Publisher: IISI - International Institute for Socio-Informatics
ISSN 1861-4280

international reports on socio-informatics

volume 19 issue 2
2022

Exploring Human-Centered AI in Healthcare: A Workshop Report

Guest Editors:

Nazmun Nisat Ontika
Sheree May Saßmannshausen
Hussain Abid Syed
Aparecido Fabiano Pinatti de Carvalho

Table of contents

The 'international reports on socio-informatics' are an online report series of the International Institute for Socio-Informatics, Bonn, Germany. They aim to contribute to current research discourses in the fields of 'Human-Computer-Interaction' and 'Computers and Society'. The 'international reports on socio-informatics' appear at least two times per year and are exclusively published on the website of the IISI.

Impressum

IISI - International Institute for Socio-Informatics
Stiftsgasse 25
53111 Bonn
Germany

fon: +49 228 6910-43
mail: iisi@iisi.de
web: http://www.iisi.de

Exploring Human-Centered AI in Healthcare: A Workshop Report

Nazmun Nisat Ontika, Sheree May Saßmannshausen, Hussain Abid Syed, Aparecido Fabiano Pinatti de Carvalho
Institute for Information Systems, University of Siegen, Germany
{nazmun.ontika, sheree.sassmannshausen, hussain.syed, fabiano.pinatti}@uni-siegen.de

Abstract. As a technique of improving the quality of life, AI has the potential to take a significant part in healthcare worldwide. However, in order to facilitate the widespread use of AI systems, we must first better comprehend the influence of AI on the healthcare sector. To create an acceptable intelligent system for healthcare, a comprehensive evaluation of ethically driven design, technology that effectively addresses human intellect, and human aspects of design is required. Our two-day workshop at the European Conference on CSCW in 2022 focused on Human-centered AI in the healthcare domain. In the workshop, we brought together researchers and practitioners in health informatics to accelerate conversations about developing usable and efficient intelligent systems that are more understandable and reliable for users.

1 Introduction

AI (Artificial Intelligence) is constantly transforming the human world's social and economical spheres. AI has advanced astonishingly in the previous decades, bringing a revolution in practically every key aspect of existence. Perhaps the most significant advantage of AI is that it frees up people to perform more important, innovative, and inventive work by doing a lot of the monotonous and

time-consuming activities in many sectors. And this is possible because of the great potential of AI, which is successfully combining human ingenuity with technological efficiency.

AI developments in all sectors are expanding by leaps and bounds. Additionally, AI is in great demand in healthcare right now since it is altering the industry, and AI adoption is transforming into a norm in several medical sectors. AI has been gradually developed and implemented in practically every field of medicine, from primary care to rare diseases, emergency medicine, biomedical research, and public health (Lekadir et al., 2022).

While medical institutions are receptive to implementing AI technologies, adoption is currently confined to certain departments and teams. Medical organizations and practitioners would have a difficult time dealing with AI if it does not integrate effortlessly into their present infrastructure, or even worse if it adds more complications. As we argued before, any new technology can be difficult to develop, and even more difficult to gain trust when having a strong infrastructure such as healthcare, where physicians need to make immediate choices with foreseeably many further implications (Ontika et al., 2022). A European Commission study report supports this claim by saying that the deficit of trust in AI-driven decision support systems is also impeding wider adoption, and concerns related to the integration of new technology into current practices are among the primary obstacles noted by relevant stakeholders in the EU Member States (PwC, 2021). Besides, numerous endeavours to design usable systems for physicians fail due to insufficient task analysis, in which essential user needs are either not discovered or their importance is undervalued (Preim & Hagen, 2011). Other key obstacles to greater AI adoption in healthcare are the absence of a human-centered approach while developing the systems, the complexity and unreliability of the final applications, the failure to include people in the development loop, and the lack of explainability for practitioners (Abdul et al., 2018).

However, one of the primary reasons why the previous two waves of AI failed was their inability to meet human requirements. In the third wave, AI system designers started to investigate many human characteristics, such as AI's ethics, interpretability, and comprehensibility to fulfill human demands and offer a pleasant user experience in a number of situations (Xu, 2019). Hence, to transform AI from just being technological to also humanistic, we need human-centered AI

(HAI or HCAI). Human-AI systems operating jointly, rather than alone, have a great potential for high effectiveness (Ahuja, 2019; Wilson & Daugherty, 2018). Moreover, humans and AI earnestly increase one another's complementary qualities using collaborative intelligence: the former's leadership, teamwork, creativity, and social skills, and the latter's speed, scalability, and quantitative capabilities (Wilson & Daugherty, 2019). HAI empowers individuals to perceive, think, create, and act in novel ways by integrating powerful user experiences with embedded AI approaches to support systems that users demand (Li, 2018; Robert et al., 2020).

Effective HAI solutions need a thorough examination of the ethically oriented design, technology that properly represents human intelligence, and human factors design (Xu, 2019). Furthermore, HAI systems should amplify human capabilities enabling individuals in extraordinary ways while maintaining human control (Shneiderman, 2020). In the context of healthcare, HAI design should research human variables and uncover medical acceptability hurdles to promote a transformational human-AI collaborative relationship centered on trustworthy AI. Incorporating human-centered and user-centered methods across the AI development phase will allow for the creation of AI algorithms that help define the requirements and values of healthcare professionals, as well as the identification and mitigation of possible risks at a preliminary phase (PwC, 2021). With the transition to human-centered AI in healthcare, work concentrating on community health workers and other frontline healthcare professionals will be critical in driving the field ahead and assuring that AI treatments function fairly globally (Okolo, 2022).

Incorporating HAI into healthcare effectively is a significant venture with constraints that entail a multi-disciplinary approach combining specialists from HCI, AI, healthcare, psychology, and social sciences. Our workshop addressed important HAI concerns, enabling optimal human-machine integration by enhancing the trustworthiness between humans and technology. We discussed ways to assure that AI applications focus on the end-user, put humans in the loop, and emphasize human values in a responsible manner. We explored different prototyping and evaluation techniques; also, how we could integrate the context of use with real user needs and usage scenarios into task analysis methods; and how all these could help make new strategies to improve the overall user experience in the healthcare context.

To explore various approaches to HAI and to develop a strategy for future scientific investigations on healthcare solutions we finalized four research questions for our workshop.

- What are the existing human-centered approaches for designing an AI-based medical diagnosis?
- How and when are end-users integrated into the development process of AI systems?
- How is it possible to make AI decisions comprehensible, fair and transparent to the end-users?
- What is the role of visualization in XAI and/or healthcare systems?

2 Workshop Course and Results

This workshop attracted more than 20 researchers from different disciplines but all of them had experience working in the field of health informatics. Five researchers couldn't join because of health or personal issues. Moreover, due to the corona situation, ten participants were on-site in Coimbra, Portugal and five persons were online and participated via Zoom. In this workshop, we discussed seven contributions from the domain of healthcare and artificial intelligence in the context of CSCW and HCI. Five of the contributions were position papers from researchers and two contributions were real use cases from the industry. In this report, we bring these seven contributions of the workshop and additionally a commentary by Jina Huh-Yoo, one of the participants of the workshop, about our main discussion points and raised questions in the workshop.

- The first contribution was from Rouncefield, Procter, and Tolmie titled "Trust. Professional Vision, and Diagnostic Work". Rob Procter presented some interesting empirical materials from their ongoing research about the everyday work in the Pathology Lab, as well as some design issues associated with developing AI systems intended to support 'trusted' processes of detection and diagnosis. He emphasized certain pathologists' actions that were rooted in a set of 'professional vision' and 'professional trust' for how to carry out the everyday diagnostic practice. He also stated that while thinking about how to build trust, it could be beneficial to examine concerns of 'professional trust'.

- In the second contribution, Christiane Grünloh presented their paper about “Balancing data-hungriness of AI and the workload of manual data collection” authored by Christiane Grünloh, Eline te Braake, Marian Hurmuz, Stephanie Jansen-Kosterink. She discussed their qualitative study with patients and healthcare professionals focusing on data collection of patients. She spoke on patients' attitudes, expectations, and experiences with healthcare data gathering. She remarked that patients' manual data collecting produces a response burden that must be balanced against their disease burden.
- “Towards Human-Centered AI: Learning from Current Practices in Radiology” by Nazmun Nisat Ontika, Sheree May Saßmannshausen, Hussain Abid Syed, Aparecido Fabiano Pinatti de Carvalho, Volkmar Pipek was the third contribution. Sheree May Saßmannshausen and Nazmun Nisat Ontika presented their first insights from an empirical study that explored current practices of radiologists in diagnosing prostate cancer. They addressed the design and decision gaps in the present process, and the need for human-centered AI in designing an explainable and trustworthy solution.
- The fourth contribution titled “Using Signals to Support Trust Building in Clinical Human-AI Collaboration” was from Naja Kathrine Kollerup, Mikael B. Skov, Niels van Berkel. Naja Kathrine Kollerup described the potential use of Relational Signalling Theory for designing trust-building signals in Human-AI interaction which can support the development of human-centered AI in healthcare. She noted that a system should impart trust in its users when acceptable, but it should also be able to alert when its suggestions are less reliable.
- Federico Cabitza and Andrea Campagner gave us the fifth contribution named “Human-AI collaboration protocols to assess real-world options of medical”. Andrea Campagner shared the result of their study that investigated the influence that AI-based decision aids exert on human decision-makers in the medical domain through the concept of human artificial intelligence collaboration protocol (HAI-CP). Their user studies were in favour of a certain interaction protocol where the use case was human-first in decision-making rather than putting AI first.

- Astrid Chow presented a use case from her industrial point of view called “A Service Design Approach for AI-Supported Clinical Tools: Collaborating with Interdisciplinary Care Teams & Patients to Provide and Leverage SDOH Data” as the sixth contribution. She discussed how a "service design blueprint" might assist in expediting the collection of SDOH (Social and Behavioral Determinants of Health) data as patients traverse the complex process of getting health care.
- Our last contribution was another practical use case presented by Nils Blaumer titled “PAIRADS: Interaction of humans and technology rethought”. He shared the companies’ insights about the interaction between humans and technology from their ongoing project, where they are building a demonstrator for radiologists to detect and diagnose prostate cancer using artificial intelligence for everyday radiology.

Prior to the workshop, the position papers were acquired and shared with everyone. Aside from the position papers, the schedule and workshop structure were provided to the participants in advance of the session. Considering the pandemic situation, we offered our workshop in a hybrid manner split into two days for three hours per day. On the first day, we started with a short introduction round followed by a presentation of our workshop plan highlighting the research questions. Three research contributions were presented in detail with an initial discussion about the topic presented after each presentation. All participants entered their questions and remarks on the shared Miro board, which were discussed in depth as a collaborative brainstorming session. We also tried to link different concerns focusing on the research questions. The first day of the workshop was completed with the reporting on the first day's discussion draft and the second day's plan. On the second day, we began with a short overview of the previous day's events and told the participants about the day's agenda. Two research articles and two use cases from the industry were presented. We followed the pattern of collecting ideas and questions on the Miro board, having the initial discussion after each presentation, and having a collaborative brainstorming session where we discussed the contributions presented on the second day, and also tried to bring back the conversation from the day before and try to combine our thoughts towards answering the questions we mentioned before.

Keeping our four research questions in mind, we created a mindmap to sort all the discussion points from both days according to their similarity and reference to the individual research questions. Our final mindmap can be seen in figure 1.

