AI and Machine Learning in Clinical Medicine: Generative and Interactive Systems at the Human-Machine Interface

Fateme Nateghi Haredasht
Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA
Email: fnateghi@stanford.edu

Dokyoon Kim
Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA
Email: dokyoon.kim@pennmedicine.upenn.edu

Joseph D. Romano
Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA
Email: joseph.romano@pennmedicine.upenn.edu

Geoff Tison
Center for Biosignal Research, University of California, San Francisco San Francisco, CA, USA
Email: geoff.tison@ucsf.edu

Roxana Daneshjou
Department of Biomedical Data Science, Stanford University Stanford, CA, USA
Email: roxanad@stanford.edu

Jonathan H. Chen
Department of Medicine and Center for Biomedical Informatics Research, Stanford University Stanford, CA, USA
Email: jonc101@stanford.edu

Artificial Intelligence (AI) technologies are increasingly capable of processing complex and multi-layered datasets. Innovations in generative AI and deep learning have notably enhanced the extraction of insights from both unstructured texts, images, and structured data alike. These breakthroughs in AI technology have spurred a wave of research in the medical field, leading to the

creation of a variety of tools aimed at improving clinical decision-making, patient monitoring, image analysis, and emergency response systems. However, thorough research is essential to fully understand the broader impact and potential consequences of deploying AI within the healthcare sector.

Keywords: Artificial Intelligence, clinical medicine, decision support systems, large language models.

1. Introduction

The integration of Artificial Intelligence (AI) into clinical medicine continues to expand at a rapid pace, promising transformative changes across diagnostics, treatment planning, and patient monitoring [1], [2], [3], [4], [5]. While AI technologies offer remarkable capabilities in analyzing extensive and complex medical data sets, their real-world application necessitates robust frameworks that support explainability and generalizability. These aspects are crucial for building trust among clinicians and patients alike, ensuring that AI-driven interventions are both understandable and applicable across diverse clinical environments. This session showcases pioneering research that addresses these needs, highlighting innovative solutions that aim to set new standards in the deployment of AI tools in medicine. The emergence of large language models (LLMs) and other sophisticated AI systems has propelled forward our ability to interpret and utilize medical data, promising significant improvements in patient outcomes [6], [7]. However, the deployment of such technologies must be accompanied by stringent evaluations to confirm their effectiveness and safety in real-world clinical settings [8], [9]. This includes understanding their potential biases, operational limitations, and their overall impact on clinical decision-making processes.

This year's session at the 2025 Pacific Symposium on Biocomputing (PSB), titled AI in Clinical Medicine: Towards Explainable and Generalizable AI Systems, concentrates on the latest advancements in AI that not only enhance clinical effectiveness but also prioritize transparency, adaptability, and ethical implementation in healthcare settings. Here, we highlight the accepted submissions for this session and set the stage for a discussion of AI's role in revolutionizing medical practice, emphasizing the need for solutions that are not only technically proficient but also ethically sound and universally beneficial. As AI continues to permeate the healthcare landscape, this session provides a critical examination of both its achievements and the challenges that lie ahead in its journey from experimental algorithms to essential clinical tools.

2. Artificial Intelligence in Clinical Medicine

2.1. *AI for Clinical Decision Support and Medical Workflows*

AI has increasingly been integrated into clinical decision-making processes, providing support for tasks such as diagnostics, treatment planning, and risk prediction [10]. Decision support systems that incorporate AI can process vast amounts of clinical data in real time, offering clinicians enhanced insights into patient care [11]. These tools are particularly valuable in settings where time and precision are critical, such as emergency departments, oncology, and intensive care units.

Prince et al. 2025 present a visual analytics framework aimed at evaluating interactive AI systems in pediatric brain tumor diagnosis [12]. Their work underscores the importance of understanding how AI can support clinicians in decision-making and improve clinical workflows.

Bedi et al. 2025 introduce QUEST-AI, an innovative LLM-based system designed to generate and refine USMLE-style exam questions [13]. By automating this process, their system promises to reduce the time and cost involved in medical education.

Rao et al. 2025 tackle the challenge of error generation in radiology reports with CX-REGen, a system that uses LLMs to create synthetic errors in chest X-ray reports to improve AI model training [14].

