The Role of Large Language Models in HCI for Personal Health Research

Lena Mamykina
om2196@cumc.columbia.edu
Columbia University
New York, New York, United States

Xuhai Xu
xx2489@columbia.edu
Columbia University
New York, New York, United States

Abstract

Technologies for personal health have been an area of active research in the HCI community for several decades. In recent years, the introduction of machine learning and artificial intelligence paved the way for a new generation of technologies for personal health. Specifically, large language models can enable human-like communication between individuals and AI on a variety of topics, including personal health, and are already actively explored in nutritional and mental health counseling. There are, however, many open questions as to appropriate and ethical introductions of LLMs in such sensitive contexts as personal health. Moreover, there are open questions as to the role of these technologies in HCI for personal health research. In this provocation, we outline some of these open questions, with the hope of starting this discussion in the HCI for health community.

ACM Reference Format:
Lena Mamykina and Xuhai Xu. 2025. The Role of Large Language Models in HCI for Personal Health Research . In *Proceedings of (CHI '25 Workshop on Envisioning the Future of Interactive Health).* ACM, New York, NY, USA, 4 pages.

1 Introduction

The use of technology for personal health has a long and rich history, evolving from simple self-tracking tools to sophisticated digital health interventions. Early systems focused primarily on personal informatics, enabling individuals to collect, visualize, and interpret their own health data [6, 15]. Over time, the field has shifted toward greater reliance on computational inferences, predictions, and recommendations, allowing for more proactive and intelligent health management [18, 21, 27].

Recent advances in large language models (LLMs) have introduced new opportunities in this space, sparking growing interest in their potential applications for personal health [3, 12, 16]. A rapidly emerging body of research explores how models like ChatGPT can be leveraged in domains such as nutrition counseling [24, 30], mental health support [13, 23, 28], and chronic disease management [20, 22]. Beyond personal health, this enthusiasm is reflected in a broader trend across multiple domains, as evidenced by the increasing number of studies in human-computer interaction (HCI) exploring LLM-driven solutions under the umbrella of "Can ChatGPT do X?"

CHI '25 Workshop on Envisioning the Future of Interactive Health, Yokohama, Japan

However, these optimistic expectations come with significant concerns. Researchers and practitioners are raising critical questions about the reliability, safety, and ethical implications of LLM-powered chatbots in sensitive health contexts [8, 11]. Unlike traditional health technologies, LLMs introduce unique risks, including misinformation, over-reliance, and challenges in ensuring accountability [7, 14]. As these models become more integrated into personal health technologies, the HCI community must carefully consider their appropriate role and impact in shaping the future of digital health and how both existing and new conferences in the field can promote rigorous and ethical research on LLMs in personal health.

In this paper, we explore key open questions at the intersection of LLMs, HCI, and personal health, outlining areas where future research and discussions can make meaningful contributions to the field. Our goal is to critically examine both the potential and limitations of LLMs in personal health contexts, fostering a dialogue on responsible and impactful innovation in this rapidly evolving landscape. We wish to articulate and discuss potential aspects of the academic work that are desirable for this proposed venue and valuable for the community.

2 An Incomplete Category List of Recent LLM-related Work in Personal Health Research

As a brief overview, we first provide a high-level, non-exhaustive list of directions that researchers have explored with LLMs in HCI for personal health.

2.1 LLM as a Health Detection / Prediction Method

LLMs have increasingly been applied as predictive tools in personal health. Recent research has demonstrated that LLMs can analyze personal health data—including self-reported information and sensor readings—to generate detections or predictions about health outcomes [12]. For instance, by integrating textual self-tracking data with quantitative metrics from wearables, various LLMs are leveraged to predict a broad range of outcomes, such as sleep quality, risk for chronic conditions, and mental health conditions, to name just a few [5]. Some recent studies show that fine-tuning domain-specific LLMs can further improve their predictive accuracy, aligning their performance with expert assessments [12, 13, 28].

2.2 LLM as Part of the Intervention Method

Innovative intervention systems have begun incorporating LLMs directly into behavioral health applications. These systems typically leverage conversational agents to deliver personalized health advice or support public health interventions [9]. Emerging literature shows applications in many fields, such as nutrition counseling [29], mental well-being support [23], and personal activity coaching [10], where chatbots powered by LLMs facilitate interactive, adaptive conversations that can mimic supportive therapy sessions. These interventions not only provide immediate, on-demand guidance but also integrate multimodal data (e.g., from wearables) to tailor recommendations in real time [25].

