RESEARCH



Artificial intelligence integration in healthcare: perspectives and trends in a survey of U.S. health system leaders



Shan Guleria¹, Janet Guptill², Ishmeet Kumar², Mia McClintic³ and Juan C. Rojas^{1,4*}

Abstract

Background The healthcare sector is rapidly integrating artificial intelligence-derived predictive models (AIDPM) to enhance clinical decision support, operational efficiency, and patient experiences. However, research on management strategies for AIDPM acquisition, deployment, and governance remains limited. This study examines changes in AIDPM integration and governance since 2021, with a particular focus on large language models and health equity considerations.

Results Our survey of health system leaders achieved a 49% response rate (32/65). While 84% of institutions reported using AIDPM in clinical practice, only 53% had established dedicated teams for these models. Compared to 2021, there was a significant increase in representation from experts in clinical informatics, operations, and quality improvement on AIDPM teams. Most organizations (41%) primarily purchased AIDPM from external vendors. Support for integrating large language models into healthcare practices was unanimous among respondents. The principal obstacles to AIDPM adoption included regulatory concerns, data security, workflow integration, and clinician acceptance. A large majority (72%) of respondents supported government regulation of AI in healthcare. While 76% of organizations reported having a team member dedicated to health equity, ethicists and diversity leaders were underrepresented on AIDPM teams (18%). Organizations reported various efforts to promote health equity, but involvement of frontline clinicians in AIDPM development and its impact on health equity was significantly less common.

Conclusions Clinical adoption of AIDPM faces challenges due to the absence of established best practices. Health system leaders strongly support federal regulations for AI in healthcare. These regulations could provide quality and safety standards. The study highlights the need for developing evaluation guidelines, especially for large language models. It also reveals a lack of uniform involvement of frontline clinicians and equity experts in AIDPM governance. Their involvement could increase adoption and trust of these new AI tools. Future research should assess healthcare systems' adherence to emerging regulations and best practice frameworks. This research should emphasize patient safety and health equity. These findings underscore the urgent need for a comprehensive roadmap. This roadmap would guide the responsible implementation of AIDPM in healthcare settings.

*Correspondence: Juan C. Rojas juan_rojas@rush.edu Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Keywords Prediction models, Artificial intelligence, AI implementation, Machine learning, Large language models, Health systems

Background

The integration and analysis of physical, economic, and behavioral data available has been termed the "Fourth Industrial Revolution" by Klaus Schwab in 2015 [1]. Since then, many economic sectors have made progress in data utilization [2]. In healthcare, there is a growing effort to provide better-informed care using artificial intelligencederived predictive models (AIDPMs). These tools combine various AI techniques with historical healthcare data to make predictions about risk and support multiple aspects of healthcare delivery. AIDPMs are being employed for a multitude of tasks, including enhancing clinical decision support by providing risk assessments and treatment recommendations, improving operational efficiency in healthcare facilities, optimizing patient experiences and engagement throughout their healthcare journey, and guiding clinician-patient interactions to potentially reduce burnout [3–5]. Each of these applications has the potential to advance healthcare delivery and outcomes. Despite this potential, there is a lack of research into the management strategies healthcare organizations employ for the acquisition, deployment, and oversight of AIDPM, as well as the governance practices that ensure their transparent and equitable use [6]. Since our initial study in 2021, the healthcare landscape has shifted significantly in the post-COVID pandemic recovery phase, with evolving technologies reshaping priorities for AI adoption in healthcare. This study seeks to survey these changes and ascertain current trends in the integration and governance of AIDPM. It aims to build upon our previous work, examining how these systems are managed within healthcare organizations and the broader implications for clinical practice [7].

