A.I. & Healthcare – Autumn 2025

Figure 1 NHS Lanarkshire Logo

For further help or information, contact - amanda.minns@lanarkshire.scot.nhs.uk

Contents

AI & AHPs

1. Hoffman, Jane, Rachel Wenke, Rebecca L. Angus, Lucy Shinners, Brent Richards and Laetitia Hattingh. **Overcoming barriers and enabling artificial intelligence adoption in allied health clinical practice: A qualitative study.** Digital health. 2025.Vol.11 pp20552076241311144.

Artificial intelligence (AI) has the potential to revolutionise healthcare. If the implementation is successful it has the potential to improve healthcare outcomes for patients and organisations. Little is known about the perceptions of allied health professionals (AHPs) towards AI in healthcare. This study investigated barriers and enablers to AI implementation in the delivery of healthcare from the AHPs perspective. Qualitative methodology informed by behaviour change theory using focus groups with AHPs at a health service in Queensland, Australia. Twenty-four barriers and 24 enablers were identified by 25 participants across four focus groups. Barriers included: lack of AI knowledge, explainability challenges, risk to professional practice, negative impact on professional practice, and role replacement. Enablers include AI training and education, regulation, reputation, understanding the healthcare benefits of AI and engaging clinical champions. AHPs have concerns about the impact and trustworthiness of AI and the readiness of organisations to support its use. Organisations must take a proactive approach and adopt targeted and multifaceted strategies to address barriers. This may include workforce upskilling, clear communication of the benefits of AI use of local champions and ongoing research.

2. Raghunathan, Kalpana, Meg E. Morris, Tafheem A. Wani, et al. **Using artificial intelligence to improve healthcare delivery in select allied health disciplines: a scoping review protocol**. BMJ Open. 2025.Vol.15(3), ppe098290.

Methods to adopt artificial intelligence (AI) in healthcare clinical practice remain unclear. The potential for rapid integration of AI-enabled technologies across healthcare settings coupled with the growing digital divide in the health sector highlights the need to examine AI use by health professionals, especially in allied health disciplines with emerging AI use such as physiotherapy, occupational therapy, speech pathology, podiatry and dietetics. This protocol details the methodology for a scoping review on the use of AI-enabled technology in sectors of the allied health workforce. The research question is ‘How is AI used by sectors of the allied health workforce to improve patient safety, quality of care and outcomes, and what is the quality of evidence supporting this use?’Methods and analysisThe review will follow the Joanna Briggs Institute scoping review guidelines. Databases will be searched from 17 to 24 March 2025 and will include PubMed/Medline, Embase, PsycINFO and Cummulative Index to Nursing and Allied Health Literature databases. Dual screening against inclusion criteria will be applied for study selection. Peer-reviewed articles reporting primary research in allied healthcare published in English within the last 10 years will be included. Studies will be evaluated using the Quality Assessment with Diverse Studies tool. The review will map the existing literature and identify key themes related to the use of AI

in the disciplines of physiotherapy, occupational therapy, speech pathology, podiatry and dietetics.Ethics and dissemination. No ethics approval will be sought, as only secondary research outputs will be used. Findings will be disseminated through peer-reviewed publication and presentations at workshops and conferences.

AI & cancer

3. Eisenstein, Michael. **How AI is helping to boost cancer screening.** Nature. 2025.Vol.640(8060), ppS62–S64.

Cancer blood tests and AI-powered scans look promising for quicker and more accurate detection of disease. Cancer blood tests and AI-powered scans look promising for quicker and more accurate detection of disease.

4. Goh, Serene, Rachel Sze Jen Goh, Bryan Chong, et al. **Challenges in Implementing Artificial Intelligence in Breast Cancer Screening Programs: A Systematic Review and Framework for Safe Adoption (Preprint).** Journal of medical Internet research. 2025.Vol.27(4), ppe62941.

Artificial intelligence (AI) studies show promise in enhancing accuracy and efficiency in mammographic screening programs worldwide. However, its integration into clinical workflows faces several challenges, including unintended errors, the need for professional training, and ethical concerns. Notably, specific frameworks for AI imaging in breast cancer screening are still lacking. This study aims to identify the challenges associated with implementing AI in breast screening programs and to apply the Consolidated Framework for Implementation Research (CFIR) to discuss a practical governance framework for AI in this context. Three electronic databases (PubMed, Embase, and MEDLINE) were searched using combinations of the keywords "artificial intelligence," "regulation," "governance," "breast cancer," and "screening." Original studies evaluating AI in breast cancer detection or discussing challenges related to AI implementation in this setting were eligible for review. Findings were narratively synthesized and subsequently mapped directly onto the constructs within the CFIR. A total of 1240 results were retrieved, with 20 original studies ultimately included in this systematic review. The majority (n=19) focused on AI-enhanced mammography, while 1 addressed AI-enhanced ultrasound for women with dense breasts. Most studies originated from the United States (n=5) and the United Kingdom (n=4), with publication years ranging from 2019 to 2023. The quality of papers was rated as moderate to high. The key challenges identified were reproducibility, evidentiary standards, technological concerns, trust issues, as well as ethical, legal, societal concerns, and postadoption uncertainty. By aligning these findings with the CFIR constructs, action plans targeting the main challenges were incorporated into the framework, facilitating a structured approach to addressing these issues. This systematic review identifies key challenges in implementing AI in breast cancer screening, emphasizing the need for consistency, robust evidentiary standards, technological advancements, user trust, ethical frameworks, legal safeguards, and societal benefits. These findings can serve as a blueprint for policy makers, clinicians, and AI developers to collaboratively advance AI adoption in breast cancer screening.

5. Puri, S., M. Bagnall and G. Erdelyi. **Review and reflections on live AI mammographic screen reading in a large UK NHS breast screening unit**. Clin.Radiol. 2025.Vol.85 pp106872.

The Radiology team from a large Breast Screening Unit in the UK with a screening population of over 135,000 took part in a service evaluation project using artificial intelligence (AI) for reading breast screening mammograms. AIMTo evaluate the clinical benefit AI may provide when implemented as a silent reader in a double reading breast screening programme and to evaluate feasibility and the operational impact of deploying AI into the breast screening programme. The service was one of 14 breast screening sites in the UK to take part in this project and we present our local experience with AI in breast screening. MATERIALS AND METHODSA commercially available AI platform was deployed and worked in real time as a 'silent third reader' so as not to impact standard workflows and patient care. All cases flagged by AI but not recalled by standard double reading (positive discordant cases) were reviewed along with all cases recalled by human readers but not flagged by AI (negative discordant cases). RESULTS9,547 cases were included in the evaluation. 1,135 positive discordant cases were reviewed, and one woman was recalled from the reviews who was not found to have cancer on further assessment in the breast assessment clinic. 139 negative discordant cases were reviewed, and eight cancer cases (8.79% of total cancers detected in this period) recalled by human readers were not detected by AI. No additional cancers were detected by AI during the study. CONCLUSIONPerformance of AI was inferior to human readers in our unit. Having missed a significant number of cancers makes it unreliable and not safe to be used in clinical practice. AI is not currently of sufficient accuracy to be considered in the NHS Breast Screening Programme.

6. Roadevin, Cristina and Harry Hill. **AI interventions in cancer screening: balancing equity and cost-effectiveness**. J.Med.Ethics. 2025. ppjme–110707.

This paper examines the integration of artificial intelligence (AI) into cancer screening programmes, focusing on the associated equity challenges and resource allocation implications. While AI technologies promise significant benefits—such as improved diagnostic accuracy, shorter waiting times, reduced reliance on radiographers, and overall productivity gains and cost-effectiveness—current interventions disproportionately favour those already engaged in screening. This neglect of non-attenders, who face the worst cancer outcomes, exacerbates existing health disparities and undermines the core objectives of screening programmes.Using breast cancer screening as a case study, we argue that AI interventions must not only improve health outcomes and demonstrate cost-effectiveness but also address inequities by prioritising non-attenders. To this end, we advocate for the design and implementation of cost-saving AI interventions. Such interventions could enable reinvestment into strategies specifically aimed at increasing engagement among non-attenders, thereby reducing disparities in cancer outcomes. Decision modelling is presented as a practical method to identify and evaluate these cost-saving interventions. Furthermore, the paper calls for greater transparency in decision-making, urging policymakers to explicitly account for the equity implications and opportunity costs associated with AI investments. Only then will they be able to balance the promise of technological innovation with the ethical imperative to improve health outcomes for all, particularly underserved populations. Methods such as distributional cost-

effectiveness analysis are recommended to quantify and address disparities, ensuring more equitable healthcare delivery.

7. Sorensen, Luke E., Vanessa A. Moore, Jared L. Gregston, et al. **The utility of artificial intelligence in gastrointestinal oncology: A systematic review of randomized controlled trials**. Cancer. 2025.Vol.131(18), ppe70043–n/a.

In the field of gastrointestinal oncology, the development of novel artificial intelligence (AI) processes may help with multiple aspects of cancer care delivery. However, a comprehensive understanding of the current utility of AI in gastrointestinal oncology is lacking. The authors conducted searches in the following databases: MEDLINE (Ovid), Embase (Ovid), and CINAHL (Cumulative Index of Nursing and Allied Health) Ultimate (EBSCO). The analysis focused on publication trends and outcomes of randomized controlled trials (RCTs) that used AI to manage gastrointestinal malignancies. From our initial search retrieval of 3730 studies, 27 RCTs (with a total of 29,895 patients) were identified that met inclusion criteria. The first RCT was published in 2019, followed by five in 2020, four in 2021, six in 2022, and 11 in 2023. Colorectal malignancies comprised the majority of the literature (23 of 27 studies; 85%), with other studies focused on gastric cancer (three of 27 studies; 11%) and hepatocellular carcinoma (one of 27 studies; 4%). Of the included RCTs, 25 (93%) had a primary outcome focused on lesion/cancer detection throughout the gastrointestinal tract using endoscopy or ultrasound, with others focused on algorithmic-based AI assistance with postoperative pain management or histologic diagnosis. Overall, 22 of 27 studies (81%) met their primary end point with a statistically significant result. In this systematic review, the authors observed a recent increase in the number of RCTs focused on AI within the field of gastrointestinal oncology and identified specific areas in which AI is being used. Findings from this work should help to inform further investigations to develop and test innovative AI uses, enhance care delivery, and improve patient outcomes. The authors reviewed randomized controlled trials (RCTs) to assess the utility of artificial intelligence (AI) in gastrointestinal cancer and found that, despite a rise in observational studies, RCTs remain limited. However, existing RCTs suggest promising benefits of AI in endoscopic evaluation, imaging, histologic diagnosis, and algorithmic-based AI assistance with postoperative pain management.

8. Wilkoff, Marni H., Nicholas R. Piniella and Rashmi Advani. **Can Artificial Intelligence Create an Accurate Colonoscopy Bowel Preparation Prompt?**. Gastro hep advances. 2025.Vol.4(2), pp100566.

Colorectal cancer is the third most common cancer in the United States, with colonoscopy being the preferred screening method. Up to 25% of colonoscopies are associated with poor preparation which leads to prolonged procedure time, repeat colonoscopies, and decreased adenoma detection. Artificial intelligence (AI) is being increasingly used in medicine, assessing medical school exam questions, and writing medical reports. Its use in gastroenterology has been used to educate patients with cirrhosis and hepatocellular carcinoma, answer patient questions about colonoscopy and provide correct colonoscopy screening intervals, having the ability to augment the patient–provider relationship. This study aims at assessing the accuracy of a ChatGPT-generated precolonoscopy bowel preparation prompt. A nonrandomized cross-sectional study assessing the perceptions of an AI-generated colonoscopy preparation prompt was conducted in a large multisite

quaternary health-care institution in the northeast United States. All practicing gastroenterologists in the health system were surveyed, with 208 having a valid email address and were included in the study. A Research Electronic Data Capture survey was then distributed to all participants and analyzed using descriptive statistics. Overall, 91% of gastroenterologist physicians determined the prompt easy to understand, 95% thought the prompt was scientifically accurate and 66% were comfortable giving the prompt to their patients. Sixty four percent of reviewers correctly identified the ChatGPT-generated prompt, but only 32% were confident in their answer. The ability of ChatGPT to create a sufficient bowel preparation prompt highlights how physicians can incorporate AI into clinical practice to improve ease and efficiency of communication with patients when it comes to bowel preparation.