2.3 LLM as Role-Playing for Intervention Evaluation or Education

Beyond conventional evaluation metrics, LLMs have been employed as simulated participants to role-play patient scenarios during the testing of health interventions or delivery health education [17, 19]. In this capacity, LLMs assume the role of a patient, generating realistic responses and behaviors based on a predefined health profile or intervention scenario. This approach enables researchers to observe how an intervention performs in a controlled yet dynamic dialogue setting, effectively "stress-testing" the system's ability to handle a wide range of patient inputs without the immediate need for human subjects [1, 31]. By acting as virtual patients, these models allow for iterative refinement of intervention strategies—identifying potential communication breakdowns or misinterpretations—and provide valuable insights into the robustness and adaptability of digital health tools before broader clinical trials are initiated.

2.4 LLM as an Analysis Tool to Facilitate Research

Finally, LLMs have transformed research methodologies in personal health by serving as advanced analysis tools that synthesize large-scale qualitative data [2, 4]. They are now routinely employed to review and summarize patient narratives, clinical notes, and even social media discussions, enabling researchers to efficiently identify trends, biases, and emerging health concerns. Although there has been some ongoing debate about the reliability and potential bias of this research methodology, the advanced text-mining capability reduces the manual burden of literature reviews and data coding [4, 26]. Additionally, LLMs assist in drafting and refining research protocols, proposals, and publications, ensuring that insights derived from big data are rapidly translated into actionable research outcomes [11]. Together, these diverse applications underscore the transformative potential of LLMs in enhancing personal health monitoring, intervention delivery, evaluation processes, research methodologies, and more applications.

3 What's next with LLM for HCI & Personal Health?

Despite the promising capabilities of LLMs demonstrated through recent work, there is a long list of open questions and challenges that are yet to be explored. Below, we propose a set of provocative questions that we hope can generate a discussion within the research community and establish some directions for future research and publications that investigate the use and applicability of LLMs for personal health.

3.1 What mechanisms are needed to ensure alignment between LLM-based chatbots and individuals' needs and expectations in the context of personal health?

For LLM-based chatbots to be truly effective in personal health contexts, they must align with individuals' diverse needs, expectations, and health goals. Achieving this alignment requires a combination of personalization mechanisms, adaptability, and user feedback loops. Personalization can be achieved through context-aware models that take into account an individual's health history, preferences, and lifestyle while maintaining user privacy and control over data. Furthermore, adaptive dialogue strategies can help these models tailor their responses based on users' evolving needs, ensuring they remain relevant and supportive over time. Incorporating human-in-the-loop approaches, where expert review or reinforcement learning with human feedback (RLHF) helps refine chatbot behavior, can also enhance alignment. Additionally, transparency in chatbot operations—such as explaining reasoning, citing sources, or indicating uncertainty—can help users calibrate their trust and expectations, reducing the risk of over-reliance or misinterpretation of AI-generated health guidance.

3.2 What mechanisms can help to leverage the generative capabilities of LLMs while establishing appropriate safeguards to ensure adherence to evidence-based health communication patterns?

While LLMs possess powerful generative abilities, their open-ended nature raises concerns about hallucinations, misinformation, and deviations from evidence-based health communication. To mitigate these risks, effective safeguard mechanisms must be in place to ensure adherence to medical guidelines and best practices and investigating appropriate safeguards is an important direction for future research. For example, this research can investigate the possibility of constraining LLM outputs using retrieval-augmented generation (RAG), where responses are anchored to verified medical literature, clinical guidelines, or trusted health sources. Hybrid models combining LLMs with rule-based systems can further enhance reliability by ensuring that certain medical queries are addressed through predefined expert-vetted pathways. Additionally, algorithmic auditing and bias detection tools can help identify and mitigate potential risks of misleading or harmful advice. To reinforce evidence-based communication, chatbots can be equipped with explainability features, such as linking to authoritative sources, providing disclaimers, and prompting users to consult healthcare professionals when necessary. Designing LLM interactions that balance creativity with accountability is crucial in maintaining both engagement and trust in personal health applications.

3.3 How do individuals interpret LLM-based chatbots for personal health, and what interaction techniques can help them arrive at appropriate and accurate interpretations of LLM capabilities?

A major challenge in deploying LLM-based chatbots for personal health lies in how users interpret their responses, capabilities, and limitations. Unlike traditional medical tools, LLMs generate responses dynamically, which can lead to misconceptions about their reliability, expertise, or authority. Users may struggle to differentiate between AI-generated insights and clinically verified information, which underscores the need for effective interaction design strategies. An example of a mechanism that addresses this challenge is progressive disclosure, where the chatbot clarifies its role, limitations, and confidence level in its responses. Interactive affordances, such as citations, confidence scores, or direct comparisons between AI-generated and evidence-backed answers, can help users make more informed judgments. Additionally, dialogic techniques—such as prompting users to reflect on chatbot-generated advice or verify information with external sources—can encourage critical engagement rather than blind acceptance. By leveraging user-centered design principles, developers can create interaction mechanisms that help individuals navigate AI-generated health information responsibly and accurately.

