HealthAIDE: Developing an Audit Framework for AI-Generated Online Health Information

Tahir Hameed[a]
Girard School of Business, Merrimack College, North Andover, MA 01845, U.S.A.

Keywords: Online Health Information, Large Language Models, Generative AI, AI Auditing, Responsible AI.

Abstract: Online health information (OHI) encompasses a wide range of public-facing content, such as information on symptoms, diseases, medications, and treatments, while online medical information (OMI) involves more specialized and regulated content, including clinical trial data, surgical procedures, and medical research. OMI generation and dissemination is held to stringent standards for accuracy, transparency, and explainability, whereas OHI often requires information-seekers to independently evaluate credibility and relevance of the information. The rise of generative AI or large language models (LLMs) has exacerbated this disparity, as LLMs are primarily applied to public-domain OHI without sufficient safeguards, leaving users vulnerable to misinformation, bias, and non-transparent outputs. This paper presents a systematic literature survey on the usage of AI and LLMs in OHI, highlighting focus areas and critical gaps in developing a robust framework for auditing AI-generated health information. The proposed HealthAIDE Framework defines four key pillars for oversight: reliability and accuracy, trust and acceptance, security and safety, and equity and fairness. A short but systematic review of AI-driven health information literature reveals areas of stronger focus, such as accuracy and trust, and weaker focus areas, such as misuse prevention and transparency. Addressing these gaps through comprehensive audits will enable responsible evolution of AI-driven health information systems.

1 INTRODUCTION

Online Health Information (OHI) searches have become a primary resource for individuals seeking to check symptoms, explore treatment options and their efficacy, and locate healthcare providers or clinics (Hameed, 2018). With the rise of pre-trained large language models (LLMs), OHI seeking, its use, and corresponding health behavior are undergoing a transformation (Clark et al., 2024; Yan et al.,2024). Traditionally, users conducting OHI searches on general-purpose or specialized healthcare search engines faced issues such as information overload, difficulty in distinguishing credible sources, and a lack of personalization (Swar et al.,2017; Freeman et al., 2020). They were less confident about the information received and therefore conscientious when using it. While generative-AI platforms address some of these issues by offering personalized, dynamic, and conversational interfaces, they also introduce new complications while amplifying some existing problems. For instance, misinformation generated by AI hallucinations in definitive tones can lead to misuse of OHI or undesirable behavior (Jin et al., 2023, Shen et al., 2024). User trust also suffers due to a lack of transparency about the sources of information, making it difficult for users to verify its accuracy. Additionally, the risks of potential misuse of sensitive user data increase as open queries are made on the internet rather than secure electronic health records or clinical databases, raising concerns about privacy and ethical use. These issues not only magnify pre-existing concerns but also raise further dilemmas around consent, privacy, and the accuracy of recommendations (Shi et al.,2024).

AI audits assess AI systems to ensure their reliability, ethical integrity, and compliance with regulations and standards. These audits aim to identify and address potential risks associated with AI, such as bias, security vulnerabilities, privacy concerns, and transparency issues. By analyzing an AI system's data, algorithms, and outputs, auditors can uncover risks, assess the impact of the systems, and recommend improvements to enhance

[a] https://orcid.org/0000-0002-6824-6803

Hameed, T.
HealthAIDE: Developing an Audit Framework for AI-Generated Online Health Information.
DOI: 10.5220/0013187100003911

In *Proceedings of the 18th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 2: HEALTHINF*, pages 173-183
ISBN: 978-989-758-731-3; ISSN: 2184-4305

trustworthiness and accountability (Falco et al., 2021; Li et al., 2024, Mökander, 2023).

AI audits are essential for fostering trust in AI systems and ensuring compliance with governmental regulations such as the General Data Protection Regulation (GDPR) in the EU, the Health Insurance Portability and Accountability Act (HIPAA), and the California Consumer Privacy Act (CCPA) in the US (Forcier et al., 2019; Mulgund et al., 2021). However, the rapid development of generative-AI systems has complicated the definition and scope of assessments in an ever-evolving information landscape. Proposed initiatives, such as the Algorithmic Accountability Act and the AI Bill of Rights in the US, along with the EU AI Act, represent significant efforts to guide the design, use, and deployment of AI systems, focusing on protecting citizens' rights (Blumenthal-Barby, 2023; Veale & Zuiderveen, 2021). These regulations aim to pre-emptively identify potential ethical, legal, and operational risks prior to deployment and ensure ongoing monitoring of AI systems post-implementation.

Even though OHI seeking is one of the fastest-developing areas leveraging generative AI and LLM-driven platforms, it also has the potential to cause significant harm through the dissemination of misinformation. Despite this, AI audits are not yet fully adopted or deeply considered in this domain. While substantial efforts are underway to support the curation, auditing, and sharing of Online Medical Information (OMI)—including biomedical content, medical databases, and repositories used by professionals—similar mechanisms are notably absent in the OHI domain for patients and caregivers (Li and Goel, 2024)

This paper introduces Health-AIDE, a preliminary framework designed to audit AI-generated OHI. Section 2 begins by exploring the rapidly changing landscape of OHI, followed by a scoping review of existing AI auditing literature to identify the critical components of an AI audit framework. Section 3 provides a systematic literature review of peer-reviewed studies on generative AI in the OHI domain. It evaluates the strengths and weaknesses of the current AI-driven OHI systems and applications against the proposed HEALTH-AIDE framework. Finally, Section 4 draws conclusions and discusses future directions for improving OHI auditing practices.

