Workflow Integration of Artificial Intelligence in Clinical Practice

Doctoral thesis

to obtain a doctorate (PhD)

from the Faculty of Medicine

of the University of Bonn

Katharina Wenderott

from Paderborn

2025

Written with authorization of

the Faculty of Medicine of the University of Bonn

First reviewer: Prof. Dr. Matthias Weigl

Second reviewer: Prof. Dr. Dr. Jens Kleesiek

Day of oral examination: 03.06.2025

From the Institute for Patient Safety

Table of Contents

List of Abbreviations 5

Definitions 6

1. Abstract 10

2. Introduction and Aims with References 11

2.1. Theoretical Background 12

2.1.1. Sociotechnical Systems Perspective – Work System Model 12

2.1.2. Conceptual Model of Workflow Integration 14

2.1.3. Technology Acceptance Model 15

2.2. Empirical Background 16

2.3. Aims 18

2.4. References 19

3. Publications 31

3.1. Publication 1: Integration of Artificial Intelligence into Sociotechnical Work Systems – Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review 32

3.2. Publication 2: Effects of Artificial Intelligence Implementation on Efficiency in Medical Imaging – A Systematic Literature Review and Meta-Analysis 43

3.3. Publication 3: Facilitators and Barriers to Implementing AI in Routine Medical Imaging: Systematic Review and Qualitative Analysis 60

3.4. Publication 4: Prospective Effects of an Artificial Intelligence-Based Computer-Aided Detection System for Prostate Imaging on Routine Workflow and Radiologists' Outcomes 84

3.5. Publication 5: Radiologists' Perspectives on the Workflow Integration of an Artificial Intelligence-Based Computer-Aided Detection System: A Qualitative Study 93

3.6. Publication 6: Workflow Analysis and Evaluation of a Next-Generation Phenotyping Tool: A Qualitative Study of Face2Gene 104

4. Discussion with References **114**

4.1. Strengths and Limitations of this Dissertation 117

4.2. Implications for Practice and Research 119

4.3. Conclusion 120

4.4. References 121

5. Acknowledgements **131**

6. Statement **132**

List of Abbreviations

AI	Artificial Intelligence
CAD	Computer-Aided Detection System
CE	Conformité Européenne
ENA	Epistemic Network Analysis
EU	European Union
FDA	US Food and Drug Administration
HCP	Health Care Professional
ML	Machine Learning
SEIPS	System Engineering Initiative for Patient Safety
SPO	Structure-Process-Outcome
TAM	Technology Acceptance Model

Definitions

AI tool/AI solution	A technology such as a software or device "that relies upon an AI/ML component to serve its purpose" (Cruz Rivera et al., 2020, p. 1353).
Algorithm	"Mathematical model responsible for learning from data and producing an output" (Vasey et al., 2022, p. 930).
Artificial Intelligence	"The ability of computers to perform tasks that normally require human intelligence“ (Ahmad et al., 2021, p. 1).
Barrier	"A barrier is defined as any factor that limits or restricts the integration or use of the AI system" (Wenderott et al., 2022, p. 3).
Clinical Setting	"Relating to the observation and treatment of actual patients rather than in silico or scenario-based simulations" (Vasey et al., 2022, p. 930).
Efficiency	"Resources used in relation to the results achieved. Typical resources include time, human effort, costs and materials" (International Organization for Standardization, 2018). Results in healthcare can be either intermediate (e.g. patients treated, waiting time) or final health outcomes (e.g. lives saved, quality adjusted life years) (Palmer & Torgerson, 1999).
Facilitator	"A facilitator is defined as any factor that promotes or expands the integration or use of the AI system in the workflow" (Wenderott et al., 2022, p. 3).
Healthcare Professionals	They "maintain health in humans through the application of the principles and procedures of evidence-based medicine and caring“ (World Health Organization, 2013, p. 57). „A minimal list would include physicians, nurses, nurse practitioners, physician's assistants, pharmacists, social