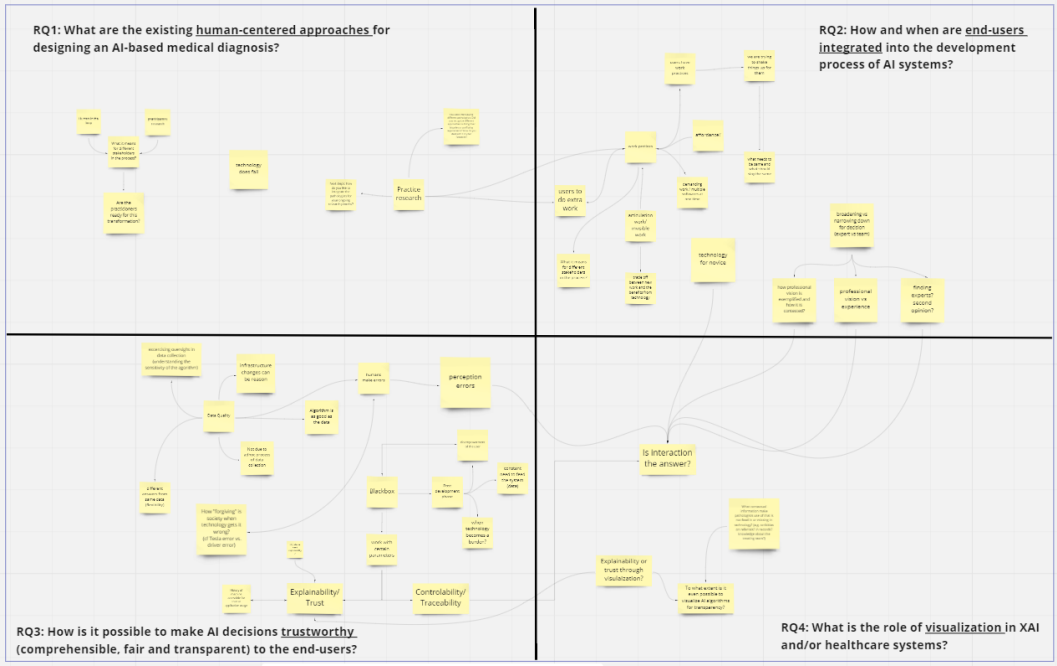


Figure 1: Mind Map of the Brainstorming about four Research Questions (created with Miro)

The connection between the discussion points shows that the four research questions strongly relate to each other and are interconnected. According to the first research question, we discussed different existing human-centered approaches such as participatory research, practice research, human-in-the-loop, etc., for designing an AI-based medical diagnosis while pointing out different stakeholders of that system and considering their perspectives. The integration of end-users into the development of AI systems was addressed by the second research question. Considering the high workload and stress in the medical domain we talked about the trade-off between the invisible work of the stakeholders for adapting to new systems and the measurable benefits of the systems for the stakeholders. We also argued on elaborating or narrowing down the explanation of the decision data generated by AI according to the expertise level of the users (e.g., 'technology for the novice'). As a part of the third research question, we talked about several attributes of the data such as explainability, understandability, acceptability, quality, sensitivity, accessibility, privacy, security, flexibility, accountability,

controllability, trustworthiness, and so on. We also raised questions about how forgiving is society when technology gets the detection/ diagnosis wrong. These thoughts lead directly to the fourth research question about the role of visualization in XAI. We debated the possibility of having better explainability or trust through visualization, but we also thought about to what extent is it even possible to visualize AI algorithms.

In the end, we reported briefly on the various conversations and conclusions in our plenary session, and we finished the workshop by making arrangements for future collaboration.

It was a good opportunity to discuss with various researchers and practitioners from the industry the meaning and approaches of human-centered AI in healthcare through the workshop. Since we have discussed multiple projects, our discussions were enriched with different use cases while creating some ideas for our ongoing and future projects. Moreover, the open questions that were raised from our discussion will also influence the involved researchers to instigate shortening the gap in Medical AI and human-centered design.

3 Acknowledgements

We would like to thank the workshop co-organizers for sharing the research initiative toward human-centered AI in healthcare with us. Furthermore, we would like to thank the participants of our workshop for sharing their current research with us and contributing to such an insightful discussion. The workshop organizers from the University of Siegen would like to acknowledge the financial support from the Bundesministerium für Bildung und Forschung - BMBF through the PAIRADS project (funding code: 16SV8651; https://pairads.ai).

4 References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. Conference on Human Factors in Computing Systems - Proceedings, 2018-April. https://doi.org/10.1145/3173574.3174156.

Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. PeerJ, 2019(10). https://doi.org/10.7717/peerj.7702.

Lekadir, K., Quaglio, G., Garmendia, A. T., & Gallin, C. (2022). Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. EPRS (European Parliamentary Research Service).

Li, F.-F. (2018). How to make A.I. that's good for people. The New York Times (March 7, 2018). https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html.

Okolo, C. T. (2022). Optimizing human-centered AI for healthcare in the Global South. Patterns, 100421. https://doi.org/10.1016/j.patter.2021.100421.

Ontika, N., Syed, H., Saßmannshausen, S., Hr Harper, R., Chen, Y., Park, S., Grisot, M., Chow, A., Blaumer, N., Fabiano, A., De Carvalho, P., & Pipek, V. (2022). Exploring Human-Centered AI in Healthcare: Diagnosis, Explainability, and Trust. https://dl.eusset.eu/bitstream/20.500.12015/4409/1/ws06.pdf.

Preim, B., & Hagen, H. (2011). HCI in Medical Visualization. In Scientific Visualization: Interactions, Features.

PwC. (2021). Study on eHealth, Interoperability of Health Data and Artificial Intelligence for Health and Care in the European Union. Lot 2: Artificial Intelligence for health and care in the EU. EUROPEAN COMMISSION.

Robert, L. P., Bansal, G., & Lütge, C. (2020). ICIS 2019 SIGHCI Workshop Panel Report: Human-Computer Interaction Challenges and Opportunities for Fair, Trustworthy, and Ethical Artificial Intelligence. AIS Transactions on Human-Computer Interaction, 12(2), 96-108.

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. AIS Transactions on Human-Computer Interaction, 109–124. https://doi.org/10.17705/1thci.00131.

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. In Harvard Business Review (Vol. 2018, Issue July-August).

Wilson, H. J., & Daugherty, P. R. (2019, April 4). How Humans and AI Are Working Together in 1,500 Companies. Harvard Business Review. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.

Xu, W. (2019). Toward human-centered AI: A Perspective from Human-Computer Interaction. Interactions, 26, 42 - 46.

A Commentary: Trust, Power, and Expectations Surrounding the AI in Healthcare

Jina Huh-Yoo
Drexel University, Philadelphia PA, USA
jinahuhyoo@drexel.edu

Abstract. As artificial intelligence (AI) becomes prevalent in various dimensions of healthcare technologies, researchers in health informatics, human-computer interaction, computer-supported collaborative work, and engineering, as well as practitioners of the healthcare field gathered to discuss AI's potential and challenges in healthcare at ECSCW 2022. In this commentary, I summarize our discussions and suggest questions to think about moving forward.

1 Commentary: Trust, Power, and Expectations of the Stakeholders

Jiang et al. (Jiang et al., 2017) illustrated a road map of clinical data generation to natural language processing data enrichment, to machine learning data analysis, to clinical decision making (Figure 2, left side in white color). As a CSCW community, our workshop participants discussed the various stakeholders involved in this road map of AI applications in healthcare—patients, practitioners, and data workers. Furthermore, we discussed the potential and challenges of the design strategies, implications, and practices to improve the different stakeholders' interaction with AI applications in healthcare, such as explainable AI, innovative data collection methods, and trust-building design strategies (Figure 2, right side in orange color).

AI applications in healthcare and their efficacy are largely driven by the quality of the data (e.g., electronic medical records (EMR), imaging data, genetic data, electrophysiological data, etc.) that train various machine learning mechanisms (de Hond et al., 2022). Data are generated through clinical activities, such as screening, diagnosis, and treatment. Practitioners also manually generate clinical notes, which can be processed through natural language processing and fed back into the dataset, which can again then be used to train the algorithms (Apell & Eriksson, 2021).

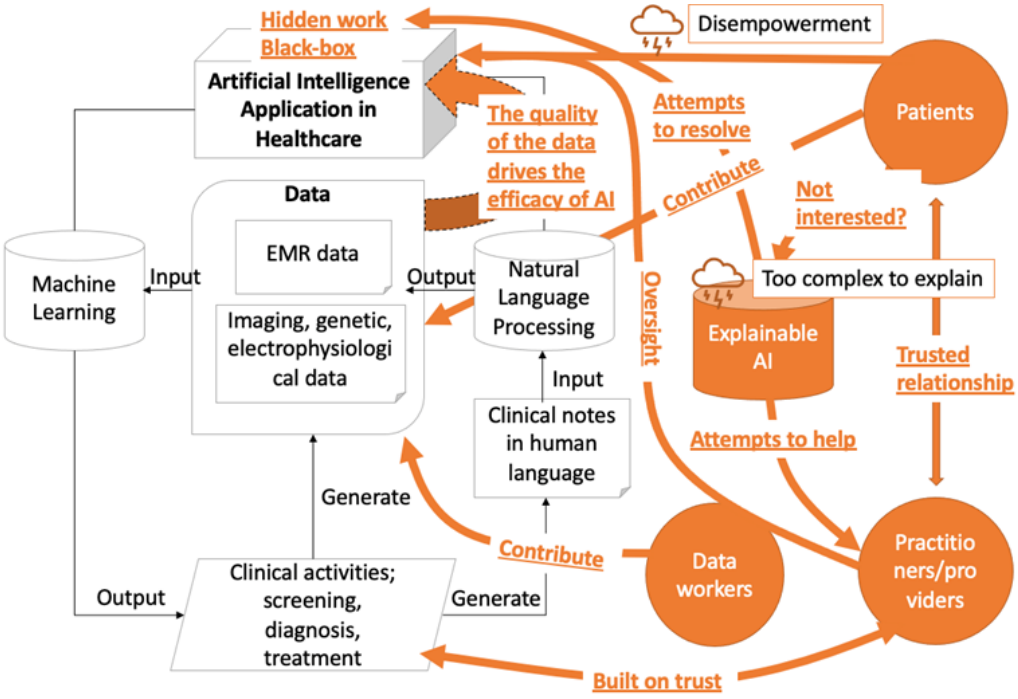


Figure 2: Left shows the road map of AI applications in healthcare that Jiang et al. (Jiang et al., 2017) have mapped out. On the right is how our workshop added discussions as a CSCW community, and what stakeholders are involved in enhancing users' interaction with the AI applications in healthcare.

From the perspective of the CSCW community, one of the things we discussed in the workshop is how various stakeholders who interface with AI applications in healthcare can be supported. A large area of interest in the community we discussed was Explainable AI (XAI) (Wells & Bednarz, 2021; Linardatos, et al., 2020; Vilone & Longo, 2020). The aspects of XAI that we discussed were mainly about its attempts to resolve the challenges around the hidden work and black box nature of AI supporting healthcare by making the process more transparent, traceable, trustworthy, and understandable (Nazar et al., 2021). XAI can support patients to set up expectations and potentially have control over the outcomes of the AI; for practitioners, it can help practitioners to give oversight over AI, for instance (Loh et al., 2022). While the area is still actively being developed, we saw the potential as well as challenges in making AI explainable. That is, it might be too complex to explain, or that the stakeholders might not be interested because there is too much information. An open question would be to understand how much information should be disclosed and with how much scaffolding depending on user profiles, characteristics, and interests in each context.

We discussed the analogy of Informed Consent processes in human subjects' research regulations over providing transparency to AI applications and providing stakeholders the power to change it. Informed consent rarely gives substantial power for people to decline to enroll in the study or make medical decisions (Nijhawan et al., 2013; Beauchamp, 2011). Similarly, data use agreements that consumers agree to when installing a new mobile app serve similar purposes—at that point of consenting to the data use, people have already decided to participate or sign up for the product or event. It becomes a formality at that point to be informed about the risks and benefits. Thus, without accurately understanding the risks, people might preliminarily consent to participate in the event.

Trust emerged multiple times as a critical concept in our workshop. Existing clinical practices have established trust. Or, at least, trust is how determines a large part of the clinical practices' efficacy (Graham, 2015). Patients trust healthcare practitioners to do good work in helping them get better. When a third-party 'thing' that has previously not been around, such as AI, intervenes in this tightly established trust relationship between the clinical practice and the patients, the consequences can be significant (Apell & Eriksson, 2021). Relatedly, literature discussed a similar term, 'otherware' (Hassenzahl et al. 2021; Laschke et al., 2020), which refers to the new paradigm of AI becoming an integral part of social interaction surrounding the technology. The question is how XAI can provide patients and practitioners the trust they need in validating its role in this already complex social space of healthcare.

A further complicating factor here is that we often do not know whether an AI is trustworthy because of its black-box nature. Thus, even though XAI succeeds to make stakeholders trust AI, it may result in an unethical practice of deceiving people to trust something they should not have. Regardless, people either overly trust AI or overly distrust AI regardless of the truth (Schmidt et al., 2020). We are stuck in this messy situation of first, needing to accurately assess and present the efficacy and potential biases and errors of an AI, which is hard, and second, being able to present it in ways that people would accurately perceive its trustworthiness. Trust is a hard-earned perception that people accumulate over long periods of time (Jacovi et al., 2021). If AI wants to gain trust, it will need to establish relationships over a long period of time, frequently showing its trustworthiness. This characteristic is a distinct one that we often overlook in evaluating the trustworthiness of an AI application- long-term use of the AI

application, its ability to continue to evolve with people, learning from mistakes, and improving its efficacy over time. Trust is not gained through a single encounter, but through multiple, past historical experiences (von Eschenbach, 2021).

Data workers can be anyone who is at any phase of the data collection, data manipulation, model training, or evaluation of the AI application (Muller et al., 2019). We discussed what kind of information would these workers be exposed to all day and what consequences it might have (Lee et al., 2015). If the algorithm only needs humans' help in those areas that mainly consist of things that may be traumatic, negative, or biased, the knowledge that the workers are accumulating and the career they are building are majorly impacted by the algorithms' ability to discern certain parts of the story.

We discussed only a small piece of the big picture that is surrounding the AI applications of healthcare through the three stakeholders: patients, practitioners, and data workers, and how they may be impacted through the concepts around trust, power, and expectations.

