paced and complex world of healthcare.

Furthermore, established AGI models may inadvertently perpetuate existing biases present in the data used for training (14). Consequently, the outcomes generated by these models might be biased, leading to the reinforcement of stereotypes and potentially causing harm to certain demographics (15). For instance, if a model's training data contains biased information about a specific ethnic group, the model could produce advice that is detrimental to patients from that group.

In short, while AGI models like ChatGPT hold the potential to revolutionize the healthcare sector, it is essential to recognize and address the potential risks they pose. Proper oversight and regulation are necessary to ensure the integration of AGI technologies into the healthcare system is both safe and effective. By addressing these risks, AGI can be harnessed to significantly improve patient care, medical knowledge, and healthcare processes, ultimately benefiting society as a whole.

We divide the ensuing discussion into 6 sections. Each section addresses specific subsets of potential harms and risks posed by large language models such as ChatGPT (see Figure 1 for an overview of this study). We aim to provide a comprehensive yet succinct summary of this subject to spur broader discussion and insights into the future of medicine in the era of AGI. In addition, in Section 7, we discuss the potential risks of AGI models (including ChatGPT) in the context of radiology.

2 Truth or dare? Inaccurate medical advice from Dr. ChatGPT

The potential for AGI models, such as ChatGPT, to provide inaccurate medical advice is a critical concern within the medical community (16, 17). As these systems generate content based on vast amounts of data (1, 2, 5), it is possible that the information provided could be misleading or incorrect (18). This raises concerns about the trustworthiness of AGI-generated medical advice, as well as the potential risks to patients who rely on such guidance. This discussion has been ongoing since (19) the introduction of smaller pre-trained models (such as BERT or T5 (20)) to healthcare NLP, but it has become a more prominent concern with the rise of LLMs.

For instance, consider a scenario where a patient consults ChatGPT for recommendations on managing their diabetes. ChatGPT might provide advice based on outdated guidelines, which could lead to inappropriate dietary recommendations or incorrect medication dosages. If the patient were to follow this advice, they might experience adverse effects, such as uncontrolled blood sugar levels or complications arising from improper treatment. For example, in 2018, the American College of Physicians (ACP) issued a new recommendation to control type 2 diabetes patients' A1C level to between 7% to 8% instead of a previous target of below 7%, since research indicated that reducing the A1C too low through medication did not prevent macrovascular complications yet might lead to substantial harms (21). If ChatGPT or similar models were trained on outdated standards, their responses might lack current medical validity.

Another example concerns the potential for ChatGPT to generate advice that does not consider individual patient factors. Clinical decision-making often requires a nuanced understanding of a patient's medical history, concurrent health conditions, and potential contraindications. An AGI model like ChatGPT, without direct access to a patient's medical records, may generate advice that is unsuitable or even harmful to the patient. For instance, a patient suffering from both hypertension and kidney disease may receive medication advice that is appropriate for managing hypertension but exacerbates their kidney condition (22), since ChatGPT is not aware of the full spectrum of the patient's problems.

Additionally, the rapidly evolving nature of medical knowledge presents challenges for AGI models. With new research findings continuously emerging, it is vital that AGI models are updated regularly to ensure that their advice remains aligned with the latest evidence-based guidelines. However, the lag between the publication of new research and its integration into AGI models may result in patients receiving advice that is no longer considered best practice. Indeed, both ChatGPT and GPT-4 were only trained on data up to September 2021 (23) and consequently have no knowledge of more recent developments.
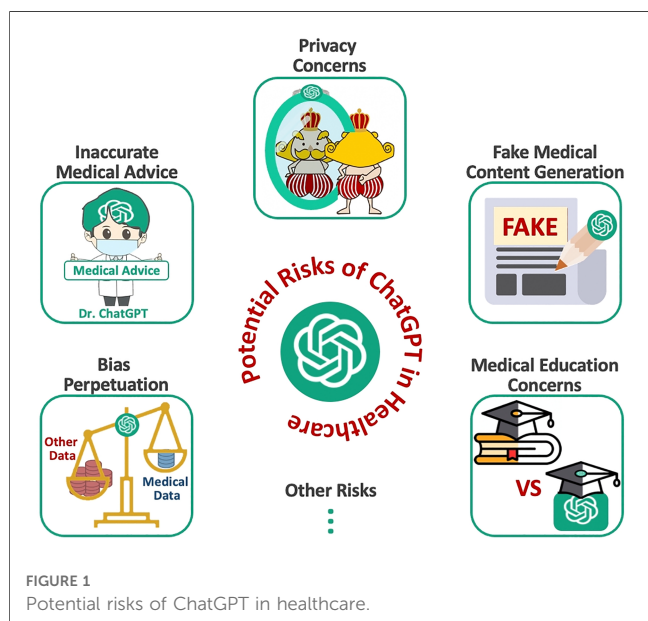

FIGURE 1
Potential risks of ChatGPT in healthcare.