As one of the earlier papers in this area, this work aims to provide a foundation for developing Health-AIDE into a comprehensive, scalable AI Auditing framework also including the soft side of appropraite communication with the patients and their caregivers. By addressing the pressing need for AI audits in the OHI domain, we hope to pave the way for safer, more reliable, and ethically governed generative AI-driven health information systems.

2 BACKGROUND & SCOPE OF AI AUDITING IN ONLINE HEALTH INFORMATION

2.1 Emergence of LLM and Generative AI OHI Platforms

LLMs such as OpenAI's GPT series, Meta’s LLama and Google's Bard have transformed OHI platforms by enabling conversational, personalized, and context-aware responses to user queries. These models utilize huge datasets of medical literature and publicly available information, to generate detailed responses tailored to individual needs. (Shen et al., 2024; Yan et al., 2024). Unlike traditional search engines, LLMs excel in synthesizing information from multiple sources and providing it in user-friendly formats. No wonder, generative AI based OHI platforms offers significant benefits, including improved accessibility, efficiency, and interactivity. Users can ask complex health-related questions and receive coherent explanations, making it easier to understand medical concepts. These tools have also shown promise in underserved regions, where limited access to healthcare professionals makes reliable online information crucial. Despite these advantages, the adoption of LLM-driven platforms has raised ethical concerns about data privacy and the trustworthiness of AI-generated advice, such as diagnostics and treatment recommendations (Cocci et al., 2024). Therefore, like all other sectors, there is a strong need to ensure LLM-based OHI platforms operate responsibly under the oversight of robust regulatory frameworks (Mesko and Topol, 2023).

2.2 Auditing Generative AI and LLM-Based Information and Systems, Towards a Framework

Auditing has long been a critical governance mechanism to ensure that information systems and data are managed in compliance with technical, legal, and ethical requirements established by manufacturers, industry organizations, and governments. Information systems auditing has matured as a professional discipline with established procedures and records (Champlain, 2003). However, the rapid evolution of generative AI and LLMs in recent years has outpaced the development of appropriate governance measures, leaving gaps in defining and implementing effective management practices.

Weidinger et al.'s (2022) taxonomy of LLM risks highlights several critical issues, including the perpetuation of discrimination and biases, inadvertent information hazards such as data leaks, malicious use such as fraudulent scams, and environmental harms caused by excessive computing power requirements, among others. However, the two most prevalent hazards they identify in generative AI are distorted human-machine interactions and misinformation hazards. The former pertains to users overestimating the capabilities of LLMs, leading to their unsafe or inappropriate use. The latter poses significant risks as less-informed users may consume misleading information, resulting in harm and potentially eroding public trust in AI-generated content.

Mökander et al. (2023) proposed a three-layered approach to auditing the outputs of LLM-based systems, aiming to ensure their effectiveness from technical, social, and legal perspectives. Their framework emphasizes equal attention to mitigating social and ethical risks associated with AI systems. The authors recommend three types of audits: governance audits for LLM providers, model audits conducted before the release of pre-trained LLMs, and application audits for specific scenarios where LLMs are deployed. This comprehensive approach suggests that AI system audits should evolve to address not only technological and procedural aspects but also complementary areas that ensure the responsible use and long-term impact of AI systems on users, societies, and the natural environment (Mökander et al., 2023).

We discuss most critical aspects of AI oversight to develop a basic framework for auditing AI in healthcare sector. At first, ensuring the accuracy and correctness of generated content is a critical factor in the reliability of any AI system. Techniques such as prompt engineering, querying, and probing serve as robust methods for generating content and comparing it against established benchmarks. However, content must also be assessed on qualitative and semantic aspects, including fluency, coherence, and relevance, which significantly contribute to overall user satisfaction. (Davis et al., 2023). Reliability of AI systems outputs involves their broader applicability i.e. scalability as well as continuous learning to stay on top of current knowledge. These aspects of reliability, accuracy and access are primarily addressed through technology and algorithmic audits, which have consistently been the most prevalent and central mechanism for evaluating AI systems.

AI systems often rely on vast amounts of sensitive and personal data. Ensuring that this data is protected from unauthorized access, breaches, or cyberattacks is fundamental to maintaining the integrity of AI systems and safeguarding user privacy (Bala et al., 2024). Robust data security measures prevent data leaks, tampering, or misuse throughout the data lifecycle— from collection and storage to processing and deployment. These measures include encryption, access controls, regular security testing, and secure data handling practices. Security audits help identify vulnerabilities, ensure compliance with data protection regulations, and verify that AI systems are equipped with the necessary safeguards to protect sensitive data (Nankya et al., 2024).