We should continue to think about how we might establish the role of AI differently depending on the context of the healthcare problems—whether it will be assistive, augmented, or autonomous. In what areas of healthcare can we trust AI to be autonomous without harm, and in what areas can we trust AI to be only assistive? How do we assess potential risks of harm? How do we assess its trade-off to benefits? How would these relationships and definitions change as AI evolves and people's perceptions change? What new kinds of data workers would emerge? How would power dynamics change consequently between AI and data workers, between patients and practitioners, and practitioners vs. AI applications?

2 References

Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 624–635. https://doi.org/10.1145/3442188.3445923.

Beauchamp T. L. (2011). Informed consent: its history, meaning, and present challenges. Cambridge quarterly of healthcare ethics: CQ: the international journal of healthcare ethics committees, 20(4), 515–523. https://doi.org/10.1017/S0963180111000259.

de Hond, A., Leeuwenberg, A. M., Hooft, L., Kant, I., Nijman, S., van Os, H., Aardoom, J. J., Debray, T., Schuit, E., van Smeden, M., Reitsma, J. B., Steyerberg, E. W., Chavannes, N. H., & Moons, K. (2022). Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ digital medicine, 5(1), 2. https://doi.org/10.1038/s41746-021-00549-7.

Graham, J. L., Shahani, L., Grimes, R. M., Hartman, C., & Giordano, T. P. (2015). The Influence of Trust in Physicians and Trust in the Healthcare System on Linkage, Retention, and Adherence to HIV Care. AIDS patient care and STDs, 29(12), 661–667. https://doi.org/10.1089/apc.2015.0156.

Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, U Rajendra Acharya, (2011–2022). Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade. Comput Methods Programs Biomed. (2022) 107161.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present, and future. Stroke and vascular neurology, 2(4), 230–243. https://doi.org/10.1136/svn-2017-000101.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy (Basel, Switzerland), 23(1), 18. https://doi.org/10.3390/e23010018.

M. Nazar, M. M. Alam, E. Yafi and M. M. Su'ud, (2021), "A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare with Artificial Intelligence Techniques," in IEEE Access, vol. 9, pp. 153316-153348, 2021, doi: 10.1109/ACCESS.2021.3127881.

Marc Hassenzahl, Jan Borchers, Susanne Boll, Astrid Rosenthal-von der Pütten, and Volker Wulf. 2020. Otherware: how to best interact with autonomous systems. interactions 28, 1 (January - February 2021), 54–57. https://doi.org/10.1145/3436942.

Matthias Laschke, Robin Neuhaus, Judith Dörrenbächer, Marc Hassenzahl, Volker Wulf, Astrid Rosenthal-von der Pütten, Jan Borchers, and Susanne Boll. (2020). Otherware needs Otherness: Understanding and Designing Artificial Counterparts. In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20). Association for Computing Machinery, New York, NY, USA, Article 131, 1–4. https://doi.org/10.1145/3419249.3420079.

Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. (2019). How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 126, 1–15. https://doi.org/10.1145/3290605.3300356.

Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. (2015). Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). Association for Computing Machinery, New York, NY, USA, 1603–1612. https://doi.org/10.1145/2702123.2702548.

Nijhawan, L. P., Janodia, M. D., Muddukrishna, B. S., Bhat, K. M., Bairy, K. L., Udupa, N., & Musmade, P. B. (2013). Informed consent: Issues and challenges. Journal of advanced pharmaceutical technology & research, 4(3), 134–140. https://doi.org/10.4103/2231-4040.116779.

Petra Apell & Henrik Eriksson (2021) Artificial intelligence (AI) healthcare technology innovations: the current state and challenges from a life science industry perspective, Technology Analysis & Strategic Management, DOI: 10.1080/09537325.2021.1971188. 1–15.

Philipp Schmidt, Felix Biessmann & Timm Teubner (2020) Transparency and trust in artificial intelligence systems, Journal of Decision Systems, 29:4, 260-278, DOI: 10.1080/12460125.2020.1819094.

Vilone, G., & Longo, L. (2020). Explainable Artificial Intelligence: A Systematic Review. arXiv. https://doi.org/https://arxiv.org/abs/2006.00093v4.

von Eschenbach, Warren J. (2021). Transparency and the Black Box Problem: Why We Do Not Trust AI. Philosophy and Technology 34 (4):1607-1622.

Wells, L., & Bednarz, T. (2021). Explainable AI and Reinforcement Learning-A Systematic Review of Current Approaches and Trends. Frontiers in artificial intelligence, 4, 550030. https://doi.org/10.3389/frai.2021.550030.

Trust, Professional Vision, and Diagnostic Work

Rob Procter[1], Peter Tolmie[2], Mark Rouncefield[3]
[1]Department of Computer Science, Warwick University and Alan Turing Institute for Data Science and AI
rob.procter@warwick.ac.uk
[2]Wirtschaftsinformatik und Neue Medien, Universität Siegen
Peter.Tolmie@uni-siegen.de
[3]Department of Computing, Lancaster University
m.rouncefield@lancaster.ac.uk

Abstract. In this paper, we consider some empirical materials from our ongoing research into forms of everyday detection and diagnosis work in healthcare settings, and how these relate to issues of trust, trust in people, technology, processes, and data.

1 Introduction

> "One of the basic conditions of any constitutive practice is a mutual commitment to rules of engagement in that practice – that is, all parties to the interaction must understand that they are engaged in the same practice, must be competent to perform the practice, must actually perform competently and assume this also of the others." (Watson 2009)

This paper considers some empirical materials from our ongoing research into forms of everyday detection and diagnosis work in healthcare settings, and how these relate to issues of trust, trust in people, technology, processes, and data. The setting for this research, and the collection of these materials, was several pathology labs where pathologists are involved in examining biopsies (i.e., small samples of tissue mounted on slides and then digitized) for various forms of cancer. The pathologists were interviewed about their working practices, the impact of the change from glass to digital imaging, and their expectations of how the introduction of AI tools would be able to assist them.

Our data points to some of the features of the everyday work of pathologists, as well as some design issues associated with developing AI systems intended to support 'trusted' processes of detection and diagnosis. There are evidently numerous issues of trust in the healthcare domain, and as technology and organizational culture change issues of trust have become more complex (Kipnis

1996). Luhmann (1990) argues that the different conceptualizations of trust generally fail to pay attention to the social process of trust production, and the ways in which trust is accomplished in social interaction.

We are interested in this paper in explicating some aspects of trust within medical domains; how this trust is achieved and supported in mundane work; the extent to which trust is supported by technology and the varied ways in which trust thereby impacts decision-making. The digitized images used in pathology are part of a trusted process, a division of labor, in which one set of actions initiates others. Knowing what the images represent provides for their use, and their 'trustability' involves knowing about the work that produced it, and what it means within the activity and the organization. What the images mean, what they refer to, and what they indicate, have to do with their place within the organizational setting since they represent and display, and make 'accountable' organizational activities.

The work of pathologists involves the employment of particular and complex perceptual skills to find what may seem small features or irregularities in the complex visual environment of a digitized image, and interpretative skills to classify them appropriately as being 'suspicious', 'cancerous', and thereby worthy of further diagnostic investigation. Whilst such work is clearly 'routine' for practitioners, as Giddens (1984) reminds us, such routine work requires a detailed analysis of its accomplishment:

> "It is a major error to suppose that these phenomena need no explanation, that they are simply repetitive forms of behavior carried out 'mindlessly'. On the contrary, as Goffman (together with ethnomethodology) has helped to demonstrate, the routinized character of most social activity is something that must be 'worked at' continually by those who sustain it in their day-to-day conduct." (Giddens 1984: 86).

The accomplishment we are especially interested in is the accomplishment of trust in the process of everyday work – not least if we are intending to design, build or evaluate AI diagnostic technologies intended to support the work of healthcare professionals such as pathologists.

2 Everyday Work in the Pathology Lab

As with many other forms of diagnostic work the practice of 'reading' and interpreting these digital images (like the original Pathology Lab practice of 'reading' glass slides through a microscope) calls for the exercise of a subtle,

learned combination of reasoning, knowledge, and skill. In previous work (Hartswood et al. 2000, 2002) we have considered such practices as constitutive of some form of 'professional vision' – "socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group" (Goodwin 1994: 606). As we have observed in other diagnostic settings, diagnostic work in the Pathology Lab involves, requires even, sets of tried and tested repertoires of 'manipulations' that are an integral part of the embodied practice of uncovering or realizing phenomena in the images – of revealing cancerous cells. These manipulations and the accompanying diagnosis are examples of 'midenic' reasoning, that is reasoning, assessing, evaluating, diagnosing, and making judgments, whilst engaged in the actual flow of activity "reasoning in the midst of reasoning about what we are doing" (Livingstone 2008: 9). We illustrate our argument with extracts from interviews with six pathologists. These lasted typically one hour and were recorded and transcribed.

In this first interview extract the pathologist is talking about the processes followed in 'looking' at an image and making a decision; the way they manipulate the image to help their decision-making process – "really focusing on getting the information I want, as you say, to either be able to quickly dismiss that area which had caught my eye initially or to is thinking OK, now this is something":

I. How do you organize looking at an image is there a particular way you do it?

*P1: So, for me personally I really focus on the low power first of all. So, first of all, to make sure that I'm happy we've got a complete image, and also it gives me a sense of where I might need to focus on so rather than going up into a medium or higher power in sort of the first field, I don't do that. I tend to scan the whole image on quite a low power first to get a feel for where I might need to go in and look at it on a higher power. And then I might sort of rescanning quite quickly on a higher power and just go into those areas where I want to focus on. So, for a much higher power. So, like I might scan the whole thing on the equivalent of like a * 2 and then sort of start to look around on a * 4 but do that quite quickly because I know where I want to go and then quickly go into sort of * 10 * 20. That sort of thing. To really focus on getting the information I want, as you say, to either be able to quickly dismiss that area which had caught my eye initially, or to be thinking OK, now this is something and you know these are the steps I need to take. This is a tumor or whatever.*

*P1: … there are some cases where it all looks a little bit different, and you just must take your time and probably scan on a higher power. Like maybe * 10 and then you know be going up and down up and down a bit more, but particularly with excision cases where you have got quite a bit of normal tissue as well as possibly a tumor, it is easier to sort of scan on low power at and then just go in and out where you need to. There is particularly like a small biopsy because you have got less material to make a diagnosis from, you are probably going to be spending longer per image than you would on a bigger case because you have got to get more information from that one image.*

At the same time, such activities are embedded in a set of clearly understood professional expectancies, what might be termed 'professional vision' that includes ideas about 'trust' and accountability concerning exactly how to go about everyday diagnostic work. This requires that features that are found or discovered

in the digitized image are put into an appropriate professional and organizational formulation – what a feature might be and how it might be treated. This is what Goodwin terms 'professional vision' where the activities of the individual are viewed relative to some particular and professional set of expectancies:

> 'The relevant unit for the analysis of the intersubjectivity at issue here is thus not these individuals as isolated entities but (. . .) a profession, a community of competent practitioners, most of whom have never met each other but nonetheless expect each other to be able to see and categorize the world in ways that are relevant to the work, tools, and artifacts that constitute their profession' (Goodwin 1994: 615).

For example, here is one pathologist talking about some of the ways in which the digitized images are manipulated in order to develop trust in what is being seen: and then a pathologist discusses his basic 'mindset' as he goes about the configuration of the equipment – here are professionals talking about how they need to set up the tools of their trade, in order to facilitate their best use and the kinds of question that occur as they do so – "you're not cruising ... you already have a plan":

I: OK, so what sort of manipulations would that involve?
P1: I mean, so there is the very basic manipulation, such as just moving the image around the screen so that you can look at different parts of the image. And obviously zooming in and out with magnification and then an added benefit of the digital is you have measuring tools, so you can very accurately measure, for example, the size of a tumor or the distance of a tumor to a surgical margin. And provides accurate information.
P3: … when you are looking at slide the slide or an image you are not cruising, kind of trying to find whatever you will come through. You will have, depending on what the type of biopsy or type specimen you have, you already have a plan you have a set of questions you are answering mentally while you are going through, and it is, this is something we try to explain to the computer scientists, we're still kind of in the process. I think they are getting the idea because trying to get you to know someone of the projects about artificial intelligence. So, from the clinical background, clinical training, and the experience when you look at the biopsy, for example, you know these are the questions, you know, the clinical history. You know these are the things I need to be looking for in each specimen. These are different sets of questions and before that even if you look for adequacy you look for representation and so on. But it kind of getting to the nutshell there are a set of questions you kind of look for their answers, while when you are looking at the slides these are related to the type of the biopsy the clinical history, and of course the presenting features on the slide itself. Because sometimes you look at the biopsy kind of things are unexpected. So, your mental kind of pathway kind of changes accordingly as well.

