*Hubert Dariusz Zając (2022): It takes a village to raise an AI system - realising AI potential in healthcare. In: Proceedings of the 20th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centred Computing on the Design of Cooperation Technologies - Doctoral Colloquium, Reports of the European Society for Socially Embedded Technologies (ISSN 2510-2591), DOI: 10.48340/ecscw2022_dc02*



# It takes a village to raise an AI system - realising AI potential in healthcare

Hubert Dariusz Zając
University of Copenhagen
*Contact Author: hdz@di.ku.dk*

## Doctoral research overview

The leading question of my PhD is: *How to sustainably realise AI potential in healthcare?* Sustainability here is "the capacity to endure" (Lago et al. (2015)) that specifically considers the relation between social and technical dimensions. I aim to use the AI4XRAY project to investigate the many ways, we can transform AI from a concept that fires the imagination, to a tool or a method that has a real and enduring impact on the collaborative medical work. In order to achieve that, I frame my research with four questions that can be answered within the context and the timeline of the AI4XRAY project. The answers will provide a concrete contribution to the understanding of sustainable AI realisation in healthcare. The four research questions are:

1. *What are the challenges faced during real-world use of AI-based medical systems and how can we mitigate them?* I posed this question at the beginning of my PhD after an initial literature analysis revealed no existing compilations of qualitative accounts of medical AI-based systems usage in real-world settings.

2. *How do domain experts and data scientists negotiate ground truth?* This question took shape after following a highly collaborative process of

negotiation, mutual learning, and co-design of ground truth for the final AI-based system.

3. *What are the work practices of radiologists and clinicians using x-rays in Denmark, Kenya, and Thailand?* And *What are the opportunities for AI to support the work of radiologists and clinicians in chest x-ray assessment in Denmark, Kenya, and Thailand?* This question exploits the core advantage and challenge of the AI4XRAY project, namely creating AI-based software for the diverse collaborative work paradigms of radiologists and clinicians from different countries, cultures, and economies.

The methodological approach I use to answer the above questions is rooted in Participatory Design (PD) (Kensing and Blomberg (2004)), Ethnography, Grounded Theory (Strauss and Corbin (1994)) and Abductive Grounded Theory (Rahmani and Leifels (2018)). The Participatory Design principles guide my work within the project. I am engaging with diverse groups of stakeholders and our development team during all the stages of software development to boost their agency over the outcome, deepen understanding of each others work practices and needs, and eventually co-create a solution. I use ethnographic methods like situated observations and qualitative interviews to help understand the complexity of work in the medical domain. It is crucial for me as a researcher and part of the AI4XRAY group to experience first-hand the collaborative work of medical professionals, learn about different types of knowledge, and dependencies existing in different medical settings. At last, to analyse collected data, I rely on Grounded Theory when exploring new topics and processing recorded materials from a new area. On the other hand, when analysing data from an already established domain, I use existing theories to inform my analysis following the Abductive Grounded Theory practice.

## Current progress

During the first year of my PhD, I focused on three activities. A systematic review of qualitative evaluations of medical AI implementations in the real world, learning about work practices around chest x-ray handling in Danish hospitals, and participating in the collaborative work of ground truth design. In the following, I elaborate on each of these activities.

### Systematic review

We completed a manuscript of a systematic review and submitted it to one of the major HCI journals. The motivation behind the review was to learn about real-world challenges faced when using AI-based systems in complex settings of collaborative medical work. There was no previous compilation of articles of such scope.
Based on the analysis of the included studies, we introduced several grounded

concepts to ease comparison and discussion of AI-related phenomena, which is one of the difficulties afflicting AI development (Girardin and Lathia (2017); Yang et al. (2019); Kayacik et al. (2019)). Among others, we focused on:

1. study outcome to explain the difference between evaluations of systems at various stages of development;

2. development approach - cohesive and discrete - to highlight the different ways an AI model can be developed and incorporated in medical software;

3. a conceptual framework of ten activities of AI medical development to provide a basis for comparison and discussion of AI development process;

4. intended use of medical AI to provide understandable categories capturing the breadth of medical AI applications.

