Use of Artificial Intelligence in Medical Research: A SWOT Analysis

Jha JP, Amgain K, Joshi LR

Karnali Academy of Health Sciences,

Jumla, Karnali, Nepal.

Corresponding Author

Jay Prakash Jha

Department of Clinical Physiology,

Karnali Academy of Health Sciences,

Jumla, Karnali, Nepal.

E-mail: jay@kahs.edu.np

Citation

Jha JP, Amgain K, Joshi LR. Use of Artificial Intelligence in Medical Research: A SWOT Analysis. *Kathmandu Univ Med J*. 2024;88(4):464-70.

ABSTRACT

As artificial intelligence (AI) tools are shaping our work, this article discusses a nuanced SWOT analysis, focusing on the applications of artificial intelligence in the area of medical research. It aims to evaluate the applications of artificial intelligence tools in medical research, discussing their implications for researchers, journals and the scientific community, addressing the growing concerns of using artificial intelligence tools in research and publication, evaluating its potential risks while harnessing the transformative potential.

The analysis is complemented by a qualitative review of online resources, articles, blogs, interviews and podcasts, elucidating the prevailing themes in artificial intelligence-related considerations.

The strengths highlight artificial intelligence's capacity to accelerate research processes, particularly in diagnostics, drug production and data analysis. On the other hand, the weaknesses underscore concerns related to interpretability, biases, and ethical considerations, urging caution in artificial intelligence reliance. Opportunities arise in the form of explainable artificial intelligence, inclusive data practices, and enhanced model validation, while threats include issues of bias, privacy, overreliance and human exploitation. Such issues can be mitigated by collaboration from multiple experts and policymakers.

The current state of artificial intelligence raises concerns about data quality, bias, transparency and ethical issues in its development and deployment. There is a need for collaborative efforts to establish ethical frameworks, regulations, and sustainability practices. A balanced approach, positioning AI as a collaborator that enhances human insights and creativity is recommended.

KEY WORDS

Artificial intelligence, Bias, Ethics, Medical research

INTRODUCTION

Artificial Intelligence (AI) simulates human-like intelligence in computer systems through software-coded heuristics, empowered by its unparalleled capacity to analyze vast datasets and generate insights beyond human capabilities. Recent advancements, particularly in language-based tools like ChatGPT, Gemini and Bing have attracted widespread popularity, thus increasing the productivity across various sectors. However, the issues of potential biases, ethical implications, and intellectual property rights associated with AI-generated content raises concern about its use in research and other scientific fields. Besides, AI is also reshaping in the content creation and publication aspects; these impacts remain insufficiently addressed in our current context. Notably, organizations such as ICMJE, WAME, and COPE, along with some journals are recognizing the need for updated recommendations and policies.[1,2] This growing awareness calls for many authors and organizations to establish guidelines for the ethical use of AI in medical research.

This article explores the applications of AI tools in medical research, discussing their implications for researchers, journals and the scientific community. It specifically examines the strengths and weaknesses of AI in medicine, explores opportunities for improvement, and addresses its potential threats.

The research used a descriptive synthesis of contents from diverse sources such as relevant research articles, online

blogs, interviews and podcasts from the experts in the field of AI and/or medicine. A thematic approach has been utilized to analyze them to determine the main themes and concerns related to AI in terms of SWOT analysis, describing the inherent strengths and weakness of current AI tools, along with the best opportunity of utilizing them in advancing the medical research, and common threats that are possible in its use.

Strengths

It is evident that AI tools, with its multifaceted approach, can significantly accelerate healthcare and research. It is already being applied in various medical fields such as diagnosis, management, drug production, and education.[3] In healthcare, AI tools are being employed to analyse medical records, genetic data and imaging scans.[4] An AI tool can be customized to meet the specific needs of individual patients, giving effective personalized medical management. Moreover, it facilitates seamless integration of data from diverse sectors within and across institutes, enhancing the comprehensiveness of health services and data analysis. Lastly, AI tools have large potential to support diversity, equity and inclusion in the society as well as in workplace, promising an inclusive healthcare and research advancement.[5,6]

A recent pilot study showed that GPT4, the latest language model by OpenAI, can outperform humans in diagnosis of clinical case challenges.[7,8] Other apparent strengths of AI tools include efficiency, cost-effectiveness, predictability to future trends and benefits in telemedicine.[9,10] Given these obvious strengths of AI, we will not delve in detail in this article.