The diagnosis appears then as a social process that is achieved by a specific 'community of practice' and to which its members are accountable. Being a competent practitioner involves being able to accountably distinguish between what is 'normal' and what is 'abnormal' in a digital image or a microscope slide and understanding the range of manipulations and shared professional interactional practices that make what is 'normal' or 'abnormal' witnessable and accountable. "... if I am looking at a piece of lung tissue, I would expect to find

almost big empty spaces lined by the alveoli or the lung epithelial tissue ... If instead of that I am finding solid areas rather than empty spaces, if I am finding all these spaces filled with either cells or any other abnormal material, then I do know that it is abnormal." It is, as Garfinkel et al. point out, the 'intertwining of worldly objects and embodied practices' (1981, p.165) that produces recognizable and accountable diagnoses and decisions.

I: What kinds of things are you looking for, and does that depend on the suspected cancer?
P2: Yes, so, if I am looking at a piece of tissue, the first thing I would try and decide is what the tissue represents, what normal site, or what normal anatomy I can see in that issue. So, if as often is the case, you would find a bit of residual normal tissue, I will identify that as the pancreas or lung whatever maybe I will try identifying that and then look for pieces of tissue that do not fit into what I would call his normal morphology for that area. So, if I am looking at a piece of liver and I can see normal liver tissue, I will keep looking at the rest of the tissue in that biopsy. And as soon as I find the focus that looks different from that. That is when I start looking at it closer and determining if that is neoplastic or meaning cancer or something that is totally unrelated and just in inflammation or an abscess. Things like that.
P2: ... if I am looking at a piece of lung tissue, I would expect to find almost big empty spaces lined by the alveoli or the lung epithelial tissue. If you if instead of that I am findings solid areas rather than empty spaces if I am finding all these spaces filled with either cells or any other abnormal material, like 15? Then I do know that it is abnormal. Similarly in the colon when I am looking at a piece of intestine, I am trying to see if I can identify all the normal different layers of the colon. Now the moment I find a layer or group of cells within a particular layer that looks abnormal, I would then zoom in on their particular focus to find out what that is.

Diagnosis should then be regarded as a material, collaborative process involving technologies, expert skills to manipulate the technology to obtain the view that aids their decision-making process, and careful and sensitive collaborative engagement with the materials. Some diagnostic activity requires what might be regarded as rational everyday knowledge; some demands specific 'scientific' epistemic practices of measurement, representation, and calculation. As Heritage (1984) argues, this kind of analysis.

> "Vividly demonstrates that where sociological research encounters institutional domains in which values, rules, or maxims of conduct are overtly invoked, the identification of these latter will not provide an explanatory terminus for the investigation. Rather their identification will constitute the first step of a study directed at discovering how they are perceived exemplified, used, appealed to, and contested." (Heritage 1984)

In these interview extracts we discover "how they are perceived exemplified, used, appealed to and contested" as we hear a skilled practitioner talking about how 'professional vision' develops over time, how expertise is acquired and used, the importance of 'context' for the exercise of professional vision, and how this can impact on the kind of decisions that are made:

P1: ... so as we are going through our training, the trainees will describe all these features. And the more experience you get, you sort of cut to the chase more and you will not include all of that. I

will include things like that if I found it very difficult to come to a decision. If I think that there is potential. Not that I am wrong, but I am giving. I am giving an indication of how sure I am about something. So, for example, you know it might be difficult. I might have shown it to several colleagues. We might have been thinking OK, this could be positive. It might not be positive. These are the things we have considered, and this is the conclusion we have come to. But if it is a really, straightforward case, even if I have had some of the same thought processes if I have been very, it's been very easy and quick to sort of dismiss that and get to the crux of the matter. Then I would not put all of that in a report.

And the vital importance of 'context':

I: When you say 'out of context' can you just expand a little bit on what you mean by that?
P3: Yeah, I mean, for example, if you have someone who had some chemotherapy, for example, and you look at their bone marrow so the bone marrow will show changes which if you don't know or you don't appreciate the chemotherapy change, you could interpret that as another disease, for example, as the patient has some sort of a bone marrow disease which can have a lot of different consequences, so this is very, kind of, you know, just one example. .., I'll tell you what for example, EBV infection for example, which is mononuclear, infectious Mono nucleosis and it can cause lymph node enlargement. And sometimes the clinicians do not know that the patient had this infection, so you get the lymph node and if you look at this lymph node biopsy, see it looks like high-grade lymphoma. And if you go and tell the clinician this is high-grade lymphoma this patient will get extensive treatment and could kill the patient.

Being part of a professional 'community of practice' also requires a willingness, an acceptance, and a set of collaborative organizational procedures to have opinions and decisions challenged and rejected or confirmed – through the process of 'second opinions'. It requires an acknowledgment that whilst practitioners effectively 'trust' their skills, there are limits to individual skill, particularly with 'difficult cases'.

I: OK, what would make for a difficult case? ... perhaps you could give an example of a case that is difficult to diagnose.
P5: I mean there was an endometrium that I did over the weekend where it just looked like the stroma or the bit that's not glands looked expanded in some way, and it just did not look like a normal menstrual endometrium if you like. So, I sent that for a second opinion. And often with breast stuff, it is not the obvious cancers they are fine, and the obvious benign aren't the problem in pathology in general, it is those in-between bits when is it a typical or a bit off really? Or is it not? You know it is those areas that tend to be very difficult. You know, a barn door cancer, fine, you know, fibroadenoma, like benign conditions. Fine, it is those in-betweens that are an issue.

And so, there are 'second opinions':

I: OK, so how do you go about inviting someone to have a look at that case?
P3: Yes, depending on which subspecialty, so for example, if I am looking at lung, I will ask X, I'll send him an email with the case number X, can you please have a look at this case? I can tell him what I am thinking, or I can tell him what I'm worried about and in this scenario, I have annotated the digital slides as well, and sometimes if he's online I can invite him and go through the case together. So, we can have some sort of a discussion, live discussion, and, for example, in the gastrointestinal pathology case I can ask Y or Z, or W for their opinion.

This is where we anticipate a concern with 'trust' can be useful, in a way that 'professional vision' perhaps might not be, in unpacking some of the important features and characteristics of everyday rules, values, and conduct. 'Professional vision' provides a gloss for shared activities and practices that perhaps need to be

more thoroughly understood. Notably, the term tells us little about trust and the role of trust in everyday working practice. As the empirical data suggests, working collaborations with colleagues and within organizational structures obviously presuppose some form or form of trust. It is not only individuals that must be trusted but also, and inevitably, organizational processes and procedures – like 'second opinions' – as well as different tools and data that permeate and mediate relationships and enable (or disable) trust. 'Trust' is clearly a difficult topic (though frequently treated unproblematically as a mere 'resource') and has produced various philosophical and social science concepts and theories (Luhman 2018, Gambetta 2000) identifying the grounds of trust in an individual's reputation, performance, and appearance with a range of different relational and cultural dimensions, none of which are necessarily adequately encompassed in the notion of 'professional vision'.

Two ideas emerge from our current empirical work and will provide a focus for future explorations and analysis. Firstly, that trust is not merely concerned with individuals, with processes or technologies but also 'trusted data'. Trusted data can be an important factor in fostering trust between workers. The interest lies in the process whereby data, like technology or our colleagues, becomes 'trusted data'. Secondly, the 'temporality' of data should be acknowledged; understandings of trust, who can be trusted, and what constitutes a 'trustable' procedure or 'trustable' data, change over time and is especially relevant in the continued maintenance of trust. Future work will explore these ideas further using some of Garfinkel's ideas about trust (Watson 2009), whereby trust becomes 'a phenomenon of ordinary membership', considering the expectancies of trust that precede interaction and those aspects of trust that emerge as part and parcel of the production and accomplishment of everyday professional work:

> "One of the basic conditions of any constitutive practice is a mutual commitment to rules of engagement in that practice – that is, all parties to the interaction must understand that they are engaged in the same practice, must be competent to perform the practice, must actually perform competently and assume this also of the others." (Watson 2009)

3 Technology, Diagnosis, and Trust

Our research on diagnostic work and ideas about trust is formulated and presented in the belief that a detailed understanding of everyday trust and diagnostic practices should be a precursor to the design and redesign of AI detection and diagnosis technologies. We are concerned with understanding the impact such

tools might have on the situated, collaborative practical activity of detection and diagnosis that we have observed. How might such tools mesh with current diagnostic practices? What future practices should be developed? In terms of providing diagnostic support, what design features of the technology might cause people to 'trust' or 'mistrust' it? Our pathologists had some clear ideas about the value of AI and technology in general and the kind of things, in terms of both organization and decision-making, that might be important to trust the technology.

P3: I think AI should, it would have several uses. One is, I think workload management and distribution. I think some of the problems in kind of pathology reporting would be solved by having a better kind of workflow ... AI would be kind of, you know, finding solutions really, you know, like hospitals, struggling with some workload that there will be automatic kind of reallocation of cases without delays. Opinions would be easier to get. So more streamlining and efficiency ... And in view of the shortages of consultant pathologists, so AI would be helpful in screening kind of filtering out all the normal you know which maybe do not need to be looked at and flagging up cases ... The other thing is AI can be used as tools applications on your diagnostic screen. For example, can pick up some tumors will need, for example, counting of mitosis for grading or some measurements, so AI can help you too in terms of it will help with reproducibility because it is a machine, so it will be quite consistent.

P5: ... it depends how far we go, doesn't it, with the AI? And I suppose it will be over time as well, won't it? So it might be that like say we get a load of bowel biopsies, you know, large bowel biopsies. And a lot of them will be normal. So maybe if the system could say you know it thinks these are normal, and then you just look at them quickly, just confirm so these are confirmatory, and I think we will probably have to go through that stage before we allow an electronic or an AI system to say it is benign and just sign out the report itself. I think that might take a while to be approved and stuff, and it will probably take an increasing level of confidence, will not it and they will have to be demonstrations that you know there is not a tiny bit of carcinoma, in it or whatever you know, and so that sort of thing would be useful.

Our research suggests that when considering how trust might be achieved it may be useful to consider some issues of 'professional trust'. As our pathologists suggest they are trusted to act in a professional way, and any contestation of a decision must be 'accountable' – reasons, professional reasons, must be provided. Trust here is not a binary value, but rather it is fine-grained and is an ongoing social accomplishment as part of the work and the demonstration of competence. The work of the Pathology Lab whilst it has its individual components also has a profoundly social character and, as our pathologists suggest, we should consider the possible impact of technology on these working arrangements and practices. This is related to the wider issue of how technology should be designed for and deployed in healthcare settings. As technology becomes ubiquitous in healthcare settings, and as the technology becomes wrapped up in the many complexities of organizational working, so the challenges of systems design correspondingly increase, since the 'design problem' becomes not merely the design and development of new healthcare technologies but the integration of IT systems with existing and developing professional work practices and beliefs, including those pertaining to 'trust'.

4 References

Gambetta, D. (2000). Can we trust Trust? Trust: Making and breaking cooperative relations, Department of Sociology, University of Oxford, pp. 213-237.

Garfinkel, H, Lynch, M., and Livingston, E. (1981). The Work of a Discovering Science Construed with Materials from the Optically Discovered Pulsar, Philosophy of the Social Sciences, vol. 11, pp. 131-158.

Giddens, A. (1984). The Constitution of Society. Berkeley, CA: University of California Press.

Goodwin, C. (1994). Professional Vision, American Anthropologist, vol. 96, pp. 606-633.

Goodwin, C. (2000). Practices of Seeing: Visual Analysis: An Ethnomethodological Approach, in T. van Leeuwen and C. Jewitt (eds): Handbook of Visual Analysis, London: Sage, pp.157-82.

Hartswood, M., Procter, R. (2000). Computer-aided Mammography: A Case Study of Error Management in a Skilled Decision-Making Task, Topics in Health Information Management, vol. 20, no. 4, pp. 38-54.

Hartswood, M., Procter, R., Rouncefield, M., and Slack, R. (2002). Performance Management in Breast Screening: A Case Study of Professional Vision and Ecologies of Practice, Journal of Cognition, Technology, and Work, vol. 4, no. 2, pp. 91-100.

Kipnis, D. (1996). Trust and Technology. In: Kramer R, Tyler T. (eds). Trust in organizations, frontiers of theory and research. London: Sage, pp. 39-50.

Livingston, E. (2008). Ethnographies of reason. Ashgate Publishing, Ltd.

Luhmann N. (1990). Familiarity, confidence, trust, problems, and alternatives, in Gambetta D, (ed). Trust, making, and breaking cooperative relations. Oxford: Basil Blackwell. Available online at www.sociology.ox.ac.uk/trust book. html

Luhmann, N. (2018). Trust and power. John Wiley & Sons.

Watson, R. (2009). Constitutive practices and Garfinkel's notion of trust: Revisited, Journal of Classical Sociology, 9(4), pp. 475-499.

Balancing data-hungriness of AI and the workload of manual data collection

Christiane Grünloh, Eline te Braake, Marian Hurmuz, Stephanie Jansen-Kosterink
eHealth Department, Roessingh Research and Development, The Netherlands
Biomedical Signals and Systems Group, University of Twente, The Netherlands
{c.grunloh, e.tebraake, m.hurmuz, s.jansen}@rrd.nl

Abstract. Artificial intelligence systems need big data sets for the development of models or machine learning algorithms. This data has to be collected, which is not always possible to do automatically. Manual data collection in healthcare is often performed by patients or their caregivers, which adds workload to their existing disease burden. This paper reports on a qualitative study with Dutch patients and healthcare professionals focusing on data collection of patients living with chronic obstructive pulmonary disease. Both groups were concerned about the response burden of patients filling in questionnaires on a daily basis, who have limited energy at their disposal, to begin with.