In the realm of medical diagnostics, the potential for ChatGPT to misinterpret symptoms or overlook critical information could lead to diagnostic errors. For example, a patient may present with symptoms that align with multiple diagnoses. The model might generate advice based on the most common condition, while failing to consider a rare but more serious alternative diagnosis. Such oversight could have serious consequences for the patient, who may not receive the correct treatment in a timely manner.

Overall, there is a significant risk for large language models to produce ungrounded or unverified medical advice. It is necessary to raise awareness of this new challenge in the AGIlandscape. We encourage efforts to instill correct knowledge into models like ChatGPT or establish guardrails that moderate generated medical content.

3 The emperor's new clothes: privacy concerns

Large language models (LLMs) like ChatGPT offer impressive capabilities (24–28), but they also come with significant privacy implications that need to be carefully addressed. One particular risk is the potential for models to inadvertently leak details from the data they were trained on. While this is a concern for all LLMs, there are additional challenges if a model trained on private data (8) were to become publicly accessible.

Datasets used to train language models can be substantial (1, 29), often reaching hundreds of gigabytes, and they draw from various sources and domains (30–34). Consequently, even when trained on public data, these datasets can contain sensitive information, such as personally identifiable information (PII) including names, phone numbers, and addresses. This raises concerns that a model trained on such data could inadvertently expose private details in its output. It is crucial to identify and minimize the risks of such leaks and develop strategies to address these concerns with future models. This has long been a concern in applying language models to healthcare (35). Prior to the advent of LLMs, BERT-based models were typically combined with differential privacy training and federated learning strategies to better protect privacy in healthcare applications (36).

Privacy and data protection regulation compliance is another significant concern associated with LLMs. These models have the capacity to "memorize" personal information, putting it at risk of being discovered by other users or potential attackers (37, 38). This ability to retain and potentially reveal personal information calls for robust measures to ensure data privacy and prevent unauthorized access.

The use of LLMs in healthcare (25, 26) has also raised privacy and security concerns, particularly regarding sensitive medical information (39, 40). Clinical notes, encompassing physician consultations, nursing assessments, lab results, and more, are often stored in free-text formats that may include identifiable or confidential patient information. Unauthorized access to this information poses significant risks to patient confidentiality and privacy. Regulations such as the U.S. Health Insurance Portability and Accountability Act (HIPAA) mandate the removal of re-

identifying information from medical records before dissemination to preserve patient confidentiality (41, 42). Researchers are actively exploring ways to mitigate these concerns by employing data masking techniques to conceal sensitive data and prevent unauthorized access (8, 43).

In conclusion, while LLMs offer impressive capabilities, the associated privacy concerns must be diligently addressed. Minimizing data leaks, protecting against unauthorized access, and implementing privacy-preserving techniques are crucial steps toward ensuring the responsible and ethical use of LLMs in safeguarding patient privacy and maintaining data confidentiality.

4 All that glitters is not gold: fake medical content generation

Large language models have been increasingly used in fake medical content generation due to their ability to generate text that mimics (10) the writing style and language of medical professionals. These models are based on machine learning algorithms that have been trained on vast amounts of text data, which enables them to generate text that is both coherent and informative. Large language models can also be used to generate fake medical content on a wide range of topics, including diagnosis, treatment, and prognosis. This can be especially useful for medical writers and content creators who need to produce a high volume of content quickly and efficiently.

The use of LLMs for generating fake medical content has become more widespread due to their impressive ability to produce convincing and coherent text that mimics human writing (44). The potential for LLMs, such as ChatGPT, to generate fake medical content raises several ethical concerns, particularly regarding patient safety and informed consent (45, 46). The dissemination of fake medical content may lead to false diagnosis, inappropriate treatments, and further medical complications, causing harm to patients. Additionally, the spread of such fake medical content can have serious public health consequences, as it can fuel the promotion of unproven treatments or products that may be ineffective or even dangerous (16).

