Mitigating issues in Healthcare AI projects

Hubert Dariusz Zając
hdz@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Finn Kensing
kensing@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

ABSTRACT

Artificial Intelligence (AI) promises great prospects. Already today, it is utilised in many areas of society. However, so far, it did not fundamentally improve healthcare software. The accounts of successful AI deployments are sparse, which contrasts with the number of publications describing recent advancements in medical AI software. We reviewed papers that describe medical professionals experiences with AI, users' involvement, and obstacles encountered during the development process. We report on the nomenclature discrepancies and propose a preliminary description of an AI-based medical software development process. We observed that end-users were not involved uniformly across different development stages of the reviewed papers. We suggest how these differences may be linked to some of the observed challenges. Based on the described experiences, we advocate for a more thorough and uniform usage of human-centred and participatory methods during the design, development, and deployment of AI-based Healthcare software.

CCS CONCEPTS

• **Software and its engineering → Software development methods**.

KEYWORDS

Artificial intelligence, AI, development process, user involvement, healthcare informatics

1 INTRODUCTION

Artificial Intelligence-based systems in healthcare are gaining much attention due to their disruptive potential. In the last decade, the number of publications that describe state of the art artificial intelligence (AI) models for healthcare solutions has grown exponentially. However, examples of successful deployment of AI in medical settings remain sparse, as the healthcare sector has yet to implement AI-based systems on a large scale.

One of the reasons for the dramatically low uptake of AI-based systems may be that the creators of such systems fail to address social and organisational aspects of their software [4] Rather developers often seem to take a techno-centric approach to the development, which stands in opposition to the best software engineering practices [8]. Moreover, software developed with disregard for sociotechnical aspects is often doomed to fail [6].

In response to the weaknesses of a narrow techno-centric software development process, Muller and Kuhn [9] presented an approach that expanded the techno-centric understanding, which has become one of the core themes of Participatory Design. Similarly, researchers have reported that software development projects benefit from including sociotechnical considerations [1].

Researchers [5, 10] have pointed out the low number of publications that focus on social aspects of AI implementation in healthcare. Such literature has started to emerge only in recent years and has primarily reported on real-world implementations, medical professionals' experiences, and human-centred onboarding [3, 4, 10, 12–14]. Additionally, the majority of the available studies have only focused on a small subset of a development process [5]. Studies that describe the entirety of a development process [10] are even sparser.

For this position paper, we analysed qualitative studies that: (1) have described any stage of an AI-based medical software development process; and (2) have reported on any sociotechnical activities. We extracted and grouped types of user involvement in the selected studies. This allowed us to derive a preliminary description of a development process of AI-based medical software. It comprises eight relevant stages described in the results section below.

A comparison of users' involvement in the AI-based medical software development projects described by the reviewed studies revealed two under-reported areas. Only two [3, 10] out of six papers reported on involving end-users at the *Problem Assessment* stage. Similarly, only one study [10] considered users' input when developing an ML model. Such disproportions contrast with the consensus regarding the desired degree to which users should be involved in the development process.

In addition to reports on users' involvement, we extracted challenges described in the studies and suggested a link between their origin and decisions made throughout the course of the projects. We observed a correlation between stages that did not include user input and the reported challenges. While we do not report on causality, we did not find similar challenges in projects that reported user involvement. We recommend more thorough user involvement in

CHI'21 Workshop: Realizing AI in Healthcare: Challenges Appearing in the Wild, May 2021, Yokohama, Japan