AI & clinical decision making

9. Awad, Nadia Hassan Ali, Wafaa Aljohani, Mai Mohammed Yaseen, Wafaa Hassan Ali Awad, Randa Ahmed Abou Elala Said Ahmed and Heba Mohammed Alanwer Ashour. **When Machines Decide: Exploring How Trust in AI Shapes the Relationship Between Clinical Decision Support Systems and Nurses' Decision Regret: A Cross-Sectional Study.** Nurs.Crit.Care. 2025.Vol.30(5), ppe70157.

Artificial intelligence (AI)-based Clinical Decision Support Systems (AI-CDSS) are increasingly implemented in intensive care settings to support nurses in complex, time-sensitive decisions, aiming to improve accuracy, efficiency and patient outcomes. However, their use raises concerns about emotional consequences, particularly decision regret, which may arise when clinical judgement or outcomes are unfavourable. Trust in AI may play a key role in shaping nurses' responses to AI-guided decisions. To examine the relationship between nurses' reliance on AI-CDSS, decision regret and trust in AI, with a focus on the moderating role of trust in the association between AI-CDSS reliance and decision regret. A cross-sectional correlational design was used. A convenience sample of 250 intensive care unit (ICU) nurses completed validated instruments: the Healthcare Systems Usability Scale (HSUS) for AI-CDSS reliance, the Decision Regret Scale (DRS) and the Trust in AI Scale. Descriptive statistics, Pearson's correlations, multiple linear regression and moderation analysis were conducted. A total of 250 ICU nurses participated in the study out of 400 approached, yielding a response rate of 62.5%. Nurses reported moderate levels of AI-CDSS reliance ($M = 78.6$, $SD = 12.4$), decision regret ($M = 38.5$, $SD = 14.8$) and trust in AI ($M = 13.9$, $SD = 3.2$). AI-CDSS reliance was negatively correlated with decision regret ($r = -0.42$, $p < 0.01$) and positively with trust in AI ($r = 0.51$, $p < 0.01$). Regression analysis showed that both AI-CDSS reliance ($\beta = -0.36$) and trust in AI ($\beta = -0.24$) significantly predicted reduced regret ($R = 0.27$, $p < 0.001$). Trust moderated the relationship, strengthening the negative association between reliance and regret. Greater reliance on AI-CDSS is associated with lower decision regret among ICU nurses, especially when trust in AI is high. Trust enhances emotional acceptance and supports effective AI integration. Building trust in AI-CDSS among nurses is essential for minimising emotional burden and optimising decision-making in critical care.

10. Choi, Jeff. **Artificial intelligence in surgery research: Successfully implementing AI clinical decision support models**. J Trauma Acute Care Surg. 2025.Vol.99(4), pp518–521.

Artificial intelligence (AI) in surgery literature typically encompasses decision support models that aim to help clinicians make better decisions. Many studies report developing and validating models, yet few models are implemented at the bedside. Exceedingly few models achieve their intended goal upon implementation. While the TRIPOD-AI and DECIDE-AI guidelines outline separate reporting standards for the development/validation, and staged implementation of AI models, respectively, this article outlines how future implementation should be considered at the outset before model development. Building on lessons from high-performing AI decision support models that faced challenges upon implementation, we will discuss study design consideration for building trustworthy and actionable AI clinical decision support models that can cross the database-to-bedside gap and become successfully implemented. This article outlines why downstream implementation must be considered at the outset before developing AI clinical decision support models, and highlights considerations for building models that can be successfully implemented at the bedside.

11. Giebel, Godwin Denk, Pascal Raszke, Hartmuth Nowak, et al. **Problems and Barriers Related to the Use of AI-Based Clinical Decision Support Systems: Interview Study**. Journal of medical Internet research. 2025.Vol.27(27-28), ppe63377.

Digitalization is currently revolutionizing health care worldwide. A promising technology in this context is artificial intelligence (AI). The application of AI can support health care providers in their daily work in various ways. The integration of AI is particularly promising in clinical decision support systems (CDSSs). While the opportunities of this technology are numerous, the problems should not be overlooked. This study aimed to identify challenges and barriers in the context of AI-based CDSSs from the perspectives of experts across various disciplines. Semistructured expert interviews were conducted with different stakeholders. These included representatives of patients, physicians and caregivers, developers of AI-based CDSSs, researchers (studying AI in health care and social and health law), quality management and quality assurance representatives, a representative of an ethics committee, a representative of a health insurance fund, and medical product consultants. The interviews took place on the web and were recorded, transcribed, and subsequently subjected to a qualitative content analysis based on the method by Kuckartz. The analysis was conducted using MAXQDA software. Initially, the problems were separated into "general," "development," and "clinical use." Finally, a workshop within the project consortium served to systematize the identified problems. A total of 15 expert interviews were conducted, and 309 expert statements with reference to problems and barriers in the context of AI-based CDSSs were identified. These emerged in 7 problem categories: technology (46/309, 14.9%), data (59/309, 19.1%), user (102/309, 33%), studies (17/309, 5.5%), ethics (20/309, 6.5%), law (33/309, 10.7%), and general (32/309, 10.4%). The problem categories were further divided into problem areas, which in turn comprised the respective problems. A large number of problems and barriers were identified in the context of AI-based CDSSs. These can be systematized according to the point at which they occur ("general," "development," and "clinical use") or according to the problem category ("technology," "data," "user," "studies," "ethics," "law," and "general"). The problems identified in this work should be further investigated. They can be used as a basis for deriving solutions to optimize development, acceptance, and use of AI-based CDSSs.

12. Kim, Bohye, Katie Ryan and Jane Paik Kim. **Assessing the impact of information on patient attitudes toward artificial intelligence-based clinical decision support (AI/CDS): a pilot web-based SMART vignette study**. J.Med.Ethics. 2025.Vol.51(8), pp541–549.

It is increasingly recognised that the success of artificial intelligence-based clinical decision support (AI/CDS) tools will depend on physician and patient trust, but factors impacting patients' views on clinical care reliant on AI have been less explored. Objective This pilot study explores whether, and in what contexts, detail of explanation provided about AI/CDS tools impacts patients' attitudes toward the tools and their clinical care. Methods We designed a Sequential Multiple Assignment Randomized Trial vignette web-based survey. Participants recruited through Amazon Mechanical Turk were presented with hypothetical vignettes describing health concerns and were sequentially randomised along three factors: (1) the level of detail of explanation regarding an AI/CDS tool; (2) the AI/CDS result; and (3) the physician's level of agreement with the AI/CDS result. We compared mean ratings of comfort and confidence by the level of detail of explanation using t-tests. Regression models were fit to confirm conditional effects of detail of explanation.ResultsThe detail of explanation provided regarding the AI/CDS tools was positively related to respondents' comfort and confidence in the usage of the tools and their perception of the physician's final decision. The effects of detail of explanation on their perception of the physician's final decision were different given the AI/CDS result and the physician's agreement or disagreement with the result.ConclusionsMore information provided by physicians regarding the use of AI/CDS tools may improve patient attitudes toward healthcare involving AI/CDS tools in general and in certain contexts of the AI/CDS result and physician agreement.

13. Rom, Harel, Ori Peleg, Yovel Rom, Anat Mirelman, Gaddi Blumrosen and Inbal Maidan. **Remote clinical decision support tool for Parkinson's disease assessment using a novel approach that combines AI and clinical knowledge**. BMC Med.Inform.Decis.Mak. 2025.Vol.25(1), pp294–7.

Background Early diagnosis of Parkinson's disease (PD) can assist in designing efficient treatments. Reduced facial expressions are considered a hallmark of PD, making advanced artificial intelligence (AI) image processing a potential non-invasive clinical decision support tool for PD detection. This study aims to determine the sensitivity of image-to-text AI, which matches facial frames recorded in home settings with descriptions of PD facial expressions, in identifying patients with PD. Methods Facial image of 67 PD patients and 52 healthy-controls (HCs) were collected via standard video recording. Using clinical knowledge, we compiled descriptive sentences detailing facial characteristics associated with PD. The facial images were analyzed with OpenAI's CLIP model to generate probability scores, indicating the likelihood of each image matching the PD-related descriptions. These scores were used in an XGBoost model to identify PD patients based on the total, motor, and facial-expression item of the MDS-UPDRS, a common scale for assessing disease severity. Results The image-to-text AI technology showed the best results in identifying PD patients based on the facial expression item (AUC = 0.78 ± 0.05), especially for those with 'mild' facial symptoms (AUC = 0.87 ± 0.04). The motor MDS-UPDRS score followed (AUC = 0.69 ± 0.05), while the total MDS-UPDRS score showed the lowest performance in identifying PD patients (AUC = 0.59 ± 0.05). PD matching probabilities between facial images and sentences

revealed significant correlations across all MDS-UPDRS components ($r > 0.23$, $p < 0.0001$). Conclusions Our results demonstrate the feasibility of using advanced AI in a clinical decision support tool for PD diagnosis, suggesting a novel approach for home-based screening to identify PD patients. This method represents a significant innovation, transforming clinical knowledge into practical algorithms that can serve as effective screening tools. Clinical trial number MOH_2023-04-16_012535.

14. Zhang, Shan, Shu Ding, Wei Cui, Xiangyu Li, Jun Wei and Ying Wu. **Evaluating the effectiveness of a clinical decision support system (AI-Antidelirium) to improve Nurses' adherence to delirium guidelines in the intensive care unit**. Intensive & critical care nursing. 2025.Vol.87 pp103933.

To evaluate the impact of Artificial Intelligence Assisted Prevention and Management for Delirium (AI-AntiDelirium) on improving adherence to delirium guidelines among nurses in the intensive care unit (ICU). Between November 2022 and June 2023, A cluster randomized controlled trial was undertaken. A total of 38 nurses were enrolled in the interventional arm, whereas 42 nurses were recruited for the control arm in six ICUs across two hospitals in Beijing, comparing nurses' adherence and cognitive load in units that use AI-AntiDelirium or the control group. The AI-AntiDelirium tailored delirium preventive or treated interventions to address patients' specific risk factors. The adherence rate of delirium interventions was the primary endpoint. The other endpoints were adherence to risk factors assessment, ICU delirium assessment, and nurses' cognitive load. The repeated measures analysis of variance was utilized to explore the influence of time, group, and time × group interaction on the repeated measurement variable (e.g., adherence, cognitive load). A cumulative total of 1040 nurse days were analyzed for this study. The adherence to delirium intervention of nurses in AI-AntiDelirium groups was higher than control units (75 % vs. 58 %, $P < 0.01$). When compared to control groups, AI-AntiDelirium was found to be significantly effective in both decreasing extraneous cognitive load ($P < 0.01$) and improving germane cognitive load ($P < 0.01$). This study supports the effectiveness of AI-AntiDelirium in enhancing nurses' adherence to evidence-based, individualized delirium intervention and also reducing extraneous cognitive load. A nurse-led systemshould be applied by nursing administrators to improve compliance with nursing interventions among ICU nurses. 10.1016/j.iccn.2024.103933

AI & copilot

15. Arslan, Banu, Mehmet Necmeddin Sutasir and Ertugrul Altinbilek. **Performance of Microsoft Copilot in the Diagnostic Process of Pulmonary Embolism**. The western journal of emergency medicine. 2025.Vol.26(4), pp1030–1039.

