RESEARCH STATEMENT Michelle Hoang Nguyen

Over the past six years of my PhD, I have focused on integrating data science solutions into clinical systems through collaborative research with key domain stakeholders. By prioritizing stakeholder perspectives, I have developed more usable and effective tools to address healthcare delivery challenges. In the process of this research, I have grown increasingly interested in how artificial intelligence (AI) tools have influenced the lives of both patients and providers, shaping the ways patients receive care and how providers practice. As generative AI is further embedded into our collective consciousness (and in many ways, becomes a new collaborator in our lives), the goal for my research is to design adaptive, human-centered AI systems to better mitigate potential harm and maximize benefit for all users.

However, achieving these goals in healthcare is challenging. We must adhere to validated clinical standards while also effectively serving patients and clinicians whose preferences vary widely, informed by values that are not right or wrong, but deeply held and legitimate. My long-term goal is to study how we may create usable, practical AI solutions for patients and providers, informed by pluralistic alignment methods. To mitigate potential harms, I aim to experiment with constraining health-focused generative AI tools through post-training alignment and knowledge integration. To maximize benefit, I commit to stakeholder-engaged research practices throughout the research lifecycle, from design and instrument development to deployment and evaluation.

Training and prior research

During my Ph.D., I have received training in biomedical data science and clinical informatics at Johns Hopkins, where I have been afforded the opportunity to collaborate regularly and manage research with diverse teams of clinical genetics professionals, computer scientists, and informaticists in the pursuit of creating new modalities to augment clinical workflows. I have been awarded NIH/NCATS TL1 Clinical and Translational Science Research Program and NIH/NIGMS T32 Computational Medicine pre-doctoral fellowships in support of this work.

Characterizing and augmenting clinical genetics care through structured and unstructured electronic health record (EHR) data. The EHR is a rich source of structured and unstructured information for understanding patient risk for disease, disease trajectories, and clinical practice, but the data is oftentimes messy and difficult to extract. I have designed and developed machine learning and clinical natural language processing algorithms to maximize usable information from the EHR to understand and support clinical genetics practice. These include efficient and novel knowledge-integrated approaches that span from rule-based methods to LLM-enabled tools, intentionally designed to align with genetic medicine knowledge. My work has included the detection of clinically useful phenotypic features from clinical notes of patients suspected of rare Mendelian disorders [1], the automated extraction of efficiency measures from genetic counseling notes, reducing annotation time by 99% [2], and knowledge-graph augmented extraction of family health history from diverse clinical and consumer data sources [3].

Developing consumer-facing digital health chatbots. I am committed to empowering consumers to understand and act on their health by designing and testing consumer-facing digital health tools, particularly chatbots. I have investigated *KIT,* a novel consumer-facing flow-based chatbot for family history (FHx) data collection, assessing the usability, engagement and report usefulness [4]. Building upon user feedback from this study, I have additionally explored consumers' and clinical genetics professionals' preferences for an LLM-enabled patient-facing FHx report. From preliminary findings, we have uncovered there are differential preferences towards family health history report guidance, motivating future work to create personalized educational material that satisfies heterogeneous patient and health provider perspectives. I have also led a student team to develop *Strolr,* a retrieval-augmented-generation chatbot to answer consumer health pregnancy queries, with the design informed by semi-structured interviews of key pregnancy and clinical stakeholders [5]. For both projects, I have generated quantitative and qualitative insights on how to develop clinical context-driven LLM tools that improve health information utility for consumers. Gaps in health information access disproportionately affect patient groups belonging to demographic minorities and those with limited health literacy. By improving health information collection and access through guided, informative chatbots, these tools can lower barriers for diverse populations to engage with their health.

Through this stakeholder-engaged research approach, from interviewing genetic counselors about family health history collection and documentation to testing chatbot designs with pregnant users, I deeply believe that effective AI safety in healthcare requires both technical robustness and responsiveness to the diverse values of the people these systems serve.

Future work and goals

Through my work in the clinical domain, I have become deeply invested in and concerned with how LLMs can be designed with more robust knowledge integration to be safer and more trustworthy for users in high-risk scenarios. Because of this, I aim to build upon my prior work by further research training in alignment and post-training research. For my future work, my goal is to study the design of human-aligned, adaptive LLM-enabled tools by explicitly incorporating human knowledge through retrieval prompting strategies and post-training alignment with knowledge graphs, while remaining attentive to pluralistic user needs through user studies with patients and providers.