1 Introduction

A common issue in artificial intelligence (AI) systems is the need for big data sets which must be collected, analyzed and learned from (Adadi, 2021). This data-hungriness has consequences that have raised concerns in terms of the sustainability of the development and use of AI systems (van Wynsberghe, 2021). Besides the resources needed for data processing, data first must be collected which is not always possible to do automatically (e.g., by sensors). In healthcare, patients are increasingly asked to provide reports about their health, quality of life, or functional status (so-called patient-reported outcomes, or PROs, Weldring and Smith, 2013). These real-world data are often used to monitor symptoms and progress or adapt treatment but can also be used in AI systems to predict exacerbation, and progression or to provide personalized virtual coaching (e.g., op den Akker et al., 2021). This position paper focuses on the data collection for AI systems in healthcare that is done manually by people and the workload that is demanded of them. We aim to (1) report on the attitude, expectations, and experience of patients and healthcare professionals regarding data collection, and

(2) open the discussion of balancing the need for large data sets with the burden of manual data collection.

2 Methods

Semi-structured interviews and several workshops were conducted with patients living with COPD and healthcare professionals (HCPs) in the Netherlands. This position paper focuses on results related to participants' preferences and attitudes regarding the data collection process. Workshops and interviews were recorded, transcribed, and then processed in coded form. The study was approved by the Institutional Review Board of Medisch Spectrum Twente (MST, number K21-20). All participants gave their written informed consent prior to starting the study.

3 Results

In total, 11 patients with COPD participated in the studies, two participated in both the interviews and workshops (N=6 female, age ranging from 63 to 80, mean 68.1 years). Furthermore, N=22 HCPs (N=17 female) from private practice and from the department of pulmonary medicine of MST with backgrounds in pulmonology, psychiatry, cardiology, internal medicine, physiotherapy, and nursing participated in interviews (N=7) and two workshops (N=12 per workshop, some attended both).

It was discussed which areas patients would like to keep track of. Patients mentioned three parameters: oxygen saturation, heart rate (HR), and blood pressure. Some participants mentioned having a pulse oximeter to measure their oxygen saturation. Patients stated to manually keep track of things (e.g., medication intake) or collect data by means of a wearable (e.g., Fitbit). While some patients are enthusiastic about this automatic way of data collection, others had negative experiences when using a Fitbit for some time: "I found that so irritating, then you think, I'm on a 100 HR and then I think 'What did I do?' You are just going to worry more." [Pat9]. When discussing data collection via questionnaires patients were reluctant. "Completing online questionnaires? Not every day! Terrible! Once a week is enough for me." [Pat1]. A reason given by another participant for this reluctance was the energy it would take to collect the data and complete the online questionnaire. One participant indicated that they do

not want to write down things like their mood as the benefit was not clear: “To keep a diary with: ‘today is a 6, tomorrow it is...’? I do not think that such things will help me.” [Pat7WS]. The HCPs in the workshops discussed that some COPD patients already collect some data (e.g., weight, oxygen saturation, body temperature). In current clinical practice, however, there is little time to go through these notes with the patients. As many of these measures were not considered to be useful, the HCPs were concerned about the time and energy spent on collecting the data. Unlike patients with diabetes who must act on measurements daily, this was different for COPD patients. The HCPs emphasized that for COPD patients there should be a balance between keeping track of things and not focusing too much on certain details and parameters. Furthermore, some patients do not want to measure all sorts of things as they do not want to be reminded of their disease all the time. According to the HCPs, special attention should be paid that the data collection does not add to their disease burden, is also used for educational purposes to support self-management, and that it also considers people with low levels of literacy.

4 Discussion

Most patients and HCPs were positive about the idea of collecting data by using sensors to gain more insights into the disease. The manual collection of data using questionnaires raised some concerns, regarding the time and energy needed to fill in questionnaires. Both groups emphasized that patients with COPD must manage their limited energy and that data collection should not unnecessarily add to their burden or keep them mentally occupied with their disease. The content of questionnaires was also considered crucial, as one not only collects information but also implicitly communicates something when asking questions. Through data collection, patients might be encouraged to focus on certain details, which, however, might not be relevant to them.

While it should be applauded that patients and caregivers are increasingly involved in research, attention should be paid to the workload that is demanded of them, which often comes in addition to their disease burden. The effort required to answer a questionnaire has been termed ‘response burden’, which is affected by factors like length, the density of sampling, the cognitive load required, layout, and interface of the survey (Rolstad et al., 2011). This creates a tension between

demandingness and inclusivity, in that participatory research, might be too demanding and thus unfeasible, or too uninclusive and thus unfair (Jongsma & Friesen, 2019). Furthermore, there is also the risk of selection bias, if "those who are attracted to citizen science and have the time, energy and technology to participate, and are undeterred by privacy and other concerns, are not representative of the population under study" (Majumder & McGuire, 2020).
A limitation of this study is the difficulty of asking hypothetical questions related to a person's willingness of filling in a questionnaire in the future. Without knowing the lengths of the questionnaire, how difficult the questions are, and the format and/or usability of the survey application, patients might have assessed their willingness too optimistically or too cautious. For example, if a questionnaire is short and well designed, a person might not mind filling this in daily. However, even if the questionnaire is short, attention should be paid to the content, as the burden may be quite high when a patient is asked to complete multiple questionnaires that reflect similar concepts (Rolstad et al., 2011).

5 Conclusion

Manual data collection by patients creates a response burden that needs to be balanced against their disease burden and the limited energy patients have at their disposal. When automatic data collection via sensors is not possible, the data collection instrument needs to be designed in a way that pays attention to multiple factors that affect the response burden.

6 Acknowledgements

This work is supported by the RE-SAMPLE project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 965315. The authors thank all participants for their time, input, and support in the studies; S. van Putten for her work in the early phase of the project; C. Bucsán, M. Brusse-Keizer, and A. Lenferink for their support with recruitment and local ethical approval at MST.

7 References

Adadi, A. (2021): 'A survey on data-efficient algorithms in big data era'. Journal of Big Data, vol. 8, no. 1, pp. 24.

Jongsma, K. and P. Friesen (2019): 'The Challenge of Demandingness in Citizen Science and Participatory Research'. The American Journal of Bioethics, vol. 19, no. 8, pp. 33–35.

Majumder, M. A. and A. L. McGuire (2020): 'Data Sharing in the Context of Health-Related Citizen Science'. The Journal of Law, Medicine & Ethics, vol. 48, no., pp. 167–177. PMID: 32342743.

op den Akker, H., M. Cabrita, and A. Pnevmatikakis (2021): 'Digital Therapeutics: Virtual Coaching Powered by Artificial Intelligence on Real-World Data'. Frontiers in Computer Science, vol. 3, no. 117.

Rolstad, S., J. Adler, and A. Rydén (2011): 'Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis'. Value in Health, vol. 14, no. 8, pp. 1101–1108.

van Wynsberghe, A. (2021): 'Sustainable AI: AI for sustainability and the sustainability of AI'. AI and Ethics, vol. 1, no. 3, pp. 213–218.

Weldring, T. and S. M. S. Smith (2013): 'Article Commentary: Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs)'. Health Services Insights, vol. 6, pp. HSI.S11093.

Towards Human-Centered AI: Learning from Current Practices in Radiology

Nazmun Nisat Ontika, Sheree May Saßmannshausen, Hussain Abid Syed, Aparecido Fabiano Pinatti de Carvalho, Volkmar Pipek
Institute for Information Systems, University of Siegen, Germany
{nazmun.ontika, sheree.sassmannshausen, hussain.syed, fabiano.pinatti, volkmar.pipek}@uni-siegen.de

Abstract. In this paper, we will present the first insights of an empirical study carried out to explore current practices of radiologists in diagnosing prostate cancer and identify the design space for potential assistive AI systems to be used in such a context. Based on data collected over more than three months of observation, we go on to discuss the relevance of human-centered AI for the design of explainable and trustworthy solutions.

1 Introduction

It is a fact that more and more there are insufficient specialist physicians to fulfill the increasing demands for healthcare (IHS Markit Ltd., 2021). In Germany, there are only about 8.792 radiologists, which is only 2,24% of all doctors (Ärztestatistik, 2018). In contrast, there is a huge frequency of radiological examinations in Germany, and it vastly outnumbered its European peers in terms of case volume, more than twice the figures reported in several other countries (C. Silvestrin, 2016). The discrepancy between available radiologists and the need for radiological examinations reflects the high workload. They usually have only about 20 minutes per report resulting in high pressure and loss of quality of the findings. The increased psychosocial workload of radiologists might have an impact on individual health and, as a result, worse patient care quality (Oechtering, T. H. et al., 2020). All these data clearly depict the radiologist deficit in Germany, which might cause congestion in significant segments of the healthcare sector. It is, therefore, sensible to think of possible solutions to support these professionals in their work, to help them to be more productive and efficient. Incorporating AI into radiology can possibly assist these physicians in this regard, by helping to make informed choices easily and quickly and can help to overcome the bottleneck.

Nevertheless, to be able to devise useful and usable solutions, it is key to apply user-centered approaches to design. Not only that but paying attention to practices is also a determinant to identifying the design space for useful solutions, which can be ultimately appropriated by the use and support of the development of new and/or improved practices (Wulf et al., 2015).

In this paper, we report on the initial phase of a Design Case Study (Wulf et al., 2015) carried out for the design of an active learning system to support radiologists in their diagnosis by using image recognition, which will be able to detect and locate anomalies. Through visualizations, the decision of the AI system will be made more understandable and transparent, which will lead to a fair and responsible perception of the system-human decisions (Lee, 2018). The radiologists will decide whether they agree with the result or not and give feedback to the AI system. It is envisaged that the training of the AI system will be supported by delivering feedback to the radiologists about the result. This human-in-the-loop model will help improve the AI system in the long term. Following the premises of the Grounded Design paradigm (Rohde et al., 2017), we have the continuous participation of radiologists in the design, assessment, and use of the envisaged AI system for our research. The design considers real decision criteria and practices of the radiologists by adapting them into the AI system. These criteria and practices have been collected in an ethnographically informed contextual study. In the following, we introduce the preliminary findings of this study.

2 Methodology

The Contextual Inquiry (Holtzblatt and Jones 2017) was conducted over a period of three months with a total of eight inquiry sessions, which were subsequently enriched by two interviews. Each observation session lasted between 3 to 4 hours and focused mainly on the current work, experiences, interactions, and challenges faced by the participants. The primary role focused on this project is the role of the radiologist. However, medical technical radiological assistants (MTRA) were also observed as additional stakeholders.

The following empirical data collection was performed:

- Five Contextual Inquiry sessions (observations with ad hoc interviews) with two radiologists (One chief radiologist, and one senior physician)

- Three Contextual Inquiry sessions with medical technical radiological assistants
- Two interviews with two radiologists (One chief radiologist, and one senior physician)

3 Empirical Results

Before the in-depth analysis of the gathered data in our ongoing project, we opted for an agile method, conducting the pre-analysis by actively listening to the recordings of the contextual inquiries in the group of 5-6 researchers and AI developers. The initial results from the pre-analysis of our empirical data collection provide insights into challenges in current work practice, artifacts used, as well as design potentials derived from them.

Currently, many and various AI systems are being developed to detect several kinds of cancer, including breast cancer, lung cancer, and prostate cancer. However, despite the improvement in detection, the new systems are hard to operate even for expert radiologists. From our observation, we have learned that radiologists operate in a demanding and complicated workplace. They must execute many software programs on multiple displays at the same time. That is why, rather than adding a layer to their practice, we must incorporate new AI applications within their current infrastructure and systems, making information conveniently accessible at their disposal. Considering radiologists in limited quantity in several regions of the world and confronted pace with fast workloads, having an automated assistant would undoubtedly benefit them. Hence, we are not only talking about the AI tools performing the automation accurately but also being easily operable, explainable, trustworthy, and integrable with existing systems.

On average there are 1 billion radiology exams each year and unfortunately, there can be 40 million radiologist errors while spending 20-30 minutes for each case (Brady, 2016). Furthermore, according to a study conducted over the previous 70 years, the discrepancy between the two radiologists is extremely steady over the periods (Schmid, 2021). In our observation, we have also noticed different opinions from radiologists looking at the same MRI scan. Moreover, we have observed that different radiology centers gave different results for the same patient at the same time. The fact that various radiologists might reach various outcomes

even with the same clinical specimens is a concern in today's prostate cancer diagnoses, which implies that treatment choices are made on unclear information. While disparities in assessment do not naturally imply wrong, knowing the roots of such disparities and their importance can aid in guiding effective action to reduce and regulate these discrepancies when they can be managed and comprehend them when they cannot.