To address the risks associated with fake medical content generation using LLMs, several strategies have been proposed (10). One approach is to develop advanced algorithms and tools that can detect fake medical content and prevent its spread. These tools could leverage machine learning and natural language processing techniques to analyze and validate the authenticity of medical content. Another approach is to increase awareness and education among healthcare professionals and the public about the risks associated with fake medical content. By promoting critical thinking and media literacy, individuals can be better equipped to identify fake medical content and make informed decisions about their health.

While LLMs have the potential to transform the content generation of medical data, their misuse for generating fake medical content raises significant ethical concerns. The development of advanced tools to detect and prevent the spread of fake medical content, along with increased awareness and

education, can help mitigate the risks associated with the use of LLMs in medical content generation.

5 Veritas vos liberabit? AIGC knowledge compromises medical education

The potential misuse of ChatGPT and similar AGI models in medical education raises significant ethical concerns. It is plausible that medical students and trainees could employ these technologies to complete assignments unethically, misrepresenting their actual knowledge and skills through AGI-generated content (AIGC) (47, 48). Such a scenario would not only result in a diminished educational experience for these students but could also jeopardize patient care when they transition into professional practice.

Moreover, relying on AGI tools to generate content might undermine students' capacity for learning, problem-solving, and generalization from known examples (49, 50). The acquisition of these essential skills is paramount for medical professionals, who must be able to navigate complex situations and adapt to new challenges. If AGI-generated content becomes a crutch for students, they may fail to develop the critical thinking abilities necessary for successful medical practice.

By bypassing the rigorous process of learning and self-discovery, students risk hindering their cognitive growth and reducing their aptitude for medical problem-solving. In the long term, this could lead to a workforce of medical professionals ill-equipped to handle the intricacies of their field, ultimately compromising the quality of healthcare and public trust in the healthcare system.

From another perspective, ChatGPT and similar models offer unique opportunities to democratize medical education that is previously not accessible to the public or medical students in disadvantaged regions. Open medical education offers noticeable benefits in avoiding unwanted treatment (51, 52), making informed decisions (51), promoting effective patient self-management (53) and achieving better clinical outcomes (54). Indeed, ChatGPT enables any audience to quickly source medical information that is previously inaccessible (55), which has a significant social impact, especially for communities and individuals that benefit from the mass dissemination of medical knowledge.

It is crucial to establish guidelines and policies governing the use of AGI models in medical education to mitigate risks and ensure that their potential benefits are harnessed without sacrificing the integrity of the learning process.

6 One lie leads to another: bias perpetuation

ChatGPT is a general-purpose LLM that is not specialized for medical problems, even though it can perform better than many models specifically fine-tuned on medical knowledge (56). It is trained on large amounts of real Internet data, which might not be adequately representative of human diversity and could contain pre-existing bias. According to the latest AI Index Report issued by Stanford University, the probability and danger of bias will develop as the size and capabilities of large language models keep growing (57). Previous models, such as BERT, also attracted concerns over bias in their training data (58). Indeed, it is unavoidable for ChatGPT to contain inherent bias. The wide popularity of ChatGPT further exacerbates existing problems.

ChatGPT has shown bias against specific groups since its training data contains racial and sexist stereotypes (59). The issues of underrepresentation, overrepresentation, and misrepresentation in web data are likely introducing various kinds of bias into ChatGPT. The outputs containing bias typically manifest in nuanced representations, which makes it difficult to recognize and correct bias and toxicity (60–62).

Also, fine-tuning ChatGPT on historical medical data has the potential to introduce or exacerbate bias that already exists within the data. For example, clinical practice biases, such as under-testing of marginalized communities, can impact the underlying clinical data and introduce bias during future training (63).

In addition, implicit bias from healthcare professionals can manifest in clinical notes, including segments of diagnoses and treatment decisions (64). ChatGPT and GPT-4 might introduce these new biases into downstream applications if such notes are used for training.

OpenAI has released plugins for images and will certainly develop multimodal foundational models in the future (e.g., GPT-4 will have capabilities to process images in the near future). But fairness research indicates that the combination with additional information or modality may not necessarily improve performance and is likely to bring about new unfairness and bias (65). For example, in CLIP, a language-vision model, historical race and gender bias are reinforced (66).