Similarly, harm prevention and misuse prevention are essential components in ensuring that AI-generated content does not lead to negative consequences for individuals or society (Ellaham et al., 2020). AI systems must be designed to avoid producing outputs that could cause physical, psychological, or social harm, such as promoting harmful behaviors, spreading misinformation, or enabling discrimination. Harm prevention strategies include incorporating ethical guidelines into the AI's design, continuously monitoring outputs for unintended negative effects, and ensuring that AI models are trained to recognize and mitigate harmful content. Misuse prevention focuses on safeguarding against malicious or unethical use of AI systems, such as using AI for fraud, manipulation, or the creation of harmful content like deepfakes. To effectively manage these risks, safety audits are required to assess how well AI systems prevent harm and misuse. These audits ensure that safeguards are in place, that ethical considerations are followed, and that AI systems are used responsibly, reducing the risk of harm to individuals and society (Shneiderman, 2020).

Equity and fairness in AI systems rely heavily on systematically evaluating the technical and organizational practices surrounding their development and deployment. This requires mechanisms to identify, mitigate, and monitor biases in datasets, models, and outputs while adhering to ethical principles, regulatory requirements, and industry standards that promote fairness and prevent discrimination (Rajkomar, 2018, Ueda et al., 2024). Ensuring accountability structures within AI providers is essential to integrating equity and fairness considerations into decision-making processes at all levels. Transparency in algorithmic design, inclusivity of training data, and representativeness of stakeholder engagement efforts are critical aspects that must be prioritized. Furthermore, effective monitoring post-deployment is necessary to detect and address unintended biases or inequities. To comprehensively address these requirements, process and governance audits emerge as a vital need, providing a structured approach to evaluate and reinforce the equitable design and implementation of AI systems, ensuring positive outcomes for all users and communities.

The AI auditing literature can clearly be organized into the following major aspects that must be monitored and assessed to develop and deploy reliable, trustworthy, secure, and inclusive AI systems. A well-structured Figure 1 outlines the proposed framework for AI audits, specifically tailored for the online health information sector, while remaining broadly applicable to any information domain.

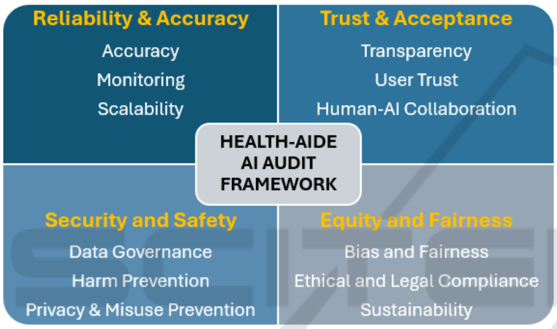

Figure 1: Health-AIDE; A preliminary framework for Auditing AI-generated health information.

3 A SYSTEMATIC REVIEW OF OHI DRIVEN BY GENERATIVE AI

ChatGPT was launched on November 30, 2022. While research on transformer models and preliminary applications existed prior, the release of ChatGPT marked a significant turning point, driving public access to advanced LLMs and generative AI systems. Recognizing this milestone, we conducted a systematic literature review of articles published on the use of LLMs and generative AI in the domain of OHI. The aim was to evaluate the recorded advancements and align them with the core dimensions of the proposed Health-Aide framework, assessing the implications and gaps in AI-driven OHI systems.

3.3 Methods

To that end, our literature review followed a structured and systematic process to ensure relevance and depth in analysing AI-generated OHI.

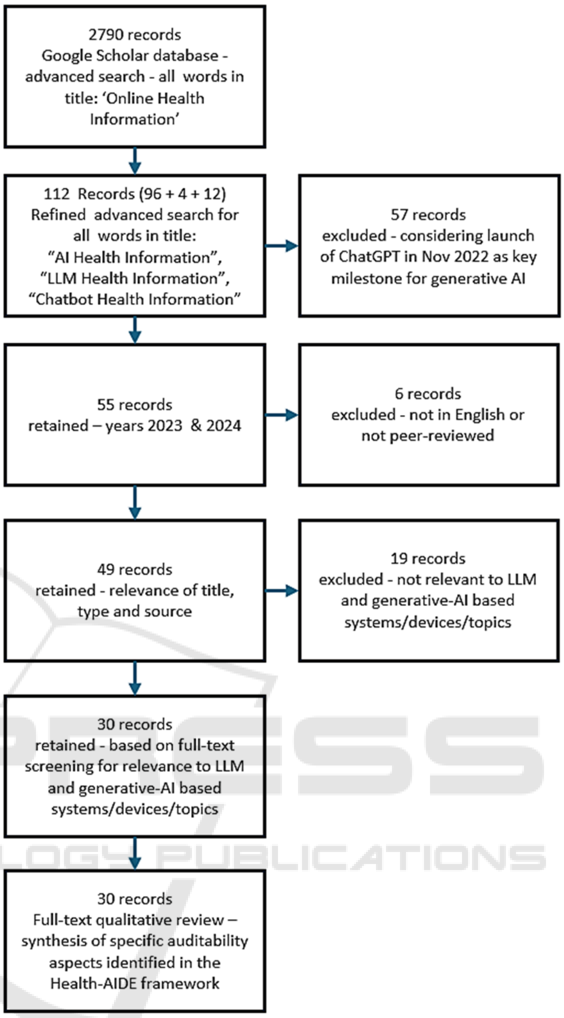
2790 records
Google Scholar database - advanced search - all words in title: 'Online Health Information'
112 Records (96 + 4 + 12)
Refined advanced search for all words in title:
"AI Health Information",
"LLM Health Information",
"Chatbot Health Information"
57 records
excluded - considering launch of ChatGPT in Nov 2022 as key milestone for generative AI
55 records
retained – years 2023 & 2024
6 records
excluded - not in English or not peer-reviewed
49 records
retained - relevance of title, type and source
19 records
excluded - not relevant to LLM and generative-AI based systems/devices/topics
30 records
retained - based on full-text screening for relevance to LLM and generative-AI based systems/devices/topics
30 records
Full-text qualitative review – synthesis of specific auditability aspects identified in the Health-AIDE framework

Figure 2: Articles selection for systematic literature review.