3.4 How do LLM-based chatbots for personal health evolve over time and what are thier long-term trajectories in different scenarios?

The capabilities and roles of LLM-based chatbots in personal health are expected to evolve significantly over time, shaped by advances in AI, regulatory frameworks, and user adoption patterns. Initially, these chatbots may serve as supplementary tools—offering basic informational support, symptom checking, or mental health assistance. However, as models improve in contextual reasoning, multimodal understanding, and personalization, they may transition toward more proactive and adaptive health companions that assist users in long-term health management, lifestyle coaching, and behavioral interventions. The regulatory landscape will also play a crucial role in shaping their evolution, determining the extent to which LLMs can integrate with electronic health records (EHRs), telemedicine, and clinical decision support systems. Furthermore, different usage scenarios—ranging from self-care applications to clinician-augmented interactions—may develop along divergent trajectories, each with distinct risks and benefits. Understanding these long-term dynamics is essential for anticipating future opportunities, ethical challenges, and research directions in the intersection of AI and personal health.

3.5 How to maintain a continuous and rigorous evaluation of LLM-based health applications?

Maintaining continuous and rigorous evaluation of LLM-based personal health applications is essential to ensure their safety, efficacy, and reliability in dynamic healthcare environments. This necessitates the development of comprehensive evaluation frameworks that encompass both pre-deployment assessments and ongoing post-deployment monitoring. Pre-deployment evaluations should rigorously test models using diverse datasets to identify potential biases and validate performance across various demographic groups. Post-deployment, continuous monitoring systems are also crucial to detect performance shifts, data drift, or emergent biases, thereby safeguarding against potential degradations over time. Implementing real-time LLM auditing mechanisms can facilitate the prompt identification and rectification of issues, ensuring that systems remain aligned with clinical standards and patient safety requirements. Moreover, fostering interdisciplinary collaboration among AI developers, healthcare professionals, and regulatory bodies is vital to establishing standardized evaluation protocols and promoting transparency in LLM-driven healthcare solutions. Such collaborative efforts are essential for the creation of robust guidelines and best practices, enhancing the trustworthiness and effectiveness of LLM-based applications in personal health contexts.

3.6 What are the ethical implications of deploying LLMs in personal health contexts, and how can ethical frameworks be established to guide responsible development and use?

The deployment of LLMs in personal health raises ethical questions related to autonomy, consent, and the potential for AI to influence personal health decisions. Establishing ethical frameworks that define the boundaries of AI intervention in personal health is imperative. These frameworks should address issues such as informed consent, transparency in AI decision-making processes, and the right to opt-out of AI-driven recommendations. Engaging stakeholders—including ethicists, healthcare providers, patients, and policymakers—in the development of these guidelines will ensure that diverse perspectives are considered, promoting the responsible integration of LLMs into personal health domains.

4 Conclusion

In this position paper, our goal was to raise some open questions regarding the appropriate use of LLMs in applications for personal health. The summary and questions above are meant as provocative and illustrative, rather than as an exhaustive list of different research directions in this area. We believe that establishing clear criteria for relevant and appropriate research on LLMs in personal health will be important for any publishing venue, including new ones as well as well-established ones. There is a considerable risk of publication venues flooded with submissions simply describing ways to adapt LLM-based chatbots to specific tasks, which may lack research insights that can move our community forward. We look forward to discussing these questions with workshop participants.

References

[1] Siyuan Chen, Mengyue Wu, Kenny Q. Zhu, Kunyao Lan, Zhiling Zhang, and Lyuchun Cui. 2023. LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation. doi:10.48550/arXiv.2305.13614 arXiv:2305.13614 [cs].

[2] Robert Chew, John Bollenbacher, Michael Wenger, Jessica Speer, and Annice Kim. 2023. LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding. doi:10.48550/arXiv.2306.14924 arXiv:2306.14924 [cs].

[3] Justin Cosentino, Anastasiya Belyaeva, Xin Liu, Nicholas A Furlotte, Zhun Yang, Chace Lee, Erik Schenck, Yojan Patel, Jian Cui, Logan Douglas Schneider, et al. 2024. Towards a personal health large language model. *arXiv preprint arXiv:2406.06474* (2024).