Since our last study, the use of AIDPM in healthcare has become more widespread, and in the intervening years, two significant factors have emerged that warrant closer examination. First, despite the aforementioned benefits of AIDPM, there are growing concerns that these tools may negatively impact health equity, the pursuit of which is defined by the World Health Organization as "giving special attention to the needs of those at greatest risk of poor health, based on social conditions" [8]. Among these concerns are the risk of AI-based predictions being less accurate for minority groups and the risk that these predictions could exacerbate existing disparities or create new ones [9-13]. Given potential impacts on health disparities, there is increasing recognition of the need to identify sources of bias and strategies to mitigate it at various stages of model development and implementation [14]. While there have been many such recommendations for healthcare organizations involved in developing internal models and acquiring vendorsourced solutions [15], it is less clear to what extent these strategies are being put into clinical and administrative practice [16]. This study aims to assess the degree of focus on health equity as well as the specific methods being employed by healthcare organizations across the United States to foster equity in the context of AIDPM.

Second, large language models (LLMs), a unique type of artificial intelligence system able to process and generate human-like text, have rapidly been adopted by both technology experts and healthcare professionals [17]. These sophisticated models, trained on vast amounts of textual data including medical literature, can mimic natural human language abilities across a wide range of topics [17]. In healthcare, LLMs have already demonstrated their usefulness by generating medical notes and answering patients' medical questions [18, 19]. The practical application of LLMs within clinical settings, as opposed to their experimental use in research, is poorly studied. Additionally, how health systems are implementing or planning to implement LLMs, the specific use cases they target, and concerns regarding data security with LLMs all remain largely unexplored [20, 21]. Considering this changing landscape of AI in healthcare, this new survey focuses on use cases for AIDPM in clinical practice, with a particular focus on LLMs and on the integration of health equity best practices into the various stages of AI model deployment.

Methods

Study design

This cross-sectional study was conducted in collaboration with the Scottsdale Institute (SI), a not-for-profit organization consisting of 65 healthcare systems committed to identifying and sharing best practices within information technology and innovation [22]. The survey instrument was based on a survey used previously by Rojas and colleagues in 2021 [7]. Their survey was originally developed with input from a representative sample of key stakeholders to collect information regarding predictive analytic team roles and current model use cases. For this study, we retained many of the original questions and added new sections focusing on novel AIDPM and LLM use cases, as well as implementation strategies that consider health equity. The new questions included 5 regarding large language models (e.g., Please provide examples of planned use cases for a large language model in your healthcare organization) and 9 regarding health equity (e.g., Does your organization have a team member or members whose focus is health equity with regards to AIDPM?). No alterations were made to the 6 retained questions from the 2021 survey [7]. The expanded survey was tested on a few representative leaders, and no further changes were needed prior to dissemination. The final version of the survey used for this study can be found in the supplementary files (Additional File 1). To address our research question, we surveyed healthcare leaders with the most local knowledge of predictive analytics activities at the healthcare system member sites of The Scottsdale Institute (SI).

The survey was emailed by SI leadership to healthcare executives at their member locations, requesting a response from the person with the most local knowledge of AI governance. Reminders were sent at 2 weeks and 1 week prior to survey closing. Responses were collected in the period between June 21, 2023, and November 30, 2023, utilizing Research Electronic Data Capture (RED-Cap) [23]. This long collection period accompanied by subsequent reminders aimed to maximize response rates and ideally minimize non-response bias [24]. The Rush University Medical Center Institutional Review Board reviewed the study protocol and issued an exempt determination for this research. Informed consent regarding the details of survey distribution and analysis was provided at the beginning of the survey. Data were manually reviewed after collection to ensure that incomplete survey responses were not included in the final analysis.

Statistical analysis

Both the Chi-squared test and z-statistics were chosen for their suitability in addressing the specific research questions posed in this study [25]. The Chi-squared test allowed for evaluation of changes in categorical distributions, while z-statistics provided a robust method for comparing proportions across the two time periods surveyed [26]. Given that multiple comparisons were made within and between the 2021 and 2023 datasets, efforts were made to account for the risk of Type I errors and control the false discovery rate [27]. For Chi-squared comparisons, the alpha was adjusted using the Bonferroni procedure, dividing the alpha of 0.05 by the number of comparisons, with *p*-values only being considered statistically significant if less than this adjusted alpha [28]. The Benjamini-Hochberg method was chosen for z-statistic comparisons as it provides a balance between detecting true effects and minimizing the number of false positives while being less conservative than other approaches given our smaller sample size [29]. Throughout the manuscript,

p-values are listed as significant only if they remained significant after accounting for a false discovery rate of 0.05, and any *p*-value greater than 0.05 was considered non-significant. Reporting of results adheres to the CROSS checklist, a standard for survey reports available on the EQUATOR Network website [30].