Initially, a Google Advanced Search was conducted using keywords like Online Health Information, yielding 2,790 results. To refine the focus, additional searches with terms such as AI Health Information, LLM Health Information, and Chatbot Health Information narrowed the results to 112. To ensure timeliness, only articles published after the launch of ChatGPT in late 2022 were considered, further reducing the scope to 55. After reviewing article types and sources, only English peer-reviewed journal articles and conference proceedings were retained, resulting in 49 relevant papers. Full-text reviews were then conducted, focusing on AI auditing frameworks, ultimately identifying 30 highly relevant sources for analysis. The process is outlined in Figure 2.

Each paper was reviewed thoroughly to analyze the scope of the system or device capturing or generating personal or clinical health information. Features, opportunities, challenges, and issues discussed in the articles were systematically mapped to a table aligned with the dimensions of the Health-Aide Framework. This mapping provided a preliminary understanding of innovation trends, highlighting where AI-driven OHI generation and sharing are prioritized. The analysis offers insights into key development areas, challenges, and gaps, forming a foundation for refining the framework further.

At this stage, the focus was on identifying priorities and areas of emphasis without assigning weights or conducting detailed comparisons. A more rigorous analysis will be conducted in the next phase using a systematic literature selection process through the Web of Science database. This follow-up will enable weighted assessments, deeper comparisons, and a broader understanding of how these align with the Health-AIDE Framework dimensions.

A table comprehensivley summarizing the mapped observations is included in Appendix A (See Table 3). This table highlights the features, opportunities, and issues focused in each health AI/LLM-based system or device with respect to the Health-AIDE framework. It also provides a clear and structured overview of the findings of the literature review, offering a visual representation of current innovation and research priorities in AI-based OHI systems. It defintley serves as a valuable resource for identifying trends and gaps for further exploration in other research and future phases of this paper.

4 ANALYSIS, DISCUSSION & CONCLUSIONS

Since this is a preliminary framework and an initial attempt at identifying the current focus areas, we began by calculating the frequencies of each auditable aspect of health information. All the counts are presented in Table 1.

Surprisingly, Human-AI Collaboration emerged as the most emphasized focus area. This indicates that both users and developers are deeply focused on understanding and improving interactions with AI systems in the health domain. Following this is *Accuracy*, a critical priority for any information system and particularly essential in healthcare, where reliable and precise data is foundational.

Table 1: On-going innovation and research areas w.r.t components of the AI/LLM-generated Online Health Information Audit Framework - Health-AIDE.

		Count of Discussions (Implementations, Opportunities or Challenges)
Reliability and Accuracy	Accuracy	15
	Currency/Updates Monitoring	9
	Scalability	5
Trust and Acceptance	Transparency	4
	User Trust	13
	Human-AI Collaboration	17
Security and Safety	Data Governance	12
	Safety and Harm Prevention	13
	Security and Misuse Prevention	4
Equity and Fairness	Bias and Fairness	11
	Ethical and Legal Compliance	4
	Environmental Sustainability	1

User Trust and Harm Prevention ranked next, reflecting growing concerns about misinformation and its potential dangers to individuals and society. These aspects highlight the need to mitigate risks associated with the dissemination of inaccurate or harmful content in AI-driven systems.

At the lower end, aspects such as Sustainability, Misuse Prevention, Ethical and Legal Compliance, and Transparency received comparatively less attention. This is somewhat unexpected, given that Data Governance is reasonably well-addressed, yet related areas like Security and Misuse Prevention remain underexplored. This disparity underscores potential gaps in the prioritization of critical aspects in AI-driven health information systems development that require further investigation and emphasis.

A co-occurrence map provides deeper insights into how various aspects are interlinked in research, innovation, and oversight. For instance, Human-AI Collaboration strongly co-occurs with Bias and Fairness, suggesting that as efforts progress to improve human-AI interactions, there is a growing recognition of the need for equity and fairness in benefiting from health information systems.

Addressing biases in design and implementation is crucial to ensuring inclusivity and equitable access to AI-driven health solutions.