	workers, dietitians, physical and occupational therapists, and medical technologists“ (Institute of Medicine, 2009, p. 118).
Human Factors	“The scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and methods to design in order to optimize human well-being and overall system performance" also called Ergonomics (Carayon et al., 2020; International Ergonomics Association, n.d.).
Machine Learning	“A field of computer science concerned with the development of models/algorithms which can solve specific tasks by learning patterns from data, rather than by following explicit rules. It is seen as an approach within the field of AI” (Cruz Rivera et al., 2020, p. 1353).
Medical Devices	“Products with a medical purpose, intended by the manufacturer for human use” (Federal Institue for Drugs and Medical Devices, Germany, n.d.).
Medical Imaging	“Medical imaging is a broad term that encompasses a variety of techniques or processes to create visual representations of the body for diagnosis, treatment, and management of diseases. […] medical imaging includes a variety of other specialties where images are acquired in order to help clinicians better understand the characteristics and mechanisms underlying disease processes that cannot be acquired simply by viewing a patient, analyzing lab reports, or asking them about their symptoms” (Krupinski, 2016, p. 545).
Medical Purposes	“Purposes that are intended to treat, diagnose, cure, mitigate, or prevent disease or other conditions” (U.S. Food and Drug Administration, 2019, p. 2).

Perceived Usefulness	“The degree to which a person believes that using a particular system would enhance his or her job performance" (Davis, 1989, p. 320).
Perceived Ease of Use	“The degree to which a person believes that using a particular system would be free of effort" (Davis, 1989, p. 320).
Routine Care	Actual clinical routine care how patients are being treated without any evaluation ongoing and “in which the decisions made have a direct effect on patient care” (Vasey et al., 2022, p. 930).
Sociotechnical Perspective	“Sociotechnical perspectives focus on understanding the interaction between two interrelated systems, the technical system and the social system, within a particular environmental context” (Whetton & Georgiou, 2010, p. 221).
Stakeholder	“Person or organization that can affect, be affected by, or perceive themselves to be affected by a decision or activity” (International Organization for Standardization, 2018).
Usability	“Extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use” (International Organization for Standardization, 2018).
User	“Person interacting with the AI system to inform their decision-making [or care procedure]. This person could be a healthcare professional or a patient” (Vasey et al., 2022, p. 930).
Workflow	“The flow of people, equipment (including machines and tools), information, and tasks, in different places, at different levels, at different timescales continuously and discontinuously, that are used or required to support the goals of the work domain” (Carayon et al., 2012, p. 509).

Workflow Integration	“The technology is seamlessly incorporated within the work system elements (i.e. people, tasks, other technologies/tools, physical environment, organization) and their interactions over time (i.e. process), specifically considering the temporal order in which work is accomplished and the point in time in which the technology will be used. This includes how the technology fits within the sequence and flow of tasks, people, information, and tools/technologies; at the individual, team, and organizational level; across different scopes of the patient journey” (Salwei et al., 2021, p. 3).
Work System	“The work system represents the context within which organizational members perform their assigned work” (Jasperson et al., 2005, p. 535).
Workload	“The relationship between a group or individual human operator and task demands. […] it is the volume of work expected of a person”(Campbell et al., 2013).

1. Abstract

Artificial intelligence (AI) is increasingly being integrated into healthcare to support clinicians, reduce workloads, and improve workflow efficiency. AI is particularly beneficial in image-based and data-driven medical fields due to its pattern recognition capabilities. While extensive research has been done to explore the potential of AI under experimental conditions, the knowledge about the intricacies of its real-world implementation in clinical settings remains scarce. Consequently, a human factors approach that considers healthcare's complexity as a sociotechnical system is essential.

In this dissertation an examination of the AI integration into clinical workflows is presented, offering a comprehensive view of human-AI interaction in healthcare's complex environment. The theoretical background of this work is the System Engineering Initiative for Patient Safety Model, with the related Conceptual Model of Workflow Integration and the Technology Acceptance Model. Three research projects – one systematic review and two use cases – are presented in this dissertation.

The systematic review involved an assessment of AI's impact on efficiency, clinician outcomes, and workflows in medical imaging, revealing a positive effect on time for tasks and leading to the identification of different AI-augmented workflows. A novel framework was also introduced to categorize the level of AI implementation in the studies. The first considered use case involved an AI implementation in a radiology department, where the AI tool under study was not yet fully integrated into the routine workflow. While the users who participated in the study initially had a positive attitude, the poor fit and longer reading times for complex cases led to workarounds and frustration. The second considered use case, a fully implemented AI tool in human genetics, highlighted that usability and organizational factors were key to successful adoption, as most users incorporated the tool into their daily routines.