Results

The response rate was 49% (32/65), defined as the number of responses received divided by the total number of institutions that received the survey. The plurality of respondents were chief medical information officers (41%, 13/32), with other roles including chief information officer (16%, 5/32), chief analytics officer (16%, 5/32), and chief medical officer (3%, 1/32) being less represented. While 84% (27/32) of the institutions reported utilizing AIDPM in clinical practice, only 53% (17/32) established a team responsible for AIDPM. Furthermore, only 30% (8/27) of these teams operate with a dedicated budget. The presence of specialized AIDPM teams in 2023 (53%, 17/32) did not show significant changes from 2021 (64%, 16/25, *p*-value = 0.40), and the same was true for the presence of dedicated budgets in 2023 (70%, 19/27) compared to 2021 (76%, 19/25, *p*-value = 0.64). In contrast, teams in 2023 exhibited a significant increase in representation from experts in clinical informatics, clinical operations, and quality improvement when compared to 2021 (p-values significant when adjusted for false discovery rate of 0.05), as shown in Figure 1A. Despite this shift in team composition, the scope of responsibilities remained largely the same (Figure 1B).

The survey also assessed the acquisition, deployment, and regulatory considerations of AIDPM. For buying versus building their own AIDPM, both strategies were represented, but the majority in 2021 (44%, 11/25) and in 2023 (41%, 11/27) primarily bought AIDPM from external vendors and built relatively few of their own (p-value > 0.05). When asked about the most-used category of AIDPM at their respective institutions, 41% (11/27) indicated business-facing models (e.g., billing, throughput, scheduling), 37% (10/27) indicated image recognition models and the minority at 22% (6/27) indicated clinical decision support models. There was unanimous support for integrating LLMs into healthcare practices. Of the respondents, 44% (14/32) intend to implement LLMs in collaboration with an electronic health record (EHR) vendor, while 25% (8/32) are planning independent implementations. However, 31% (10/32) acknowledged support for LLMs but reported no concrete plans for their adoption within their organizations. For proposed use cases, managing physician inboxes and summarizing patient histories were noted by 35% (14/40) of respondents.



Fig. 1 Figure 1A. Role representation within dedicated teams for developing and deploying artificial intelligence-derived predictive models (AIDPM). Experts in clinical informatics, clinical operations, and quality improvement were significantly more represented in 2023 compared to 2021 (*p*-value < 0.05). Ethicists and leaders in diversity, equity, and inclusion were not assessed in 2021 but had relatively low representation compared to other roles in 2023. Figure 1B. Responsibility breakdown for dedicated AIDPM teams was similar in 2023 compared to 2021 (*p*-values all > 0.05), except for health system AIDPM governance, which was not assessed in 2021

The principal obstacles to AIDPM adoption were regulatory concerns, data security, workflow integration, and clinician acceptance, as illustrated in Figure 2A. Specific causes for clinician reluctance included alert fatigue, perceived threats to professional autonomy, and liability issues, detailed in Figure 2B. Reflecting on these obstacles and risks, 72% (23/32) of respondents supported government regulation of AIDPM in healthcare, with the majority (41%, 13/32) suggesting that the Food and Drug Administration should oversee this regulation.

Regarding health equity, 76% (13/17) of organizations reported having a team member dedicated to health equity. However, ethicists and leaders in diversity, equity, and inclusion (DEI) were critically under-represented on AIDPM teams at 18% (3/17). It remains unclear what the precise training and roles of those team members dedicated to health equity are, if not in the areas of DEI or bioethics. Respondents consistently reported efforts to promote health equity, including assembling groups of diverse stakeholders, analyzing data for evidence of socioeconomic or racial bias, and evaluating AIDPM for impact on health equity (Figure 3). However, far less common (p-value < 0.001) was the involvement of frontline clinicians in the development of AIDPM and its impact on health equity (Figure 3).