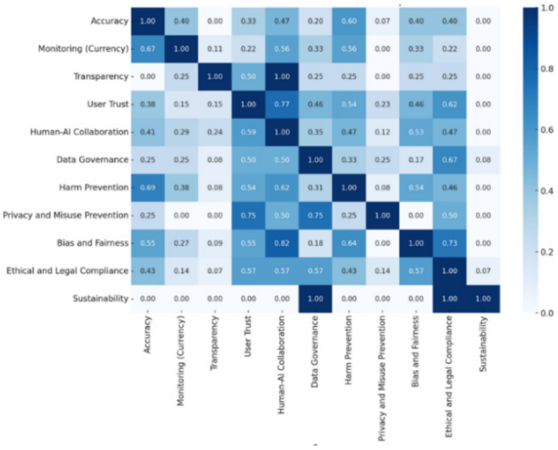

Figure 3: Co-Occurrence map of auditability aspects of AI/LLM-generated online health information.

Additionally, Data Governance and User Trust strongly co-occur with Privacy and Misuse Prevention, highlighting that privacy concerns and potential data breaches are central to maintaining trust in AI-based health information systems. Robust privacy protections and safeguards against misuse are essential for fostering confidence and reliability in these systems.

Transparency and Human-AI Collaboration also consistently co-occur, emphasizing the users' need to understand the sources of information and the explainability of how complex content is generated and organized. Transparency is critical in fostering trust, as it provides clear insights into how AI systems work and ensures that outputs are explainable and intuitive. In health information systems, this is particularly important as decisions based on AI-generated content can directly impact individual well-being. Transparency is foundational for building trust and enabling meaningful collaboration between humans and AI.

In conclusion, these observations demonstrate that existing and emerging laws, regulations, technical guidelines, and process standards must be complemented by interconnected and synergistic requirements to ensure that information generated by AI systems is trustworthy, safe, secure and equitable. Addressing these aspects holistically is vital to fostering user confidence and minimizing risks in AI-driven health information systems.

A scan of existing or emerging industry standards, audit toolkits, and national-level regulations reveals an increasing emphasis on addressing the identified auditability aspects for AI-driven software and information systems in general (see Table 2). However, compliance with these auditing standards and regulations has yet to gain widespread acceptance in the OHI sector,. This gap highlights the need for greater efforts to enforce audit frameworks in OHI domain for ensuring accountability, transparency, and trust in AI-driven health information systems.

Table 2: AI standards, regulations and audit toolkits w.r.t to OHI auditability aspects identified in HealthAIDE.

		AI Standards, Regulations and Auditing Toolkits
Reliability and Accuracy	Accuracy	ISO/IEC 23894:2023, ISO/IEC 25010, NIST AI RMF, EU AI Act, FDA SaMD
	Currency/ Updates Monitoring	
	Scalability	
Trust and Acceptance	Transparency	IEEE 7001-2021, ISO/IEC TR 24028, ANSI/CTA-2090, NIH DISCERN, Algorithmic Accountability Act
	User Trust	
	Human-AI Collaboration	
Security and Safety	Data Governance	GDPR, EU AI Act, HIPPA, CCPA, FedRAMP, HL7, FHIR, IEEE 7010-2020
	Safety and Harm Prevention	
	Security and Misuse Prevention	
Equity and Fairness	Bias and Fairness	WHO Ethics Guidelines, DoD Ethical/Responsible AI Guidelines, Ada Toolkit, ISO/IEC TR 24027:2021
	Ethical and Legal Compliance	
	Environmental Sustainability	

This study represents a preliminary, smaller survey and the development of an initial framework for auditing AI-generated health information. Future work aims to expand on this research by incorporating a larger dataset of papers from the Web of Science database and engaging expert reviews. This expanded approach will address technical, process, and interdisciplinary issues at a deeper level, while also considering the details of above-noted and emerging regulations, technical standards and professional codes to create a more comprehensive and robust framework.

REFERENCES

Alhendawi, K. M. (2024). Task-technology fit model: Modelling and assessing the nurses' satisfaction with health information system using AI prediction models. *International Journal of Healthcare Management*, 17(1), 12-24.

Alkhalaf, M., Yu, P., Yin, M., & Deng, C. (2024). Applying generative AI with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *Journal of Biomedical Informatics*, 104662.

Amin, M. S., Johnson, V. L., Prybutok, V., & Koh, C. E. (2024). An investigation into factors affecting the willingness to disclose personal health information when using AI-enabled caregiver robots. *Industrial Management & Data Systems*, 124(4), 1677-1699.

Aparicio, E., Nguyen, Q., Doig, A. C., Gutierrez, F. X. M., Sah, S., Mane, H., ... & He, X. (2023, November). Development of rosie the chatbot: A health information intervention for pregnant and new mothers of color. In APHA 2023 Annual Meeting and Expo. APHA.

Arda, S. (2024). Taxonomy to Regulation: A (Geo) Political Taxonomy for AI Risks and Regulatory Measures in the EU AI Act. *arXiv preprint arXiv:2404.11476*.

Ascorbe, P., Campos, M. S., Domínguez, C., Heras, J., & Reinares, A. R. T. (2023). prevenIA: a Chatbot for Information and Prevention of Suicide and other Mental Health Disorders. In *SEPLN (Projects and Demonstrations)* (pp. 26-30).

Bala, I., Pindoo, I., Mijwil, M. M., Abotaleb, M., & Yundong, W. (2024). Ensuring security and privacy in Healthcare Systems: a Review Exploring challenges, solutions, Future trends, and the practical applications of Artificial Intelligence. *Jordan Medical Journal*, *58*(3).