The work for this dissertation combined various study designs across different medical specialties. By identifying multiple variations of AI-facilitated clinical workflows, it was emphasized that the context of AI implementation is often unique and AI implementation requires local adaptation. Moreover, recommendations drawing upon identified facilitators and barriers are proposed which should be considered for safe and effective future implementation processes of AI in clinical care.

2. Introduction and Aims with References

Technological advances such as the wide-spread availability of artificial intelligence (AI), such as generative AI, are not only changing our daily lives, but are also being applied across various industries and public sectors, including healthcare. Given the increasing number of patients who are seeking treatment due to an aging society and the corresponding high work demands for health care professionals (HCPs), the healthcare sector is the subject of many promises regarding the introduction of AI (Ahmad et al., 2021; He et al., 2019; Wenderott et al., 2023). The uptake of AI-driven technological solutions in healthcare is associated with many potential benefits, such as the reduction of documentation burden, increase in medication safety, better accessibility of expertise, or advances in personalized medicine (Graafsma et al., 2024; Huang et al., 2024; Lee et al., 2024; van Leeuwen et al., 2021).

In some publications and from opinion leaders the threat that AI could replace clinicians in the future has been expressed, but in more recent work the perspective has been supported that AI will rather augment clinicians in their work (Ahuja, 2019; Langlotz, 2019; Pinto Dos Santos et al., 2019; Topol, 2019; Wong et al., 2019). That AI holds a large potential for augmenting the healthcare sector is also demonstrated by the growing number of patents of AI-enabled medical devices that has been registered in the US and EU as well as a rise in FDA approvals (Aboy et al., 2023; Joshi et al., 2024). While the number of available AI-enabled medical devices is constantly growing, in the past the majority of studies have been focused on the technological features of these solutions, whereas little research has involved the evaluation of their actual impact in clinical practice (Han et al., 2024; Wolff et al., 2021; Yin et al., 2021). The evaluations undertaken have predominantly involved the testing of AI solutions only with retrospective data or in laboratory settings, with the obtained results not necessarily being transferable to the complexities of real-world clinical settings (Han et al., 2024; Park et al., 2020; Widner et al., 2023).

2.1. Theoretical Background

2.1.1. Sociotechnical Systems Perspective – Work System Model

Work in healthcare is characterized by high uncertainty, dynamic changes, and multiple interacting people or elements, therefore creating a highly complex sociotechnical work system (Carayon, 2006; Carayon et al., 2014). To effectively analyze healthcare interventions, such as the introduction of AI in clinical settings, it is crucial to assess their impact within this intricate work system. The scientific discipline which involves studying work systems and their interactions is Human Factors.

A model that is grounded in Human Factors and provides a useful conceptual background for understanding these complex interactions is the work system model of healthcare, also called the Systems Engineering Initiative for Patient Safety (SEIPS) model (Carayon et al., 2006, 2014). In the SEIPS model (Fig. 1) five key elements are identified that form the work system (Carayon et al., 2006, 2014):

1. Person: The individual or group, which is the center of the work system, such as a patient or caregiver.
2. Tasks: The physical or psychological demands that are placed on individuals.
3. Tools and Technologies: Instruments like medical devices or electronic patient records, which are used in the work system.
4. Physical Environment: Factors like lighting or noise in the workplace that affect the individuals.
5. Organization: The hospital or clinic where individuals work or seek care.

These five elements are affected by the external environment, such as regulations or policies. Additionally, the model integrates the Structure-Process-Outcome (SPO) Model of Healthcare Quality (Carayon et al., 2014; Donabedian, 1978). In this sequential model it is illustrated how the work system, through specific processes or workflows like clinical care, results in outcomes for both patients and caregivers (Carayon et al., 2012). Previously, workflows were defined as a linear sequence of tasks aimed at achieving a specific goal. Carayon et al. (2012) expanded this definition to highlight the dynamic interaction of work system elements, describing a workflow as the “flow of people, equipment (including machines and tools), information, and tasks, in different places, at different levels, at different time scales continuously and discontinuously, that are used or required to support the goals of the work domain" (Carayon et al., 2012, p. 509). Additionally, the SEIPS model has also been proven useful in identifying work system barriers and facilitators (Carayon & Perry, 2021; Hoonakker et al., 2013; Hose et al., 2023; Xie et al., 2014), which are factors that hinder or promote the performance or fulfillment of certain tasks or processes with regard to reaching specific goals (Wooldridge et al., 2020). By identifying facilitators and barriers which are associated with different work system elements, a deeper understanding of the system performance can be achieved (Wooldridge et al., 2020). While traditionally, factors that affected a process or outcome were categorized either as a facilitator or barrier, Hoonakker et al. (2017) have proposed a more nuanced view, suggesting that there are different dimensions, instead of fixed categories, where the same factor within a dimension can be both a barrier and a facilitator, depending on the unique implementation context.