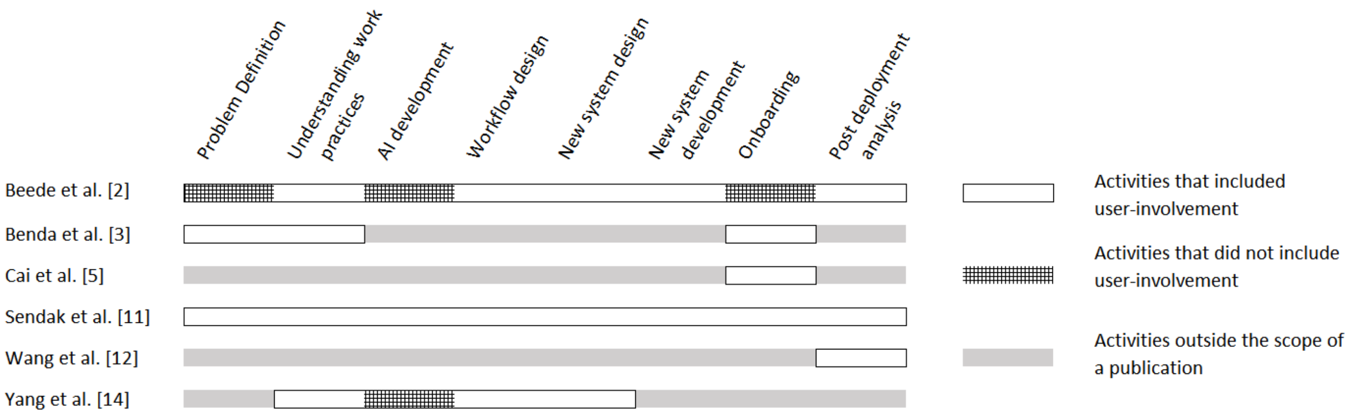

Figure 1: Reported users' involvement in medical AI development.

AI-based healthcare software development – not only to benefit from the best practices and thereby mitigate potential issues, but also to harness the full potential of AI in healthcare.

2 RESULTS

We conducted a manual search for publications describing qualitative studies of AI development projects in a medical setting. We found eight publications describing six distinct AI applications. In order to understand why, how, and when users were involved, we extracted and coded conducted activities, goals behind them, participants, and time of execution. Using the codes, we performed a thematic analysis [7]. It was a crucial step, as many of the authors report on the same activities using different naming conventions and different levels of detail. Based on the discovered themes, we have proposed a preliminary high-level development and implementation process of an AI-based healthcare system.

2.1 Preliminary unified development process

We divided the process into eight development stages. We depicted them in a linear fashion in Figure 1 however, some of the stages overlap, span longer periods, or are revisited throughout the projects. Due to the lack of consensus among researchers regarding the process nomenclature, we have proposed our preliminary division and naming based on the results of the thematic analysis. Authors use their own, on many occasions, internal definitions of phases and activities, which makes them laborious to compare, quantify, and assess across different projects. Our description of each of the stages can be found in Table 1.

Subsequently, we annotated areas of focus for each of the reviewed publications. Stages that were outside the scope of a publication were marked in solid grey. Next, we assessed when in the course of the projects, team members included users' input. All these areas were left white with a black border. Stages that were relevant to the study but did not involve users were marked with a checked pattern. The distribution of users' involvement across various stages can be seen in Figure 1.

2.2 Reported challenges

Next, we screened the selected articles for reported challenges. We coded issues, obstacles, and problems reported by the authors. Subsequently, we performed a thematic analysis [7] to assess, whether the challenges generalised across the projects or were distinct and unique to each of them.

The analysis revealed that the great amount of findings were common across the projects. Similar issues arose despite different domains, backgrounds, implementation sites, and functions. We grouped the coded issues into challenges that can be found in Table 2. For this position paper, we decided to focus on a subset of faced challenges instead of reporting all the encountered issues. We primarily selected challenges preventing the realisation of the AI's full potential within the project's context.

3 DISCUSSION

We have analysed a preliminary selection of articles describing expectations, experiences, and opinions concerning the use of AI-based software in medical practice. The analysis focused on the development process and user involvement. We have observed that (1) users are not adequately involved across all development stages; (2) available studies do not report uniformly on all of the development stages, sometimes leaving out information about types of performed activities. These two points suggest that optimal development processes of AI-based software have not yet become well understood, described, and integrated into best practices among AI researchers and developers.

3.1 AI development and users' involvement

The available studies tend to under-report on the *AI development* stage. In two out of the three articles that described a full development process of an AI system [2, 14], model creation was not described. Instead, the AI models were given and presumably developed independently. We hypothesise that it could have lead to the following challenges described by our colleagues (see Table 2). (1) *AI constraints*; (2) *Processing real-world data*. Both of the challenges were discovered through *Post deployment analysis*, which proves

Stage	Goal
Problem definition	Clarifying requirements. Defining the goal of the project.
Understanding work practices	Establishing thorough understanding of existing work practices, process, systems, regulations etc.
AI development	Designing, developing, and implementing an AI model capable of supporting the project's goal. Testing on prospective data.
Workflow design	Designing new work practices around the software incorporating the AI model.
New system design	Designing new system or integrating the AI model in an existing system. Focusing on usability, UX, UI, and translation of the AI model's outputs.
New system development	Testing integration of the AI model with existing systems or implementing a new one. On-site prospective testing.
Onboarding	Deploying and implementing the new solution in the organisation. Preparing educational materials, establishing processes supporting affected personnel.
Post deployment analysis	Gathering quantitative and qualitative data from the deployment sites to plan further improvements and assess performance in a real-world setting.