Blumenthal-Barby, J. (2023). An AI bill of rights: Implications for health care AI and machine learning—A bioethics lens. *The American Journal of Bioethics*, *23*(1), 4-6.

Champlain, J. J. (2003). *Auditing information systems*. John Wiley & Sons.

Chang, C. W., Hu, M., Ghavidel, B., Wynne, J. F., Qiu, R. L. J., Washington, M., ... & Yang, X. (2024). An LLM-Based Framework for Zero-Shot De-Identifying Flexible Text Data in Protected Health Information Enabling Potential Risk-Informed Patient Safety. *International Journal of Radiation Oncology*, Biology, Physics, 120(2), e518.

Clark, O., Reynolds, T. L., Ugwuabonyi, E. C., & Joshi, K. P. (2024, June). Exploring the Impact of Increased Health Information Accessibility in Cyberspace on Trust and Self-care Practices. In *Proceedings of the 2024 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems* (pp. 61-70).

Cocci, A., Pezzoli, M., Lo Re, M., Russo, G. I., Asmundo, M. G., Fode, M., ... & Durukan, E. (2024). Quality of information and appropriateness of ChatGPT outputs for urology patients. *Prostate cancer and prostatic diseases*, *27*(1), 103-108.

Davis, R., Eppler, M., Ayo-Ajibola, O., Loh-Doyle, J. C., Nabhani, J., Samplaski, M., ... & Cacciamani, G. E. (2023). Evaluating the effectiveness of artificial intelligence–powered large language models application in disseminating appropriate and readable health information in urology. The Journal of urology, 210(4), 688-694.

Ellaham, S., Ellahham, N., & Simsekler, M. C. E. (2020). Application of artificial intelligence in the health care safety context: opportunities and challenges. *American Journal of Medical Quality*, *35*(4), 341-348.

Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., ... & Yeong, Z. K. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence*, *3*(7), 566-571. Freeman, J. L., Caldwell, P. H., & Scott, K. M. (2020). The role of trust when adolescents search for and appraise online health information. *The Journal of Pediatrics*, 221, 215-223.

Forcier, M. B., Gallois, H., Mullan, S., & Joly, Y. (2019). Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers?. *Journal of Law and the Biosciences*, *6*(1), 317-335.

Hameed, T. (2018). Impact of Online Health Information on Patient-physician Relationship and Adherence; Extending Health-belief Model for Online Contexts. In *HEALTHINF* (pp. 591-597).

Harrington, C. N., & Egede, L. (2023, April). Trust, comfort and relatability: Understanding black older adults' perceptions of chatbot design for health information seeking. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-18).

Hatamlah, H. (2024). Adolescents' online health information seeking: Trust, e-health literacy, parental influence, and AI-generated credibility. *International Journal of Data and Network Science*, 8(2), 809-822.

Hong, S. J., & Cho, H. (2023). Privacy management and health information sharing via contact tracing during the COVID-19 pandemic: A hypothetical study on AI-based technologies. *Health Communication*, 38(5), 913-924.

Jin, Q., Leaman, R., & Lu, Z. (2023). Retrieve, summarize, and verify: how will ChatGPT affect information seeking from the medical literature?. *Journal of the American Society of Nephrology*, 34(8), 1302-1304.

Jin, E., & Eastin, M. (2024). Towards more trusted virtual physicians: the combinative effects of healthcare chatbot design cues and threat perception on health information trust. *Behaviour & Information Technology*, 1-14.

Kunlerd, A. (2024). Developing an Innovative Health Information Service System: The Potential of Chatbot Technology. Suan Sunandha Science and Technology Journal, 11(2), 61-69.

Latt, P. M., Aung, E. T., Htaik, K., Soe, N. N., Lee, D., King, A. J., ... & Fairley, C. K. (2024). Real-World

Evaluation of Artificial Intelligence (AI) Chatbots for Providing Sexual Health Information: A Consensus Study Using Clinical Queries.

Li, S., Mou, Y., & Xu, J. (2024). Disclosing Personal Health Information to Emotional Human Doctors or Unemotional AI Doctors? Experimental Evidence Based on Privacy Calculus Theory. *International Journal of Human–Computer Interaction*, 1-13.

Li, Y., & Goel, S. (2024). Making it possible for the auditing of AI: A systematic review of AI audits and AI auditability. *Information Systems Frontiers*, 1-31.

Liew, T. W., Tan, S. M., Yoo, N. E., Gan, C. L., & Lee, Y. Y. (2023). Let's talk about Sex!: AI and relational factors in the adoption of a chatbot conveying sexual and reproductive health information. *Computers in Human Behavior Reports*, 11, 100323.

Link, E., & Beckmann, S. (2024). AI at everyone's fingertips? Identifying the predictors of health information seeking intentions using AI. *Communication Research Reports*, 1-11.

Liu, J., Wang, J., Huang, H., Zhang, R., Yang, M., & Zhao, T. (2023, October). Improving LLM-Based Health Information Extraction with In-Context Learning. In *China Health Information Processing Conference* (pp. 49-59). Singapore: Springer Nature Singapore.