Redundant work steps are identified in the transfer of diagnostic results to the report, which has the potential for automation. Furthermore, manual calculations are performed, such as the calculation of prostate size, prostate volume as well as PSA (prostate-specific antigen) density, which takes a lot of time. Moreover, radiologists use their mobile phones for several calculations. A more efficient solution could be to automate these calculations by image recognition and automatic transmission of the patient's PSA value.

We have remarked on several possible reasons that could result in a faulty diagnosis. One of the big reasons is several manual inputs including manual calculation and manual transferring of data among different systems by radiologists and MTRA. Furthermore, the tedious workload, tiresome work stress, and constant interruptions by other administrative work (such as phone calls, paperwork, etc.) could also lead to a questionable diagnosis.

The radiologists mentioned in their interviews with us that they are very satisfied with the current system, but they are using it for the last 10 years and comparing it with the system they were using 10 years ago, but it is most likely that the old system is very outdated. Hence, the assessment of the satisfaction with the current systems of the users remains questionable.

Interestingly when the radiologists were asked about the need for a possible AI system and if they will trust the system, they answered positively and mentioned there is scope for automation and it will definitely help them in their everyday work, and they will trust the system if it is able to prove its efficiency and accuracy.

4 Further Development: Design Scope

User interface design is also a critical aspect to consider for any product development. Healthcare physicians, as users in general, want user interfaces that

are simple to operate yet aesthetic, as well as intriguing and encouraging (Wang et al., 2021).

We have gained important insights about the current diagnosis of prostate cancer from our pre-analysis, and we have identified some areas for design implications in the upcoming phase of our research.

Different visualization approaches using artificial intelligence will be explored, such as outlining the prostate, showing the different areas (e.g., the peripheral or transitional zone of a prostate), or for example, showing abnormalities that could indicate possible tumors or carcinomas.

Regarding the current working practice, concrete procedures and classifications have been identified that lead to the assessment of the severity of potential prostate carcinoma. These findings could also be transformed into concrete requirements for the AI system. The entire workflow from preparation/examination, through the main diagnosis, to debriefing with a second radiologist and reporting was identified and will be examined for further requirements as part of the thematic analysis.

From our pre-analysis, it is evident that there are design and decision gaps in the current procedure, but we need to make this diagnosis system standardized to deal with divergences in the assessment of radiological images and act appropriately. We aim at addressing this issue in our next phase of research.

5 Acknowledgments

The authors would like to thank their project partners from the MVZ Jung-Stilling in Siegen for their participation. The authors would also like to acknowledge the financial support from the Bundesministerium für Bildung und Forschung - BMBF through the PAIRADS project (funding code: 16SV8651; https://pairads.ai) as well as the financial support through the German Research Foundation (DFG) – Project number 262513311 – SFB 1187 Media of Cooperation.

6 References

Ärztestatistik zum 31. Dezember 2018 Bundesgebiet gesamt. (n.d.). Retrieved from https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Statistik 2018/Stat18AbbTab.pdf.

Brady, A. P. (2016). Error and discrepancy in radiology: inevitable or avoidable? Insights into Imaging, 8(1), 171–182. https://doi.org/10.1007/s13244-016-0534-1.

C. Silvestrin. (2016). Europe's Looming Radiology Capacity Challenge A Comparative Study. Retrieved from https://www.telemedicineclinic.com/wp-content/uploads/2016/11/Europes_looming_radiology_capacity_challenge-A_comparitive_study.pdf.

Holtzblatt, K. and Jones, S. (2017) 'Contextual inquiry: A participatory technique for system design', in Participatory Design: Principles and Practices, pp. 177–210. doi: 10.1201/9780203744338.

IHS Markit Ltd., A. of A. M. C. (2021). The Complexities of Physician Supply and Demand: Projections From 2019 to 2034.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1), 205395171875668. https://doi.org/10.1177/2053951718756684.

Oechtering, T. H., Panagiotopoulos, N., Völker, M., Lohwasser, S., Ellmann, S., Molwitz, I., Storz, C., Winther, H., Eisenblaetter, M., Antoch, G., Schönberg, S. O., Barkhausen, J., Anton, F., Neumann, S., Layer, G., Doerfler, A., Koerber, F., Wessling, J., Wucherer, M., & Raspe, M. (2020). Work and Training Conditions of German Residents in Radiology - Results from a Nationwide Survey Conducted by the Young Radiology Forum in the German Roentgen Society. Arbeits- und Weiterbildungsbedingungen in der Facharztweiterbildung Radiologie – Ergebnisse einer bundesweiten Weiterbildungsumfrage durch das Forum Junge Radiologie in der Deutschen Röntgengesellschaft. RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin, 192(5), 458–470. https://doi.org/10.1055/a-1047-1075.

Rohde, M., Brödner, P., Stevens, G., Betz, M. and Wulf, V. (2016) 'Grounded Design: A Praxeological IS Research Perspective', Journal of Information Technology, 32(2), pp. 163–179. doi: 10.1057/jit.2016.5.

Schmid, A. M., Raunig, D. L., Miller, C. G., Walovitch, R. C., Ford, R. W., O'Connor, M., Brueggenwerth, G., Breuer, J., Kuney, L., & Ford, R. R. (2021). Radiologists and Clinical Trials: Part 1 The Truth About Reader Disagreements. Therapeutic Innovation & Regulatory Science, 55(6), 1111–1121. https://doi.org/10.1007/s43441-021-00316-6.

Wang, D., Wang, L., & Zhang, Z. (2021). Brilliant ai doctor in rural clinics: Challenges in ai powered clinical decision support system deployment. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3411764.3445432.

Wulf, V., Müller, C., Pipek, V., Randall, D., Rohde, M. and Stevens, G. (2015) 'Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies', in Wulf, V., Schmidt, K., and Randall, D. (eds) Designing Socially Embedded Technologies in the Real-World. London, UK: Springer London, pp. 111–150. doi: 10.1007/978-1-4471-6720-4_7.

Using Signals to Support Trust Building in Clinical Human-AI Collaboration

Naja Kathrine Kollerup, Mikael B. Skov, Niels van Berkel
Department of Computer Science, Aalborg University
nkka@cs.aau.dk

Abstract. Artificial Intelligence (AI) has the technological potential to transform healthcare by assisting medical personnel in their everyday workflow. For successful collaboration and adoption of AI technology, end-users need to trust the AI system. In this paper, we outline the use of Relational Signalling Theory, an established theory on Human-Human trust building, as a conceptual lens for designing trust-building signals in Human-AI interaction. We argue that the use of a theoretical foundation in the design and evaluation of interactions supports the development of Human-Centered AI in healthcare.

1 Introduction

Artificial Intelligence (AI) systems take an increasingly larger role in assisting humans in clinical decision-making (Oh et al., 2018; Yang et al., 2020). For AI systems to be used successfully in clinical practice requires AI systems to collaborate and align with day-to-day clinical practice (Wang et al., 2020). Thus, integrating AI systems as a collaborative partner in human workflow, especially in high-stake workflows encountered in healthcare, requires humans to put trust in the AI support system (van Berkel et al., 2022; Vereschak et al., 2021).

Current AI systems often face distrust, which can result in underestimation of the AI's capabilities, disuse, increased user workload, or deterioration of the performance of the system (Okamura and Yamada, 2020). In this article, we argue that in order to design AI systems that evoke trust among their end-users, we first need to understand how humans build trust. We first briefly summarise the Relational Signalling Theory (RST) (Six et al., 2010) and how it is used to describe trust-building behavior in human-to-human relationships. Secondly, we outline how the concept and techniques of relational signals potentially can have profound implications for trust-building in Human-AI teams. Finally, we present examples within the healthcare domain of qualities AI systems should acquire to facilitate trust building.

Through our existing understanding of human-to-human trust-building, we outline concrete takeaways for Human-AI trust-building and motivate a novel research direction for the CSCW and HCI communities in relation to clinical Human-AI collaboration.

2 Relational Signalling Theory

RST is a theory proposed by Lindenberg (2000). RST is grounded in two basic assumptions: First, human behavior is goal-directed, and to explain the social context, one must pay attention to the goals of individuals. Second, human behavior is context-dependent (Six et al., 2010). Relational signals are signs in the behavior of the trustee (*i.e.*, the party aiming to create a trust), where the trustor (*i.e.*, the party assessing the trustee) considers two distinct aspects; Does the trustee show signs in the behavior of interest in maintaining a relationship in the future (ability dimension of trustworthiness), and does the trustee show signs in the behavior of having the competences to perform according to the expectations? (internal dimension of trustworthiness).

Lindenberg (2000) distinguishes between three master frames of operation: the hedonic frame, the gain frame, and the normative or solidarity frame. The first two frames are ego-oriented and serve one's own interest, whereas the third frame is alter-oriented, which means that one will show concern for the other individual (Six et al., 2010). People will look for signs in the behavior of another individual indicating stability in the solidarity frame, and moreover, to which degree the individual is interested in maintaining a relationship in the future.

3 Signalling types

A better understanding of the signal types emitted from AI systems (the trustee) supports the design of systems that are trusted by clinician and patient alike (the trustor) and helps to answer the question: 'What qualities should the AI system have in order for humans to trust it as a collaborative partner in clinical care?' To answer this question, we need to understand not only the frame of operation but consider the signals as an incorporated part of trust building. We draw upon the Signalling theory (Donath, 2007) concerned with understanding why specific

signals are reliable and others are not. Donath identifies several signals that fall under two categories: Assessment and Conventional signals.

- *Assessment signals*, which are costly to fake and therefore considered honest and reliable signals. The quality they signal is 'wasted' in the production and therefore tends to be expensive to produce (Shami et al., 2009) and is challenging to fake. For example, lifting a heavy weight sends a reliable signal of strength – a weaker person simply cannot do it (Donath, 2007).
- *Conventional signals*, which are cheaper to produce and therefore considered less reliable and open to deception. To give an example, one may choose to use a deceptive picture of oneself on social media. This will result in loose of meaning, and these signals become unreliable (Shami et al., 2009).

Building on these established notions of trust-building in Human-Human interaction, we propose the use of RTS to frame Human-AI interactions. Through the aforementioned signal types from signaling theory, we can conceptualize, design, and study the signals that affect a trustor's belief or behavior (Lampe et al., 2007). Especially within the healthcare domain, trust-building is essential for both clinicians and patients. We next outline how signals can be embedded in Human-AI collaboration scenarios in clinical care to enhance trust.

4 Using signals in healthcare AI systems

The qualities the AI system should acquire to facilitate trust-building for medical practitioners and patients depend on the signals being sent from the AI system. These qualities can almost be anything, *e.g.*, honesty, and reliability (Donath, 2007).

As an example of using signals in a healthcare context, we describe an AI-powered computer vision system designed to identify moles from melanomas. To gain the user's trust, the AI system can present several assessment signals. For example, the system can present alternative considerations made in its assessment and detail why these considerations did not end up being the final assessment. Alternatively, the system can highlight which specific elements of the image are deemed suspicious. These are costly signals (computational power, added

explanations) that we hypothesize would increase end-user trust. Conventional signals, such as presenting the outcome of the analysis without any type of explanation, are less likely to support trust building in Human-AI interactions and come at a lower cost to the AI system.

5 Conclusion and future work

In this position paper, we have provided an overview of relational signaling theory and argued for its use in trust-building in clinical Human-AI collaboration. Given the importance of trust in the healthcare domain, it is critical that we design systems that instill trust in users where appropriate but are also able to highlight to its users when its recommendations are less trustworthy and should receive extra scrutiny. We call for future work to empirically study the degree to which assessment signals and conventional signals support trust-building in end-users of Human-AI systems.

6 Acknowledgements

This work is supported by the EXPLAIN-ME project of Digital Research Centre Denmark (DIREC) under Innovation Fund Denmark.

7 References

Donath, J. (2007): 'Signals in Social Supernets'. J. Computer-Mediated Communication, vol. 13, pp. 231–251.

Lampe, C. A., N. Ellison, and C. Steinfield (2007): 'A Familiar Face (Book): Profile Elements as Signals in an Online Social Network'. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 435–444.

Lindenberg, S. (2000): 'It Takes Both Trust and Lack of Mistrust: The Workings of Cooperation and Relational Signaling in Contractual Relationships'. Journal of Management and Governance, vol. 4, pp. 11–33.

Oh, C., J. Song, J. Choi, S. Kim, S. Lee, and B. Suh (2018): I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence, p. 1–13.

Okamura, K. and S. Yamada (2020): 'Adaptive trust calibration for human-AI collaboration'. PLOS ONE, vol. 15, pp. e0229132.

Shami, N. S., K. Ehrlich, G. Gay, and J. T. Hancock (2009): 'Making Sense of Strangers' Expertise from Signals in Digital Artifacts'. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 69–78.