Weaknesses

Along with the rising promises of AI, we must also be aware of their inherent weaknesses that can affect the validity, fairness and ethical conduct of a research. Currently, the state of AI hype has far exceeded the state of AI science, commonly observed in areas like digital art, media creation, future prediction and patient care, where AI's potential benefits are overstated. Following are some of the inherent weakness of any of the currently available AI tools.

Lack of Interpretability: The current nature of AI mechanism is like a black-box which means that we are unable to explain how an AI system arrives at its decisions. This issue limits its interpretability while trying to validate results or make critical decisions based on AI-generated findings.

Bias and ethical concerns: AI tools can inherit the flaws from the data used to train them in its results. If the training data does not include people from diverse groups, the AI system can perpetuate or even amplify the inherent biases, affecting the accuracy and fairness of the research findings.[11-14] A recent review by Navarro et al. has uncovered that 87% of the machine learning-based prediction models have high risk of bias due to factors like small study size, poor handling of missing data and failure to deal with overfitting.[15]

Although there is rising trend in research on AI application in medicine, its ethical issues have largely been ignored. AI chatbots have been revealed to have serious risks and also has potential to generate non-original or scientifically inaccurate texts, which can lead to scientific misconduct.[16-21]

Lack of Creativity and accuracy: Overdependence on AI tools may hinder the user's creativity and contextual understanding which is important in research. They are best suited for pattern recognition and optimization tasks but might not excel in tasks that necessitate human-like reasoning and intuition. Research has shown that ChatGPT's statistical advice can be inaccurate and based on faulty assumptions, leading to poor reproducibility and potentially fabricated or falsified results.[22] Language tools like ChatGPT may provide short version of text, but not meaningful summaries, as they do not truly grasp key points, merely generating text based on statistical patterns. As noted by The New Yorker, ChatGPT functions like a "blurry JPEG of the web," replicating information without fully comprehending it.[23] Such limitations highlight the need for human oversight in tasks requiring in-depth analysis and interpretation, such as in medical applications. This over-reliance not only impacts the quality of decision-making but also raises concerns about the potential displacement of human roles, which is further explored in the next section.

Opportunities

There are some methods we can implement to mitigate above weaknesses of AI and try to preserve its benefits. Some are discussed below.

Explainable AI (XAI): XAI is a set of tools that elucidate machine learning model predictions, aiming to make AI more transparent and understandable to humans.[24,25] This makes us possible to retain intellectual oversight over AI systems despite it being kind of "black-box". As AI is being used in critical operations like healthcare and finances, XAI is gaining attention in the research.

Data inclusion and model validation: As the data used to train the AI model is the basis of its performance, diverse and inclusive training data is important and can be used to address the AI related issues like bias and lack of generalizability. Historically, the volume of data was thought to be the best factor affecting the performance of AI model.[26] However, newer researches in MIT have consistently shown that the quality of data is equally, if not more, important than the quantity of data, and that inclusive data sources can enhance its quality.[27] For this, collaborative efforts, data-sharing initiatives and guidelines for inclusive data collection seem pertinent. The initiatives by International training center of international labour organization and World economic forum aim to promote

equity in AI to develop human-centered AI tools for shared benefits to all.[28,29]

Open-source models allow researchers to share their code and models with others, which can help collaboration and cross-validity of the models, thus improving their quality.[30] WHO has an open bank of anonymized global health data in Global Health Observatory.[31] Another example is US National Cancer Institute Genomic Data Commons which provides assess to genomic and clinical data.[32]

Enhancing ethical frameworks: Establishing ethical frameworks for AI involves collaboration between researchers, policymakers and regulatory bodies. Developing clear guidelines, informed consent processes and stringent data protection measures are essential. For example, European Union has released framework for the ethical design and development of AI, establishing seven key requirements for a trustworthy AI system including human agency, technical robustness, and privacy protection.[33,34] Similarly, UNESCO's global standard on AI ethics, called the Recommendation on the Ethics of Artificial Intelligence, is adopted by all 193 member states in November 2021.[35] These frameworks from the organizations represent global efforts that reflect collective commitment to ensuring the ethical design, development and deployment of AI technologies worldwide.