Introduction: Patients with pulmonary embolism (PE) often present with non-specific signs and symptoms mimicking other conditions and complicating diagnosis. In this study we aimed to evaluate the performance of an artificial-intelligence tool, Microsoft Copilot, in the diagnostic process of PE, using clinical data including demographics, complaints, and vital signs. Methods: We conducted this study using 140 clinical vignettes, including 70 patients with and 70 patients without PE. The vignettes were derived from published case reports within the last 10 years. We used Copilot for its free GPT-4 integration to analyze clinical data and answer two questions after each vignette. We compared Copilot's ability to identify PE within the top 10 differential diagnoses, and its ability to predict the risk of PE

when compared to the use of the Wells score by two independent investigators. Results: Copilot correctly included PE in the differential diagnosis in 94.3% of cases by listing it within the top 10 conditions. Risk assessment by Copilot yielded significantly higher levels in patients with PE (P.05). Copilot demonstrated better discriminatory power than the Wells score in risk assessment of PE (area under the curve 0.713 vs 0.583), with statistical significance (P<0.001 vs P=.091). Sensitivity, specificity, positive predictive value, and negative predictive value for discriminating between the combination of low- and intermediate- vs high-risk categories were 34%, 97.1%, 92.3%, and 59.6%, respectively. Conclusion: This study explores the potential of Copilot as a tool in clinical decision-making, demonstrating a high rate of correctly identifying PE and improved performance over the Wells score. However, further validation in larger populations and real-world settings is crucial to fully realize its potential. 10.5811/WESTJEM.24995

16. Ramchandani, Rashi, Eddie Guo, Michael Mostowy, et al. **Comparison of ChatGPT-4, Copilot, Bard and Gemini Ultra on an Otolaryngology Question Bank**. Clinical otolaryngology. 2025.Vol.50(4), pp704–711.

ABSTRACT Objective To compare the performance of Google Bard, Microsoft Copilot, GPT-4 with vision (GPT-4) and Gemini Ultra on the OTO Chautauqua, a student-created, faculty-reviewed otolaryngology question bank. Study Design Comparative performance evaluation of different LLMs. Setting N/A. Participants N/A. Methods Large language models (LLMs) are being extensively tested in medical education. However, their accuracy and effectiveness remain understudied, particularly in otolaryngology. This study involved inputting 350 single-best-answer multiple choice questions, including 18 image-based questions, into four LLMS. Questions were sourced from six independent question banks related to (a) rhinology, (b) head and neck oncology, (c) endocrinology, (d) general otolaryngology, (e) paediatrics, (f) otology, (g) facial plastics, reconstruction and (h) trauma. LLMs were instructed to provide an output reasoning for their answers, the length of which was recorded. Results Aggregate and subgroup analysis revealed that Gemini (79.8%) outperformed the other LLMs, followed by GPT-4 (71.1%), Copilot (68.0%), and Bard (65.1%) in accuracy. The LLMs had significantly different average response lengths, with Bard ($\bar{x}$ = 1685.24) being the longest and no difference between GPT-4 ($\bar{x}$ = 827.34) and Copilot ($\bar{x}$ = 904.12). Gemini's longer responses ($\bar{x}$ =1291.68) included explanatory images and links. Gemini and GPT-4 correctly answered image-based questions (n = 18), unlike Copilot and Bard, highlighting their adaptability and multimodal capabilities. Conclusion Gemini outperformed the other LLMs in terms of accuracy, followed by GPT-4, Copilot and Bard. GPT-4, although it has the second-highest accuracy, provides concise and relevant explanations. Despite the promising performance of LLMs, medical learners should cautiously assess accuracy and decision-making reliability. 10.1111/coa.14302

AI & data

17. Alamri, Malak, Mamoona Humayun, Khalid Haseeb, Naveed Abbas and Naeem Ramzan. **AI-Powered Adaptive Disability Prediction and Healthcare Analytics Using Smart Technologies.** Diagnostics (Basel). 2025.Vol.15(16), pp2104.

Background: By leveraging advanced wireless technologies, Healthcare Industry 5.0 promotes the continuous monitoring of real-time medical acquisition from the physical environment. These systems help identify early diseases by collecting health records from

patients' bodies promptly using biosensors. The dynamic nature of medical devices not only enhances the data analysis in medical services and the prediction of chronic diseases, but also improves remote diagnostics with the latency-aware healthcare system. However, due to scalability and reliability limitations in data processing, most existing healthcare systems pose research challenges in the timely detection of personalized diseases, leading to inconsistent diagnoses, particularly when continuous monitoring is crucial. Methods: This work propose an adaptive and secure framework for disability identification using the Internet of Medical Things (IoMT), integrating edge computing and artificial intelligence. To achieve the shortest response time for medical decisions, the proposed framework explores lightweight edge computing processes that collect physiological and behavioral data using biosensors. Furthermore, it offers a trusted mechanism using decentralized strategies to protect big data analytics from malicious activities and increase authentic access to sensitive medical data. Lastly, it provides personalized healthcare interventions while monitoring healthcare applications using realistic health records, thereby enhancing the system's ability to identify diseases associated with chronic conditions. Results: The proposed framework is tested using simulations, and the results indicate the high accuracy of the healthcare system in detecting disabilities at the edges, while enhancing the prompt response of the cloud server and guaranteeing the security of medical data through lightweight encryption methods and federated learning techniques. Conclusions: The proposed framework offers a secure and efficient solution for identifying disabilities in healthcare systems by leveraging IoMT, edge computing, and AI. It addresses critical challenges in real-time disease monitoring, enhancing diagnostic accuracy and ensuring the protection of sensitive medical data. 10.3390/diagnostics15162104

18. Fogel, Paul and George Luta. **On board with COMET to improve omics prediction models: AI for healthcare data**. Nat Mach Intell. 2025.Vol.7(2), pp168–169.

The performance of omics prediction models can be significantly improved by combining limited patient proteomic data with widely available electronic health records. 10.1038/s42256-025-00990-3

19. Ruta, Michael R., Tony Gaidici, Chase Irwin and Jonathan Lifshitz. **ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research**. Journal of medical Internet research. 2025.Vol.27(4), ppe63550.

ChatGPT, a conversational artificial intelligence developed by OpenAI, has rapidly become an invaluable tool for researchers. With the recent integration of Python code interpretation into the ChatGPT environment, there has been a significant increase in the potential utility of ChatGPT as a research tool, particularly in terms of data analysis applications. This study aimed to assess ChatGPT as a data analysis tool and provide researchers with a framework for applying ChatGPT to data management tasks, descriptive statistics, and inferential statistics. A subset of the National Inpatient Sample was extracted. Data analysis trials were divided into data processing, categorization, and tabulation, as well as descriptive and inferential statistics. For data processing, categorization, and tabulation assessments, ChatGPT was prompted to reclassify variables, subset variables, and present data, respectively. Descriptive statistics assessments included mean, SD, median, and IQR calculations. Inferential statistics assessments were conducted at varying levels of prompt specificity ("Basic," "Intermediate," and "Advanced"). Specific tests included chi-square, Pearson correlation, independent 2-sample t test, 1-way ANOVA, Fisher exact, Spearman

correlation, Mann-Whitney U test, and Kruskal-Wallis H test. Outcomes from consecutive prompt-based trials were assessed against expected statistical values calculated in Python (Python Software Foundation), SAS (SAS Institute), and RStudio (Posit PBC). ChatGPT accurately performed data processing, categorization, and tabulation across all trials. For descriptive statistics, it provided accurate means, SDs, medians, and IQRs across all trials. Inferential statistics accuracy against expected statistical values varied with prompt specificity: 32.5% accuracy for "Basic" prompts, 81.3% for "Intermediate" prompts, and 92.5% for "Advanced" prompts. ChatGPT shows promise as a tool for exploratory data analysis, particularly for researchers with some statistical knowledge and limited programming expertise. However, its application requires careful prompt construction and human oversight to ensure accuracy. As a supplementary tool, ChatGPT can enhance data analysis efficiency and broaden research accessibility.

AI & disease detection

20. Hathaway, Quincy A. and Yashbir Singh. **Advancing Early Detection of Chronic Obstructive Pulmonary Disease Using Generative AI**. Radiology.Artificial intelligence. 2025.Vol.7(5), ppe250555. 10.1148/ryai.250555

21. Lei, Chaoyu, Kang Dang, Sifan Song, et al. **AI-assisted facial analysis in healthcare: From disease detection to comprehensive management**. Patterns (New York, N.Y.). 2025.Vol.6(2), pp101175.

Medical conditions and systemic diseases often manifest as distinct facial characteristics, making identification of these unique features crucial for disease screening. However, detecting diseases using facial photography remains challenging because of the wide variability in human facial features and disease conditions. The integration of artificial intelligence (AI) into facial analysis represents a promising frontier offering a user-friendly, non-invasive, and cost-effective screening approach. This review explores the potential of AI-assisted facial analysis for identifying subtle facial phenotypes indicative of health disorders. First, we outline the technological framework essential for effective implementation in healthcare settings. Subsequently, we focus on the role of AI-assisted facial analysis in disease screening. We further expand our examination to include applications in health monitoring, support of treatment decision-making, and disease follow-up, thereby contributing to comprehensive disease management. Despite its promise, the adoption of this technology faces several challenges, including privacy concerns, model accuracy, issues with model interpretability, biases in AI algorithms, and adherence to regulatory standards. Addressing these challenges is crucial to ensure fair and ethical use. By overcoming these hurdles, AI-assisted facial analysis can empower healthcare providers, improve patient care outcomes, and enhance global health. Imagine a world in which a simple photograph of one’s face could offer insights that could assist doctors in detecting and managing certain diseases earlier. This possibility has emerged through the integration of artificial intelligence (AI) into facial analysis. Several health conditions leave subtle clues on our faces, indicating that AI can learn to recognize with remarkable accuracy. AI can help with early disease detection, guide treatment decisions, and track health over time by analyzing facial features. This approach has the potential to significantly improve patient care and public health worldwide. However, for this technology to be truly beneficial, we must address important issues, including data privacy, model interpretability, accuracy and fairness of AI algorithms, and compliance with digital health regulations. If

these obstacles can be effectively resolved, then AI-driven facial analysis could emerge as a valuable tool in healthcare, offering advanced medical insights that are both accessible and impactful. AI is changing healthcare by enabling facial analysis for health monitoring and disease screening and management. This review highlights the potential of AI-assisted facial analysis to detect subtle facial phenotypes linked to health disorders. While promising, adoption requires addressing key challenges, including privacy, interpretability, accuracy, bias, and regulation. By overcoming these hurdles, this technology can benefit patient care, offering a non-invasive, cost-effective, and user-friendly approach to improving global health outcomes.

22. Shafi, Majid, Shabia Shabir, Sami Jan, et al. **The role of artificial intelligence in detecting avian influenza virus outbreaks: A review**. Open veterinary journal (Tripoli, Libya). 2025.Vol.15(5), pp1880–1894.

Avian influenza remains a significant threat to the global poultry industry and public health, necessitating rapid and accurate diagnostic methods. Traditional diagnostic techniques, such as serological assays and polymerase chain reaction-based methods, have proven effective, but they often lack the speed and predictive capability required for early intervention. The integration of artificial intelligence (AI) has revolutionized avian influenza detection by using machine learning models for early disease prediction and AI-driven imaging for accurate diagnosis. Additionally, AI-enhanced molecular diagnostic techniques and biosensors significantly increase the sensitivity and specificity of detecting poultry diseases. The combination of big data analytics and AI enables real-time monitoring, which improves forecasting of outbreaks and response strategies. By integrating data from various sources, such as genetic, environmental, and epidemiological information, AI enhances the early detection and risk assessment of diseases. Additionally, AI models are becoming essential for predicting how diseases might spread from animals to humans, which helps prevent infections. However, challenges such as data biases, ethical concerns, and the need for standardized protocols must be addressed to ensure responsible AI deployment. As technology progresses, AI is poised to revolutionize the management of avian influenza, providing a proactive and data-informed method for controlling diseases, ultimately protecting the health status.

AI & education

23. Arries, C., A. Patel, S. Pandey, N. Keller and K. Ta. **Integrating Artificial Intelligence (AI) in Medical Education: A Pilot Training First-Year Medical Students.** Academic pathology. 2025.Vol.12(3), pp100188. 10.1016/j.acpath.2025.100188

24. Liu, Yang, Chujun Shi, Liping Wu, et al. **Development and Validation of a Large Language Model–Based System for Medical History-Taking Training: Prospective Multicase Study on Evaluation Stability, Human-AI Consistency, and Transparency**. JMIR medical education. 2025.Vol.11 ppe73419.