Fostering research and development to detect, mitigate, and prevent bias, toxicity, and other undesirable behaviors in large language models like ChatGPT and GPT-4 are crucial for a responsible AI future. By actively pursuing these objectives, we can ensure that these powerful tools serve as inclusive, unbiased, and beneficial resources for users across diverse backgrounds. This pursuit not only safeguards the ethical foundations of AI but also greatly enhances its potential to positively impact society.

7 Potential risks of artificial general intelligence (AGI) models in radiology

Inaccurate medical advice: A substantial risk associated with AGI models like ChatGPT involves the provision of inaccurate medical advice, particularly in the nuanced field of radiology. For instance, when interpreting a radiology report indicating the presence of a small pulmonary nodule, the model might suggest watchful waiting based on outdated guidelines, overlooking recent research that indicates a higher malignancy risk requiring more active intervention.

Privacy Concerns: Privacy is a paramount concern in the use of AGI models. For example, patients seeking guidance for

understanding a radiology report could unwittingly disclose sensitive health data like a past diagnosis of breast cancer or a family history of genetic disorders. Potential security vulnerabilities in the wider AGI system could expose this sensitive data to misuse or unauthorized access.

Fake Medical Content Generation: The potential for AGI models to generate misleading or false medical content presents a significant risk. An individual with malevolent intent could misuse AGI tools to fabricate radiology reports, falsely indicating the presence or absence of a medical condition, such as fabricating a report showing a clean bill of health when the actual scan revealed lung nodules indicative of early-stage cancer.

Compromise of Medical Education: AGI models have the potential to unintentionally undermine the quality of medical education. For example, a medical trainee could become overly reliant on a multimodal AGI model for interpreting brain MRI scans for assignments. This over-reliance could deprive them of crucial learning experiences and inadvertently foster an environment of plagiarism and excessive dependence on automation. Consequently, this may lead to the undesired outcome of inadequately trained professionals tasked with handling complex or ambiguous clinical situations.

Bias Perpetuation: Lastly, AGI models can unintentionally propagate existing biases in medical data. For example, if the data used to train ChatGPT and GPT-4 over-represents Caucasian individuals, the models might be less adept at interpreting radiology reports concerning conditions more prevalent in other ethnic groups, such as the higher incidence of sarcoidosis in African American populations (67).

8 Conclusion

In conclusion, while the emergence of large language models such as ChatGPT offers promising prospects for revolutionizing healthcare, addressing the potential risks and challenges is of paramount importance. Future research should focus on developing robust methods to ensure the accuracy, reliability, and privacy compliance of medical contents generated by AGI, as well as monitoring and mitigating the biases that may be introduced during the training of these models. Additionally, guidelines and regulations should be established to govern the

use of AGI models in medical education, promoting their responsible use and preserving the integrity of the learning process.

By acknowledging and addressing these challenges, Artificial General Intelligence can be harnessed to revolutionize patient care and healthcare, ultimately benefiting society as a whole and significantly promoting national health. The interdisciplinary collaboration between AGI researchers, medical professionals, ethicists, and policy-makers will play a crucial role in shaping the future of medicine in the era of AGI, ensuring its safe and effective integration into healthcare systems worldwide.

Author contributions

Conceptualization: ZL, XL, WL, DS and TL. Methodology: ZL, XL and TL. Writing – original draft: ZL, LZ, XY, ZW, CC and HD. Writing – review and editing: NL, JL, XL, WL, DS and TL. Supervision: XL, WL and TL. All authors contributed to the article and approved the submitted version.

Conflict of interest

DS was employed by Shanghai United Imaging Intelligence Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors DZ, XL, WL, JL declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Liu Y, Han T, Ma S, Zhang J, Yang Y, Tian J, et al. Summary of ChatGPT/GPT-4 research, perspective towards the future of large language models. arXiv Preprint arXiv:2304.01852 (2023).

2. Zhao L, Zhang L, Wu Z, Chen Y, Dai H, Yu X, et al. When brain-inspired AI meets AGI. arXiv preprint arXiv:2303.15935 (2023).