Environmental Sustainability: To mitigate the environmental impact of AI, several initiatives have been introduced under the concept of "Green AI." These efforts aim to reduce the energy consumption and carbon footprint associated with AI systems. Google has successfully used machine learning to optimize data center cooling, resulting in a 40% reduction in energy usage.[36] With appropriate engineering, prompt tuning and model fine-tuning, the model can optimize hardware usage, reducing the carbon footprint in adapting foundation models for certain tasks. Strategies such as quantization of numeric data (reducing the precision of numbers in a model), distillation (compressing a large model into a smaller one), client-side caching (storing some data on the user's device); along with using specialized hardware (like in-memory computing and analogue computing) are techniques used to make AI models more efficient.[37]

Threats

While the benefits of AI tools in research are undeniable, their adoption also poses challenges and real life consequences. Let's explore external challenges that could practically impact the utilization of AI in the medical field.

Data Subject - amplification of ethical issues: Traditionally, a researcher would collect relevant data from human subject by visiting them. Taking consent and addressing ethical aspects were easy. In today's era of data processing, our personal and professional data is decoupled from human subject, as they are easily accessible from various sources. This shift from "human subject" to "data subject" is a major source of the ethical problems, especially in relation to AI. Apart from the obvious unethical practices and privacy invasion, this changing dynamics of research and work amplifies the bias related to AI implementation.

The current technological development in low income countries like Nepal is much lower than in high income countries. This skewness becomes significant in AI development. More than half of the datasets used for clinical AI originate from either the US or China, and majority of the authors are male.[38] Health contexts of high-income countries (HICs) are significantly different from our context, which could lead to the unreflective application of AI systems.[38] Addressing these concerns is crucial not only for ethical reasons but also to mitigate the threat in the reliability of AI systems.

Lack of Standardization: The rapid growth of AI tools has outpaced the development of comprehensive regulations and standards. This deficiency may impede the seamless integration of AI into medical research. On the other hand, in the absence of effective regulation, the utility of research findings and conclusions generated by unregulated AI systems in the medical domain may be compromised. To balance these polar issues, time and resources need to be invested in such implementation processes, with collaboration among healthcare sectors, regulatory bodies and other stakeholders to establish effective standards, a notion shared by other publishers also.[39]

Pressure to Publish: The relentless pressure on researchers to publish papers has created a breeding ground for unethical practices within the academic community, and language-based AI models can be severely utilized to fabricate any article. This "publish or perish" phenomenon has led to a surge in the production of substandard papers (called paper mills), fuelled by the fear of falling behind in the competitive academia and to safeguard their academic standing.[19,40-42] Such practice raises a significant threat to the integrity of scientific research; it not only compromises the quality of research, but also undermines the credibility of scholarly publications. Eventually, this erodes the already fragile public trust in the scientific community.[43]

Overreliance on AI: As discussed above, overreliance on AI is its weakness; which can pose a distinct threat to the research process. Medical research is not just about crunching the numbers among the complex health data, it is about the lives and stories of real people. Recent meta-analysis on diagnosis of deep learning models and health-care professionals indicates that AI-based diagnosis is no more accurate than healthcare professionals, as they still report algorithmic suggestions even when they're incorrect.[44] This limits the depth of understanding provided by human experts-creativity, intuition, contextually relevant question and decision.

Malicious Use: Internet security faces a new horizon of challenge with the advent of AI. It is of public concern

and not specific to health. According to a recent US-based report, the four common types of cyber attacks are data poisoning, evasion, abuse and privacy attack.[45] As a practical example of its abuse, an attacker can use machine learning to understand a victim's personal life by looking into their social media activity. This leads to identity theft, loss of reputation or finances, or other serious consequences.