Ma, Y., Achiche, S., Pomey, M. P., Paquette, J., Adjtoutah, N., Vicente, S., ... & MARVIN chatbots MARVIN chatbots Patient Expert Committee. (2024). Adapting and evaluating an AI-Based Chatbot through patient and stakeholder engagement to provide information for different health conditions: Master Protocol for an Adaptive Platform Trial (the MARVIN Chatbots Study). *JMIR Research Protocols*, 13(1), e54668.

McMahon, E., Fetters, T., Jive, N. L., & Mpoyi, M. (2023). Perils and promise providing information on sexual and reproductive health via the Nurse Nisa WhatsApp chatbot in the Democratic Republic of the Congo. *Sexual and Reproductive Health Matters,* 31(4), 2235796.

Mendel, T., Nov, O., & Wiesenfeld, B. (2024). Advice from a Doctor or AI? Understanding Willingness to Disclose Information Through Remote Patient Monitoring to Receive Health Advice. *Proceedings of the ACM on Human-Computer nteraction*, 8(CSCW2), 1-34.

Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ digital medicine, 6(1), 120.

Mökander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. *Digital Society*, *2*(3), 49.

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2023). Auditing large language models: a three-layered approach. *AI and Ethics*, 1-31.

Mulgund, P., Mulgund, B. P., Sharman, R., & Singh, R. (2021). The implications of the California Consumer Privacy Act (CCPA) on healthcare organizations: Lessons learned from early compliance experiences. *Health Policy and Technology*, *10*(3), 100543.

Nankya, M., Mugisa, A., Usman, Y., Upadhyay, A., & Chataut, R. (2024). Security and Privacy in E-Health Systems: A Review of AI and Machine Learning Techniques. *IEEE Access*.

Nyarko, A. J. (2024). Exploring Ghanaian Tertiary Students' Perceptions Towards AI as a First-Hand Source of Health Information for Diagnosis and Self-Medication. *Journal of Health Informatics in Africa*, 11(1), 64-76.

Ono, G. N., Obi, E. C., Chiaghana, C., & Ezegwu, D. (2024). Digital Divide and Access: Addressing Disparities in Artificial Intelligence (Ai) Health Information for Nigerian Rural Communities. *Social Science Research*, 10(3).

Park, J., Singh, V., & Wisniewski, P. (2023). Supporting youth mental and sexual health information seeking in the era of artificial intelligence (ai) based conversational agents: Current landscape and future directions. Available at SSRN 4601555.

Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of internal medicine*, *169*(12), 866-872.

Raddatz, N., Kettinger, W. J., & Coyne, J. (2023). Giving to Get Well: Patients' Willingness to Manage and Share Health Information on AI-Driven Platforms. *Communications of the Association for Information Systems*, 52(1), 1017-1049.

Rezaee, Z., Homayoun, S., Poursoleyman, E., & Rezaee, N. J. (2023). *Giving to Get Well: Patients' Willingness to Manage and Share Health Information on AI-Driven Platforms.* Global Finance Journal, 55.

Sakriwattana, K. (2024). Factor affecting intention to use chatbot for health information. *Procedia of Multidisciplinary Research*, 2(7), 5-5.

Shen, S. A., Perez-Heydrich, C. A., Xie, D. X., & Nellis, J. C. (2024). ChatGPT vs. web search for patient questions: what does ChatGPT do better?. *European Archives of Oto-Rhino-Laryngology*, 281(6), 3219-3225.

Shi, X., Liu, J., Liu, Y., Cheng, Q., & Lu, W. (2024). Know where to go: Make LLM a relevant, responsible, and trustworthy searchers. *Decision Support Systems*, 114354.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *10*(4), 1-31.

Swar, B., Hameed, T., & Reychav, I. (2017). Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. *Computers in human behavior*, *70*, 416-425.

Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., ... & Naganawa, S. (2024). Fairness of artificial intelligence in healthcare: review and recommendations. *Japanese Journal of Radiology*, *42*(1), 3-15.

Vaira, L. A., Lechien, J. R., Abbate, V., Allevi, F., Audino, G., Beltramini, G. A., ... & De Riu, G. (2024).

Validation of the QAMAI tool to assess the quality of health information provided by AI. medRxiv, 2024-01.

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97-112.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., ... & Gabriel, I. (2022, June). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 214-229).

Xiao, Z+3:50., Liao, Q. V., Zhou, M., Grandison, T., & Li, Y. (2023, March). Powering an ai chatbot with expert sourcing to support credible health information access. In *Proceedings of the 28th international conference on intelligent user interfaces* (pp. 2-18).

Yan, Y., Hou, Y., Xiao, Y., Zhang, R., & Wang, Q. (2024). KNOWNET: Guided Health Information Seeking from LLMs via Knowledge Graph Integration. *IEEE Transactions on Visualization and Computer Graphics*.

Yiannakoulias, N. (2024). Spatial intelligence and contextual relevance in AI-driven health information retrieval. *Applied Geography*, 171, 103392.

Yin, R., & Neyens, D. M. (2024). Examining how information presentation methods and a chatbot impact the use and effectiveness of electronic health record patient portals: An exploratory study. *Patient Education and Counseling*, 119, 108055.