3. Li J, Dada A, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy, systematic review. medRxiv 2023-03 (2023).

4. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, et al. Artificial intelligence, machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol*. (2018) 15(3):504–8. doi: 10.1016/j.jacr.2017.12.026

5. Liu Z, He M, Jiang Z, Wu Z, Dai H, Zhang L, et al. Survey on natural language processing in medical image analysis. *Zhong Nan Da Xue xue Bao. Yi Xue Ban*. (2022) 47(8):981–93. doi: 10.11817/j.issn.1672-7347.2022.220376

6. Sallam M. ChatGPT utility in healthcare education, research,, practice: systematic review on the promising perspectives, valid concerns. In: *Healthcare*. Vol. 11. MDPI (2023). p. 887.

7. Haupt CE, Marks M. AI-generated medical advice—GPT and beyond. *JAMA*. (2023) 329(16):1349–50. doi: 10.1001/jama.2023.5321

8. Liu Z, Yu X, Zhang L, Wu Z, Cao C, Dai H, et al. DeID-GPT: zero-shot medical text de-identification by GPT-4. arXiv preprint arXiv:2303.11032 (2023).

9. Vaishya R, Misra A, Vaish A. ChatGPT: is this version good for healthcare and research? *Diabetes Metab Syndr: Clin Res Rev*. (2023) 17(4):102744. doi: 10.1016/j.dsx.2023.102744

10. Liao W, Liu Z, Dai H, Xu S, Wu Z, Zhang Y, et al. Differentiate ChatGPT-generated and human-written medical texts. arXiv preprint arXiv:2304.11567 (2023).

11. Mangaokar N, Pu J, Bhattacharya P, Reddy CK, Viswanath B. Jekyll: attacking medical image diagnostics using deep generative models. In: *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE (2020). p. 139–57.

12. Thurzo A, Strunga M, Urban Ráta, Surovková J, Afrashtehfar KI. Impact of artificial intelligence on dental education: a review and guide for curriculum update. *Educ Sci.* (2023) 13(2):150. doi: 10.3390/educsci13020150

13. Hosseini M, Gao CA, Liebovitz DM, Carvalho AM, Ahmad FS, Luo Y, et al. An exploratory survey about using ChatGPT in education, healthcare, and research. medRxiv 2023-03 (2023).

14. Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z. ChatGPT and a new academic reality: artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *J Assoc Inf Sci Technol.* (2023):1–23. doi: 10.2139/ssrn.4389887

15. Ferrara E. Should ChatGPT be biased? Challenges, risks of bias in large language models. arXiv preprint arXiv:2304.03738 (2023).

16. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT, other large language models are double-edged swords. *Radiology.* (2023) 307(2):e230163. doi: 10.1148/radiol.230163

17. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. *J Esthet Restor Dent.* (2023) 35 (7):1098–102. doi: 10.1111/jerd.13046

18. Lee H. The rise of ChatGPT: exploring its potential in medical education. *Anat Sci Educ.* (2023):1–6. doi: 10.1002/ase.2270

19. Merine R, Purkayastha S. Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics. In: *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE (2022). p. 567–74.

20. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J Mach Learn Res.* (2020) 21(1):5485–551. doi: 10.5555/3455716.3455856

21. Smetana GW, Nathan DM, Dugdale DC, Burns RB. To what target hemoglobin A1c level would you treat this patient with type 2 diabetes? Grand rounds discussion from Beth Israel deaconess medical center. *Ann Intern Med.* (2019) 171(7):505–13. doi: 10.7326/M19-0946

22. Pugh D, Gallacher PJ, Dhaun N. Management of hypertension in chronic kidney disease. *Drugs.* (2019) 79:365–79. doi: 10.1007/s40265-019-1064-1

23. Lopez-Lira A, Tang Y. Can ChatGPT forecast stock price movements? return predictability and large language models. arXiv preprint arXiv:2304.07619 (2023).

24. Holmes J, Liu Z, Zhang L, Ding Y, Sio TT, McGee LA, et al. Evaluating large language models on a highly-specialized topic, radiation oncology physics. arXiv preprint arXiv:2304.01938 (2023).

25. Ma C, Wu Z, Wang J, Xu S, Wei Y, Liu Z, et al. ImpressionGPT: an iterative optimizing framework for radiology report summarization with ChatGPT. arXiv preprint arXiv:2304.08448 (2023).

26. Wu Z, Zhang L, Cao C, Yu X, Dai H, Ma C, et al. Exploring the trade-offs: Unified large language models vs local fine-tuned models for highly-specific radiology NLI task. arXiv preprint arXiv:2304.09138 (2023).

