Multi-stakeholder Co-design and Trust Calibration as Emerging Grand Challenges in Visual Analytics for Healthcare: Findings from VAHC 2025

Gabriela Morgenshtern, MSc.[1,2], Alessio Arleo, PhD[3], David Gotz, PhD[4], Robert Laramee, PhD[5], Sylvia Miksch, PhD[6], Renata Raidou, PhD[6], Jürgen Bernard, PhD[1,2]
[1]University of Zurich, Zurich, ZH; [2]Digital Society Initiative, Zurich, ZH; [3]Eindhoven University of Technology, Eindhoven, NL; [4]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA [5]University of Nottingham, Nottingham, UK; [6]TU Wien, Vienna, AT

Introduction

Visual Analytics in Healthcare (VAHC) sits at the intersection of interactive visual data analytics and clinical informatics. Over sixteen iterations of the workshop series at AMIA and IEEE VIS, and two major JAMIA publications on challenge areas [1, 2], the community has maintained longstanding awareness of open problems: integrating visualization into clinical workflows, scaling tools across heterogeneous user populations, and establishing evaluation standards commensurate with clinical evidence requirements. Here, we draw directly from the expertise of 40 active visualization and clinical informatics researchers who participated in a recent round-table session on the Grand Challenges of VAHC, held at IEEE VIS 2025. In learning health systems (LHS), clinical practice data feed back to iteratively refine decision support, making it an adaptive, human-facing intervention [3, 4]. This raises two challenges for the design and governance of LHS: that co-design must serve stakeholders who both produce and consume the system's data, requiring methodological investment beyond current design study practice; and that trust calibration must track a model whose confidence shifts with accumulated outcomes, requiring visualization-specific operationalization to address uncertainty representation, risk communication, and user trust.

Methods

Pre-Workshop Submission Coding. To ground these challenges in the community's priorities, we coded the 15 accepted VAHC 2025 submissions against a controlled vocabulary of 20 thematic codes, derived inductively from the submissions[1], and further informed by prior VAHC workshop proceedings [2]. Coding was performed independently by the first author and subsequently reviewed by all co-authors, with disagreements resolved by consensus. Similarity-based clustering grouped the 15 submissions into three themes, each characterized by a subset of the codes: ***Technical Innovation vs. Clinical Reality*** (*LLM applications, LLM-supported visualization, human-centered AI (HCAI), clinical policy moderated through software intervention, decision-making support*); ***Human-Centered Scalable VAHC*** (*user task characterization, evaluation methodology, personal healthcare, clinical healthcare, wearable data, genomics, scalable data analysis*); and ***From Foundations to Actionable Insights*** (*visualization fundamentals, exploratory data analysis support, insight-based workflow, reasoning under uncertainty*). Every submission falls within at least one theme, and theme labels include the codes most characteristic of each cluster. The remaining four codes (*athletics support, dimensionality reduction, explanations of modeling results, and multimodal support*) were judged overly specific and retained for analytical provenance, but not used to inform the discussion prompts of the themes. Our codes and themes are available for exploration as an interactive visualization[2].

Workshop Execution. Each theme was formulated as an open discussion prompt. At the VAHC workshop at IEEE VIS 2025, 40 attendees engaged in 90-minute structured group discussions, with one discussion group assigned to each theme. Participants had expertise in visual analytics, health visualization, clinical informatics, and related areas. They contributed domain knowledge and used physical sticky notes to raise challenges and research directions. Three authors designed the discussion prompts and moderated the groups. Additional note-takers recorded the group discussions. Participants were informed that empirical measures from the session would be converted into a research writeup. The analysis aggregated discussion outputs without attributing statements to individual participants.

Post-Workshop Reflexive Thematic Analysis. Following the workshop, each moderating pair conducted reflexive thematic analysis of the empirical material, consisting of group notes and sticky-note outputs, using the pre-workshop coding as contextual material. The full author team then collaboratively analyzed all three discussion sets, iteratively identifying convergences across themes. This process revealed two cross-cutting topics as main results.