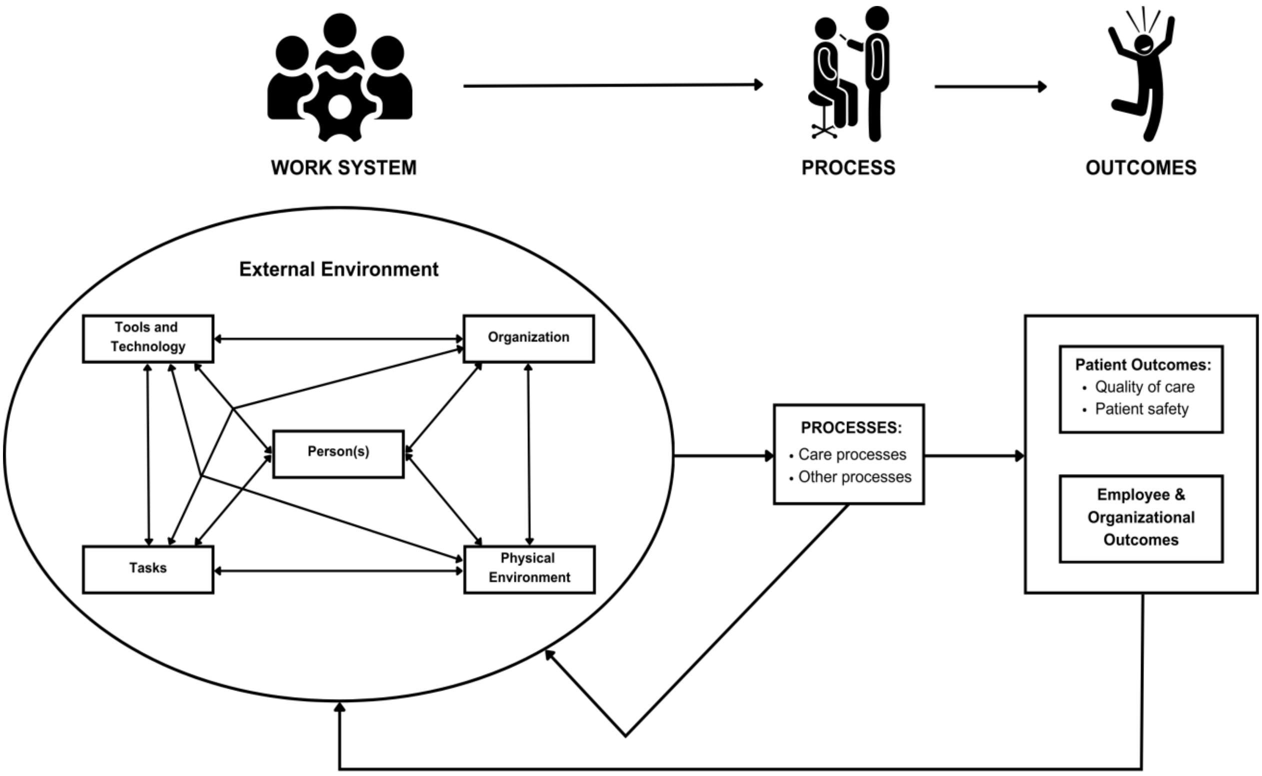


Fig. 1: Work System Model of Healthcare, also called the Systems Engineering Initiative for Patient Safety (SEIPS) Model (Carayon et al., 2006, 2014)

2.1.2. Conceptual Model of Workflow Integration

Based on the SEIPS model, one model that has involved the specific addressing of the integration of novel technologies in healthcare is the Conceptual Model of Workflow Integration (Fig. 2) by Salwei et al. (2021). In this model it is described how the introduction of a technology can reshape the work system, involving the novel technology as a new, distinct element that interacts with the other work system elements. The authors describe workflow integration as "system interactions and how well the new technology fits (or does not fit) within the temporal flow of work (i.e. process). A technology is integrated in the workflow if the system interactions resulting from the introduction of the new technology fit in the flow of work" (Salwei et al., 2021, p. 4). By also integrating the SPO-model, according to Salwei et al. (2021), the workflow integration of a technology results in work system outcomes. For instance, with regard to HCPs, positive outcomes may include increased acceptance of the new technology or reduced workload, whereas negative outcomes could involve workarounds or frustration.