27. Zhong T, Wei Y, Yang L, Wu Z, Liu Z, Wei X, et al. ChatABL: abductive learning via natural language interaction with ChatGPT. arXiv preprint arXiv:2304.11107 (2023).

28. Dai H, Liu Z, Liao W, Huang X, Wu Z, Zhao L, et al. ChatAUG: leveraging ChatGPT for text data augmentation. arXiv preprint arXiv:2302.13007 (2023).

29. Liao W, Liu Z, Dai H, Wu Z, Zhang Y, Huang X, et al. Mask-guided bert for few shot text classification. arXiv preprint arXiv:2302.10447 (2023).

30. Rezayi S, Dai H, Liu Z, Wu Z, Hebbar A, Burns AH, et al. ClinicalRadioBERT: knowledge-infused few shot learning for clinical notes named entity recognition. In: *Machine Learning in Medical Imaging: 13th International Workshop, MLMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings.* Springer (2022). p. 269–78.

31. Rezayi S, Liu Z, Wu Z, Dhakal C, Ge B, Zhen C, et al. AgriBERT: knowledge-infused agricultural language models for matching food and nutrition. In: De Raedt L, editor. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22.* International Joint Conferences on Artificial Intelligence Organization (2022). p. 5150–6. AI for Good.

32. Gao L, Biderman S, Black S, Golding L, Hoppe T, Foster C, et al. The pile: an 800 gb dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027 (2020).

33. Cai H, Liao W, Liu Z, Huang X, Zhang Y, Ding S, et al. Coarse-to-fine knowledge graph domain adaptation based on distantly-supervised iterative training. arXiv preprint arXiv:2211.02849 (2022).

34. Liu Z, He X, Liu L, Liu T, Zhai X. Context matters: a strategy to pre-train language model for science education. arXiv preprint arXiv:2301.12031 (2023).

35. Basu P, Roy TS, Naidu R, Muftuoglu Z, Singh S, Mireshghallah F. Benchmarking differential privacy and federated learning for bert models. *arXiv* [preprint]. (2021). doi: 10.48550/arXiv.2106.13973

36. Basu P, Roy TS, Naidu R, Muftuoglu Z, Singh S, Mireshghallah F. Benchmarking differential privacy and federated learning for bert models. arXiv preprint arXiv:2106.13973 (2021).

37. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al. Extracting training data from large language models. *arXiv* [preprint]. (2021). doi: 10.48550/arXiv.2012.07805

38. Ye J, Maddi A, Murakonda SK, Bindschaedler V, Shokri R. Enhanced membership inference attacks against machine learning models. Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (2022). p. 3093–106. doi: 10.1145/3548606.3560675

39. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev: Comput Stat.* (2021) 13(6):e1549. doi: 10.1002/wics.1549

40. Urbain J, Kowalski G, Osinski K, Spaniol R, Liu M, Taylor B, et al. Natural language processing for enterprise-scale de-identification of protected health information in clinical notes. In: *AMIA Annual Symposium Proceedings.* Vol. 2022. American Medical Informatics Association (2022). p. 92.

41. HHS Office for Civil Rights. Standards for privacy of individually identifiable health information. Final rule. *Fed Regist.* (2002) 67(157):53181–273. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/standards-privacy-individually-identifiable-health-information/index.html

42. Ahmed T, Al Aziz MM, Mohammed N. De-identification of electronic health record using neural network. *Sci Rep.* (2020) 10(1):1–11. doi: 10.1038/s41598-020-75544-1

43. Sharma N, Anand A, Singh AK. Bio-signal data sharing security through watermarking: a technical survey. *Computing.* (2021) 103(9):1–35. doi: 10.1007/s00607-020-00881-y

44. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to fake it: limited responses, fabricated references provided by ChatGPT for medical questions. *Mayo Clin Proc Digital Health.* (2023) 1(3):226–34. doi: 10.1016/j.mcpdig.2023.05.004

45. Angelis LD, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE. ChatGPT, the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health.* (2023) 11:1166120. doi: 10.3389/fpubh.2023.1166120

46. Patel SB, Lam K, Liebrenz M. ChatGPT: friend or foe. *Lancet Digit Health.* (2023) 5:E102. doi: 10.1016/S2589-7500(23)00023-7

