Designing Conversational Agents for Supporting Clinicians' Exploration of Care Quality Data

HAMZAH ZIADEH

As clinicians struggle to improve their care [11, 18, 21], decision support systems such as conversational assistants (CAs) can support clinicians in exploring care quality data to identify improvements [1, 15]. CAs can allow for specifying actions (e.g. creating a line plot) in natural language to simplify interactions compared to traditional interfaces which required clinicians to learn how to use an interface [7, 12, 22]. While expert users (e.g. clinicians, data scientists, etc.) reported CAs as easier and faster to use compared to dashboards [1], search engines [3], and databases [7, 12, 22], only few studies investigated designing CAs for clinicians when exploring care quality data [1, 2]. However, these studies typically explored clinicians interactions with CAs in limited lab settings and only evaluated task performance (e.g. time for completing task) [1, 2].

When exploring care quality dashboards, clinicians often feel overwhelmed with the number of indicators and lacked the statistical knowledge to explore data [6, 18, 21]. While CAs can suggest actions for data exploration, clinicians need to understand the provenance of insights produced from the data [10, 19] (i.e. the data sources, aggregation methods, actions taken, reasoning, etc. [5]). To communicate provenance, some CAs employed step-wise validation in which users stated a goal (e.g. creating a prediction model) and validated CAs' data exploration actions towards that goal [7, 16]. This required CAs to state every action and ask for feedback from users to allow for changes [7, 16]. While studies hypothesised that step-wise validation can increase users' sense of agency (feeling of control over the data exploration) and CAs' explainability (ability to explain the findings behind the suggested insight), it can also increase the workload on users [7, 14, 19]. Moreover, many clinicians lack the knowledge to validate data exploration actions or gaining bias towards suggested actions [4, 8, 19].

To investigate supporting clinicians during data exploration, previous studies [4, 6, 21] used goal setting theory which describes the process and requirements for users to create goals from exploring data [13]. However, recent reviews hypothesised that goal setting theory does not align with needs of clinicians [4]. For example, goal setting theory aims to describe how an individual can explore self-tracked data to set personal goals and motivate actions towards change [13]. On the other hand, clinicians explore data to facilitate group discussions with colleagues for goal setting and brainstorm organisational changes to provide better care [6, 17, 21]. Typically, clinicians have awareness of their shortcomings and already feel motivated to improve regardless of goal setting [6, 9, 21].

In my future work I aim to explore using alternative frameworks for analysing clinicians' data exploration to design CAs. For example, CP-FIT [4] describes the cyclical process of gathering, exploring, and disseminating care quality data in hospitals. Additionally, the knowledge generation loop model defines concrete phases and problems involved in exploring complex data sets such as those found in care quality registries [20]. By further understanding the tasks and motivations of data exploration, I aim to design, evaluate, and iterate on a CA that can support clinicians.

REFERENCES

[1] Taqdir Ali, Jamil Hussain, Muhammad Bilal Amin, Musarrat Hussain, Usman Akhtar, Wajahat Ali Khan, Sungyoung Lee, Byeong Ho Kang, Maqbool Hussain, Muhammad Afzal, Hyeong Won Yu, Ubaid Ur Rehman, Ho-Seong Han, June Young Choi, and Arif Jamshed. 2020. The Intelligent Medical Platform: A Novel Dialogue-Based Platform for Health-Care Services. *Computer* 53, 2 (Feb. 2020), 35–45. https://doi.org/10.1109/mc.2019.2924393

Author's address: Hamzah Ziadeh, hazi@create.aau.dk.

[2] Vidisha Bhatt, Juan Li, and Bikesh Maharjan. 2021. DocPal: A Voice-based EHR Assistant for Health Practitioners. In *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*. IEEE. https://doi.org/10.1109/healthcom49281.2021.9399013

[3] Timothy W Bickmore, Dina Utami, Robin Matsuyama, and Michael K Paasche-Orlow. 2016. Improving Access to Online Health Information With Conversational Agents: A Randomized Controlled Experiment. *Journal of Medical Internet Research* 18, 1 (Jan. 2016), e1. https://doi.org/10.2196/jmir.5239

[4] Benjamin Brown, Wouter T Gude, Thomas Blakeman, Sabine N van der Veer, Noah Ivers, Jill J Francis, Fabiana Lorencatto, Justin Presseau, Niels Peek, and Gavin Daker-White. 2019. Clinical performance feedback intervention theory (CP-FIT): a new theory for designing, implementing, and evaluating feedback in health care based on a systematic review and meta-synthesis of qualitative research. *Implementation Science* 14, 1 (2019), 1–25.