We categorised reported challenges and provided an overview of their origins and consequences. Based on the described categories we synthesised three sociotechnical challenges unique to medical AI that directly affect the work of HCI and CSCW researchers. Solving the right problem, designing the right solution, and balancing authority and accountability. To solve these challenges, we proposed three recommendations:

1. employing existing conceptual frameworks within the targeted domain to deepen our understanding and gain new viewpoints following advises of Blomberg (1993);

2. embracing and exercising HCI methods and practices throughout the entirety of the AI development process;

3. expanding the design space and moving outside of the beaten track, following the discussion by Bjørn and Boulus-Rødje (2015).

Investigating work practices

Belated by the global pandemic of COVID-19, we started ethnographic work at the main hospital in Denmark in the late spring of 2021. The primary focus of that work was to understand the collaborative work that involved the use of chest x-rays and explore which areas of work could benefit from AI support. We aimed to develop an understanding of the work practice of radiologists and ordering clinicians in Denmark to subsequently challenge it and compare through an ethnographic study in collaborating hospitals in Kenya and Thailand.
Due to the specificity of the AI4RAY project, we started from the work of radiologists. We observed five highly specialised thoracic radiologists and three resident radiologists over the course of a week. Each of the observations was extended with a semi-structured qualitative interview. Every consecutive interview was informed by previous observations and interviews. Additionally, we observed

the process of taking x-rays and interviewed three radiographers from the same hospital.
However, had we focused only on radiologists and stopped the activities at that point, we would have observed only a section of the work involving the handling of chest x-rays. We established contact with six clinicians from the same hospital who ordered the highest number of x-rays within a month. All of them worked at one of the following three departments: heart medicine department, thoracic surgery department, intensive care department. We conducted qualitative interviews, informed by the previous activities, to understand the early stages of "x-ray life" and collaboration between the different medical professionals through the medium of x-ray or otherwise.
We learned that the work at the main hospital is highly specialised and doctors employed at that hospital are experts in their domain. It is specialised to such an extent that most of the ordering clinicians we interviewed complete their assessment of taken chest x-rays and take action without waiting for an evaluation from the radiology department. It resulted in the majority of the x-rays becoming a mundane and unwanted part of practice on the radiologists' side. To understand better the relation between specialisation, chest x-rays, and collaboration, we intend to repeat similar work at regional hospitals. This work is currently in progress.

## Designing ground truth

I conducted participative observations of the collaborative work of radiologists and data scientists on designing ground truth. In order to obtain a high-quality medical dataset that can serve as the ground truth for an ML algorithm, our team requires a vast number of high-quality labels linked to radiology reports and chest x-rays. However, obtaining such labels provided by subject matter experts, in this case, radiologists, is a resource-intensive process (Fort (2016)). Moreover, it is often perceived as a necessity, and the articulation work that enables creating labelled datasets is considered banal and obvious by data practitioners (Feinberg (2017)). These viewpoints certainly add to the practice of not documenting work by data scientists (Zhang et al. (2020)), which may further result in making it invisible and impossible to inspect at the later stages of AI development (Muller et al. (2021)).
During the collaboration between radiologists and data scientists, they negotiated and designed a label structure that is currently used by subject matter experts to label a selection of x-rays that is going to be used as ground truth by the future AI. Moreover, we designed and implemented a custom tool for labelling of x-rays and radiological reports, as the number of specialised tools supporting such a process is very low (Chen et al. (2019)). The results of that collaboration will have a direct impact on the final capabilities of the future system and the way it will be used. However, the design decisions taken during this articulation work are severely underreported and underresearched (Muller et al. (2021)).

## Future work

The next steps in my research include unpacking the collaborative process of creating a label structure. I will base this analysis on sensemaking (Weick and Sutcliffe (2005)) and focus on the tensions between the two groups, their goals, and motivations. I plan to expand on the labelling process suggested by Fort (2016), through investigation of the collaborative design process that took place before the act of labelling and its implications for the final system. I will analyse artefacts collected during that process - digital notes and memos from the meetings, email communication, audio recordings of the meetings. I will use the acquired knowledge to inform a series of qualitative interviews with team members working on the ground truth design. To assess the outcome of the collaboration between radiologists and data scientists, I am evaluating the labelling tool through qualitative interviews and observations of labelling performed by a group of external radiologists.
Subsequently, I am planning on deepening my understanding of work practices involving x-rays at Danish hospitals. Simultaneously with the development of the first versions of the AI model, I will shift from an ethnographic approach to participatory design to involve relevant stakeholders in the work on the system. I will introduce early prototypes to their clinical practice to evaluate our assumptions and provide an opportunity for collaborative work on future versions.
Equipped with the knowledge gathered in Danish hospitals, I plan to conduct ethnographic and participatory work in Kenya and Thailand. It is essential for the project to compare the Danish work practices with the ones of radiologists and ordering clinicians in Kenya and Thailand to search for similarities and differences that can result in different needs towards an AI-based prioritisation system. I plan to co-design with relevant stakeholders in the three focal countries to make sure their needs, goals, and concerns are voiced and taken into consideration in the new system.