History-taking is crucial in medical training. However, current methods often lack consistent feedback and standardized evaluation and have limited access to standardized patient (SP) resources. Artificial intelligence (AI)-powered simulated patients offer a promising solution; however, challenges such as human-AI consistency, evaluation stability, and transparency remain underexplored in multicase clinical scenarios. This study aimed to develop and

validate the AI-Powered Medical History-Taking Training and Evaluation System (AMTES), based on DeepSeek-V2.5 (DeepSeek), to assess its stability, human-AI consistency, and transparency in clinical scenarios with varying symptoms and difficulty levels. We developed AMTES, a system using multiple strategies to ensure dialog quality and automated assessment. A prospective study with 31 medical students evaluated AMTES's performance across 3 cases of varying complexity: a simple case (cough), a moderate case (frequent urination), and a complex case (abdominal pain). To validate our design, we conducted systematic baseline comparisons to measure the incremental improvements from each level of our design approach and tested the framework's generalizability by implementing it with an alternative large language model (LLM) Qwen-Max (Qwen AI; version 20250409), under a zero-modification condition. A total of 31 students practiced with our AMTES. During the training, students generated 8606 questions across 93 history-taking sessions. AMTES achieved high dialog accuracy: 98.6% (SD 1.5%) for cough, 99.0% (SD 1.1%) for frequent urination, and 97.9% (SD 2.2%) for abdominal pain, with contextual appropriateness exceeding 99%. The system's automated assessments demonstrated exceptional stability and high human-AI consistency, supported by transparent, evidence-based rationales. Specifically, the coefficients of variation (CV) were low across total scores (0.87%-1.12%) and item-level scoring (0.55%-0.73%). Total score consistency was robust, with the intraclass correlation coefficients (ICCs) exceeding 0.923 across all scenarios, showing strong agreement. The item-level consistency was remarkably high, consistently above 95%, even for complex cases like abdominal pain (95.75% consistency). In systematic baseline comparisons, the fully-processed system improved ICCs from 0.414/0.500 to 0.923/0.972 (moderate and complex cases), with all CVs ≤1.2% across the 3 cases. A zero-modification implementation of our evaluation framework with an alternative LLM (Qwen-Max) achieved near-identical performance, with the item-level consistency rates over 94.5% and ICCs exceeding 0.89. Overall, 87% of students found AMTES helpful, and 83% expressed a desire to use it again in the future. Our data showed that AMTES demonstrates significant educational value through its LLM-based virtual SPs, which successfully provided authentic clinical dialogs with high response accuracy and delivered consistent, transparent educational feedback. Combined with strong user approval, these findings highlight AMTES's potential as a valuable, adaptable, and generalizable tool for medical history-taking training across various educational contexts.

25. Merkebu, Jerusalem and Anita Samuel. **Humanizing AI training for health professions educators**. Med.Teach. 2025. pp1–4.

Health professions educators grapple with profound emotional burdens-technology-related distress, gnawing self-doubt, and a chilling fear of obsolescence-when urged to integrate generative AI (GenAI) into their teaching. Prevailing faculty development often overlooks this critical affective dimension, focusing on technical skills while leaving anxieties unaddressed. A six-week course, 'AI in Health Professions Education,' was developed, grounded in empathetic course design and trauma-informed pedagogy, to create a supportive and safe learning environment. Core principles of safety, transparency, collaboration, and empowerment were woven into the course. This involved validating initial emotional responses, fostering psychological safety for experimentation, instructor transparency about AI use and their own anxieties, and offering learners a choice to restore agency. Addressing affective readiness is paramount; technical competence in AI cannot flourish in the absence of acknowledged fear. Centering on human emotion and autonomy,

using customizable and psychologically safe strategies, effectively dismantles globally shared anxieties, such as the fear of inadequacy, making AI adoption feasible and empowering individuals across diverse contexts. Future efforts will include structured research into affective learning outcomes, such as the emotional trajectory of tech confidence, and the development of a toolkit for empathetically designed AI training globally, particularly in low-resource settings.

26. Veseli, Enis, Mojtaba Mehrabanian and Nour Ammar. **The potential of artificial intelligence in the early detection of systemic diseases during routine dental care**. Br.Dent.J. 2025.Vol.239(3), pp168–174.

Objective Artificial intelligence (AI) is transforming healthcare by significantly enhancing diagnostic capabilities across various disciplines. This review explores AI's potential to support dentists in identifying systemic diseases and conditions, such as skin cancer, autism spectrum disorder, hypertension, diabetes, halitosis, osteoporosis and carotid artery calcification, during routine dental care. It highlights the role of AI-based technologies in early detection and improved patient outcomes.Methods A literature search was conducted via PubMed, Google Scholar and an extensive hand search. Eligible publications were those that reported on novel AI applications for early systemic disease detection and which could be employed in the dental setting. Three researchers independently conducted research for relevant literature and extracted the data. Studies were included in consensus.Results In total, 42 eligible articles were included and their findings were summarised. Applications included AI-driven image analysis for skin cancer and tongue colour changes in diabetes, breath analysis for halitosis-linked conditions, and radiographic detection of calcifications and osteoporosis. The reviewed studies demonstrated that AI exhibits significant potential in improving diagnostic accuracy for systemic diseases observable during dental examinations, enabling timely referrals and interventions and improved healthcare integration.Conclusion The findings suggest that AI can assist dentists in identifying systemic diseases beyond oral health concerns. Integrating AI-driven diagnostic tools into routine dental care has the potential to enhance early detection of systemic conditions, reducing the burden on healthcare systems and improving patient outcomes. While the results are promising, further research is needed to refine AI technologies and validate their clinical utility. AI should be seen as a complementary tool, augmenting but not replacing the expertise of dental and medical professionals.

27. Syeda, Laveeza H., Zoya Batool, Zeeshan Hayder and Shabana Ali. **Medical undergraduate students' awareness and perspectives on artificial intelligence: A developing nation's context**. BMC Med.Educ. 2025.Vol.25(1), pp1060–10.

Artificial intelligence (AI) is reshaping healthcare, yet its integration into medical education remains limited. This study assesses undergraduate healthcare students' knowledge and perceptions of AI, its applications, challenges, and the need for AI education in healthcare curricula. Methods A cross-sectional study was conducted at Riphah International University from August to October 2023, involving 939 undergraduate students from medical, dental, pharmacy, nursing, and physical therapy disciplines. Data was collected using a validated questionnaire and analyzed using IBM SPSS Version 26. Inferential statistical test such as The Kruskal-Wallis H and Mann-Whitney U tests were applied to compare AI knowledge and perceptions across disciplines and genders. Results Results demonstrated moderate AI knowledge, with significant differences across disciplines (p = 0.039). BDS students had the

highest AI knowledge, while nursing students scored the lowest. Most students (77%) attended AI-related talks, but only 11.8% had formal AI training. Perceptions toward AI's role in patient care were generally positive, with 73.6% believing AI could aid in patient documentation and 68.7% supporting its role in selecting health interventions. Concerns were raised about AI's impact on job displacement, ethical challenges, and feasibility in developing countries. Despite this, 78.8% supported AI integration into medical curricula, and 82.2% endorsed AI training as part of medical education. Conclusion Undergraduate healthcare students recognize AI's potential in medicine but express concerns about ethical implications and job displacement. The findings highlight the need for structured AI education in medical curricula to bridge knowledge gaps and prepare future healthcare professionals for AI-driven practice.

28. Syversen, Aron, Oliver Umney, Lewis Howell, et al. **"How can we involve Patients?" - Students' perspectives on embedding PPIE into a doctoral training centre for AI in medical diagnosis and care**. Res Involv Engagem. 2025.Vol.11(1), pp77–9.

Plain English summary Artificial intelligence (AI) promises to transform healthcare research. However, patients and the public are still not widely involved or engaged within this research area. There is a growing recognition of the importance of incorporating Patient and Public Involvement and Engagement (PPIE) earlier into researcher training. Doctoral training programmes train and support cohorts of PhD students all within a similar research field and therefore may provide the perfect environment to train researchers in PPIE. This paper describes and evaluates the PPIE activities and training within the Centre for Doctoral Training (CDT) in Artificial Intelligence for Medical Diagnosis and Care ("AI-Medical"), at the University of Leeds in the United Kingdom. Authored primarily by PhD candidates from the AI-Medical CDT, it provides an overview of the PPIE activities conducted by students in the CDT between 2021 and 2024. The paper includes first-hand accounts of student experiences, evidenced by quotes, and reflects on these experiences whilst also sharing key learning outcomes. The paper also reflects on the suitability, difficulties, and benefits of including PPIE activities as part of doctoral training programmes, which both develop research leaders of the future and support the students in completing their PhDs. This is particularly important given the current lack of examples incorporating PPIE into AI research projects. It also offers some actionable recommendations for integrating PPIE into future PhD research, whether in other PhD training programmes or within individual research projects. Although written from the viewpoint of the PhD students, this paper will be of interest to patients and the public too, given the increasing use and exploration of AI in health research and therefore the need for the involvement of patients and the public in that work. Artificial intelligence (AI) in healthcare is a rapidly developing research field, but there is limited evidence that patients and public are widely engaged or involved with its progression. Alongside this, there is a growing recognition of the importance of incorporating Patient and Public Involvement and Engagement (PPIE) earlier into researcher training. Doctoral training programmes (centres) may provide the perfect environment to address both issues. This paper describes and evaluates Patient and Public Involvement and Engagement (PPIE) activities within the Centre for Doctoral Training (CDT) in Artificial Intelligence for Medical Diagnosis and Care ("AI-Medical"), at the University of Leeds in the United Kingdom. Authored primarily by PhD candidates from the AI-Medical CDT, it gives an overview of the PPIE activities conducted within the CDT, including accounts of first-hand experiences, supported by quotes and reflections from students. It also shares key learning

outcomes and makes actionable recommendations for integrating PPIE into future PhD programmes and individual research projects. These insights highlight both the successes and challenges of embedding PPIE in healthcare-focused AI research projects in a doctoral training centre.

AI & ethics

29. Barton, Antonia, Emily Hudson, Emma Lavery and Simon Clark. **8371 AI is not coming for your job (yet): especially, if that includes writing useful patient information leaflets.** Arch.Dis.Child. 2025.Vol.110 ppA231–A232.

We have been promised that AI will make our lives easier. We assessed three open AI platforms' generation of parental information leaflets for paediatric conditions.MethodWe performed a structured qualitative analysis using the 'Suitability Assessment of Materials (SAM) framework' for health-related educational resources. SAM consists of 6 criteria: 'content'(information to help solve problems); 'literacy demand'(common words are used); 'graphics'(simple appropriate drawings); 'layout and typography'(type size 12 point etc); 'learning stimulation and motivation'(complex topics are subdivided into small parts); and 'cultural appropriateness'(examples present the culture in positive ways').We randomly sourced NHS information leaflets for parents in 13 paediatric conditions (gastroenteritis; jaundice; asthma; eczema; scarlet fever; UTI; meningitis; cleft lip and palate; fever; abdominal pain; headache; sepsis; and chickenpox) from hospitals in England regions, Scotland and Wales; no leaflets from Northern Ireland were found. Using a random number generator, 2 leaflets from England (North), 2 from England (South), 1 from Wales, and 1 from Scotland were sourced for each condition, finding 73 NHS leaflets. Using the prompt: 'Create an information leaflet for a parent of a child with insert condition]','ChatGPT™','Google Gemini™', and 'Microsoft Copilot AI™' were each asked to produce a leaflet for each condition making 39 AI leaflets.The SAM scores for NHS's and the 3 AI's leaflets were compared.ResultsWe found that only 1 in 5 Trusts/Health Boards produced a leaflet available for each condition.Overall NHS leaflets scored better with a mean (95% confidence intervals of the mean) 77.7%(75.6–79.9). The overall score for AI leaflets was 67.4%(63.8–70.8)(p<0.0001).The NHS leaflets scored statistically significantly higher than AI for:'Content' 88.4%(85.8%-91.0%) versus 75.0%(72.3–77.7)(p<0.05);'Graphics' 65.7%(60.7–69.7) versus 46.7%(43.0–50.4)(p<0.05);'Cultural appropriateness' 57.0%(51.3–63.7) versus 37.0%(31.0–43.0)(p<0.05).The AI leaflets scored statistically significantly higher for 'Literacy demand'86.1%(83.5–88.7) versus NHS 78.0%(74.5–81.5)(p<0.05).There were no differences for 'Layout and typography' and 'Learning stimulation and motivation'.ChatGPT™ performed statistically significantly better overall than Copilot™ and Gemini™: 77.1%(71.6–82.6) versus 62.1%(59.5–64.6)(p=0.0004) versus 62.9%(57.2–68.6)(p=0.003). There was no difference in scores between Copilot™ and Gemini™ There was no difference in scores between ChatGPT™ overall and NHS leaflets, but the 'Content' score for ChatGPT™ was lower 71.5%(64.1%-79.1%)(p=0.0001).Conclusion Open AI cannot yet write parent information leaflets. Even if you used ChatGPT™ significant changes are needed for the required NHS standard.AI leaflets lacked detail and graphics; most missed at least one crucial informational element. None of the AIs generated or incorporated pictures independently.AI scored higher for literacy demand, due to consistent use of short sentences, bullet points, and basic vocabulary.More detailed AI prompts might result in creation of better leaflets. Scoring could not be blinded, but the SAM framework

should have mitigated that. As only 1 in 5 Trusts/Health Boards produced a leaflet, it might be in the best interest of patient safety and improving educational health services, to adapt another centre's leaflets rather than rely on AI. Future studies need to include user assessment.