47. King MR, ChatGPT. A conversation on artificial intelligence, chatbots,, plagiarism in higher education. *Cell Mol Bioeng.* (2023) 16(1):1–2. doi: 10.1007/s12195-022-00754-8

48. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT-reshaping medical education and clinical management. *Pak J Med Sci.* (2023) 39(2):605. doi: 10.12669/pjms.39.2.7653

49. Tlili A, Shehata B, Adarkwah MA, Bozkurt A, Hickey DT, Huang R, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments.* (2023) 10(1):15. doi: 10.1186/s40561-023-00237-x

50. Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A swot analysis of ChatGPT: Implications for educational practice and research. *Innov Educ Teach Int.* (2023):1–15. doi: 10.1080/14703297.2023.2195846

51. Chan Y, Irish JC, Wood SJ, Rotstein LE, Brown DH, Gullane PJ, et al. Patient education and informed consent in head and neck surgery. *Arch Otolaryngol–Head Neck Surg.* (2002) 128(11):1269–74. doi: 10.1001/archotol.128.11.1269

52. Tannenbaum C, Martin P, Tamblyn R, Benedetti A, Ahmed S. Reduction of inappropriate benzodiazepine prescriptions among older adults through direct patient education: the empower cluster randomized trial. *JAMA Intern Med.* (2014) 174(6):890–8. doi: 10.1001/jamainternmed.2014.949

53. Osborne RH, Spinks JM, Wicks IP. Patient education and self-management programs in arthritis. *Med J Aust.* (2004) 180(5):S23. doi: 10.5694/j.1326-5377.2004.tb05909.x

54. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter social media is an effective tool for breast cancer patient education and support: patient-reported outcomes by survey. *J Med Internet Res.* (2015) 17(7):e188. doi: 10.2196/jmir.4721

55. Biswas S. ChatGPT and the future of medical writing. *Radiology.* (2023) 307(2):e223312–00. doi: 10.1148/radiol.223312

56. Nori H, King N, Mayer McKinney S, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv preprint arXiv:2303.13375 (2023).

57. Zhang D, Maslej N, Brynjolfsson E, Etchemendy J, Lyons T, Manyika J, et al. The ai index 2022 annual report. ai index steering committee. *Stanford Institute for Human-Centered AI* (2022).

58. Adam H, Balagopalan A, Alsentzer E, Christia F, Ghassemi M. Mitigating the impact of biased artificial intelligence in emergency decision-making. *Commun Med.* (2022) 2(1):149. doi: 10.1038/s43856-022-00214-4

59. Deshpande A, Murahari V, Rajpurohit T, Kalyan A, Narasimhan K. Toxicity in ChatGPT: analyzing persona-assigned language models. *arXiv* [preprint]. (2023). doi: 10.48550/arXiv.2304.05335

60. Sallam M. The utility of ChatGPT as an example of large language models in healthcare education, research and practice: Systematic review on the future perspectives and potential limitations. medRxiv 2023-02 (2023).

61. Lawrence CD. *Hidden in white sight: how AI empowers and deepens systemic racism*. New York, NY: CRC Press (2023).

62. Asch DA. An interview with ChatGPT about health care. *NEJM Catal Innov Care Deli*. (2023) 4(2). doi: 10.1056/CAT.23.0043

63. Wiens J, Price WN, Sjoding MW. Diagnosing bias in data-driven algorithms for healthcare. *Nat Med*. (2020) 26(1):25–6. doi: 10.1038/s41591-019-0726-6

64. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: a systematic review. *BMC Med Ethics*. (2017) 18(1):19. doi: 10.1186/s12910-017-0179-8

65. Cui C, Yang H, Wang Y, Zhao S, Asad Z, Coburn LA, et al. Deep multi-modal fusion of image and non-image data in disease diagnosis and prognosis: a review. *Prog Biomed Eng (Bristol)*. (2023) 5(2):10.1088/2516-1091/acc2fe. doi: 10.1088/2516-1091/acc2fe

66. Birhane A, Prabhu VU, Kahembwe E. Multimodal datasets: misogyny, pornography, and malignant stereotypes. *arXiv* [preprint]. (2021). doi: 10.48550/arXiv.2110.01963

67. Mirsaeidi M, Machado RF, Schraufnagel D, Sweiss NJ, Baughman RP. Racial difference in sarcoidosis mortality in the united states. *Chest*. (2015) 147(2):438–49. doi: 10.1378/chest.14-1120