Six, F., B. Nooteboom, and A. Hoogendoorn (2010): 'Actions that Build Interpersonal Trust: A Relational Signalling Perspective'. Review of Social Economy, vol. 68, no. 3, pp. 285–315.

van Berkel, N., J. Opie, O. F. Ahmad, L. Lovat, D. Stoyanov, and A. Blandford (2022): 'Initial Responses to False Positives in AI-Supported Continuous Interactions: A Colonoscopy Case Study'. ACM Trans. Interact. Intell. Syst., vol. 12, no. 1.

Vereschak, O., G. Bailly, and B. Caramiaux (2021): 'How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies'. Proc. ACM Hum. -Comput. Interact., vol. 5, no. CSCW2.

Wang, D., E. Churchill, P. Maes, X. Fan, B. Shneiderman, Y. Shi, and Q. Wang (2020): 'From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People'. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. p. 1–6.

Yang, Q., A. Steinfeld, C. Rosé, and J. Zimmerman (2020): Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design, p. 1–13.

Human-AI collaboration protocols to assess real-world options of medical AI

Federico Cabitza[1,2], Andrea Campagner[1]
[1]Department of Computer Science, Systems and Communication, University of Milano-Bicocca, Milan, Italy
[2]IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
Contact Author: federico.cabitza@unimib.it

1 Introduction

We agree with what Elmore and Lee (2022) wrote in a recent editorial: "there are complex interactions between a computer algorithm output and the interpreting physician [and that the] extent to which physicians may be influenced by the many types and timings of computer cues remains unknown". We concur with this observation and feel that it can be easily generalized to any work domain where stakes are high, and decisions are knowledge intensive. To study the influence that AI-based decision aids can exert on human decision-makers, we propose the concept of human - artificial intelligence collaboration protocol (HAI-CP), which is "an integrated set of rules and policies that stipulate the use of AI-exhibiting tools by competent practitioners to perform a certain task or do a certain job"[1].

The concept of HAI-CP is proposed as a conceptual construct to describe (as well as design and evaluate) different ways in which users and their AI tools can interact to have their work done and make better decisions. For instance, different HAI-CPs determine what steps of the task should be fully automated, or kept under tight human oversight; what data are made available to the AI tool as its input; what data this tool is supposed to provide users with as its output; at what step, and in what order with respect to the work of human beings, the tool should give its output and with what kind of autonomy or control.

[1] A collaboration protocol specializes in the more general concept of interaction protocol. Although adopting the term collaboration is not a neutral choice (no terminological choice really is), we also believe that it is opportune to adopt a term that specifically concerns "work settings, that is, [work practice] under conditions of severe constraints" (Schmidt and Simonee, 1996). In so doing, we recognize that the concept of interaction is necessarily broader and capable to include any informal, entertainment, or ludic settings, and, more generally, information and knowledge retrieval activities that are not necessarily associated with a formal task or with tasks mutually associated with other tasks in the context of more complex and articulated processes.

2 Study Development and Results

We performed three user studies, involving clinicians in realistic diagnostic scenarios to see if significant differences could be observed in different HAI-CP in terms of overall effectiveness (i.e., accuracy), and, relatedly, if ram Ais in decision making can induce any detectable form of automation bias (Lyell and Coiera, 2017) or other differentiating effects. We compared strict second-opinion classes of HAI-CPs (see Figure 1), both the AI-first and the human-first configuration, in three diagnostic settings: knee lesion MRI interpretation, ECG reading, and vertebral x-ray reading, where each task sequence, or scenario, is a protocol. The results of the three studies, represented in terms of benefit diagrams (Tschandl et al., 2020), are reported in Figures 2a, 2b, 3a, 3b, and 4. To summarize our results, as can be easily observed from the diagrams, the concept of a HAI-CP enables us to compare different protocols in terms of their impact, defined as the difference in diagnostic accuracy.

We notice that AI support had a positive impact on the decision-making quality, as the protocols featuring this type of support reported higher accuracy than that reported by humans alone. By contrast, the impact of XAI support was more controversial, providing no additional improvement (or even having a detrimental effect) compared to providing AI support alone. Considering these findings, and despite their obvious limited generalizability to other work contexts and domains, we believe that the proposed concept of HAI-CP could foster research on, and have value in, the modeling and assessment of the impact (either positive or negative) that AI systems can have in real-world decision workflow.

3 Reference

Elmore, J. G. and C. I. Lee (2022): 'Artificial Intelligence in Medical Imaging—Learning From Past Mistakes in Mammography'. In: JAMA Health Forum, Vol. 3. pp. e215207–e215207.

Lyell, D. and E. Coiera (2017): 'Automation bias and verification complexity: a systematic review'.

Journal of the American Medical Informatics Association, vol. 24, no. 2, pp. 423–431.

Schmidt, K. and C. Simonee (1996): 'Coordination mechanisms: Towards a conceptual foundation of CSCW systems design'. Computer Supported Cooperative Work (CSCW), vol. 5, no. 2, pp. 155–200.

Tschandl, P., C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, et al. (2020): 'Human–computer collaboration for skin cancer recognition'. Nature Medicine, vol. 26, no. 8, pp. 1229–1234.

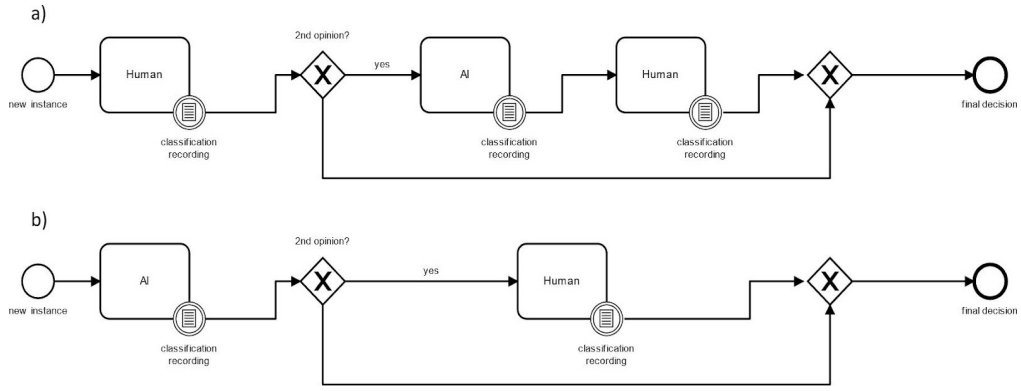
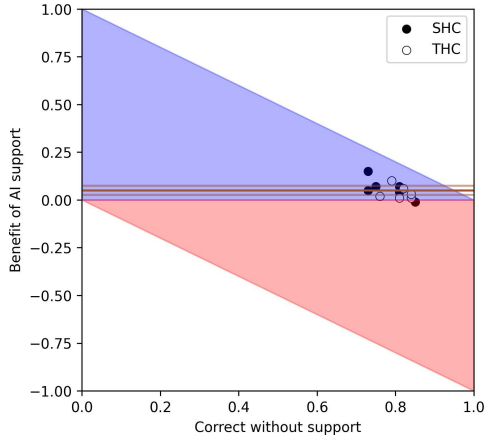
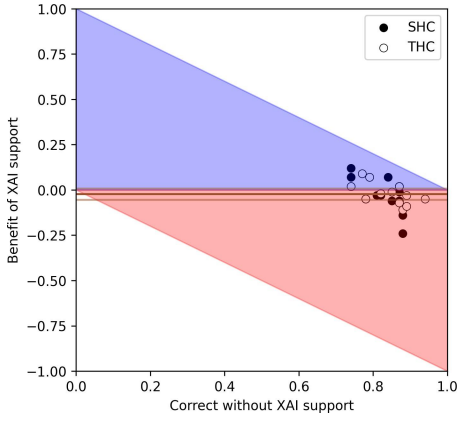


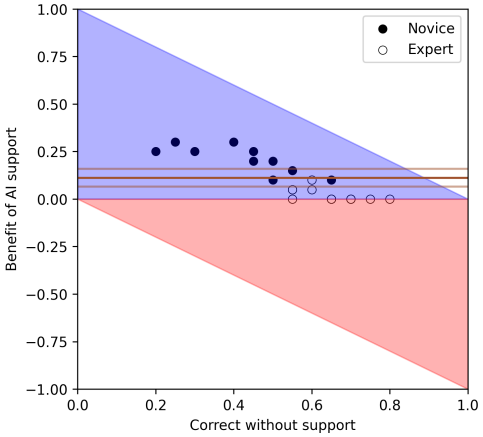
Figure 1: The interaction protocols we considered in this study. a) the strict second- opinion human-first use case. b) the strict second-opinion AI-first use case. The final decision is always up to the human decision maker. Registrations of provisional decisions (classification) allowed us to compare error rates between these two business use cases.



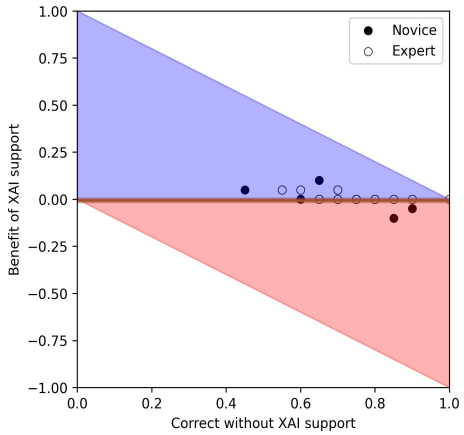
(a) Benefit diagram for the knee lesion MRI interpretation study, showing the impact of AI support. Dots represent the accuracies of the radiologists, while the brown lines represent the average difference in accuracy between the two protocols, along with the corresponding 95% confidence interval. The blue region denotes an improvement, while the red region a worsening. AI support had a positive impact, resulting in a significant increase in accuracy.



(b) Benefit diagram for the knee lesion MRI interpretation study, showing the impact of XAI support. Dots represent the accuracies of the radiologists, while the brown lines represent the average difference in accuracy between the two protocols, along with the corresponding 95% confidence interval. The blue region denotes an improvement, while the red region a worsening. XAI support had a negative impact, resulting in a decrease in accuracy, albeit not significant.



(a) Benefit diagram for the ECG reading study, showing the impact of AI support. Dots represent the accuracies of the cardiologists, while the brown lines represent the average difference in accuracy between the two protocols, along with the corresponding 95% confidence interval. The blue region denotes an improvement, while the red region a worsening. AI support had a strongly positive impact, resulting in a significant increase in accuracy.



(b) Benefit diagram for the ECG reading study, showing the impact of XAI support. Dots represent the accuracies of the cardiologists, while the brown lines represent the average difference in accuracy between the two protocols, along with the corresponding 95% confidence interval. The blue region denotes an improvement, while the red region a worsening. XAI support had no impact on average, with neither an increase nor a decrease in accuracy.

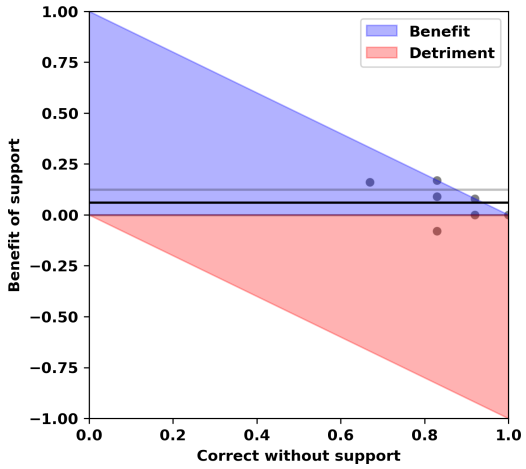


Figure 4: Benefit diagram for the vertebral x-ray reading study, showing the impact of AI support against having no AI support at all, stratified by type of AI support. In the graph, the dots represent the accuracies of the radiologists, while the black lines represent the average difference in accuracy between the two protocols, along with the corresponding 95% confidence interval. The blue region denotes an improvement, while the red region a worsening. AI support had a positive impact on average, with a significant increase in accuracy.

A Service Design Approach for AI-Supported Clinical Tools: Collaborating with Interdisciplinary Care Teams & Patients to Provide and Leverage SDOH Data

Astrid Chow
Eleanor Health
Principal Product Designer and User Research Lead
astrid.chow@eleanorhealth.com

1 Presentation Summary

Health equity means that everyone has a fair and just opportunity to be as healthy as possible. This requires acknowledging and working to reduce obstacles to health, such as poverty, discrimination, and their consequences, including powerlessness and lack of access to good jobs with fair pay, quality education and housing, safe environments, and health care (CDC, 2022).

These obstacles to health equity are considered part of a broader concept called Social and Behavioral Determinants of Health (SBDOH). SBDOH refers to “conditions in the environments where people are born, live, learn, work, play, worship, and age that affect a wide range of health, functioning, and quality-of-life outcomes and risks” (HHS, 2020).

Eleanor Health is a start-up committed to providing care to individuals affected by mental health problems, including Substance Use Disorder (SUD). Eleanor Health prioritizes health equity in its care model, which is one of its four foundational operating principles (“Core Four”) - equity, harm reduction, trauma-informed care, and team-based care (Hochuli, Butler, Larris, & Ray, 2022) (Figure 1). These principles are especially relevant because they are also important foundations for designing ethically for health care using AI.