Healey et al. 2025 introduce LLM-CGM, a benchmark for summarizing continuous glucose monitor (CGM) data using AI, with applications in enhancing diabetes management [15].

Lastly, Godeme et al. 2025 investigate the use of synthetic text for developing NLP models to support peer supporters [16]. This study demonstrates that AI-generated synthetic text can effectively augment training datasets, which enhances the fidelity of peer support tools. Their findings emphasize the utility of AI in improving both training and support mechanisms for peer-assisted health interventions.

2.2. *Improving AI Models for Critical Healthcare Tasks*

To ensure that AI models can handle the complexity and variability of healthcare data, improving their generalizability and performance is crucial. Healthcare environments are diverse, and AI models must perform well across different populations, institutions, and clinical settings. Additionally, AI tools need to handle both structured data, such as lab results and vital signs, and unstructured data, like clinical notes, to offer comprehensive support [17], [18]. Several papers in this session focus on optimizing the performance and utility of AI models in healthcare.

Shashikumar and Nemati 2025 present a comparative study of LLMs in sepsis prediction, demonstrating that smaller models can achieve performance levels comparable to larger ones, thus offering more resource-efficient solutions [19].

Wang et al. 2025 explore the use of LLMs in cancer registry coding, where AI models are applied to streamline and enhance the accuracy of reporting in real-world hospital settings [20].

Weissenbacher et al. 2025 developed an NLP-based system to evaluate the appropriateness of pediatric antibiotic prescriptions, contributing to improved antibiotic stewardship practices in healthcare [21].

2.3. *Ethical and Regulatory Considerations in AI Deployment*

The rapid development of AI in healthcare presents significant ethical and regulatory challenges. AI systems must be designed and deployed in ways that prioritize patient safety, privacy, and equity. As AI becomes more embedded in clinical workflows, it is critical to ensure that these systems comply with existing regulations and adapt to evolving legal frameworks [22]. Regulatory agencies, such as the FDA, are tasked with ensuring that AI technologies are safe, effective, and accessible to all patients [23].

Rincon et al. 2025 explore the evolving regulatory landscape in healthcare AI, focusing on how recent Supreme Court decisions could impact the authority of regulatory agencies like the FDA [24]. Their analysis highlights the potential implications of regulatory uncertainty for the healthcare industry.

In a related study, Levy et al. 2025 investigate the use of AI to predict suicide risk in veterans, integrating both structured and unstructured EHR data [25]. Their work underscores the ethical importance of using AI responsibly in sensitive areas like mental health.

2.4. *Generalizability and Validation of AI Models*

Ensuring that AI models are generalizable and can be validated across different clinical settings is essential for their widespread adoption [26]. AI systems trained on data from a single institution or population often struggle to perform well in other settings due to variations in patient demographics, clinical practices, and data collection methods. Generalizability is crucial for developing AI tools that can be deployed in diverse healthcare environments without compromising accuracy or fairness [27].

In this session, Banerjee et al. 2025 introduce a multi-site validation framework to test the robustness of radiology AI models across different populations and institutions, addressing the critical need for AI systems that can generalize beyond their training data [28].

Xiong et al. 2025 propose i-MedRAG, an iterative Retrieval-Augmented Generation (RAG) system designed to improve medical question-answering by incorporating follow-up queries, further enhancing AI's ability to handle complex clinical cases [29].

Ramwala et al. 2025 present ClinValAI, a cloud-based framework for the external validation of AI models in medical imaging, ensuring that these models meet high standards of performance and fairness [30].

Lastly, Keat et al. 2025 introduce PGxQA, a resource for evaluating the performance of LLMs on pharmacogenomic question-answering tasks [31]. This benchmark is designed to assess the ability of LLMs to provide clinically accurate information related to pharmacogenomics, which is crucial for ensuring these AI tools' safety and effectiveness when used in personalized medicine applications.

3. Conclusion

The papers presented in this session demonstrate the expanding role of AI in clinical medicine. They showcase a range of applications designed to improve diagnostic accuracy, enhance decision support, and address ethical and regulatory challenges. As AI continues to integrate into healthcare, the need for rigorous validation, regulatory oversight, and ethical deployment becomes increasingly important. These contributions highlight the promise of AI while addressing the ongoing challenges of ensuring that these systems are safe, explainable, and generalizable across diverse clinical environments.