[4] Shih-Chieh Dai, Aiping Xiong, and Lun-Wei Ku. 2023. LLM-in-the-loop: Leveraging Large Language Model for Thematic Analysis. doi:10.48550/arXiv.2310.15100 arXiv:2310.15100 [cs].

[5] Zachary Englhardt, Chengqian Ma, Margaret E. Morris, Chun-Cheng Chang, Xuhai "Orson" Xu, Lianhui Qin, Daniel McDuff, Xin Liu, Shwetak Patel, and Vikram Iyer. 2024. From Classification to Clinical Insights: Towards Analyzing and Reasoning About Mobile and Behavioral Health Data With Large Language Models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 2 (May 2024), 1–25. doi:10.1145/3659604

[6] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Osaka Japan, 731–742. doi:10.1145/2750858.2804250

[7] Joschka Haltaufderheide and Robert Ranisch. 2024. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). *NPJ Digital Medicine* 7 (July 2024), 183. doi:10.1038/s41746-024-01157-x

[8] Lujain Ibrahim, Saffron Huang, Lama Ahmad, and Markus Anderljung. 2024. Beyond static AI evaluations: advancing human interaction evaluations for LLM harms and risks. doi:10.48550/arXiv.2405.10632 arXiv:2405.10632 [cs].

[9] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3544548.3581503

[10] Matthew Jörke, Shardul Sapkota, Lyndsea Warkenthien, Niklas Vainio, Paul Schmiedmayer, Emma Brunskill, and James Landay. 2024. Supporting Physical Activity Behavior Change with LLM-Based Conversational Agents. doi:10.48550/arXiv.2405.06061 arXiv:2405.06061 [cs].

[11] Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. 2024. "I'm categorizing LLM as a productivity tool": Examining ethics of LLM use in HCI research practices. doi:10.48550/arXiv.2403.19876 arXiv:2403.19876 [cs].

[12] Yubin Kim, Xuhai Xu, Daniel McDuff, Cynthia Breazeal, and Hae Won Park. 2024. Health-LLM: Large Language Models for Health Prediction via Wearable Sensor Data. In *Proceedings of Machine Learning Research*. arXiv. http://arxiv.org/abs/2401.06866 arXiv:2401.06866 [cs].

[13] Tin Lai, Yukun Shi, Zicong Du, Jiajie Wu, Ken Fu, Yichao Dou, and Ziqi Wang. 2023. Psy-LLM: Scaling up Global Mental Health Psychological Services with AI-based Large Language Models. http://arxiv.org/abs/2307.11991 arXiv:2307.11991 [cs].

[14] Moustafa Laymouna, Yuanchao Ma, David Lessard, Tibor Schuster, Kim Engler, and Bertrand Lebouché. 2024. Roles, Users, Benefits, and Limitations of Chatbots in Health Care: Rapid Review. *Journal of Medical Internet Research* 26 (July 2024), e56930. doi:10.2196/56930

[15] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Atlanta Georgia USA, 557–566. doi:10.1145/1753326.1753409

[16] Lingyao Li, Jiayan Zhou, Zhenxiang Gao, Wenyue Hua, Lizhou Fan, Huizi Yu, Loni Hagen, Yongfeng Zhang, Themistocles L. Assimes, Libby Hemphill, and Siyuan Ma. 2024. A scoping review of using Large Language Models (LLMs) to investigate Electronic Health Records (EHRs). doi:10.48550/arXiv.2405.03066 arXiv:2405.03066 [cs].

[17] Yanzeng Li, Cheng Zeng, Jialun Zhong, Ruoyu Zhang, Minhao Zhang, and Lei Zou. 2024. Leveraging Large Language Model as Simulated Patients for Clinical Education. doi:10.48550/arXiv.2404.13066 arXiv:2404.13066 [cs].

[18] James J. Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B. Strub. 2006. Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp 2006: Ubiquitous Computing*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Paul Dourish, and Adrian Friday (Eds.). Vol. 4206. Springer Berlin Heidelberg, Berlin, Heidelberg, 261–278. doi:10.1007/11853565_16 Series Title: Lecture Notes in Computer Science.

[19] Ryan Louie, Ananjan Nandi, William Fang, Cheng Chang, Emma Brunskill, and Diyi Yang. 2024. Roleplay-doh: Enabling Domain-Experts to Create LLM-simulated Patients via Eliciting and Adhering to Principles. doi:10.48550/arXiv.2407.00870 arXiv:2407.00870 [cs].