[5] Michael Correll. 2019. Ethical Dimensions of Visualization Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3290605.3300418

[6] Laura Desveaux, Noah Michael Ivers, Kim Devotta, Noor Ramji, Karen Weyman, and Tara Kiran. 2021. Unpacking the intention to action gap: a qualitative study understanding how physicians engage with audit and feedback. *Implementation Science* 16, 1 (Feb. 2021). https://doi.org/10.1186/s13012-021-01088-1

[7] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S. Bernstein. 2018. Iris. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3173574.3174047

[8] Dipesh P Gopal, Ula Chetty, Patrick O'Donnell, Camille Gajria, and Jodie Blackadder-Weinstein. 2021. Implicit bias in healthcare: clinical practice, research and decision making. *Future Healthcare Journal* 8, 1 (March 2021), 40–48. https://doi.org/10.7861/fhj.2020-0233

[9] Wouter T Gude, Benjamin Brown, Sabine N van der Veer, Heather L Colquhoun, Noah M Ivers, Jamie C Brehaut, Zach Landis-Lewis, Christopher J Armitage, Nicolette F de Keizer, and Niels Peek. 2019. Clinical performance comparators in audit and feedback: a review of theory and evidence. *Implementation Science* 14, 1 (2019), 1–14.

[10] Richard Heeks. 2006. Health information systems: Failure, success and improvisation. *International journal of medical informatics* 75 (03 2006), 125–37. https://doi.org/10.1016/j.ijmedinf.2005.07.024

[11] Peter U Heuschmann, Marcel K Biegler, Otto Busse, Susanne Elsner, Armin Grau, Uwe Hasenbein, Peter Hermanek, Rudolf WC Janzen, Peter L Kolominsky-Rabas, Klaus Kraywinkel, et al. 2006. Development and implementation of evidence-based indicators for measuring quality of acute stroke care: the Quality Indicator Board of the German Stroke Registers Study Group (ADSR). *Stroke* 37, 10 (2006), 2573–2551.

[12] Anjali Khurana, Parsa Alamzadeh, and Parmit K. Chilana. 2021. ChatrEx: Designing Explainable Chatbot Interfaces for Enhancing Usefulness, Transparency, and Trust. In *2021 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE. https://doi.org/10.1109/vl/hcc51201.2021.9576440

[13] A Locke and G Latham. 1989. *A theory of goal setting and task performance*. Prentice-Hall, London, England.

[14] Gonzalo Gabriel Méndez, Uta Hinrichs, and Miguel A. Nacenta. 2017. Bottom-up vs. Top-down. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3025453.3025942

[15] Tom Nadarzynski, Oliver Miles, Aimee Cowie, and Damien Ridge. 2019. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *DIGITAL HEALTH* 5 (Jan. 2019), 205520761987180. https://doi.org/10.1177/2055207619871808

[16] Elnaz Nouri, Robert Sim, Adam Fourney, and Ryen W. White. 2020. Step-wise Recommendation for Complex Task Support. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM. https://doi.org/10.1145/3343413.3377964

[17] Sofie Ordies, Gwendolyne Peeters, Anouk Lesenne, Patrick Wouters, Ludovic Ernon, Kim Bekelaar, and Dieter Mesotten. 2021. Interaction between stroke severity and quality indicators of acute stroke care: a single-center retrospective analysis. *Acta Neurologica Belgica* (2021), 1–8.

[18] Matthew Quigley, Sophia Zoungas, Edward Zimbudzi, Natalie Wischer, Sofianos Andrikopoulos, and Sally E. Green. 2022. Making the most of audit and feedback to improve diabetes care: a qualitative study of the perspectives of Australian Diabetes Centres. *BMC Health Services Research* 22, 1 (Feb. 2022). https://doi.org/10.1186/s12913-022-07652-9

[19] Antoine Richard, Brice Mayag, François Talbot, Alexis Tsoukias, and Yves Meinard. 2020. What does it mean to provide decision support to a responsible and competent expert? *EURO Journal on Decision Processes* 8, 3-4 (Nov. 2020), 205–236. https://doi.org/10.1007/s40070-020-00116-7

[20] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey Ellis, and Daniel A Keim. 2015. The role of uncertainty, awareness, and trust in visual analytics. *IEEE transactions on visualization and computer graphics* 22, 1 (2015), 240–249.

[21] Vibeke Sparring, Emma Granström, Magna Andreen Sachs, Mats Brommels, and Monica E Nyström. 2018. One size fits none–a qualitative study investigating nine national quality registries' conditions for use in quality improvement, research and interaction with patients. *BMC health services research* 18, 1 (2018), 802.

[22] Arjun Srinivasan and Vidya Setlur. 2021. Snowy: Recommending Utterances for Conversational Visual Analysis. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. ACM. https://doi.org/10.1145/3472749.3474792