Discussion

Building upon our prior work, these results offer a more current understanding of trends in AIDPM use and governance. While the adoption of AIDPM is common practice among healthcare institutions, only a minority have dedicated teams and budgets for these initiatives, showing little change from 2021 despite the increased attention these tools have received in the intervening years. Also surprising is the relative lack of focus from AIDPM teams on post-deployment accuracy and safety as measured by team roles and responsibilities. It is possible that these roles are relegated to other teams or individuals, but further clarification on this point should be obtained in the future as post-deployment monitoring is a major barrier to the long-term success and safety of AIDPM.

Every respondent supported the use of LLMs in clinical practice, although organizations varied in terms of their preparedness to act on that inclination towards



Fig. 2 Figure 2A. Respondents identified their perceived most relevant barriers to the successful adoption of artificial intelligence-derived predictive models (AIDPM) into clinical practice. Barriers deemed less relevant the top 5 were left unranked. Figure 2B. Respondents identified their perceived most relevant reasons why clinicians at their institutions may be hesitant regarding adopting artificial intelligence-derived predictive models (AIDPM) into clinical practice. Barriers deemed less relevant, the top 5 were left unranked.

these models. There is increased recognition of the lack of established guidance for LLM implementation at healthcare organizations [31], and our results suggest that organizations are preparing to make the leap into clinically-focused LLMs without a clear roadmap on how best to assess these tools. As many proposed LLM use cases, including many described by our respondents, involve direct communication with patients, implementation of LLMs requires a high level of scrutiny and care, and further research is needed on how best to achieve that level of scrutiny. Particularly concerning is the lack of validated methods for monitoring the performance and accuracy of LLM output over time. It is one matter to assess the accuracy of an LLM's responses to a standardized test [32], but it is something quite different to determine whether an LLM is providing helpful communications to patients via the electronic medical record or whether it is sufficiently assisting the clinical decision making of a physician at the bedside. Future studies on this subject should more closely examine the specific use



Fig. 3 Respondents identified the frequency with which their respective institutions take actions to promote health equity at every stage of the development and deployment of artificial intelligence-derived predictive models (AIDPM) in their clinical practice. The center line divides each bar into frequently performed actions (to the right of the line) and infrequently performed actions (to the left of the line). Actions to promote health equity were consistent (p-value > 0.05) across institutions except informing frontline clinicians of the health equity impacts of AIDPM, which was significantly less commonly undertaken by institutions (p-value < 0.001)

cases for LLMs being implemented in practice and what the major successes and barriers are, particularly with regards to post-deployment model evaluation.

While clinician involvement in AIDPM deployment has increased compared to 2021, the reported number of physicians and nurses represented on AIDPM teams are still relatively low. Clinician acceptance was identified as a major barrier to AIDPM adoption in this study, and end-user involvement has been proposed as a method of improving trust in AIDPM [33]. If a major goal for institutions is increasing trust and adoption of AIDPM in clinical practice, end-user involvement represents an area for improvement. Additionally, respondents expressed interest in government regulation of AIDPM. Such regulation could be another method of promoting trust in these tools, which is an argument that both tech and healthcare leaders have been making in the public discourse on this subject [34]. There has also been a growing focus on so-called "explainable" models, which help illuminate how predictions are derived and can be applied to further assess models for bias before large-scale deployment [35]. Optimizing explainability of AIDPM may be an important factor in improving adoption due to concerns about liability and patient and clinician distrust.

Our survey also elucidated concerns regarding data security and liability when it comes to adoption of AIDPM in the healthcare setting. Many applications for these tools have become available to the general public faster than institutional security systems have been able to keep up, and there remains risk that patient personal health information could be inadvertently exposed by something as simple as a physician asking a clinical question of a publicly available LLM. Further exploration into the cyber security practices of institutions will be necessary to develop roadmaps for health systems to follow in order to best protect patients' privacy while attempting to use AIDPM to improve patient health outcomes. While some have proposed that AIDPM itself could be used to improve the cyber security of a health system, it remains to be seen that such efforts would be effective in terms of improving security or trust [36].