30. Kim, Jee Young, Alifia Hasan, Jacqueline Kueper, et al. **Establishing organizational AI governance in healthcare: a case study in Canada**. npj Digit.Med. 2025.Vol.8(1), pp522–8.

This research applies the People, Process, Technology, and Operations (PPTO) framework to develop AI governance within a large hospital system in Canada that is early in AI adoption. Stakeholder interviews identified the organization's strengths, gaps, and priorities for AI governance, providing foundational insights into the organization's readiness and needs. Co-design workshops then adapted the PPTO framework to the organization's specific context. Together, these efforts led to the creation of policies and the formation of an AI governance committee within the organization. This work demonstrates that the PPTO framework is a practical and adaptable tool for developing AI governance in real-world healthcare settings. It also addresses a critical gap in the field by generating empirical evidence of how a conceptual AI governance framework can be implemented within healthcare delivery organizations to drive organizational change.

AI & healthcare

31. Faiyazuddin, Md, Syed Jalal Q. Rahman, Gaurav Anand, et al. **The Impact of Artificial Intelligence on Healthcare: A Comprehensive Review of Advancements in Diagnostics, Treatment, and Operational Efficiency**. Health science reports. 2025.Vol.8(1), ppe70312–n/a. ABSTRACT

Background and Aims Artificial Intelligence (AI) beginning to integrate in healthcare, is ushering in a transformative era, impacting diagnostics, altering personalized treatment, and significantly improving operational efficiency. The study aims to describe AI in healthcare, including important technologies like robotics, machine learning (ML), deep learning (DL), and natural language processing (NLP), and to investigate how these technologies are used in patient interaction, predictive analytics, and remote monitoring. The goal of this review is to present a thorough analysis of AI's effects on healthcare while providing stakeholders with a road map for navigating this changing environment. Methods This review analyzes the impact of AI on healthcare using data from the Web of Science (2014–2024), focusing on keywords like AI, ML, and healthcare applications. It examines the uses and effects of AI on healthcare by synthesizing recent literature and real-world case studies, such as Google Health and IBM Watson Health, highlighting AI technologies, their useful applications, and the difficulties in putting them into practice, including problems with data security and resource limitations. The review also discusses new developments in AI, and how they can affect society. Results The findings demonstrate how AI is enhancing the skills of medical professionals, enhancing diagnosis, and opening the door to more individualized treatment plans, as reflected in the steady rise of AI-related healthcare publications from 158 articles (3.54%) in 2014 to 731 articles (16.33%) by 2024. Core applications like remote monitoring and predictive analytics improve operational effectiveness and patient involvement. However, there are major obstacles to the mainstream implementation of AI in healthcare, including issues with data security and

budget constraints. Conclusion Healthcare may be transformed by AI, but its successful use requires ethical and responsible use. To meet the changing demands of the healthcare sector and guarantee the responsible application of AI technologies, the evaluation highlights the necessity of ongoing research, instruction, and multidisciplinary cooperation. In the future, integrating AI responsibly will be essential to optimizing its advantages and reducing related dangers.

32. He, Taotao, Jing Huang, Yongheng Li, et al. **The mediation effect of AI self-efficacy between AI literacy and learning engagement in college nursing students: A cross-sectional study**. Nurse education in practice. 2025.Vol.87 pp104499.

This study investigates the mediating role of artificial intelligence self-efficacy (AISE) in the association between AI literacy and learning engagemeg nursing students in higher education. The fourth technological revolution driven by AI has profound implications for healthcare education. Insufficient AI literacy and AISE among nursing students may hinder their learning engagement, yet empirical evidence on these relationships remains limited. A cross-sectional study. A convenience sampling method was used to select 2029 nursing students from 29 colleges in China from November to December 2024. A general information questionnaire, AI literacy scale, AISE scale and learning engagement scale were used to conduct the online questionnaire survey. Data were analyzed using descriptive analysis, Pearson correlation test and Bootstrap method. Results show that AI literacy was positively correlated with AISE and learning engagement ($r = 0.462–0.435$, $P < 0.01$) and AISE was positively correlated with learning engagement ($r = 0.537$, $P < 0.01$). The indirect effect of AI literacy on learning engagement through AISE was 0.130, accounting for 20.44 % of the total effect. The AI literacy of nursing students is suboptimal; thus, enhancing foundational AI education is essential. AI literacy directly predicts learning engagement, with AISE serving as a partial mediator. Educators should integrate AI training into curricula, provide practical scenarios and foster interdisciplinary collaboration to improve AI competency and engagement. 10.1016/j.nepr.2025.104499

33. Morley, Jessica, Eleanor Barry and Lucinda Hiam. **Can a digital NHS be equitable?**. BMJ. 2025.Vol.389 ppr1317. Infrastructure and inclusion are key to the rollout of AI. 10.1136/bmj.r1317

34. Perlis, Roy. **Changing Opinions About AI in Health Care**. JAMA : the journal of the American Medical Association. 2025.Vol.334(9), pp777. 10.1001/jama.2025.12974

35. Stokel-Walker, Chris. **Artificial intelligence in medicine: How do experts think AI could transform the NHS?**. BMJ. 2025.Vol.388 ppr248.

AI presents some exciting opportunities for the NHS, but leaders need to drive the technology and not the other way round, reports Chris Stokel-Walker.

AI & learning/information

36. Pal, Avishek, Tenzin Wangmo, Trishna Bharadia, et al. **Generative AI/LLMs for Plain Language Medical Information for Patients, Caregivers and General Public: Opportunities, Risks and Ethics**.

Patient preference and adherence. 2025.Vol.19 pp2227–2249. Generative artificial intelligence (gAI) tools and large language models (LLMs) are gaining popularity among non-specialist audiences (patients, caregivers, and the general public) as a source of plain language medical information. AI-based models have the potential to act as a convenient, customizable and easy-to-access source of information that can improve patients' self-care and health literacy and enable greater engagement with clinicians. However, serious negative outcomes could occur if these tools fail to provide reliable, relevant and understandable medical information. Herein, we review published findings on opportunities and risks associated with such use of gAI/LLMs. We reviewed 44 articles published between January 2023 and July 2024. From the included articles, we find a focus on readability and accuracy; however, only three studies involved actual patients. Responses were reported to be reasonably accurate and sufficiently readable and detailed. The most commonly reported risks were oversimplification, over-generalization, lower accuracy in response to complex questions, and lack of transparency regarding information sources. There are ethical concerns that overreliance/unsupervised reliance on gAI/LLMs could lead to the "humanizing" of these models and pose a risk to patient health equity, inclusiveness and data privacy. For these technologies to be truly transformative, they must become more transparent, have appropriate governance and monitoring, and incorporate feedback from healthcare professionals (HCPs), patients, and other experts. Uptake of these technologies will also need education and awareness among non-specialist audiences around their optimal use as sources of plain language medical information.

37. Rodger, Daniel, Sebastian Porsdam Mann, Brian Earp, Julian Savulescu, Christopher Bobier and Bruce P. Blackshaw. **Generative AI in healthcare education: How AI literacy gaps could compromise learning and patient safety**.

Nurse education in practice. 2025.Vol.87 pp104461. To examine the challenges and opportunities presented by generative artificial intelligence in healthcare education and explore how it can be used ethically to enhance rather than compromise future healthcare workforce competence. Generative artificial intelligence is fundamentally changing healthcare education, yet many universities and healthcare educators have failed to keep pace with its rapid development. A discussion paper. Discussion and analysis of the challenges and opportunities presented by students' increasing use of generative artificial intelligence in healthcare education, with particular focus on assessment approaches, critical thinking development and artificial intelligence literacy. Students' widespread use of generative artificial intelligence threatens assessment integrity and may inhibit critical thinking, problem-solving skills and knowledge acquisition. Without adequate artificial intelligence literacy there is a risk of eroding future healthcare workforce competence and compromising patient safety and professional integrity. While generative artificial intelligence presents significant challenges to healthcare education, it offers great promise if used carefully with awareness of its limitations. The development of artificial intelligence literacy is crucial for maintaining professional standards and ensuring patient safety and mitigating its potentially negative impact on the formation of critical thinking skills.

38. Yun, Hye Sun and Timothy Bickmore. **Online Health Information–Seeking in the Era of Large Language Models: Cross-Sectional Web-Based Survey Study**. Journal of medical Internet research. 2025.Vol.27(7), ppe68560.

As large language model (LLM)-based chatbots such as ChatGPT (OpenAI) grow in popularity, it is essential to understand their role in delivering online health information compared to other resources. These chatbots often generate inaccurate content, posing potential safety risks. This motivates the need to examine how users perceive and act on health information provided by LLM-based chatbots. This study investigates the patterns, perceptions, and actions of users seeking health information online, including LLM-based chatbots. The relationships between online health information-seeking behaviors and important sociodemographic characteristics are examined as well. A web-based survey of crowd workers was conducted via Prolific. The questionnaire covered sociodemographic information, trust in health care providers, eHealth literacy, artificial intelligence (AI) attitudes, chronic health condition status, online health information source types, perceptions, and actions, such as cross-checking or adherence. Quantitative and qualitative analyses were applied. Most participants consulted search engines (291/297, 98%) and health-related websites (203/297, 68.4%) for their health information, while 21.2% (63/297) used LLM-based chatbots, with ChatGPT and Microsoft Copilot being the most popular. Most participants (268/297, 90.2%) sought information on health conditions, with fewer seeking advice on medication (179/297, 60.3%), treatments (137/297, 46.1%), and self-diagnosis (62/297, 23.2%). Perceived information quality and trust varied little across source types. The preferred source for validating information from the internet was consulting health care professionals (40/132, 30.3%), while only a very small percentage of participants (5/214, 2.3%) consulted AI tools to cross-check information from search engines and health-related websites. For information obtained from LLM-based chatbots, 19.4% (12/63) of participants cross-checked the information, while 48.4% (30/63) of participants followed the advice. Both of these rates were lower than information from search engines, health-related websites, forums, or social media. Furthermore, use of LLM-based chatbots for health information was negatively correlated with age (ρ=-0.16, P=.006). In contrast, attitudes surrounding AI for medicine had significant positive correlations with the number of source types consulted for health advice (ρ=0.14, P=.01), use of LLM-based chatbots for health information (ρ=0.31, P<.001), and number of health topics searched (ρ=0.19, P<.001). Although traditional online sources remain dominant, LLM-based chatbots are emerging as a resource for health information for some users, specifically those who are younger and have a higher trust in AI. The perceived quality and trustworthiness of health information varied little across source types. However, the adherence to health information from LLM-based chatbots seemed more cautious compared to search engines or health-related websites. As LLMs continue to evolve, enhancing their accuracy and transparency will be essential in mitigating any potential risks by supporting responsible information-seeking while maximizing the potential of AI in health contexts.

AI & mental health

39. Fischer, Leo, Paula Antonia Mann, Minh-Hieu H. Nguyen, et al. **AI for mental health: clinician expectations and priorities in computational psychiatry**. BMC Psychiatry. 2025.Vol.25(1), pp584–8.