References

[1] M. Khalifa and M. Albadawy, "Artificial Intelligence for Clinical Prediction: Exploring Key Domains and Essential Functions," *Comput. Methods Programs Biomed. Update*, vol. 5, p. 100148, Jan. 2024, doi: 10.1016/j.cmpbup.2024.100148.

[2] S. Bindra and R. Jain, "Artificial intelligence in medical science: a review," *Ir. J. Med. Sci. 1971 -*, vol. 193, no. 3, pp. 1419–1429, Jun. 2024, doi: 10.1007/s11845-023-03570-9.

[3] L. Weidener and M. Fischer, "Artificial Intelligence in Medicine: Cross-Sectional Study Among Medical Students on Application, Education, and Ethical Aspects," *JMIR Med. Educ.*, vol. 10, no. 1, p. e51247, Jan. 2024, doi: 10.2196/51247.

[4] H. Akhlaghi *et al.*, "Machine learning in clinical practice: Evaluation of an artificial intelligence tool after implementation," *Emerg. Med. Australas.*, vol. 36, no. 1, pp. 118–124, 2024, doi: 10.1111/1742-6723.14325.

[5] V. D. Karalis, "The Integration of Artificial Intelligence into Clinical Practice," *Appl. Biosci.*, vol. 3, no. 1, Art. no. 1, Mar. 2024, doi: 10.3390/applbiosci3010002.

[6] S. Reddy, "Evaluating large language models for use in healthcare: A framework for translational value assessment," *Inform. Med. Unlocked*, vol. 41, p. 101304, Jan. 2023, doi: 10.1016/j.imu.2023.101304.

[7] E. Ullah, A. Parwani, M. M. Baig, and R. Singh, "Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology – a recent scoping review," *Diagn. Pathol.*, vol. 19, no. 1, p. 43, Feb. 2024, doi: 10.1186/s13000-024-01464-7.

[8] C. T.-T. Chang *et al.*, "Red Teaming Large Language Models in Medicine: Real-World Insights on Model Behavior," *medRxiv*, pp. 2024–04, 2024.

[9] S. Gottlieb and L. Silvis, "How to Safely Integrate Large Language Models Into Health Care," *JAMA Health Forum*, vol. 4, no. 9, p. e233909, Sep. 2023, doi: 10.1001/jamahealthforum.2023.3909.

[10] A. F. Hernandez and C. J. Lindsell, "The Future of Clinical Trials: Artificial to Augmented to Applied Intelligence," *JAMA*, vol. 330, no. 21, pp. 2061–2063, Dec. 2023, doi: 10.1001/jama.2023.23822.

[11] R. Khera *et al.*, "AI in Medicine—JAMA's Focus on Clinical Outcomes, Patient-Centered Care, Quality, and Equity," *JAMA*, vol. 330, no. 9, pp. 818–820, Sep. 2023, doi: 10.1001/jama.2023.15481.

[12] E. W. Prince, T. C. Hankinson, and C. Gorg, "A Visual Analytics Framework for Assessing Interactive AI for Clinical Decision Support," presented at the Pacific Symposium on Biocomputing (PSB)., 2025.

[13] S. Bedi *et al.*, "QUEST-AI: A System for Question Generation, Verification, and Refinement using AI for USMLE-Style Exams," 2025. doi: 10.1101/2023.04.25.23288588.

[14] V. M. Rao, S. Zhang, J. Acosta, and P. Rajpurkar, "CX-REGen: Generating Complete, Consistent, and Clinically Relevant Errors Within Chest X-Ray Reports," presented at the Pacific Symposium on Biocomputing (PSB)., 2025.

[15] E. Healey and I. Kohane, "LLM-CGM: A Benchmark for Large Language Model-Enabled Summarization of Time Series Data from Wearable Continuous Glucose Monitors for Conversational Diabetes Management," *Pacific Symposium on Biocomputing (PSB).*, 2025.