Table 1: A preliminary stage model of AI based system development and implementation derived from reviewed literature.

how important it is to continue the studies even after a successful *Onboarding*.

Learning from other software development projects [1], we have so far proposed that more thorough user involvement at the *AI Development* stage could have lead to the discovery and mitigation of these challenges. Involved end-users may have been able to point out quality differences between the training set data and the data processed on-site. Moreover, they could have raised questions concerning the choice of input data. E.g. Sendak et al. [10] reported that the development team informed the list of AI inputs through several qualitative activities involving relevant end-users. Removing some of the constraints imposed on AI could potentially prevent *processing real-world data* challenges from occurring or it could have brought it to the attention of the development team.

3.2 Choosing the right focus

Another, even more crucial, stage of every software development process is *Problem definition*. Sendak et al. [10] mentioned conducting qualitative activities at the very beginning of their project. It was reported that thanks to the gathered information, the project's goal shifted from sepsis detection to sepsis detection and management of the treatment process. It has been discovered that the main problem that physicians were facing was not sepsis detection but following up on the treatment. Missing this subtle distinction could have led to a sub-optimal design of the new system.

On the other hand, in the project described by Beede et al. [2], the problem definition started from a bold target for retina scans set by the Thai government. Subsequently, the authors analysed work practices in clinics that performed retina scans. According to the reported observations, waiting for the first scan reading was the main obstacle to a fast referral and diagnosis. To increase the number of people screened at the clinics, project members decided to develop an AI system that decided whether to refer a patient for a hospital visit or not. However, introducing AI resulted in one additional step in the screening process that had to be performed at the clinics. Together with the *Varying resources* between the clinics and *AI constraints* challenges, the time needed for one patient to be screened not only did not decrease but in many cases increased significantly resulting in, in fact, lower number of screened patients.

We believe that challenging the potential solution, and project member's understanding of the problem through end-user involvement could have highlighted these issues. We can, undoubtedly, see the value of informing the workflow and system design through observations, interviews, and other qualitative activities. We consider that challenging the derived conclusions through another iteration of human-centred activities would result in a more effective and efficient AI-based medical software.

4 CONCLUSIONS

In this position paper we have focused on challenges encountered during and user involvement in medical AI software development processes. We included studies that involved end users, with a special emphasis on real-world applications.

Challenge	Description
Processing real world data	Reported by Beede et al. [2], the assessed algorithm was incapable of assessing approx. 20% of all of the cases due to the input quality not matching training data.
Differences between deployment sites	Benda et al.[3] raised the question of handling varying data formats used at different medical sites. A problem with accessing data from individual systems was reported in the deployed system by Wang et al.[12] Moreover, the differences may be also observed in access to basic utilities that are nowadays taken for granted i.e. Internet. Beede et al. [2] reported that the system's requirement for stable and broadband Internet connection jeopardised the screening process in few of the clinics.
AI constraints	Reported by Beede et al. [2] and Wang et al. [12], medical professionals wanted to provide additional, relevant information, or increase the quality of the input data. However, the AI model was incapable of accepting additional information.
Varying resources	In order for the AI-based system to be fully utilised, it required additional human and technical resources. Wang et al. [12] reported that clinics did not have the work force needed to collect necessary data for the AI to run. Benda et al. [3] described the situation when the output of an AI prediction would be ignored due to lack of funding for the follow up treatment. Beede et al.[2] mentioned significant financial burden imposed on the patients that were falsely diagnosed.
Solving the true problem	Beede et al. [2] reported in the motivation implementation the desire to increase the number of screened patients. However, after implementation of the AI the number of screened patients did not increase.
Trustworthiness	Deemed as required for a successful AI implementation by Sendak et al. [11] It has been found by Wang et al. [12] that low accuracy and inexplicability of results diminishes trust in the system and effectively its usefulness. Cai et al. [5] points out that the complete lack of understanding of AI is deteriorating trust. Similar statement presented Yang et al. [14] reporting that AI models should be validated in clinical studies and the results made available to medical professionals.

Table 2: Challenges reported across the studies.