Mental disorders represent a major global health challenge, with an estimated lifetime prevalence approaching 30%. Despite the availability of effective treatments, access to mental health care remains inadequate. Computational psychiatry, leveraging advancements in artificial intelligence (AI) and machine learning (ML), has shown potential for transforming mental health care by improving diagnosis, prognosis, and the personalization of treatment. However, integrating these technologies into routine clinical practice remains limited due to technical and infrastructure challenges. While ongoing computational developments will enhance AI's precision, many studies focus on its broad

potential without providing specific, clinician-informed guidance for immediate application. To address this gap and the urgent need for clinically actionable AI tools, we surveyed 53 psychiatrists and clinical psychologists to identify their priorities for AI in mental health care. Our findings reveal a strong preference for tools enabling continuous monitoring and predictive modeling, particularly in outpatient settings. Clinicians prioritize accurate predictions of symptom trajectories and proactive patient monitoring over interpretability and explicit treatment recommendations. Self-reports, third-party observations, and sleep quality and duration emerged as key data inputs for effective models. Together, this study provides a clinician-driven roadmap for AI integration, emphasizing predictive models based on ecological momentary assessment (EMA) data to forecast disorder trajectories and support real-world practice.

AI & nursing

40. Joo, Jee Young, Megan F. Liu and Mu-Hsing Ho. **Nurses' perceptions of artificial intelligence adoption in healthcare: A qualitative systematic review**. Nurse education in practice. 2025.Vol.88 pp104542.

This study aimed to systematically review and synthesize the most recent qualitative studies on frontline nurses' insights and perspectives regarding the use of artificial intelligence (AI) tools in their clinical practice in hospital settings. There is limited information on frontline nurses' perceptions, attitudes and expectations regarding the adoption of AI in healthcare. A systematic review and thematic synthesis of qualitative evidence was conducted. A systematic search was conducted across five electronic databases—CINAHL, PubMed, Web of Science, Ovid and Science Direct—to identify qualitative studies published between January 2020 and December 2024. After selecting studies, a thematic synthesis was performed. This review followed the PRISMA checklist, was registered with PROSPERO and included a quality appraisal of the retrieved studies. Nine qualitative studies published between 2022 and 2024 were included in this systematic review. The included studies were conducted in five countries at university-affiliated or tertiary hospitals. Participants included 140 frontline nurses and nurse managers with prior experience using technology tools in clinical settings. Five common themes were identified: ethical issues; increased workload; seamless and efficient patient care; reinforcement rather than replacement; and AI as a future nursing care solution The five themes identified in this review provide valuable insights into how AI tools can be integrated into current and future frontline nursing practices to enhance patient care. Nurse leaders and healthcare policymakers can use these findings to improve nursing research, facilitate the adoption of new AI tools and support their implementation in healthcare settings. •There is limited understanding of the challenges faced by frontline nurses in adopting AI tools for nursing care.•This review identified common themes regarding frontline nurses' perceptions of AI utilization in patient care.•Nurses identified AI-driven healthcare inequity as the most profound ethical concern.•Nurses are convinced that they have the intrinsic ability to provide human-centered care.•Healthcare leaders should fund more research on AI effectiveness in improving nursing management and patient outcomes.

AI & orthopaedics

41. Van Eecke, Eduard, Wouter Schroven, Maxim Vanderstappen, Simran Grewal and Michel van den Bekerom. **Appraisal of ChatGPT's responses to common patient**

questions regarding acromioclavicular joint dislocations. JSES reviews, reports, and techniques. 2025.

ChatGPT, an expanding artificial intelligence platform, is rapidly becoming a source of medical knowledge for patients. The purpose of this study is to evaluate the quality and readability of information provided by ChatGPT 4.0 in response to the most frequently asked patients' questions regarding acromioclavicular (AC) joint dislocations. Twenty-five frequently asked patient questions regarding AC joint dislocations were posed to ChatGPT 4.0. The quality and accuracy of the responses was graded by two fellowship-trained shoulder surgeons using a 5-point Likert scale. Responses were analyzed for readability using six established metrics. ChatGPT provided responses to 25 frequently asked questions, with consensus reached on 12 questions. The final average score was 4/5, reflecting good quality with minor inaccuracies. Treatment-related questions scored lower compared to other categories, with a significant difference observed between treatment and rehabilitation (P = .025). Cohen's kappa indicated poor correlation (0.085), though the percent agreement was 48%, increasing to 100% when allowing for a 1-point difference. Readability scores revealed moderate difficulty levels, suitable for a high school-level audience. ChatGPT delivers accurate and easily comprehensible information on AC joint dislocations, highlighting its potential to improve patient education. Although the model generally provides high-quality responses, its limitations in addressing treatment-related questions underscore the importance of clinician oversight. ChatGPT can therefore serve as a valuable complement to traditional patient education methods.

AI & paediatrics

42. Boztas, Asya Eylem and Esra Ensari. **Comperative analysis of three chatbot responses on pediatric primary nocturnal enuresis**. Journal of pediatric urology. 2025.

The purpose of the study was to evaluate both the accuracy and reproducibility of the answers given by ChatGPT-4o®, Gemini® and Copilot® to frequently asked questions about pediatric primary enuresis nocturna. Forty frequently asked questions about primary nocturnal enuresis were asked 2 times, one week apart, on ChatGPT-4o, Gemini and Copilot. One of each pediatric surgeon and nephrologist independently scored the answers into 4 groups: comprehensive/correct (1), incomplete/partially correct (2), a mix of accurate and inaccurate/misleading (3), and completely inaccurate/irrelevant (4). The accuracy and reproducibility of each chatbots answers were evaluated. In comparison of these most common used chatbots, the order of completely correct response rates from highest to lowest was Chat GPT-4o and followed by Copilot and Gemini. With an accuracy percentage of 92.5 %, ChatGPT-4o gave the most accurate responses of any AI chatbot. Gemini answered 50 % of questions correctly. Copilot was the weakest successful chatbot in answering questions about enuresis nocturna with 45 % of completely accurate answer ratio. Besides Copilot has a ratio of 2.5 % for completely inaccurate/irrelevant response. Reproducibility of ChatGPT-4o, Gemini and Copilots were 85 %, 77.5 %, 70 % respectively. ChatGPT-4o is more successful in providing a high percentage of accurate responses regarding nocturnal enuresis. Both patients and their parents can use it, especially for simple, low-complexity medical questions. However, it should be used alongside expert healthcare proffesional. 10.1016/j.jpurol.2025.04.031

43. Kamal, Ahmed Hassan. **AI Chatbots in Pediatric Orthopedics: How Accurate Are Their Answers to Parents' Questions on Bowlegs and Knock Knees?**. Healthcare (Basel). 2025.Vol.13(11), pp1271.

Background/Objectives: Large-language modules facilitate accessing health information instantaneously. However, they do not provide the same level of accuracy or detail. In pediatric orthopedics, where parents have urgent concerns regarding knee deformities (bowlegs and knock knees), the accuracy and dependability of these chatbots can affect parent decisions to seek treatment. The goal of this study was to analyze how AI chatbots addressed parental concerns regarding pediatric knee deformities. Methods: A set of twenty standardized questions, consisting of ten questions each on bowlegs and knock knees, were designed through literature reviews and through analysis of parental discussion forums and expert consultations. Each of the three chatbots (ChatGPT, Gemini, and Copilot) was asked the same set of questions. Five pediatric orthopedic surgeons were then asked to rate each response for accuracy, clarity, and comprehensiveness, along with the degree of misleading information provided, on a scale of 1–5. The reliability among raters was calculated using intraclass correlation coefficients (ICCs), while differences among the chatbots were assessed using a Kruskal–Wallis test with post hoc pairwise comparisons. Results: All three chatbots displayed a moderate-to-good score for inter-rater reliability. ChatGPT and Gemini's scores were higher for accuracy and comprehensiveness than Copilot's ($p < 0.05$). However, no notable differences were found in clarity or in the likelihood of giving incorrect answers. Overall, more detailed and precise responses were given by ChatGPT and Gemini, while, with regard to clarity, Copilot performed comparably but was less thorough. Conclusions: There were notable discrepancies in performance across the AI chatbots in providing pediatric orthopedic information, which demonstrates indications of evolving potential. In comparison to Copilot, ChatGPT and Gemini were relatively more accurate and comprehensive. These results highlight the persistent requirement for real-time supervision and stringent validation when employing chatbots in the context of pediatric healthcare.

44. Olawade, David B., Kusal Weerasinghe, Jennifer Teke, Maines Msiska, Stergios Boussios and Eleni Hatzidimitriadou. **Evaluating AI adoption in healthcare: Insights from the information governance professionals in the United Kingdom**. International journal of medical informatics (Shannon, Ireland). 2025.Vol.199 pp105909.

Artificial Intelligence (AI) is increasingly being integrated into healthcare to improve diagnostics, treatment planning, and operational efficiency. However, its adoption raises significant concerns related to data privacy, ethical integrity, and regulatory compliance. While much of the existing literature focuses on the clinical applications of AI, limited attention has been given to the perspectives of Information Governance (IG) professionals, who play a critical role in ensuring responsible and compliant AI implementation within healthcare systems. This study aims to explore the perceptions of IG professionals in Kent, United Kingdom, on the use of AI in healthcare delivery and research, with a focus on data governance, ethical considerations, and regulatory implications. A qualitative exploratory design was employed. Six IG professionals from NHS trusts in Kent were purposively selected based on their roles in compliance, data governance, and policy enforcement. Semi-structured interviews were conducted and thematically analysed using NVivo software, guided by the Unified Theory of Acceptance and Use of Technology (UTAUT).

Thematic analysis revealed varying levels of AI knowledge among IG professionals. While participants acknowledged AI's potential to improve efficiency, they raised concerns about data accuracy, algorithmic bias, cybersecurity risks, and unclear regulatory frameworks. Participants also highlighted the importance of ethical implementation and the need for national oversight. AI offers promising opportunities in healthcare, but its adoption must be underpinned by robust governance structures. Enhancing AI literacy among IG teams and establishing clearer regulatory frameworks will be key to safe and ethical implementation.

AI & patient opinion

45. Ogundare, Oluwatosin, Tolu Owadokun, Temitope Ogundare, Promise Ekpo, Ha Linh Nguyen and Stephen Bello. **Integrated artificial intelligence in healthcare and the patient's experience of care**. Sci Rep. 2025.Vol.15(1), pp21879–12.

Healthcare is plagued with many problems that Artificial Intelligence (AI) can ameliorate or sometimes amplify. Regardless, AI is changing the way we reason towards solutions, especially at the frontier of public health applications where autonomous and co-pilot AI integrated systems are now rapidly adopted for mainstream use in both clinical and non-clinical settings. In this regard, we present empirical analysis of thematic concerns that affect patients within AI integrated healthcare systems and how the experience of care may be influenced by the degree of AI integration. Furthermore, we present a fairly rigorous mathematical model and adopt prevailing techniques in Machine Learning (ML) to develop models that utilize a patient's general information and responses to a survey to predict the degree of AI integration that will maximize their experience of care. We model the patient's experience of care as a continuous random variable on the open interval ( ) and refer to it as the AI Affinity Score which encapsulates the degree of AI integration that the patient prefers within a chosen healthcare system. We present descriptive statistics of the distribution of the survey responses over key demographic variables viz. Age, Gender, Level of Education as well as a summary of perceived attitudes towards AI integrated healthcare in these categories. We further present the results of statistical tests conducted to determine if the variance across distributions of AI Affinity Scores over the identified groups are statistically significant and further assess the behavior of any independent distribution of AI Affinity Scores using a Bayesian nonparametric model.

46. Rafaï, Ismaël, B. Davin-Casalena, Dimitri Dubois, Thierry Blayac and Bruno Ventelou. **Public hesitancy for AI-based detection of neurodegenerative diseases in France**. Sci Rep. 2025.Vol.15(1), pp26849–9.

Recent advances in artificial intelligence (AI) have made it possible to detect neurodegenerative diseases (NDDs) earlier, potentially improving patient outcomes. However, AI-based detection tools remain underutilized. We studied individual valuation for early diagnosis tests for NDDs. We conducted a discrete choice experiment with a representative sample of the French adult population (N = 1017). Participants were asked to choose between early diagnosis tests that differed in terms of: (1) type of test (saliva vs. AI-based tests analysing electronic health records); (2) identity of the person communicating the test results; (3) sensitivity; (4) specificity; and (5) price. We calculated the weights in the decision for each attribute and examined how socio-demographic characteristics influenced them. Respondents revealed a reduced utility value when AI-based testing was involved (valuated at an average of €36.08, CI €22.13; €50.89]) and when results were communicated

by a private company (€95.15, CI €82.01; €109.82]). We interpret these figures as the shadow price that the public attaches to medical data privacy. Beyond monetization, our representative sample of the French population appears reluctant to adopt AI-powered screening, particularly when performed on large sets of personal data. However, they would be more supportive when medical expertise is associated with the tests.