Data poisoning is a novel technique of tampering with the training data used in machine learning models such that the model misreads or fails to understand the data. Training on the corrupt training data leads to security vulnerabilities and backdoor exploitation. This has been tested in connectome-based predictive model in neuroimaging where a poisoned portion of training data encourages misclassification of specific data subset. The authors suggest that the data used to train the AI del needs to be carefully evaluated and that defences need to be implemented against such data poisoning.[46]

Another way of using AI in cybercrime is two-faced model, also called deceptive or dual-use model. A preprint research explains that such models seem helpful and truthful during training and testing, but behaves differently once deployed, such as generating response that are intentionally misleading or harmful. Attempts to detect such behaviour can even make the models better at hiding their true nature. An example is deepfake medical image which are manipulated to generate real-looking, say, chest X-ray of patient to fake a serious condition.[47]

Human exploitation: As the tech industry rushes to make more lucrative products, they tend to fall in the business of using human labour resources to their products, shrouding the back story in darkness. Scholars Kate Crawford and Krystal Kauffman have shed light in the less visible aspects of AI development; human labour from South Asia and Africa is extensively involved in AI content moderation and data labelling.[48,49] This hidden workforce augments underpayment and grapples them with mental health issues due to the nature and volume of their work. Richard Mathenge, a former content moderator in Kenya, shares in Mozilla's IRL podcast the vivid psychological trauma induced by explicit content labelling for ChatGPT.[50]

These researches argue that the artificial intelligence is neither artificial nor intelligent, but rather the compound product of natural resources exploitation and intense human labour kept under the hood. They also highlight the regressive stereotypes that are baked into AI algorithms and can perpetuate existing social inequalities. These are serious issues in current society before addressing which we cannot seamlessly integrate AI into any production system.

Future direction

Looking ahead, the incorporation of AI into medical research has multiple challenges and needs a thoughtful and systematic approach. We feel that the topic of AI in general is too vast for an article and has multiple questions; which needs an open discussion involving different sectors. At present, a cautious optimistic approach would probably be the best strategy for its introduction.

Global perspective: As we explore the potential of this technology, we must prioritize ethical considerations and human rights, placing the security of sensitive medical data at the forefront. Putting ethics and human rights at the heart, WHO has published in 2021 its 6 principles for the design and use of AI: human autonomy, public safety and interest, transparency, responsibility, equity, and promoting a responsive and sustainable AI.[51]

As already explained, training of AI models in datasets as diverse as the population itself is important. This requires a holistic approach from technical, regulatory, economic and privacy standpoint involving researchers, policymakers and healthcare professionals. This insight has also been reflected by various experts and authors.[14,52,53] For example, The Physiological Society has explained the the need of including physiologists in all aspects of development and implementation of AI-based healthcare system, including its regulation.[13] Policymakers need to safeguard the ethical and beneficial integration of new technologies, ensuring universal access. We believe that rather than ignoring or blocking the use of AI, breaking down the barriers to AI utilization and promoting diversity in AI will lead to its democratization.

Individual perspective: At the user level, we need to balance the comfort of AI automation with the wisdom of human insights. AI should be seen as a collaborator, enhancing human creativity and insights. Responsibility for AI use rests with the user and an AI tool at present cannot be given responsibility of an author, as expressed by many publishers and journals.[54-56] Editors and publishers need to give clear guidelines about transparent reporting and declarations of AI tool usage by the researchers.

Nepal's Perspective: LMIC like Nepal faces unique challenges when it comes to adopting digital technologies due to its rugged geography, limited infrastructure, and relatively low internet penetration.[57,58] Despite this, an extensive use of language models is a common observation among Nepali researchers and clinicians. Currently, our effort focuses mainly in identifying the best use case for these novel tools in our setting. Firstly we can focus on capacity-building and education, particularly in the areas of data science, AI, and digital health. This can be started by training healthcare professionals and researchers on the use of technologies, as well as including them in curricula for students in these fields. We should also consider partnerships between universities, companies and government agencies to develop and test any AI-powered tools that we commonly use.