The model by Salwei et al. (2021) has also been applied to the introduction of AI in medical settings, where workflow integration has been identified as a central challenge to the introduction of AI in medical settings (Matheny et al., 2019; Salwei & Carayon, 2022; Wolff

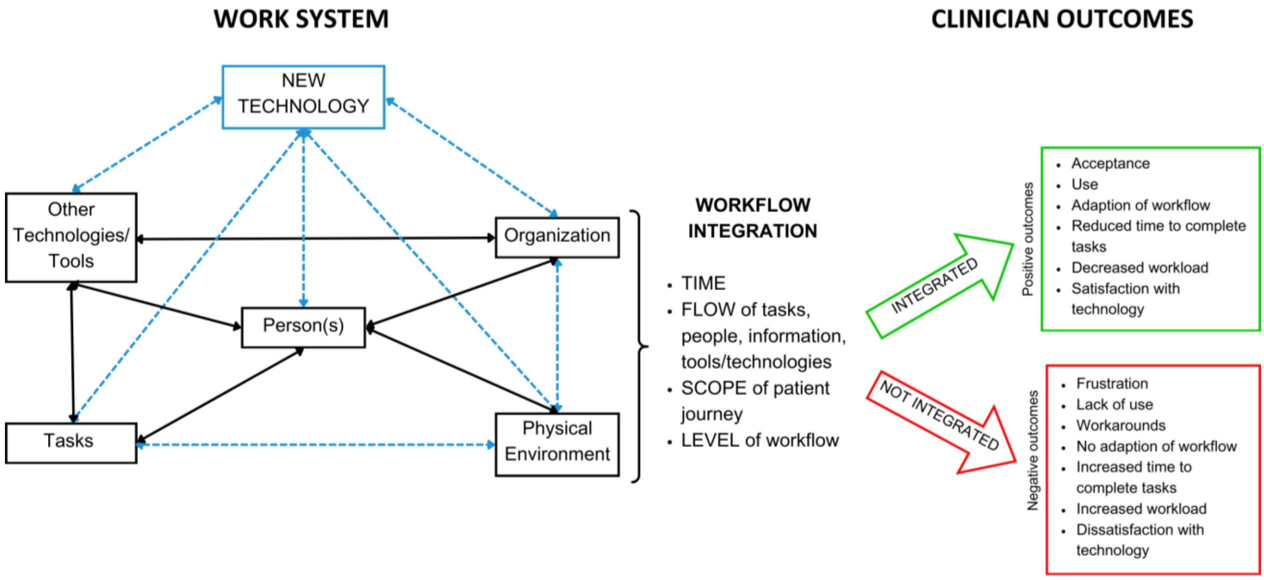


Fig. 2: Conceptual Model of Workflow Integration (Salwei et al., 2021)

et al., 2021). Furthermore, Salwei and Carayon (2022) outlined key sociotechnical considerations when applying the model to AI integration in healthcare:

1. The interaction between AI technology and other work system elements.
2. The integration of AI within the temporal aspects of work.
3. The role of AI in supporting decision-making processes while ensuring transparency and trust.