AI & public health

47. Muralidharan, Vijaytha, Madelena Y. Ng, Shada AlSalamah, et al. **Global Initiative on AI for Health (GI-AI4H): strategic priorities advancing governance across the United Nations**. npj Digit.Med. 2025.Vol.8(1), pp219–4.

The Global Initiative on Artificial Intelligence for Health (GI-AI4H), established by the World Health Organization, serves to harmonize governance standards for artificial intelligence (AI). The GI-AI4H spearheads novel on-the-ground efforts, especially in low- and middle-income countries, to advance ethical, regulatory, implementation, and operational dimensions of global governance for health AI. The GI-AI4H's efforts across the United Nations drives safe, ethical, equitable, and sustainable health AI use for the global community.

48. Mulshine, James L., Ricardo S. Avila, Mario Silva, et al. **AI integrations with lung cancer screening: Considerations in developing AI in a public health setting**. European journal of cancer (1990). 2025.Vol.220 pp115345.

Lung cancer screening implementation has led to expanded imaging of the chest in older, tobacco-exposed populations. Growing numbers of screening cases are also found to have CT-detectable emphysema or elevated levels of coronary calcium, indicating the presence of coronary artery disease. Early interventions based on these additional findings, especially with coronary calcium, are emerging and follow established protocols. Given the pace of diagnostic innovation and the potential public health impact, it is timely to review issues in developing useful chest CT screening infrastructure as chest CT screening will soon involve millions of participants worldwide. Lung cancer screening succeeds because it detects curable, early primary lung cancer by characterizing and measuring changes in non-calcified, lung nodules in the size-range from 3mm to 15 mm in diameter. Therefore, close attention to imaging methodology is essential to lung screening success and similar image quality issues are required for reliable quantitative characterization of early emphysema and coronary artery disease. Today's emergence of advanced image analysis using artificial intelligence (AI) is disrupting many aspects of medical imaging including chest CT screening. Given these emerging technological and volume trends, a major concern is how to balance the diverse needs of parties committed to building AI tools for precise, reproducible, and economical chest CT screening, while addressing the public health needs of screening participants receiving this service. A new consortium, the Alliance for Global Implementation of Lung and Cardiac Early Disease Detection and Treatment (AGILEDxRx) is committed to facilitate broad, equitable implementation of multi-disciplinary, high quality chest CT screening using advanced computational tools at accessible cost. •Artificial Intelligence has potential to facilitate chest CT screening implementation.•Chest CT screening detects lung cancer, emphysema and coronary artery calcification.•Chest CT screening involves processes from image acquisition, analysis to decision support.•AI can

optimize image measurement and facilitate radiological reporting.•Critical AI tools for screening quality needed for all.

49. Nivethitha, V., R. A. Daniel, B. N. Surya and G. Logeswari. **Empowering public health: Leveraging AI for early detection, treatment, and disease prevention in communities - A scoping review**. Journal of postgraduate medicine (Bombay). 2025.Vol.71(2), pp74–81.

ABSTRACT India's healthcare system faces substantial challenges, including a high burden of communicable and non-communicable diseases, limited access to healthcare in rural areas, and a shortage of skilled healthcare professionals. Artificial intelligence (AI) offers promising solutions to address these gaps by enhancing diagnostic accuracy, improving disease prediction, and optimizing treatment management. This scoping review examines AI's role in early detection, treatment, and disease prevention in community health settings. A comprehensive literature search was conducted in PubMed, Embase, Scopus, and Google Scholar from January 2013 to July 2024. Eligible studies focused on the application of AI in public health, emphasizing early detection, disease prevention, and treatment interventions. Data on AI models, health outcomes, and performance metrics were extracted and analyzed in line with PRISMA-ScR guidelines. Forty-eight studies were analyzed and categorized into diagnostic accuracy, disease prediction, treatment management, and clinical validation. AI-based tools, such as AIDMAN for malaria detection, demonstrated high diagnostic accuracy (95%) and AUC (0.96). Predictive models for chronic kidney disease (93% accuracy) and diabetes (91% accuracy) showed substantial promise. TB screening using AI-powered cough analysis achieved 86% accuracy. The studies also emphasized AI's role in managing chronic diseases, facilitating early interventions, and reducing healthcare burdens in resource-limited settings. AI has the potential to revolutionize healthcare delivery in India, particularly in underserved regions, by enhancing early detection and treatment. However, challenges related to data privacy, algorithmic bias, and infrastructure require attention. Continued research and policy development are essential to fully harness AI's capabilities in improving public health outcomes.

AI & radiology

50. Schlemmer, Heinz-Peter. **Navigating the AI revolution: will radiology sink or soar?** Japanese journal of radiology. 2025.

The rapid acceleration of digital transformation and artificial intelligence (AI) is fundamentally reshaping medicine. Much like previous technological revolutions, AI—driven by advances in computer technology and software including machine learning, computer vision, and generative models—is redefining cognitive work in healthcare. Radiology, as one of the first fully digitized medical specialties, is at the forefront of this transformation. AI is automating workflows, enhancing image acquisition and interpretation, and improving diagnostic precision, which collectively boost efficiency, reduce costs, and elevate patient care. Global data networks and AI-powered platforms are enabling borderless collaboration, empowering radiologists to focus on complex decision-making and patient interaction. Despite these profound opportunities, widespread AI adoption in radiology remains limited, often confined to specific use cases, such as chest, neuro, and musculoskeletal imaging. Concerns persist regarding transparency, explainability, and the ethical use of AI systems,

while unresolved questions about workload, liability, and reimbursement present additional hurdles. Psychological and cultural barriers, including fears of job displacement and diminished professional autonomy, also slow acceptance. However, history shows that disruptive innovations often encounter initial resistance. Just as the discovery of X-rays over a century ago ushered in a new era, today, digitalization and artificial intelligence will drive another paradigm shift—this time through cognitive automation. To realize AI's full potential, radiologists must maintain clinical oversight and safeguard their professional identity, viewing AI as a supportive tool rather than a threat. Embracing AI will allow radiologists to elevate their profession, enhance interdisciplinary collaboration, and help shape the future of medicine. Achieving this vision requires not only technological readiness but also early integration of AI education into medical training. Ultimately, radiology will not be replaced by AI, but by radiologists who effectively harness its capabilities.

AI & screening

51. Cold, Kristoffer Mazanti, Amaan Ali, Lars Konge, Flemming Bjerrum, Laurence Lovat and Omer Ahmad. **Bowel preparation assessment using artificial intelligence: Systematic review**. Endoscopy International Open. 2025.Vol.13

Insufficient bowel preparation is the leading cause of missed adenomas in colonoscopy. The Boston Bowel Preparation Scale (BBPS) is the most thoroughly validated and widely used scale to estimate risk of missed adenomas. Artificial intelligence (AI) could automatically quantify bowel preparation, thus reducing bias and limitations inherent in human rating. This systematic review aimed to identify, describe, and evaluate all AI-BPS systems for colonoscopy. A systematic literature review was conducted using MEDLINE, EMBASE, and SCOPUS based on three sets of terms aligned with the inclusion criteria: colonoscopy, BPS, and AI. Two reviewers independently evaluated and completed data extraction from the articles. A total of 1,449 studies were identified, with eight meeting the eligibility criteria. Six AI-BPS systems were trained on expert BBPS ratings, and two studies used a fecal-mucosal ratio. All studies compared their AI-BPS with expert BBPS ratings; two showed that their AI-BPS outperformed expert BBPS ratings, and six showed comparable performances. Three studies also demonstrated correlations with adenoma detection rates (ADRs), adenoma miss rates (AMRs), or polyp detection rates (PDRs). Only one prospective study implemented its AI-BPS, finding lower AMR in adequately prepared compared with inadequately prepared bowels. AI-BPS can standardize and outperform human bowel preparation evaluation by better correlating with expert BBPS ratings, AMR, ADR, and PDR. Further research following recommended reporting guidelines is needed to allow for cross-study comparisons and meta-analysis, which was not possible in this study due to heterogonous study design and reporting metrics.

52. Lim, Daniel Yan Zheng, Yu Bin Tan, Jonas Ren Yi Ho, et al. **Vision-language large learning model, GPT4V, accurately classifies the Boston Bowel Preparation Scale score**. BMJ Open Gastroenterol. 2025.Vol.12(1), ppe001496.

Large learning models (LLMs) such as GPT are advanced artificial intelligence (AI) models. Originally developed for natural language processing, they have been adapted for multi-modal tasks with vision-language input. One clinically relevant task is scoring the Boston Bowel Preparation Scale (BBPS). While traditional AI techniques use large amounts of data for training, we hypothesise that vision-language LLM can perform this task with fewer

examples. MethodsWe used the GPT4V vision-language LLM developed by OpenAI, via the OpenAI application programming interface. A standardised prompt instructed the model to grade BBPS with contextual references extracted from the original paper describing the BBPS by Lai et al (GIE 2009). Performance was tested on the HyperKvasir dataset, an open dataset for automated BBPS grading.ResultsOf 1794 images, GPT4V returned valid results for 1772 (98%). It had an accuracy of 0.84 for two-class classification (BBPS 0–1 vs 2–3) and 0.74 for four-class classification (BBPS 0, 1, 2, 3). Macro-averaged F1 scores were 0.81 and 0.63, respectively. Qualitatively, most errors arose from misclassification of BBPS 1 as 2. These results compare favourably with current methods using large amounts of training data, which achieve an accuracy in the range of 0.8–0.9.ConclusionThis study provides proof-of-concept that a vision-language LLM is able to perform BBPS classification accurately, without large training datasets. This represents a paradigm shift in AI classification methods in medicine, where many diseases lack sufficient data to train traditional AI models. An LLM with appropriate examples may be used in such cases.

AI & scribe technology

53. Biro, Joshua, Jessica L. Handley, Nathan K. Cobb, et al. **Accuracy and Safety of AI-Enabled Scribe Technology: Instrument Validation Study**. Journal of medical Internet research. 2025.Vol.27(3), ppe64993.

Artificial intelligence-enabled ambient digital scribes may have many potential benefits, yet results from our study indicate that there are errors that must be evaluated to mitigate safety risks.Artificial intelligence-enabled ambient digital scribes may have many potential benefits, yet results from our study indicate that there are errors that must be evaluated to mitigate safety risks. 10.2196/64993

54. Eccles, Abi, Tom Pelly, Catherine Pope and John Powell. **Unintended consequences of using ambient scribes in general practice**. BMJ. 2025.Vol.390 ppe085754.

Abi Eccles and colleagues argue that better evidence is crucial to ensure widespread adoption of ambient scribes does not compromise quality of care. 10.1136/bmj-2025-085754

55. Garcia Sanchez, Carolina, Anna Kharko, Maria Hägglund, Sara Riggare and Charlotte Blease. **Health Care Professionals' Experiences and Opinions About Generative AI and Ambient Scribes in Clinical Documentation: Protocol for a Scoping Review**. JMIR research protocols. 2025.Vol.14 ppe73602.