For implementation of AI system, LMIC should ideally have their own national dataset. For this, they can initially

focus on enhancing capacity-building and education in data science, AI and digital health. This involves training healthcare professionals and researchers in these areas and integrating it into various courses, including medical education. Additionally, partnerships between universities, companies and government agencies can facilitate the development and testing of AI-powered tools in Nepalese population. We can use cost-effective and transparent tools like open source AI frameworks (eg. TensorFlow, OpenCV and Weka); and tailor it to Nepalese health data, ensuring diversity and representation. It also needs to be efficient and considerate of the limited resources. We need to invest resources and research projects for innovative business models to reach remote communities, along with regular training to the health-workers. And finally algorithms need to be audited to identify and rectify biases. While the ideal scenario involves Nepal having its own dataset, practical constraints may necessitate accessing international expertise and technology through partnerships.

Keeping the WHO's six principles of AI use at heart, we need to publish a code of ethics for its use in healthcare and research in Nepal. Further, we can also build a platform among technology experts, policymakers and health researchers to establish an integrated community, which can be a hub for collaboration, sharing of knowledge and resources.

Implementation of AI in our setting can cover broad areas of healthcare, ranging from telemedicine in rural setting to policymaking in the center. Predictive analysis can predict the likelihood of disease outbreaks and to identify high-risk populations, which is useful to target public health interventions and improve disease surveillance. AI algorithm can also be used to manage the limited healthcare resources; for example by predicting patient flow to allocate resources accordingly. Throughout these developments, we must prioritize our citizens' privacy, ensuring that sensitive health data is not shared openly. It is now more important than ever to take action on these issues.

CONCLUSION

This study aimed to explore the current state of AI in medicine and identify key themes and concerns related to its implementation. Our qualitative analysis revealed that along with the promises in improving patient care and research, there are several challenges and limitations that need to be addressed before implementing AI.

The hype surrounding AI is likely to have exceeded its scientific capabilities, particularly in areas such as future prediction and patient care. Importantly, it raises concerns about data quality, bias, and transparency in AI development and deployment. These issues need to be addressed through rigorous testing, validation, and regulation. Thirdly, we identified the need for more interdisciplinary collaboration among various fields, including but not limited to medicine, computer science and social sciences. Finally, we advocate for the importance of ethical implications in AI system implementation in healthcare so as to ensure the patient autonomy, confidentiality and to avoid human biases.

Dwelling in the uncharted terrain of artificial intelligence, let us not merely adapt to change but actively shape the future—forging a path where ethics, collaboration and responsibility become the keystones of transformative progress, ensuring that AI complements and enhances – not replaces – human capabilities.

ACKNOWLEDGEMENT

Authors would like to thank Dr. Bishal Joshi for his valuable suggestion and feedback in the manuscript.

REFERENCES

1. Bhosale U, Nair A. Global standards for ethical peer review practices [Internet]. 2023 [cited 2024 Sept 19]. Available from: https://www.enago.com/academy/guidelines-for-ai-use-in-publications/
2. Alfonso F, Crea F. New recommendations of the International Committee of Medical Journal Editors: use of artificial intelligence. *Eur Heart J.* 2023 Aug 14;44(31):2888-2890. doi: 10.1093/eurheartj/ehad448. PMID: 37453047.
3. Liu PR, Lu L, Zhang JY, Huo TT, Liu SX, Ye ZW. Application of Artificial Intelligence in Medicine: An Overview. *Curr Med Sci*. 2021 Dec;41(6):1105-1115. doi: 10.1007/s11596-021-2474-3. Epub 2021 Dec 6. PMID: 34874486; PMCID: PMC8648557.
4. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* 2019 Nov 19;11(1):70. doi: 10.1186/s13073-019-0689-8. PMID: 31744524; PMCID: PMC6865045.
5. Valencia S, Cave R, Kallarackal K, Seaver K, Terry M, Kane SK. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. InProceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Apr 19 (pp. 1-14). Available from https://research.google/pubs/the-less-i-t
6. Chakraborty N, Mishra Y, Bhattacharya R, Bhattacharya B. Artificial Intelligence: The road ahead for the accessibility of persons with Disability. *Materials Today: Proceedings.* 2023 Jan 1;80:3757-61. DOI: 10.1016/j.matpr.2021.07.374. Available from https://www.sciencedirect
7. Beam AL, Drazen JM, Kohane IS, Leong TY, Manrai AK, Rubin EJ. Artificial Intelligence in Medicine. *N Engl J Med.* 2023 Mar 30;388(13):1220-1. doi: 10.1056/NEJMe2206291. PMID: 36988598.
8. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. *NEJM AI*. 2023 Dec 11;1(1):AIp2300031. DOI: 10.1056/AIp2300031
9. National Institute of Health. Artificial Intelligence and medical research [Internet]. U.S. Department of Health and Human Services; 2023 [cited 2024 Sept 19]. Available from: https://newsinhealth.nih.gov/2023/01/artificial-intelligence-medical-research
10. Open Access Government. What are the pros and cons of implementing AI in Healthcare? [Internet]. 2022 [cited 2024 Sept 19]. Available from: https://www.openaccessgovernment.org/what-are-the-pros-and-cons-of-implementing-ai-