The COVID-19 pandemic starkly illuminated the pervasive health disparities that remain in the United States, and it is clear that health systems themselves have a role in addressing these disparities. Amid growing concerns that AIDPM could itself widen these gaps in health equity, our results found that ethicists and leaders in diversity, equity, and inclusion are represented on very few AIDPM teams. Although respondents expressed that organizations are taking actions to prevent AIDPM from infringing on health equity, more definitive assessments of the efficacy of these measures are needed.

Limitations of our study include its small sample size and its concentration on larger, innovative healthcare systems whose leaders are SI members. These factors may limit the generalizability of our findings as smaller healthcare systems with fewer resources or less interest in novel technologies may adopt very different governance strategies than the ones highlighted by our study. Despite these limitations, our study illuminates how health systems in the United States are locally implementing AI governance and its critical intersections with their health equity missions.

Conclusions

Clinical adoption of AIDPM remains challenging, primarily due to the absence of established best practices. While there is broad support for integrating large language models in healthcare, many organizations lack concrete implementation plans, highlighting an urgent need for developing LLM evaluation guidelines. Collaborative groups such as the Coalition for Health AI (CHAI) and the Health AI Partnership (HAIP) could be well-positioned to publish frameworks that healthcare organizations can use to vet and monitor LLMs. Our study also reveals a strong interest among respondents for federal regulations on AIDPM, which would enable organizations to purchase AI-based tools from vendors with assurance of specific quality and safety standards. As these regulations and best practice frameworks are developed and implemented, future research will be crucial to assess healthcare systems' adherence to these principles, with particular emphasis on patient safety and health equity.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s44247-024-00135-3.

Additional file 1. Survey Instrument. We have removed the first two questions, which asked for respondent organization and title, because of potentially identifying response choices. The rest of the survey questions are included.

Additional file 2. REDCap Survey Data. This file contains responses from the REDCap survey. We have removed respondent job titles and replaced organization names with Organization_1, etc. to protect participant privacy.

Acknowledgements

We thank the healthcare leaders in the Scottdale Institute network for completing the survey.

Authors' contributions

Study concept and design: SG, JCR, JG, IK, MM. Acquisition of data: SG, JCR, JG, IK, MM. Analysis and interpretation of data: SG, JCR, JG, IK, MM. First drafting of the manuscript: SG. Critical revision of the manuscript for important intellectual content: SG, JCR, JG, IK, MM. Statistical analysis: SG, JCR. Study supervision: JCR. Data access and responsibility: SG, JCR had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

Not applicable.

Data availability

Survey data are provided in the supplementary information files.

Declarations

Ethics approval and consent to participate

The Rush University Medical Center Institutional Review Board reviewed the study protocol and approved the study under the exempt category of review (ID #22111801). We obtained informed consent from all study participants. Upon receiving the survey instrument sent to them via email using REDCap, participants' consent was obtained after they read the introduction and consent materials prior to starting the survey instrument. The study was conducted in accordance with the ethical guidelines stated in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Internal Medicine, Rush University Medical Center, Chicago, IL 60612, USA. ²Scottsdale Institute, Maple Grove, MN, USA. ³Center for Integrated Microbiome and Chronobiology Research, Rush University Medical Center, Chicago, IL, USA. ⁴Division of Pulmonology, Critical Care, and Sleep Medicine, Rush University Medical Center, 1520 W Harrison St Joan and Paul Rubschlager Building, 6th floor, Chicago, IL 60607, USA.

Received: 7 May 2024 Accepted: 10 October 2024 Published online: 19 November 2024