APPENDIX

Table 3: Systematic Literature Review – Mapping key features, opportunities, challenges and issues of AI systems in literature onto Health-AIDE AI Auditing framework.

			Reliability & Accuracy			**Trust and Acceptance**			**Security and Safety**			**Equity**		
	Reference	**AI Scope**	Accuracy	Monitoring (Currency)	Scalability	Transparency	User Trust	Human-AI Collaboration	Data Governance	Harm Prevention	Privacy and Misuse Prevention	Bias and Fairness	Ethical and Legal Compliance	Sustainability
1	Xiao, Z. et al., (2023)	Expert-sourced chatbot; Design, Evaluation, OHI	X	X			X	X		X		X	X	
2	Vaira et al. (2024)	OHI, Quality assessment of medical Artificial Intelligence (QAMAI) based on mDISCERN	X	X										
3	Hatamlah, H. (2024)	Parental oversight and communication to ensure reliable and trustworthy info and behavior	X				X			X				
4	Yiannakoulias, N. (2024)	Spatial dispersion and OHI	X									X	X	
5	Amin, M. S., et al., (2024).	Caregiver robots interaction, Disclosing personal health info (PHI)					X		X	X	X		X	

Table 3: Systematic Literature Review – Mapping key features, opportunities, challenges and issues of AI systems in literature onto Health-AIDE AI Auditing framework (cont.).

			Reliability & Accuracy			**Trust and Acceptance**			**Security and Safety**			**Equity**		
	Reference	**AI Scope**	Accuracy	Monitoring (Currency)	Scalability	Transparency	User Trust	Human-AI Collaboration	Data Governance	Harm Prevention	Privacy and Misuse Prevention	Bias and Fairness	Ethical and Legal Compliance	Sustainability
6	Liew, T. W., et al. (2023)	Chatbot, Youth interactions on sexual and reproductive health (SRH) information,				X	X	X	X				X	
7	Link, E., & Beckmann, S. (2024).	OHI seeking	X											
8	Li, S.et al., (2023)	AI doctors, disclosure of PHI					X	X	X		X		X	
9	Raddatz, N.et al. (2023)	Internet of Medical Things, disclosure of PHI					X		X	X			X	
10	Park, J. et al. (2023)	Sexual and reproductive chatbots, adolescents, OHI	X					X	X	X		X	X	
11	Nyarko, A. J. (2024)	Engaging with OHI for self-care purposes, Ghanian students					X	X				X	X	
12	Ono, G.N. et al. 2024	Digital gaps, AI access, Nigerian students										X	X	
13	Alhendawi, K. M. (2024).	Connected medical devices cloud network, Exchange of health info	X	X					X				X	
14	Hong, S. J.et al. (2023)	Users considering adoption of contact-tracing app., COVID-19							X		X			
15	Mendel, T. et al.2024	Remote patient monitoring devices, dis-closure of PHI					X	X	X					
16	Latt, P. M.et al. (2024)	Sexual health, prompt-tuned chatbots, Australia	X	X	X			X		X				
17	Rezaee, Z.et al. (2023)	ESG disclosures by firms, EU and others							X				X	X

Table 3: Systematic Literature Review – Mapping key features, opportunities, challenges and issues of AI systems in literature onto Health-AIDE AI Auditing framework (cont.).

			Reliability & Accuracy			**Trust and Acceptance**			**Security and Safety**			**Equity**		
	Reference	**AI Scope**	Accuracy	Monitoring (Currency)	Scalability	Transparency	User Trust	Human-AI Collaboration	Data Governance	Harm Prevention	Privacy and Misuse Prevention	Bias and Fairness	Ethical and Legal Compliance	Sustainability
18	Alkhalaf, M.et al. (2024)	Zero-shot prompt engineering, RAG, EHRs, AI Hallucinations			X	X		X						
19	Ma, Y.et al. (2024)	MARVIN Chatbots, Multiple disease diagnostics and care		X	X			X	X					
20	Liu, J.et al. (2023)	CHINA, Health information extraction, In-context learning	X	X	X									
21	Chang, C. W.et al. (2024)	Zero-shot de identification of health data	X	X					X	X				
22	Nguyen, V. C.et al. (2024)	Mental health, Misinformation , Few-shot	X							X				
23	Jin, E.et al. (2024)	Virtual Doctors/ Physicians, Chatbots	X				X	X			X			
24	Harrington, C. N. et al. (2024)	Chatbots, Gender and race associations					X	X	X			X	X	
25	Yin, R., & Neyens, D. M. (2024).	Design, Chatbots, EHRs, Patient Portals				X		X						
26	Kunlerd, A. (2024).	Line Chatbots, Diet regimens, Health		X		X	X	X		X		X		
27	Sakriwattana, K. (2024).	Chatbots, Intergenerational differences, Age gaps		X				X		X		X		
28	Aparicio, E.et al. (2023)	Chatbots, Pre-natal care, Post-partum care, Mothers of color	X				X	X		X		X	X	
29	McMahon, E.et al. (2023)	Congo, Sexual and Reproductive Health, NISA chatbot	X		X		X	X		X		X		
30	Ascorbe, P.et al. (2023)	Mental health, Suicide Preven-tion, Chatbot, WhatsApp, Spain	X					X		X		X	X	