11. Keeling G. Algorithmic Bias, Generalist Models, and Clinical Medicine. *arXiv preprint arXiv:*2305.04008. 2023 May 6. Available from https://arxiv.org/pdf/2305.04008
12. Ferrara E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources. Impacts, And Mitigation Strategies. *arXiv*. 2023. Available from https://arxiv.org/pdf/2304.07683.
13. Mackenzie A. From 'Black Box' to Trusted Healthcare Tools - Physiology's role in unlocking the potential of AI for health. The Physiological Society. 2023 [cited 2024 Sept 19]. Available from https://www.physoc.org/policy/public-health-an
14. Kaushal A, Altman R, Langlotz C. Health care AI systems are biased. Scientific American. New York: Scientific American; 2020 Nov 17 [cited 2023 Sept 19]. Available from https://www.scientificamerican.com/article/health-care-ai-systems-are-biased/
15. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ.* 2021 Oct 20;375:n2281. doi: 10.1136/bmj.n2281. PMID: 34670780; PMCID: PMC8527348.
16. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. *BMC Med Inform Decis Mak*. 2020 Jul 22;20(1):170. doi: 10.1186/s12911-020-01191-1. PMID: 32698869; PMCID: PMC7376886.
17. Khlaif ZN, Mousa A, Hattab MK, Itmazi J, Hassan AA, Sanmugam M, et al. The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. *JMIR Med Educ.* 2023 Sep 14;9:e47049. doi: 10.2196/47049. PMID: 37707884; PMCID: PMC10636627.
18. Heikkilä M. Three ways AI Chatbots are a security disaster [Internet]. MIT Technology Review; 2023 [cited 2024 Sept 19]. Available from: https://www.technologyreview.com/2023/04/03/1070893/three-ways-ai-chatbots-are-a-security-disaster
19. Piasecki J, Walkiewicz-Żarek E, Figas-Skrzypulec J, Kordecka A, Dranseika V. Ethical issues in biomedical research using electronic health records: a systematic review. *Med Health Care Philos*. 2021 Dec;24(4):633-658. doi: 10.1007/s11019-021-10031-6. Epub 2021 Jun 19. PMID: 34146228; PMCID: PMC8214390.
20. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics.* 2021 Feb 15;22(1):14. doi: 10.1186/s12910-021-00577-8. PMID: 33588803; PMCID: PMC7885243.
21. Magubane N. The hidden costs of AI: Impending energy and resource strain [Internet]. University of Pennsylvania; 2023 [cited 2024 Sept 19]. Available from: https://penntoday.upenn.edu/news/hidden-costs-ai-impending-energy-and-resource-strain
22. Ordak M. ChatGPT's Skills in Statistical Analysis Using the Example of Allergology: Do We Have Reason for Concern? *Healthcare (Basel).* 2023 Sep 15;11(18):2554. doi: 10.3390/healthcare11182554. PMID: 37761751; PMCID: PMC10530997.
23. Chiang T. ChatGPT is a blurry JPEG of the web. The New Yorker [Internet]. 2023 Feb 9 [cited 2024 Sep 19]; Available from: https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web
24. Rai A. Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science. 2020 Jan;48:137-41. DOI: 10.1007/s11747-019-00710-5.
25. Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. *Harvard Data Science Review.* 2019 Nov 22;1(2):1-9. DOI: 10.1162/99608f92.5a8a3a3d Available at https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/8
26. Ashley IN, Westreich DJ. "Big data: a revolution that will transform how we live, work, and think." *Am J Epidemiol.* 179.9 (2014): 1143-4.
27. Massachusetts Institute of Technology. Can machine-learning models overcome biased datasets? *Science Daily*. 2022 Feb 21. Available from https://www.sciencedaily.com/releases/2022/02/220221115403.htm