## Expected contributions

I plan for my PhD to contribute through a series of focused articles and studies to the broad question on how to sustainably realise AI potential in healthcare. The first contribution in the form of a systematic review is conceptualising the many sides of medical AI development and use. It enables HCI and CSCW researchers, data scientists, healthcare professionals, and medical team leaders to gain a unique outlook into commonly faced challenges during medical AI development and use, as well as recommendations on how to tackle them. The second contribution is bringing forward and unpacking the invisible work required to create medical datasets. Through disclosing tensions and motivations occurring during the collaborative process of ground truth design, I plan to directly contribute to the creation of more robust datasets and deepening understanding of the process that takes place before training an AI model. At last a comparison study of the work practices involving x-rays handling in Denmark, Kenya, and Thailand will provide

a unique insight into the collaborative work of medical professionals with x-ray examinations serving as the collaboration medium. This comparison will be further developed into a case study of varying needs towards AI prioritisation systems in radiology.

# References


Bjørn, P. and N. Boulus-Rødje (2015): 'The Multiple Intersecting Sites of Design in CSCW Research'. *Computer Supported Cooperative Work (CSCW)*, vol. 24, pp. 319–351.

Blomberg, J. (1993): 'Ethnographic field methods and their relation to design'. *Participatory design: Principles and practices*, vol. 7, pp. 123–155.

Chen, S., J. Guo, C. Wang, X. Xu, Z. Yi, and W. Li (2019): 'DeepLNAnno: a Web-Based Lung Nodules Annotating System for CT Images'. *Journal of Medical Systems*, vol. 43, no. 7, pp. 197.

Feinberg, M. (2017): 'A design perspective on data'. In: *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2017-May. pp. 2952–2963.

Fort, K. (2016): *Collaborative Annotation for Reliable Natural Language Processing: Technical and Sociological Aspects*.

Girardin, F. and N. Lathia (2017): 'When User Experience Designers Partner with Data Scientists'. In: *AAAI Spring Symposia*.

Kayacik, C., S. Chen, S. Noerly, J. Holbrook, A. Roberts, and D. Eck (2019): 'Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface'. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*.

Kensing, F. and J. L. Blomberg (2004): 'Participatory Design: Issues and Concerns'. *Computer Supported Cooperative Work (CSCW)*, vol. 7, pp. 167–185.

Lago, P., S. A. Koçak, I. Crnkovic, and B. Penzenstadler (2015): 'Framing sustainability as a property of software quality'. *Communications of the ACM*, vol. 58, pp. 70 – 78.

Muller, M., C. T. Wolf, J. Andres, M. Desmond, N. N. Joshi, Z. Ashktorab, A. Sharma, K. Brimijoin, Q. Pan, E. Duesterwald, and C. Dugan (2021): 'Designing Ground Truth and the Social Life of Labels'. pp. 1–16.

Rahmani, F. and K. Leifels (2018): 'Abductive Grounded Theory: a worked example of a study in construction management'. *Construction Management and Economics*, vol. 36, pp. 565 – 583.

Strauss, A. and J. M. Corbin (1994): 'Grounded theory methodology: An overview.'.

Weick, K. E. and K. M. Sutcliffe (2005): 'Organizing and the Process of Sensemaking'.

Yang, Q., J. Cranshaw, S. Amershi, S. T. Iqbal, and J. Teevan (2019): 'Sketching NLP: A Case Study of Exploring the Right Things To Design with Language Intelligence'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*.

Zhang, A. X., M. J. Muller, and D. Wang (2020): 'How do Data Science Workers Collaborate? Roles, Workflows, and Tools'. *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, pp. 1 – 23.