[20] Lena Mamykina, Elliot Mitchell, Pooja Desai, and David Albers. 2024. Intelligent Decision Support in Personal Health: Personalized Health Coaching in Type 2 Diabetes. In *Human Computer Interaction in Healthcare: The Role of Cognition*, Andre W. Kushniruk, David R. Kaufman, Thomas G. Kannampallil, and Vimla L. Patel (Eds.). Springer Nature Switzerland, Cham, 413–438. doi:10.1007/978-3-031-69947-4_17

[21] Elliot G. Mitchell and Lena Mamykina. 2024. Chapter 22 - Artificial Intelligence to Support Self-management and Coaching. In *Diabetes Digital Health, Telehealth, and Artificial Intelligence*, David C. Klonoff, David Kerr, and Juan C. Espinoza (Eds.). Academic Press, 275–286. doi:10.1016/B978-0-443-13244-5.00007-9

[22] Sara Montagna, Stefano Ferretti, Lorenz Cuno Klopfenstein, Antonio Florio, and Martino Francesco Pengo. 2023. Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*. 205–212.

[23] Jingping Nie, Hanya (Vera) Shao, Yuang Fan, Qijia Shao, Haoxuan You, Matthias Preindl, and Xiaofan Jiang. 2025. LLM-based Conversational AI Therapist for Daily Functioning Screening and Psychotherapeutic Intervention via Everyday Smart Devices. *ACM Trans. Comput. Healthcare* (Jan. 2025). doi:10.1145/3712299 Just Accepted.

[24] Annalisa Szymanski, Brianna L Wimer, Oghenemaro Anuyah, Heather A Eicher-Miller, and Ronald A Metoyer. 2024. Integrating Expertise in LLMs: Crafting a Customized Nutrition Assistant with Refined Template Instructions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–22. doi:10.1145/3613904.3641924

[25] Ruolan Wu, Chun Yu, Xiaole Pan, Yujia Liu, Ningning Zhang, Yue Fu, Yuhan Wang, Zhi Zheng, Li Chen, Qiaolei Jiang, Xuhai Xu, and Yuanchun Shi. 2024. MindShift: Leveraging Large Language Models for Mental-States-Based Problematic Smartphone Use Intervention. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–24. doi:10.1145/3613904.3642790

[26] Ziang Xiao, Xingdi Yuan, Q. Vera Liao, Rania Abdelghani, and Pierre-Yves Oudeyer. 2023. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *28th International Conference on Intelligent User Interfaces*. ACM, Sydney NSW Australia, 75–78. doi:10.1145/3581754.3584136

[27] Xuhai Xu, Xin Liu, Han Zhang, Weichen Wang, Subigya Nepal, Yasaman Sefidgar, Woosuk Seo, Kevin S. Kuehn, Jeremy F. Huckins, Margaret E. Morris, Paula S. Nurius, Eve A. Riskin, Shwetak Patel, Tim Althoff, Andrew Campbell, Anind K. Dey, and Jennifer Mankoff. 2023. GLOBEM: Cross-Dataset Generalization of Longitudinal Human Behavior Modeling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–34. doi:10.1145/3569485

[28] Xuhai Xu, Bingsheng Yao, Yuanzhe Dong, Saadia Gabriel, Hong Yu, James Hendler, Marzyeh Ghassemi, Anind K. Dey, and Dakuo Wang. 2024. Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 1 (March 2024), 1–32. doi:10.1145/3643540

[29] Cheng Xue, Yijie Guo, Ziyi Wang, Mona Shimizu, Jihong Jeung, and Haipeng Mi. 2024. DishAgent: Enhancing Dining Experiences through LLM-Based Smart Dishes. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '24)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3672539.3686755

[30] Zhongqi Yang, Elahe Khatibi, Nitish Nagesh, Mahyar Abbasian, Iman Azimi, Ramesh Jain, and Amir M. Rahmani. 2024. ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health* 32 (June 2024), 100465. doi:10.1016/j.smhl.2024.100465

[31] Huizi Yu, Jiayan Zhou, Lingyao Li, Shan Chen, Jack Gallifant, Anye Shi, Xiang Li, Wenyue Hua, Mingyu Jin, Guang Chen, Yang Zhou, Zhao Li, Trisha Gupte, Ming-Li Chen, Zahra Azizi, Yongfeng Zhang, Themistocles L. Assimes, Xin Ma, Danielle S. Bitterman, Lin Lu, and Lizhou Fan. 2024. AIPatient: Simulating Patients with EHRs and LLM Powered Agentic Workflow. doi:10.48550/arXiv.2409.18924 arXiv:2409.18924 [cs].