[16] J. Godeme *et al.*, "Artificial Allies: Validation of Synthetic Text for Peer Support Tools through Data Augmentation in NLP Model Development," *Pacific Symposium on Biocomputing (PSB).*, 2025.

[17] W. S. Shah, Z. Elkhwesky, K. M. Jasim, E. F. Y. Elkhwesky, and F. F. Y. Elkhwesky, "Artificial intelligence in healthcare services: past, present and future research directions," *Rev. Manag. Sci.*, vol. 18, no. 3, pp. 941–963, Mar. 2024, doi: 10.1007/s11846-023-00699-w.

[18] I. Ghebrehiwet, N. Zaki, R. Damseh, and M. S. Mohamad, "Revolutionizing personalized medicine with generative AI: a systematic review," *Artif. Intell. Rev.*, vol. 57, no. 5, p. 128, Apr. 2024, doi: 10.1007/s10462-024-10768-5.

[19] S. P. Shashikumar and S. Nemati, "A Prospective Comparison of Large Language Models for Early Prediction of Sepsis," . *Pacific Symposium on Biocomputing (PSB).*, 2025.

[20] C.-K. Wang, M.-S. Huang, Y.-H. Yang, V. S. Tseng, and H.-J. Dai, "Using Large Language Models for Efficient Cancer Registry Coding in the Real Hospital Setting: A Feasibility Study," *Pacific Symposium on Biocomputing (PSB).*, 2025.

[21] D. Weissenbacher *et al.*, "Automated Evaluation of Antibiotic Prescribing Guideline Concordance in Pediatric Sinusitis Clinical Notes," 2025. doi: 10.1101/2024.08.09.24311714.

[22] G. Karimian, E. Petelos, and S. M. A. A. Evers, "The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review," *AI Ethics*, vol. 2, no. 4, pp. 539–551, Nov. 2022, doi: 10.1007/s43681-021-00131-7.

[23] T. Gundersen and K. Bærøe, "The Future Ethics of Artificial Intelligence in Medicine: Making Sense of Collaborative Models," *Sci. Eng. Ethics*, vol. 28, no. 2, p. 17, Apr. 2022, doi: 10.1007/s11948-022-00369-2.

[24] N. Rincon, S. Gerke, and J. K. Wagner, "Implications of An Evolving Regulatory Landscape on the Development of AI and ML in Medicine," presented at the Pacific Symposium on Biocomputing (PSB)., 2025.

[25] J. Levy, M. Dimambro, A. Diallo, J. Gui, B. Shiner, and M. Levis, "Investigating the Differential Impact of Psychosocial Factors by Patient Characteristics and Demographics on Veteran Suicide Risk Through Machine Learning Extraction of Cross- Modal Interactions," presented at the Pacific Symposium on Biocomputing (PSB)., 2025.

[26] J. Yang, A. A. S. Soltan, and D. A. Clifton, "Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening," *Npj Digit. Med.*, vol. 5, no. 1, pp. 1–8, Jun. 2022, doi: 10.1038/s41746-022-00614-9.

[27] M. D. Abràmoff *et al.*, "Considerations for addressing bias in artificial intelligence for health equity," *Npj Digit. Med.*, vol. 6, no. 1, pp. 1–7, Sep. 2023, doi: 10.1038/s41746-023-00913-9.

[28] O. Banerjee *et al.*, "A Multi-Site Framework for Validating Automatic Metrics in Radiology Generative AI," *Pacific Symposium on Biocomputing (PSB).*, 2025.

[29] G. Xiong, Q. Jin, X. Wang, M. Zhang, Z. Lu, and A. Zhang, "Improving Retrieval-Augmented Generation in Medicine with Iterative Follow-up Questions," *Pacific Symposium on Biocomputing (PSB).*, 2025.

[30] O. A. Ramwala *et al.*, "ClinValAI: A framework for developing Cloud-based infrastructures for the External Clinical Validation of AI in Medical Imaging," presented at the Pacific Symposium on Biocomputing (PSB)., 2025.

[31] K. Keat *et al.*, “PGxQA: A Resource for Evaluating LLM Performance for Pharmacogenomic QA Tasks,” presented at the Pacific Symposium on Biocomputing (PSB)., 2025.