Collecting and storing SDOH information from patients is a critical part of gathering the medical and social history of patients to understand and mitigate these obstacles to provide more equitable care.

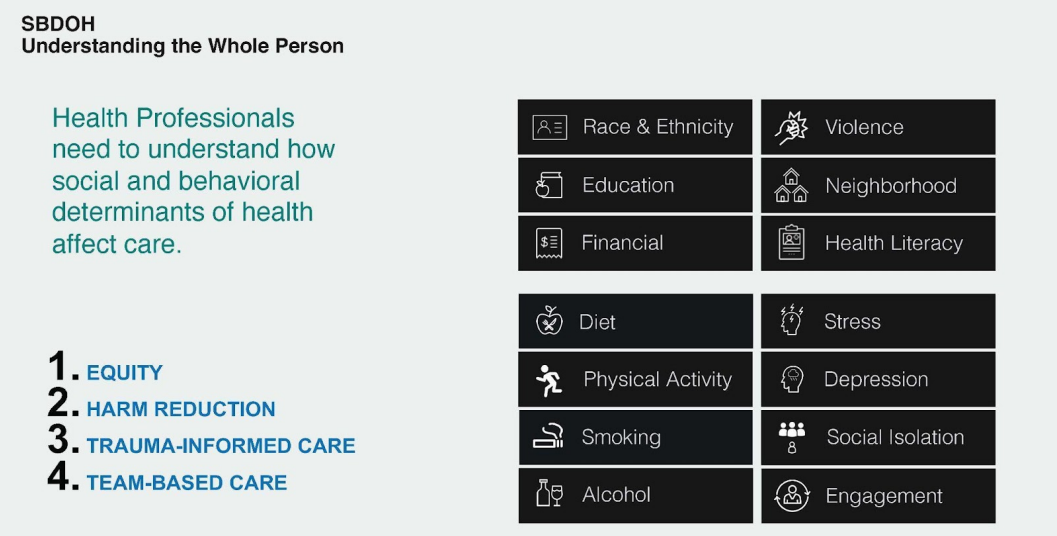


Figure 1: Four foundational operating principles ("Core Four").

However, the complexity of health care delivery in today's modern world means that clinicians often don't readily have access to a patient's SDOH information in a centralized place. Because SDOH is captured across multiple sources, care teams must often forage to find relevant information.

Service design is a way to organize the operations, services, and processes of a system to ensure it is user-centered, providing necessary support, scalability, and sustainability. This presentation highlights how a service design lens through a "service design blueprint" can help streamline the collection of SDOH information as patients navigate the complicated process of receiving health care (Barbarin & Chow, 2021) (Figure 2). Additionally, it discusses how Artificial Intelligence (AI) can support the collection of SDOH in a way that increases and retains patients' engagement in their own care, emphasizing ethical, socially conscious design practices.

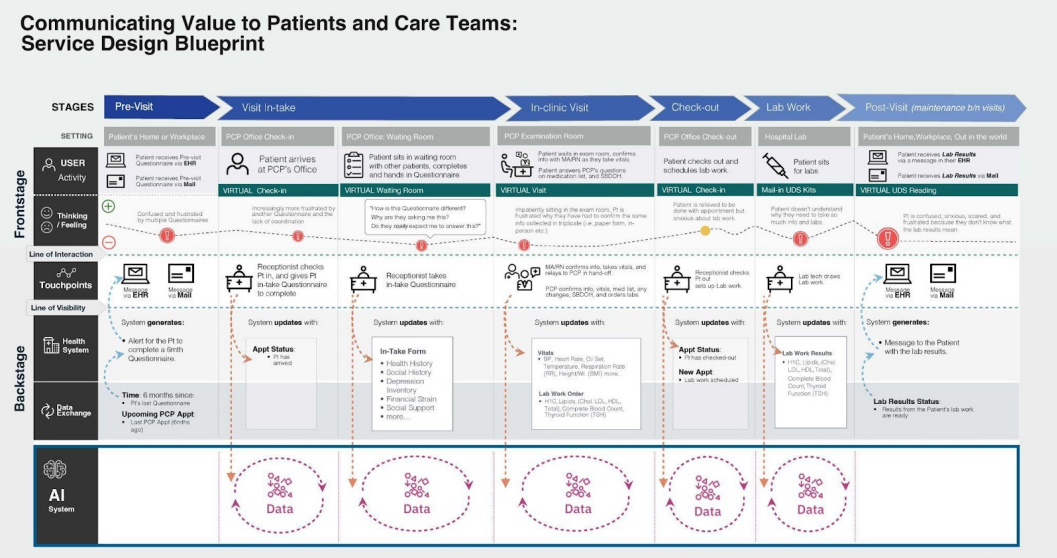


Figure 2: Service design blueprint.

2 Acknowledgements

I'd like to thank the other workshop organisers Volkmar Pipek, Richard Harper, Yunan Chen, Sun Young Park, Miria Grisot, Nils Blaumer, Aparecido Fabiano Pinatti de Carvalho, Nazmun Nisat Ontika, Sheree May Saßmannshausen, Hussain Abid Syed for inviting me to share my experiences as a designer working in the SUD and mental health start-up industry.

3 References

Barbarin, A., & Chow, A. (2021). Data Work for AI-Supported Clinical Tools: Showing Value to Encourage Patient Engagement in Providing SBDOH Data. HFES International Symposium. Human Factors and Ergonomics in Health Care.

Centers for Disease Control. National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). (2022). Health Equity. Accessed on 10/5/2022 from: https://www.cdc.gov/chronicdisease/healthequity/index.htm.

Chenault, A., Chow, A., & Wu, J. (2017). The potential ick factor: Ethical considerations for designing for healthcare".O'Reilly AI Conference San Francisco.

Cutler, A., Pribić, M., & Humphrey, L. (2018, September). Everyday Ethics for Artificial Intelligence. Retrieved from https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf.

Hochuli , M., Butler, D., Larris, L., & Ray, B. (2022). Health Equality in Action. Health Equity Council and the African American Behavioral Health Center of Excellence.

U.S. Department of Health and Human Services. (2030). Healthy People: Social Determinants of Health. Accessed on 10/5/2022 from: https://health.gov/healthypeople/priority-areas/social-determinants-health.

PAIRADS: Interaction of humans and technology rethought

Demster Bijl, Nils Blaumer, David Matuschek
Gemedico GmbH, Allendorf (Eder), Germany
hello@gemedico.com

1 AI System Summary

AI in healthcare may improve healthcare and quality of life, provide a more precise diagnosis and treatment plans, and result in overall improved patient outcomes. One of the prominent benefits of an AI System is to aid in the detection of cancer from MRI studies. In this article, we are focusing on this notion, especially for clinically significant prostate cancer. For good integration in the workflow of radiologists, it is helpful when the system outputs are of a kind that radiologists are already familiar with. Hence, it takes regular steps (e.g. recognition of the prostate) and assigns cancerous lesions and a study as a whole PI-RADS score.
Following PI-RADS 2.1 an MRI study for clinically significant prostate cancer detection should include a T1W series, an axial T2W series, at least one coronal or sagittal T2W series, a high b-value (1400+) DWI series, an ADC series, and in many cases a DCE series (Turkbey et al., 2019).
The AI System can be subdivided into three main parts, which each have their own sub-purpose, and require different images and different AI approaches. These will briefly be explained in turn.

2 Part 1: Anatomical Delineations

2.1 Motivation

Generally, there are four major zones within the normal prostate: the peripheral zone, the central zone, the transition zone, and the anterior fibromuscular stroma (Bhavsar & Verma, 2014) but we only look into two of them at the moment, because these are the relevant zones for the PI-RADS Scheme.
The purpose of the first part is to create anatomical delineations for the MRI images. Specifically, the prostate itself, and two of its zones, namely the peripheral zone (PZ) and the rest, sometimes called the central gland (CG), are delineated. Such delineations are helpful in many ways. For example, a prostate delineation allows the determination

of prostate volume and the calculation of PSA density, as well as checking for extraprostatic extensions when combined with lesion localization.
The zonal division of the prostate is essential because lesion appearance, PI-RADS score assignment, cancer frequency, and cancer outcome all differ per zone and are hence highly relevant to clinically significant prostate cancer detection. More essential features can be calculated such as PZ-PSA density.

2.2 Data

In reality, in each case, of delineations per zone made by one or more radiologists, or of semi-automated delineations, that is, delineations that were automatically generated but reviewed and, where needed, adjusted by radiologists.

2.3 Techniques

This part of the system expects an axial T2W series as input, as this is the sequence from most clearly shows the anatomy, although we are additionally experimenting with adding an ADC sequence as a second input. Its outputs consist of anatomical delineations. Such delineations can be visualized easily, e.g.in the form of contours around or coloring on the original MRI images, per object of interest.
We use techniques from semantic segmentation to perform the delineation. Within the field of semantic segmentation, we make use of Deep Learning techniques, which come either in the form of convolutional neural networks (CNNs) or Transformers.
We use techniques from semantic segmentation to perform the delineation. To explain the concept, in short, these techniques assign to each pixel in an image the class to which they belong, in our case the prostate, the PZ, and the CG. Semantic segmentation techniques do not distinguish between different instances of the same class, that is, they cannot tell different objects of the same type apart. As the MRI images only ever contain a single prostate, PZ and CG this is not a problem, fortunately. Within the field of semantic segmentation, we make use of Deep Learning techniques, which come either in the form of convolutional neural networks (CNNs) or Transformers.

3 Part 2: Lesion Detection

3.1 Motivation

The purpose of the second part is the localization of significantly cancerous lesions on the MRI images. This has clear relevance on its own and can additionally be combined with anatomical delineations for further processing.

3.2 Data

Ground truth consists, in each case, of delineations for significantly cancerous lesions made by one or more radiologists or of automatically generated ground truths that used textual radiologist reports for additional processing. Studies without significantly cancerous lesions are also included, which helps models to learn what significantly cancerous lesions do not look like.

3.3 Techniques

This part of the system expects an axial T2W series, an ADC series, and a high b-value DWI series as inputs. These are all necessary, as they are all highly relevant to clinically significant cancer detection. The ADC/DWI series is especially important for the detection of lesions in the PZ, and the T2W, especially for the detection of lesions in the transitional zone (TZ, a subregion of our CG).
We have different models that perform different types of localization, namely as bounding boxes (3D rectangular delineations) and as normal segmentations (simply delineations of an arbitrary shape, ideally the lesion shape). As it is possible for more than 1 lesion to be present in the series, further post-processing is performed to tell lesions apart. One then has a localization per lesion, which can again be visualized as a contour or coloring on the MRI images. These values additionally come with a confidence level, like an estimated probability, for the lesion being clinically significantly cancerous. Techniques come from the intersection of Deep Learning with object detection and semantic segmentation in computer vision, with postprocessing to tell the objects apart. Semantic segmentation works as before, with the classes being 'Significantly Cancerous Tissue' and its complement. Object detection works similarly but can be understood to assign classes to rectangular regions (often of varying size) instead of single pixels. Hence it is less precise, but it more naturally allows for telling objects apart, as the rectangular regions are set to capture an entire object.

4 Part 3: Lesion Classification

4.1 Motivation

The purpose of the third part of the AI system is the assignment of a PI-RADS score to lesions detected by part two of the AI system. This step helps to put the predictions of the AI system in a 'language' that the radiologists are already familiar with.

4.2 Data

The ground reality is, in each case, of delineations for significantly cancerous lesions made by one or more radiologists, PI-RADS scores that were assigned to that lesion, and one or more MRI images.

4.3 Techniques

This part of the system expects a Lesion Localization, prostate zonal delineations, and one or more series from, axial T2W series, ADC series, and a high b-value DWI series as input. As mentioned, zonal information is relevant for the PI-RADS score assignment and the choice for these image series is for the same reasons as those mentioned in part two. As part two concerns the detection of clinically significant lesions, not all types of lesions, the only plausible PI-RADS scores to assign are 3, 4, and 5. A lower than 3 is, then, implied when we do not detect a lesion.

We use two types of techniques to perform Lesion Classification. The first is Deep Learning techniques, which take the MRI series and the lesion localization as input and report back to a PI-RADS class. The other consists of an additional step: first so-called 'radiomics' features are extracted, which describe properties, e.g. longest diameter or mean color, of the lesion and the lesion area on the MRI images. These features are then forwarded to traditional machine learning techniques and neural networks (but not deep networks/ Deep Learning) for classification, again resulting in an assigned class per lesion.

5 References

B. Turkbey, A. B. Rosenkrantz, M. A. Haider, A. R. Padhani, G. Villeirs, K. J. Macura, C. M. Tempany, P. L. Choyke, F. Cornud, D. J. Margolis, et al., (2019), "Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2," European urology, vol. 76, no. 3, pp. 340–351, 2019.

Bhavsar, Anil., & Verma, Sadhna. (2014). Anatomic imaging of the prostate. BioMed research international, 2014, 728539. https://doi.org/10.1155/2014/728539.

international reports on socio-informatics