Generative artificial intelligence (GenAI) leverages large language models (LLMs) that are transforming health care. Specialized ambient GenAI tools, like Nuance Dax, Speke, and Tandem Health, "listen" to consultations and generate clinical notes. Medical-focused models, like Med-PaLM, provide tailored health care insights. GenAI's capability to summarize complex data and generate responses in various conversational styles or literacy levels makes it particularly valuable since it has the potential to alleviate the burden of clinical documentation on health care professionals (HCPs). While GenAI may prove to be helpful, offering novel benefits, it comes with its own set of challenges. The quality of the source data can introduce biases, leading to skewed recommendations or outright false information (so-called hallucinations). In addition, due to the conversational nature of chatbot responses, users may be susceptible to misinformation, posing risks to both safety

and privacy. Therefore, careful implementation and rigorous oversight are essential to ensure accuracy, ethical integrity, and alignment with clinical standards. Despite these advances, currently, no review has investigated HCPs' experiences and opinions about GenAI in clinical documentation. Yet, such a perspective is crucial to better understand how these technologies can be safely and ethically adopted and implemented in clinical practice. We aim to present the protocol for a scoping review exploring HCPs' experiences and opinions about GenAI and ambient scribes in clinical documentation. This scoping review will be carried out following the methodological framework of Arksey and O'Malley and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews) checklist. Relevant papers will be searched for in PubMed, IEEE Xplore, APA PsycInfo, CINAHL, and Web of Science. The review will include studies published between January 2023 and September 2025. Studies will be included that represent original peer-reviewed work that explores HCPs' experiences and opinions about the use of GenAI or ambient scribes for clinical documentation. Data extraction will include publication type, country, sample characteristics, clinical setting, study aim, study design, research question, and key findings. Study quality will be assessed using the Mixed Methods Appraisal Tool. The results will be presented as a narrative synthesis structured along the key themes of the evidence mapped. Data will be collated and presented in charts and tabular format. Findings will be reported in a peer-reviewed scoping review. This will be the first scoping review that considers HCPs' experiences and opinions about GenAI and ambient scribes in clinical documentation. The results will clarify how HCPs use-or avoid using-GenAI in daily health care work. This insight will help address perceived benefits, risks, expectations, and uncertainties. It may also reveal key research gaps in the field.

56. Leung, Tiffany I., Andrew J. Coristine and Arriel Benis. **AI Scribes in Health Care: Balancing Transformative Potential With Responsible Integration**. JMIR medical informatics. 2025.Vol.13 ppe80898.

The administrative burden of clinical documentation contributes to health care practitioner burnout and diverts valuable time away from direct patient care. Ambient artificial intelligence (AI) scribes—also called “digital scribes” or “AI scribes”—are emerging as a promising solution, given their potential to automate clinical note generation and reduce clinician workload, and those specifically built on a large language model (LLM) are emerging as technologies for facilitating real-time clinical documentation tasks. This potentially transformative development has a foundation on longer-standing, AI-based transcription software, which uses automated speech recognition and/or natural language processing. Recent studies have highlighted the potential impact of ambient AI scribes on clinician well-being, workflow efficiency, documentation quality, user experience, and patient interaction. So far, limited evidence indicates that ambient AI scribes are associated with reduced clinician burnout, lower cognitive task load, and significant time savings in documentation, particularly in after-hours electronic health record (EHR) work. One consistently reported benefit is the improvement in the patient-physician interaction, as physicians feel more present during a clinical encounter. However, these benefits are counterbalanced by persisting concerns regarding the accuracy, consistency, language use, and style of AI-generated notes. Studies noting errors, omissions, or hallucinations caution that diligent clinician oversight is necessary. The user experience is also heterogeneous, with benefits varying by specialty and individual workflow. Further, there are concerns about ethical and legal issues, algorithmic bias, the potential for long-term “cognitive debt” from

overreliance on AI, and even the potential loss of physician autonomy. Additional pragmatic concerns include security, privacy, integration, interoperability, user acceptance and training, and the cost-effectiveness of adoption at scale. Finally, limited studies describe adoption or evaluation of these technologies by nonphysician clinicians and health professionals. Although ambient AI scribes and AI-driven documentation technologies are promising as potentially practice-changing tools, there are many questions remaining. Key issues persist, including responsible deployment with the goal of ensuring that ambient AI scribes produce clinical documentation that supports more efficient, equitable, and patient-centered care. To advance our collective understanding and address key issues, JMIR Medical Informatics is launching a call for papers for a new section on "Ambient AI Scribes and AI-Driven Documentation Technologies." As editors, we look forward to the opportunity to advance the science and understanding of these fields through publishing high-quality and rigorous scholarly work in this new section of JMIR Medical Informatics .

57. Ong, Julene Hui Wun, Joshua Yi Min Tung, Gerald Gui Ren Sng, et al. **A pilot study using ambient artificial intelligence scribes in clinical documentation in a urology outpatient clinic**. BJU Int. 2025.Vol.136(3), pp415–417.

58. Shah, Shreya J., Trevor Crowell, Yejin Jeong, et al. **Physician Perspectives on Ambient AI Scribes**. JAMA network open. 2025.Vol.8(3), ppe251904.

Limited qualitative studies exist evaluating ambient artificial intelligence (AI) scribe tools. Such studies can provide deeper insights into ambient AI implementations by capturing lived experiences. To evaluate physician perspectives on ambient AI scribes. A qualitative study using semistructured interviews guided by the Reach, Efficacy, Adoption, Implementation, Maintenance/Practical, Robust Implementation, and Sustainability Model (RE-AIM/PRISM) framework, with thematic analysis using both inductive and deductive approaches. Physicians participating in an AI scribe pilot that included community and faculty practices, across primary care and ambulatory specialties, were invited to participate in interviews. This ambient AI scribe pilot at a health care organization in California was conducted from November 2023 to January 2024. Facilitators and barriers to adoption, practical effectiveness, and suggestions for improvement to enhance sustainability. Twenty-two semistructured interviews were conducted with AI pilot physicians from primary care (13 59%]) and ambulatory specialties (9 41%]), including physicians from community practices (12 55%]) and faculty practices (10 45%]). Facilitators to adoption included ease of use, ease of editing, and generally positive perspectives of tool quality. Physicians expressed positive sentiments about the impact of the ambient AI scribe tool on cognitive demand (16 of 16 comments 100%]), temporal demand (28 comments 62%]), work-life integration (10 of 11 comments 91%]), and overall workload (8 of 9 comments 89%]). Physician perspectives of the impact of the ambient AI scribe tool on their engagement with patients were mostly positive (38 of 56 comments 68%]). Barriers to adoption included limited functionality with non-English speaking patients and lack of access for physicians without a specific device. Physician perspectives on accuracy and style were largely negative, particularly regarding note length and editing requirements. Several specific suggestions for tool improvement were identified, and physicians were optimistic regarding the potential for long-term use of ambient AI scribes. In this qualitative study, ambient AI scribes were found to positively impact physician workload, work-life integration, and patient engagement. Key facilitators and barriers to adoption were identified, along with specific suggestions for tool

improvement. These findings suggest the potential for ambient AI scribes to reduce clinician burden, with user-centered recommendations offering practical guidance on ways to improve future iterations and improve adoption.

59. Shemtob, Lara, Azeem Majeed and Thomas Beaney. **Regulation of AI scribes in clinical practice**. BMJ. 2025.Vol.389 ppr1248. Defining responsibilities for clinical information. 10.1136/bmj.r1248

60. Stokel-Walker, Chris. **The "ambient scribe" tools listening to and summarising your doctor-patient consultations**. BMJ. 2025.Vol.389 ppr663. New AI tools that record and summarise doctor-patient conversations are becoming increasingly common. But what are they, and can they be trusted? Chris Stokel-Walker reports.

61. Patterson, James, Maya Kovacs and Caitlin Lees. **Ambient Artificial Intelligence Scribes: A Pilot Survey of Perspectives on the Utility and Documentation Burden in Palliative Medicine.** Healthcare (Basel). 2025.Vol.13(17), pp2118.

Background/Objectives: There is growing evidence to support ambient artificial intelligence (AI) scribes in healthcare to improve medical documentation by generating timely and comprehensive notes. Using the Plan–Do–Study–Act (PDSA) methodology, this study evaluated the utility and potential time savings of an ambient AI scribe, Scribeberry, (V2), in a palliative medicine outpatient setting, comparing it to the standard practice of dictation. Methods: This prospective quality improvement study was conducted at an academic medical center by two palliative medicine resident physicians. Residents documented patient visits using a freely available ambient AI scribe software program, Scribeberry, as well as using standard dictation software. Primary outcome measures included the editing time for the AI scribe and the dictating and editing times for a dictated manuscript, as well as subjective assessments of the accuracy, organization, and overall usefulness of the AI-generated clinical letters. Results: A heterogenous response was seen with the implementation of an AI scribe. One resident saw a statistically significant reduction ($p < 0.025$) in the time spent on clinical documentation, while a second resident saw essentially no improvement. The resident who experienced time savings with the ambient AI scribe also demonstrated a significant improvement in the graded organization and usefulness of the AI outputs over time, while the other resident did not demonstrate significant improvements in any of the metrics assessed over the course of this project. Conclusions: This pilot study describes the use of an ambient AI scribe software program, Scribeberry, in the community palliative medicine context. Our results showed a mixed response with respect to time savings and improvements in the organization, accuracy, and overall clinical usefulness of the AI-generated notes over time. Given the small sample size and short study duration, this study is insufficiently powered to draw conclusions with respect to AI scribe benefits in real-world contexts.

62. Razai, Mohammad S. **Artificial intelligence scribes in general practice: sacrificing the art of medicine for promised efficiency.** British journal of general practice. 2025.Vol.75(757), pp366–367.

Razai explores the growing adoption of artificial intelligence (AI) in healthcare, especially in general practice, where tools like ambient AI scribes are promoted as solutions to rising workloads and administrative burdens. These AI systems transcribe consultations and generate documentation, aiming to free clinicians for more patient-focused care. However,

the article urges caution. Previous technologies have failed to reduce workload and sometimes added complexity or disrupted continuity of care. A trial of an AI scribe showed it missed key emotional and narrative elements of the consultation, producing impersonal, jargon-filled, and sometimes inaccurate notes. She argues that GP work is deeply relational and interpretive, not a data-entry task. Clinical notes carry meaning, nuance, and context--elements AI struggles to capture. While AI tools may improve, their integration into practice should be carefully considered to avoid undermining the human, reflective nature of general practice. She advocates for a cautious, values-driven approach to AI adoption.

AI & vascular

63. Zyada, Ayman, Ayman Fakhry, Sohiel Nagib, et al. **How Well Do Different AI Language Models Inform Patients About Radiofrequency Ablation for Varicose Veins?.** Curēus (Palo Alto, CA). 2025.Vol.17(6), ppe86537.

Introduction The rapid integration of artificial intelligence (AI) into healthcare has led to increased public use of large language models (LLMs) to obtain medical information. However, the accuracy and clarity of AI-generated responses to patient queries remain uncertain. This study aims to evaluate and compare the quality of responses provided by five leading AI language models regarding radiofrequency ablation (RFA) for varicose veins. Objective To assess and compare the reliability, clarity, and usefulness of AI-generated answers to frequently asked patient questions about RFA for varicose veins, as evaluated by expert vascular surgeons. Methods A blinded, comparative observational study was conducted using a standardized list of eight frequently asked questions about RFA, derived from reputable vascular surgery centers across multiple countries. Five top-performing, open-access LLMs (ChatGPT-4, OpenAI, San Francisco, CA, USA; DeepSeek-R1, DeepSeek, Hangzhou, Zhejiang, China; Gemini 2.0, Google DeepMind, Mountain View, CA, USA; Grok-3, xAI, San Francisco, CA, USA; and LLaMA 3.1, Meta Platforms, Inc., Menlo Park, CA, USA) were tested. Responses from each model were independently evaluated by 32 experienced vascular surgeons using four criteria: accuracy, clarity, relevance, and depth. Statistical analyses, including Friedman and Wilcoxon signed-rank tests, were used to determine model performance. Results Grok-3 was rated as providing the highest-quality responses in 51.6% of instances, significantly outperforming all other models ($p < 0.0001$). ChatGPT-4 ranked second with 23.1%. Gemini, DeepSeek, and LLaMA showed comparable but lower performance. Question-specific analysis revealed that Grok-3 dominated responses related to procedural risks and post-procedure care, while ChatGPT-4 performed best in introductory questions. A subgroup analysis showed that user experience level had no significant impact on model preferences. While 42.4% of respondents were willing to recommend AI tools to patients, 45.5% remained uncertain, reflecting ongoing hesitation. Conclusion Grok-3 and ChatGPT-4 currently provide the most reliable AI-generated patient education about RFA for varicose veins. While AI holds promise in improving patient understanding and reducing physician workload, ongoing evaluation and cautious clinical integration are essential. The study establishes a baseline for future comparisons as AI technologies continue to evolve.