How can we design knowledge-grounded chatbots that balance conversational flexibility with data quality, user safety, and privacy? Chatbots, in various domains, have demonstrated the ability to elicit high levels of self-disclosure from users, and in this way, show great promise in automating data collection tasks but also present potential dangers for users. In prior work to develop *KIT*, we learned that chatbot prompts and the logical flow of questions have an essential role in the ultimate family history data completeness. In human-led conversations and interviews, humans adapt their questions based on their conversational partner in a highly flexible way, with additional probing questions or a more circuitous, conversational route. Additionally, users bring different expectations, communication styles, and values to these interactions that require consideration. How can we design chatbots that better mimic this human flexibility of additional questioning based on inferred user preferences through written cues, while maintaining the structure to capture a wide range of necessary questions, as is needed for complex data collection tasks? I am interested in how embedding explicit knowledge to inform conversational flow, user preferences, and specifying different levels of rigidity to a specified conversational flow affects the resulting data quality. Additionally, for data collection chatbots designed to collect sensitive health data, such as family history, can we ensure data privacy (through frameworks like contextual integrity) while respecting patients' varying comfort levels with disclosure and their cultural or personal values around sharing family information?

How can we build effective safeguards for LLM tools that protect users with different levels of knowledge, health literacy, and values? From clinical stakeholder interviews to test *Strolr*, we learned that simply telling users to refer to their healthcare providers for personalized guidance as part of its response to common, harmless health questions was well-intentioned, but perhaps not clinically useful. In human-to-human interactions, we adjust our guidance based on the complexity and danger of a person’s question as well as our inference about their level of expertise and knowledge. I am interested in developing adaptive safeguards for LLMs that account for user context and expertise to reduce harm, recognizing that what constitutes "helpful" guidance may differ across users with varying cultural backgrounds, health beliefs, and risk tolerances. How can we incorporate domain knowledge to assess user queries for potential harm and the severity of outcomes if users are given an incorrect answer and use that to inform the LLM response? If we present users with a chatbot’s reasoning derived from its knowledge-based decision-making, how would this influence their behavior and trust in the system?

My experience implementing generative AI systems in high-stakes clinical settings has shaped a genuine commitment to AI safety and human-aligned design. I have seen firsthand non-expert users excited about the potential of generative AI, but also wary of its implications for their privacy and safety. My research vision is to contribute to a future where human-AI interaction is built on mutual trust and interrogation, where humans of different backgrounds, experiences, and values can question AI systems and AI systems will remain reliably transparent, accountable, and responsive to that plurality.

References

[1] Yang K.K., **Nguyen M.H.**, Jelin A., Rouhizadeh M., Sobreira N., Taylor C.O. (March 2023). "Detecting Phenotypes Among Patients Suspected of Rare Mendelian Disorders." *AMIA Informatics Summit 2023*, Seattle, WA.

[2] **Nguyen M.H**., Applegate C., Murray B., Zirikly A., Tichnell C., Pendleton C., Gordon C., Yanek L.R., James C.A. Taylor C.O. (2025). Generating Real-World Evidence of Genetic Counseling Efficiency with Natural Language Processing. Journal of the American Medical Informatics Association. https://doi.org/10.1093/jamia/ocaf190.

[3] **Nguyen M.H**., Soley N., Zirikly A., Taylor C.O. Improving quality of family health history structured information retrieval with ontology-augmented large language model retrieval. [In preparation for BioNLP @ ACL 2026]

[4] **Nguyen M.H.,** Sedoc, J., & Taylor C. O. (2024). Usability, engagement, and report usefulness of chatbot-based family health history data collection: Mixed-methods analysis. Journal of Medical Internet Research. doi:10.2196/55164. http://dx.doi.org/10.2196/55164

[5] **Nguyen M.H.,** Soley N., Rattsev I., Jelin A., Taylor C.O. (November 2024). "Strolr: An LLM-enabled Chatbot to Support Pregnant Women's Quick and Easy Information Seeking from Trustworthy Sources." Systems demonstration. AMIA National Symposium 2024, San Francisco, CA.