2.1.3. Technology Acceptance Model

While in the work for this dissertation a systemic perspective has been adopted on AI implementation, relying solely on the SEIPS-based approach seemed insufficient (Salwei et al., 2021; Wenderott, Krups, Luetkens, et al., 2024). Particularly in the context of AI in healthcare, where not all users are willing to accept AI, the individual's decision to adopt and use the technology plays a critical role (Choudhury & Asan, 2022). In the Technology Acceptance Model (TAM), developed by Davis (1989), the importance of this individual-level decision is highlighted, which is essential for the successful implementation of AI. Furthermore, in the TAM (Fig. 3) it is stated that the intention to use a technology and its actual use are influenced by the individual's attitude toward this technology. This attitude consists of the perceived usefulness and the perceived ease of technology use (Davis, 1989). These perceptions are influenced by external variables which have been further defined in the second and third version of the TAM, for example including subjective norms

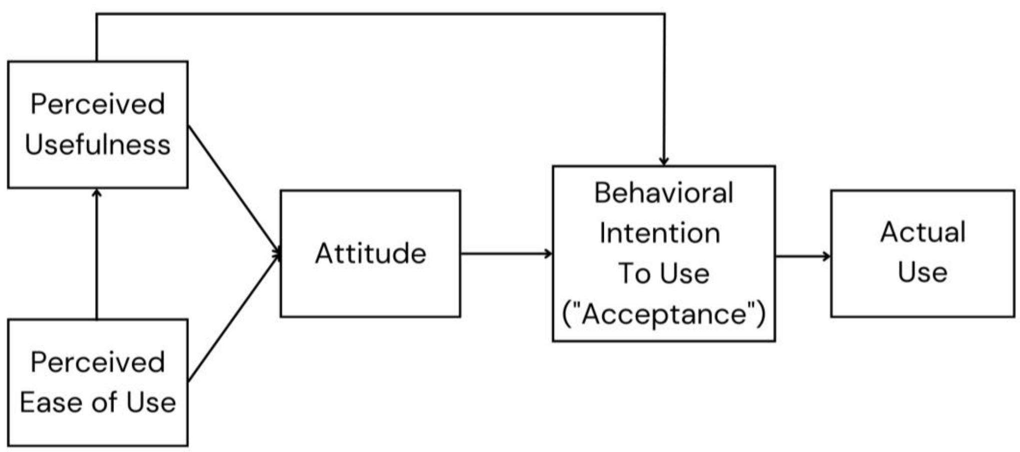


Fig. 3: Technology Acceptance Model (Davis, 1989)

(TAM 2) or computer self-efficacy (TAM 3) (Venkatesh & Bala, 2008; Venkatesh & Davis, 2000). By combining the systemic approach of SEIPS with the individual-focused TAM, a better understanding of both the organizational and individual factors that influence the successful implementation of AI in a healthcare setting can be achieved (Choudhury, 2022).

2.2. Empirical Background

In addition to the theoretical foundation of this dissertation, previous research has shaped and informed its different projects. In 2021, marking the start of this dissertation, only a few AI solutions had been translated into clinical practice (Wolff et al., 2021; Yin et al., 2021). Among the early adopters were medical specialties that rely on image-based data, such as radiology, pathology, and dermatology, as they are particularly suited to AI algorithms due to their ability to analyze patterns in data (Hosny et al., 2018; Tang et al., 2018). Additionally, these specialties face an increase in quantity and quality of medical imaging data, while the number of trained professionals is not rising at the same rate. To alleviate this growing workload AI is hoped to support medical imaging analysis by creating more efficient workflows and reducing reading times (Hosny et al., 2018; van Leeuwen et al., 2021). This potential is also mirrored by the high number of AI algorithms in these specialties that have received FDA clearance (Joshi et al., 2024). As AI technologies have also already been integrated in routine care in these specialties, the research reported in this dissertation was focused on image-based specialties.

At the same time as more AI algorithms for healthcare have been approved for the market, their potential benefits have been highlighted in numerous studies. However, a significant gap remains between research on the technological features or theoretical potential of these solutions and their practical implementation in real-world clinical care (Yin et al., 2021). This has also come to attention of human factors and ergonomics researchers, who emphasize the need for comprehensive, system-based research that addresses the ecological validity, the users, and the clinical environment (Asan & Choudhury, 2021; Lau et al., 2020). A further point underscoring the importance of such evaluations, is the observation that the efficacy of an AI tool depends not only on the accuracy of the algorithm but also on its seamless integration into the local context and smooth fit with its users (Marco-Ruiz et al., 2024). Therefore, the work for this dissertation involved an examination of the integration of AI into routine care from a human factors perspective, aiming for an evaluation beyond technical model performance.