[1]https://visualanalyticshealthcare.github.io/homepage/2025/proceedings.html
[2]https://observablehq.com/@gabmorg/amia-2026-submission-coding-and-theme-breakdown

Results

Multi-stakeholder co-design. Participants across all three groups voiced a recurring tension between the narrowness of current design approaches and the breadth of stakeholders that healthcare visualization tools must serve in practice. Some framed this as a question of access and audience, highlighting the need to account for different learning modalities, varying educational backgrounds, and the distinct pressures of decision and emotional fatigue faced by different user types. Others made the structural problem explicit: the patient-doctor dyad is "only the tip of the iceberg," with relevant stakeholders spanning patients, families, support groups, nurses, nutritionists, laboratory staff, insurers, and health system administrators. Participants further observed that workflows vary substantially by clinical domain, with little generalization across applications: dementia care requires years of continuous collaborative decision-making, oncology may be time-sensitive and emergent, and diabetes involves long-term coordination across many roles. These observations expose a methodological shortcoming: current design study methodology [5] falls short at multi-stakeholder requirement analysis, and evaluation frameworks have not kept pace with the diversity of user groups that deployed tools must serve. Adapting implementation science frameworks, such as NASSS [6], to the multi-stakeholder complexity of VAHC contexts represents a concrete direction for medical informatics research, as does investment in evaluation methodology capable of measuring tool effectiveness across heterogeneous user groups.

Trust in human-centered AI workflows. Participants across all three groups identified trust calibration as a central unresolved challenge in deploying HCAI in clinical contexts, and named VA as a potential mediator of accessibility across user types. Several noted that miscalibrated trust is not solely a patient literacy problem; even highly educated AI researchers might trust LLMs when they should not. Participants further observed that current explainable AI research is oriented toward model development and validation, whereas clinical practice relies on models that have already been peer-reviewed, creating a mismatch between what the research community produces and what clinicians expect and need for care. Participants also named AI-driven de-skilling as a concrete risk, noting that workflows designed to automate rather than augment expert reasoning erode the tacit clinical knowledge that human-in-the-loop systems depend on. The challenge, then, becomes designing VA tools that promote appropriately calibrated trust. Proposed directions included gold-standard, empirically validated techniques for communicating risk and uncertainty to heterogeneous audiences, better grounding of AI outputs in verifiable clinical evidence, and decomposing evaluation tasks to atomic, measurable benchmarks. Together, these reframe trust calibration as a testable deployment outcome, tied closely to model confidence and performance.

Conclusion

These findings position the VAHC community to shape the design, evaluation, and governance of LHS. Realizing this will require coordinated investment in multi-stakeholder design methods, trust-aware visualization frameworks, and standards for communicating uncertainty. Ongoing analysis of the workshop findings will further inform our proposed research directions toward these goals.

References

1. Jesus J. Caban and David Gotz. Visual analytics in healthcare - opportunities and research challenges. *J. Am. Medical Informatics Assoc.*, 22(2):260–262, 2015, doi:10.1093/JAMIA/OCV006.
2. Alessio Arleo, Annie T Chen, David Gotz, Swaminathan Kandaswamy, and Jürgen Bernard. Reflections on interactive visualization of electronic health records: past, present, future. *J. Am. Medical Informatics Assoc.*, 31(11):2423–2428, 10 2024, doi:10.1093/jamia/ocae249.
3. Charles P. Friedman, Adam K. Wong, and David Blumenthal. Achieving a nationwide learning health system. *Science Translational Medicine*, 2(57):57cm29, 2010, doi:10.1126/scitranslmed.3001456.
4. Yang Chen, Steve Harris, Yvonne Rogers, Tariq Ahmad, and Folkert W. Asselbergs. Nudging within learning health systems: Next generation decision support to improve cardiovascular care. *European Heart Journal*, 43(13):1296–1306, April 2022, doi:10.1093/eurheartj/ehac030.
5. Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012, doi:10.1109/TVCG.2012.213.
6. T. Greenhalgh and S. Abimbola. The NASSS framework – a synthesis of multiple theories of technology implementation. *Studies in Health Technology and Informatics*, 263:193–204, July 2019, doi:10.3233/SHTI190123.