28. International Training Centre of the International Labour Organization. Power in inclusive artificial intelligence training. [Internet]. International Training Centre of the International Labour Organization. 19 April 2021 [cited 2024 Sept 19]. Available from https://www.itcilo.org/stories/power-inclusive-artificial-intelligence-training
29. World Economic Forum. A blueprint for equity and inclusion in artificial intelligence. [Internet]. World Economic Forum. 2022 June 29 [cited 2024 Sept 19]. Available from https://www.weforum.org/publications/a-bluepri
30. Jha JP. Evaluating the Potentials of Open Source in Medical Field: Are We Ready to Embrace it? Potential of Open Source. *JKAHS*. 2023 Dec 31 [cited 2024 Sept 19];6(3):1-2. Available from https://jkahs.org.np/jkahs/index.php/jkahs/article/view/880
31. World Health Organization. Global Health Observatory (GHO). [Internet]. World Health Organization; [cited 2024 Sept 19]. Available from https://www.who.int/data/gho/
32. National Cancer Institute. Genomic Data Commons (GDC). [Internet]. National Cancer Institute; [cited 2024 Sept 19]. Available from https://portal.gdc.cancer.gov/
33. European Commission. Ethics guidelines for trustworthy AI. [Internet]. European Commission. 2019 April 8 [cited 2024 Sept 19]. Available from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
34. European Parliament. Artificial intelligence: An ethical and legal analysis. [Internet]. European Parliament; 2019. Available from https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf
35. UNESCO. Recommendation on the Ethics of Artificial Intelligence. [PDF] Paris: UNESCO; 2022. [cited 2024 Sept 19] Available from: https://unesdoc.unesco.org/ark:/48223/pf0000381137
36. Evans R, Gao J. DeepMind AI reduces energy used for cooling Google data centers by 40%. Google Blog [Internet]. 2016 Jul 20 [cited 2024 Sep 19]; Available from: https://blog.google/outre HYPERLINK "https://blog.google/outreach-initiatives/environment/deepmind-ai-reduces-energy-used-for/"ach-initiatives/environment/deepmind-ai-reduces-energy-used-for/
37. Banipal IS, Mazumder S. How to make AI sustainable. Nature [Internet]. 2024 Feb 29 [cited 2024 Sep 19]; Available https://doi.org/10.1038/d44151-024-00024-8".org/10.1038/d44151-024-00024-8
38. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. *PLOS Digital Health*. 2022 Mar 31;1(3):e0000022. DOI: 10.1371/journal.pdig.0000022 PMC9931338
39. Weissglass DE. Contextual bias, the democratization of healthcare, and medical artificial intelligence in low-and middle-income countries. *Bioethics.* 2022 Feb;36(2):201-9. DOI: 10.1111/bioe.12927, PubMed
40. Rawat S, Meena S. Publish or perish: Where are we heading? J Res Med Sci. 2014 Feb;19(2):87-9. PMID: 24778659; PMCID: PMC3999612.
41. COPE. Paper mills research. COPE: Committee on Publication Ethics [Internet]. 2022 Jun [cited 2024 Sep 19]; Available from: https://publicationethics.org/node/55256. DOI: https://doi.org/10.24318/jtbG8IHL
42. Busa I. 'Publish or perish' versus scientific collaboration in medical research. *The Oxford Scientist* [Internet]. 2022 Sep 6 [cited 2024 Sep 19]; Available from: https://oxsci.org/publish-or-perish/
43. Sutter PM. Rescuing Science: Restoring Trust in a Time of Doubt. New York: Prometheus Books; 2023. ISBN-13: 978-1538181614

44. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019 Oct;1(6):e271-e297. doi: 10.1016/S2589-7500(19)30123-2. Epub 2019 Sep 25. Erratum in: Lancet Digit Health. 2019 Nov;1(7):e334. doi: 10.1016/S2589-7500(19)30160-8. PMID: 33323251.

45. Vassilev A, Oprea A, Fordyce A, Anderson H. Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations. [PDF] Paris: NIST; 2023 Jan. [cited 2024 Sept 19] DOI: 10.6028/NIST.AI.100-2e2023 Available from: https://csrc.nist.gov/pubs/ai/100/2/e2023/final

46. Rosenblatt M, Scheinost D. Data Poisoning Attack and Defenses in Connectome-Based Predictive Models. In: Baxter, J.S.H., et al. Ethical and Philosophical Issues in Medical Imaging, Multimodal Learning and Fusion Across Scales for Clinical Decision Support, and Topological Data Analysis for Biomedical Imaging. EPIMI ML-CDS TDA4BiomedicalImaging. pp. 3-13. 2022 Sep 18. Springer, Cham. DOI: "https://doi.org/10.1007/978-3-031-23223-7_1"10.1007/978-3-031-23223-7_1

47. Hubinger E, Denison C, Mu J, Lambert M, Tong M, MacDiarmid M, et al. Sleeper Agents: Training Deceptive LLMs that Persist Through Safety Training. *arXiv preprint*, arXiv:2401.05566. 2024 Jan 10 [cited 2024 Sept 19] "https://doi.org/10.48550/arXiv.2401.05566"10.48550/arXiv.2401.05566

48. Zoe Corbyn. Microsoft's Kate Crawford: 'AI is neither artificial nor intelligent'. [internet]. 2021 June 6 [cited 2024 Sept 19]. Available from https://www.theguardian.com/technology/2021/jun/06/microsofts-kate-crawford-ai-is-neither-artificial-nor-intelligent.

49. Xiang C. AI Isn't Artificial or Intelligent. [Internet]. Vice; 2022 Dec 6 [cited 2024 Sept 19]. Available from: https://www.vice.com/en/article/wxnaqz/ai-isnt-artificial-or "https://www.vice.com/en/article/wxnaqz/ai-isnt-artificial-or-intelligent"-intelligent

50. Mozilla. The Humans in the Machine. [Internet]. IRL Podcast. 2022. Season 7, Episode 2. [cited 2024 Sept 19]. Available from: https://irlpodcast.org/season7/episode2/

51. World Health Organization. WHO issues first global report on AI in health and six guiding principles for its design and use. Geneva: WHO; 2021 [cited 2024 Sept 19]. Available from: https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use

52. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med*. 2023 Jun 14;6(1):113. doi: 10.1038/s41746-023-00858-z. PMID: 37311802; PMCID: PMC10264403.

53. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. *Patterns (N Y)*. 2021 Oct 8;2(10):100347. doi: 10.1016/j.patter.2021.100347. PMID: 34693373; PMCID: PMC8515002.

54. Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. *JAMA*. 2023;329(8):637-9. DOI:10.1001/jama.2023.1344 HYPERLINK "https://pubmed.ncbi.nlm.nih.gov/36719674/"PubMed

55. Springer Nature. European Review [Internet]. Heidelberg; c2013 [cited 2024 Sept 19]. Available from: https://link.springer.com/journal/12144/submission-guide

56. Ethical and responsible use of generative AI in scholarly research. Medium [Internet]. 2023 Jan 24 [cited 2024 Sept 19]. Available from: https://drqwrites.medium.com/ethical-and-responsible-use-of-generative-ai-in-scholarly-research-a96b7e3cf4f

57. Parajuli R, Bohara D, KC M, Shanmuganathan S, Mistry SK, Yadav UN. Challenges and opportunities for implementing digital health interventions in Nepal: A rapid review. *Front Digit Health*. 2022 Aug 25;4:861019. doi: 10.3389/fdgth.2022.861019. PMID: 36120714; PMCID: PMC9480345.

58. Dahal D, Sharma P. Exploring the Challenges and Opportunities of Implementing Artificial Intelligence in Healthcare Settings in Nepal: a literature review. *Kathmandu Univ Med J*. 2023;84(4):436-43. Available at https://kumj.com.np/issue/84/436-443.pdf.