First, we have reported that the nomenclature and the reporting fidelity varies significantly across the assessed projects. We have advocated for and proposed a preliminary stage model of an AI-based system development process to account for these discrepancies. We believe that unifying naming could support discoverability and enable easier assessment of qualitative development processes.

Second, we have assessed where and when in the studies were users involved. We have used that assessment to suggest a potential link between that lack of consistent users' involvement and the reported challenges. To account for that, we suggest a more thorough use of user-centred activities throughout all of the development stages of medical AI.

REFERENCES

[1] Muneera Bano and Didar Zowghi. 2013. User involvement in software development and system success: A systematic literature review. In *ACM International Conference Proceeding Series*. 125–130. https://doi.org/10.1145/2460999.2461017

[2] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, Laura M Vardoulakis, Fred Hersch Google, Health Google Health, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, Google Health, Laura M Vardoulakis, and Palo Alto. 2020. A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. In *Conference on Human Factors in Computing Systems - Proceedings (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376718

[3] Natalie C Benda, Lala Tanmoy Das, Erika L Abramson, Katherine Blackburn, Amy Thoman, Rainu Kaushal, Yongkang Zhang, and Jessica S Ancker. 2020. "how did you get to this number?" Stakeholder needs for implementing predictive analytics: A pre-implementation qualitative study. *Journal of the American Medical Informatics Association* 27, 5 (5 2020), 709–716. https://doi.org/10.1093/jamia/ocaa021

[4] Carrie J Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. *Conference on Human Factors in Computing Systems - Proceedings* (2 2019), 14. https://doi.org/10.1145/3290605.3300234

[5] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello Ai": Uncovering the onboarding needs of medical practitioners for human–AI collaborative decision-making. https://doi.org/10.1145/3359206

[6] Chris Clegg, Carolyn Axtell, Leela Damodaran, Barbara Farbey, Richard Hull, Raymond Lloyd-Jones, John Nicholls, Reg Sell, and Christine Tomlinson. 1997. Information technology: A study of performance and the role of human and organizational factors. *Ergonomics* 40, 9 (9 1997), 851–871. https://doi.org/10.

1080/001401397187694

[7] Helen Gavin. 2012. Thematic Analysis. In *Understanding Research Methods and Statistics in Psychology*. SAGE Publications Ltd, 273–281. https://doi.org/10.4135/9781446214565.n17

[8] Sari Kujala. 2003. User involvement: A review of the benefits and challenges. *Behaviour and Information Technology* 22, 1 (1 2003), 1–16. https://doi.org/10.1080/01449290301782

[9] Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Commun. ACM* 36, 6 (1993), 24–28.

[10] Mark Sendak, Madeleine Clare Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. 2020. "The Human Body is a Black Box": Supporting Clinical Decision-Making with Deep Learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT '20)*. Association for Computing Machinery, New York, NY, USA, 99–109. https://doi.org/10.1145/3351095.3372827

[11] Mark Sendak, William Ratliff, Dina Sarro, Elizabeth Alderton, Joseph Futoma, Michael Gao, Marshall Nichols, Mike Revoir, Faraz Yashar, Corinne Miller, Kelly Kester, Sahil Sandhu, Kristin Corey, Nathan Brajer, Christelle Tan, Anthony Lin, Tres Brown, Susan Engelbosch, Kevin Anstrom, Madeleine Clare Elish, Katherine Heller, Rebecca Donohoe, Jason Theiling, Eric Poon, Suresh Balu, Armando Bedoya, and Cara O'Brien. 2020. Real-world integration of a sepsis deep learning technology into routine clinical care: Implementation study. *JMIR Medical Informatics* 8, 7 (7 2020), 1–16. https://doi.org/10.2196/15182

[12] Ding Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Dakuo Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural China: Tensions and Challenges in AI-Powered CDSS Deployment. (1 2021). https://doi.org/10.1145/3411764.3445432

[13] Q. Waymel, S. Badr, X. Demondion, A. Cotten, and T. Jacques. 2019. Impact of the rise of artificial intelligence in radiology: What do radiologists think? *Diagnostic and Interventional Imaging* 100, 6 (6 2019), 327–336. https://doi.org/10.1016/j.diii.2019.03.015

[14] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fiting intelligent decision support into critical, clinical decision-making processes. *Conference on Human Factors in Computing Systems - Proceedings* 11 (2019), 1–11. https://doi.org/10.1145/3290605.3300468