References

- Schwab, K. The Fourth Industrial Revolution. Foreign Affairs. 2015. Available at https://www.foreignaffairs.com/world/fourth-industrial-revol ution. Accessed 7 May 2024.
- Topol EJ, Verghese A. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York, NY: Basic Books; 2019.
- Rehman A, Naz S, Razzak I. Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. Multimed Syst. 2022;28:1339–71.
- Parikh RB, Kakad M, Bates DW. Integrating predictive analytics into highvalue care: the dawn of precision delivery. JAMA. 2016;315:651.
- 5. Pearson TA, et al. Precision health analytics with predictive analytics and implementation research. J Am Coll Cardiol. 2020;76:306–20.
- Eaneff S, Obermeyer Z, Butte AJ. The case for algorithmic stewardship for artificial intelligence and machine learning technologies. JAMA. 2020;324:1397.
- Rojas JC, Rohweder G, Guptill J, Arora VM, Umscheid CA. Predictive analytics programs at large healthcare systems in the USA: a national survey. J Gen Intern Med. 2022;37:4015. https://doi.org/10.1007/ s11606-022-07517-1.
- World Health Organization. Social determinants of health. WHO. Available at https://www.who.int/health-topics/social-determinants-of-health# tab=tab_1. Accessed 7 May 2024.
- 9. Makhni S, Chin MH, Fahrenbach J, Rojas JC. Equity challenges for artificial intelligence algorithms in health care. Chest. 2022;161:1343–6.
- Berdahl CT, Baker L, Mann S, Osoba O, Girosi F. Strategies to improve the impact of artificial intelligence on health equity: scoping review. JMIR AI. 2023;2:e42936.
- 11. Artificial Intelligence in Health Care. The Hope, the Hype, the Promise, the Peril. Washington: National Academy of Medicine; 2020.
- Rojas JC, et al. Framework for integrating equity into machine learning models. Chest. 2022;161:1621–7.
- Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169:866.
- Nazer LH, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2:e0000278.
- Gichoya JW, et al. Al pitfalls and what not to do: mitigating bias in Al. Br J Radiol. 2023;96:20230023.

- 16. de Hond AAH, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med. 2022;5:2.
- 17. Thirunavukarasu AJ, et al. Large language models in medicine. Nat Med. 2023;29:1930–40.
- Nayak A, et al. Comparison of history of present illness summaries generated by a chatbot and senior internal medicine residents. JAMA Intern Med. 2023;183:1026.
- Ayers JW, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183:589.
- Nasr M, et al. Scalable Extraction of Training Data from (Production) Language Models. arXiv preprint arXiv 2023. https://doi.org/10.48550/ARXIV. 2311.17035.
- Ullah E, Parwani A, Baig MM, Singh R. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology – a recent scoping review. Diagn Pathol. 2024;19:43.
- 22. The Scottsdale Institute. The healthcare executive resource for information management. Scottsdale Institute. Available at https://www.scott sdaleinstitute.org/. Accessed 7 May 2024.
- Harris PA, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95:103208.
- Sedgwick P. Non-response bias versus response bias. BMJ. 2014;348:g2573–g2573.
- Liu H, Setiono R. Chi2: feature selection and discretization of numeric attributes. Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence. Herndon: 1995;388–91.
- Soms AP. Exact confidence intervals, based on the Z statistic, for the difference between two proportions. Commun Stat - Simul Comput. 1989;18:1325–41.
- Pollard P, Richardson JT. On the probability of making type I errors. Psychol Bull. 1987;102:159–63.
- Napierala MA. What is the Bonferroni correction? AAOS Now 40. 2012. Available at https://link.gale.com/apps/doc/A288979427/HRCA?u= anon~9fbca60d&sid=googleScholar&xid=c4be5015. Accessed 7 May 2024.
- Ferreira JA, Zwinderman AH. On the Benjamini–Hochberg method. Ann Statist. 2006;34:1827–49.
- Sharma A, et al. A consensus-based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med. 2021;36:3179–87.
- Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. Healthc Basel Switz. 2023;11:2776.
- Kung TH, et al. Performance of ChatGPT on USMLE: Potential for Alassisted medical education using large language models. PLOS Digit Health. 2023;2:e0000198.
- Shulha M, Hovdebo J, D'Souza V, Thibault F, Harmouche R. Integrating explainable machine learning in clinical decision support systems: study involving a modified design thinking approach. JMIR Form Res. 2024;8:e50475.
- Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023;6:120.
- Loh HW, et al. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). Comput Methods Progr Biomed. 2022;226:107161.
- Radanliev P, De Roure D. Disease X vaccine production and supply chains: risk assessing healthcare systems operating with artificial intelligence and industry 4.0. Health Technol. 2023;13:11–5.

Publisher's Note

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