For studying AI integration within the system of healthcare, the study by Strohm et al. (2020) served as a best practice example at the time of this dissertation's planning. They established an interview-based study involving various stakeholders, such as radiologists, data scientists, and managers, to explore the implementation of a specific AI tool in different radiology departments. They provided an in-depth analysis of an AI implementation process, but only conducted interviews after the AI system had been in use, without considering any professionals' perceptions beforehand. Strohm et al. (2020) identified barriers and facilitators for the implementation of this AI system, such as inconsistent trust and acceptance among radiologists or the anticipated benefits of the technology. However, they did not describe how the workflow actually changed during technology implementation, despite this being proposed as a key implementation factor - both as a facilitator (i.e., minimizing workflow changes) and a barrier (i.e., unstructured implementation). The importance of workflow integration has been further emphasized by researchers such as Wolff et al. (2021) and Yin et al. (2021). In combination with the Conceptual Model of Workflow Integration, their research led to the positioning of workflow integration as a central topic in this dissertation and also sparked the research on related facilitators and barriers.

A review of the literature on AI workflow integration revealed that AI can be incorporated in various ways, each with distinct implications for the resulting changes. For instance, reviewing AI-generated information and results may enhance existing tasks or introduce an additional step to the workflow (Agarwal et al., 2024). In image-reading tasks, AI can serve as a gatekeeper, presenting only flagged images to the human reader. Alternatively, AI and human analysis can occur sequentially, with either the AI or the clinician reviewing the data first. In a concurrent design, both the AI and the human reader review the data simultaneously, with the AI's findings being integrated later (Beyer et al., 2007; Dahlblom et al., 2023). These designs can impact reading time and sensitivity, therefore it is important to consider how the AI solution has been implemented into the respective workflow (Dahlblom et al., 2023; Miyake et al., 2013). Reporting guidelines such as DECIDE-AI, CONSORT-AI, and SPIRIT-AI include a detailed description of the intended use and integration of an AI tool in the clinical pathway, aiming to ensure transparency and a high quality of reporting for AI implementation studies (Cruz Rivera et al., 2020; Liu et al., 2020; Vasey et al., 2022). Whereas in some studies detailed descriptions of AI-supported workflows have been provided (e.g. Diao et al., 2022; A. E. Hassan et al., 2023), this is not yet standard. Therefore, in the work for this dissertation, the conditions and effects of different AI-supported workflows were examined and contextualized with the identified facilitators and barriers to AI implementation.

Consequently, the work for this dissertation involved the integration of a human factors perspective, which has a crucial role for the successful bridging of the gap between AI development and implementation in healthcare (The DECIDE-AI Steering Group, 2021). While incorporating human factors early in development is essential, the full impact and consideration of these factors can only be evaluated within actual clinical practice (The DECIDE-AI Steering Group, 2021; Vasey et al., 2022). As these solutions become more widely accessible, there is a growing need for thorough evaluations of AI systems' real-world implementation (Han et al., 2024; Wenderott, Krups, Zaruchas, et al., 2024; Yin et al., 2021).

2.3. Aims

To meet the previously identified need for research, the aim of the work for this dissertation was to investigate the workflow integration of AI in clinical practice. A further aim was to generate a holistic assessment of the human-AI interaction in the complex sociotechnical work system of healthcare. Due to the vast advances of AI within image-based medical specialties as well as to enable a comparison across the conducted studies, the work for this dissertation was focused on this particular area of AI application in healthcare.

To fulfill the overarching aim, the following objectives were defined:

1. Assess and synthesize the current knowledge and literature base on how AI impacts clinical workflows, their efficiency, and clinician outcomes.
2. Prospectively map the changes that occur through the integration of an AI-based application within the work system and determine the effects on healthcare professionals in a clinical use case.
3. Identify barriers and facilitators of AI integration into clinical workflows and derive best practices for future implementation scenarios.

2.4. References

Aboy, M., Price, W. N., & Raker, S. (2023). Mapping the Patent Landscape of Medical Machine Learning. *Nature Biotechnology*, *41*(4), 461–468. https://doi.org/10.1038/s41587-023-01735-6

Agarwal, R., Dugas, M., & Gao, G. (Gordon). (2024). Augmenting Physicians with Artificial Intelligence to Transform Healthcare: Challenges and Opportunities. *Journal of Economics & Management Strategy*, *33*(2), 360–374. https://doi.org/10.1111/jems.12555

Ahmad, Z., Rahim, S., Zubair, M., & Abdul-Ghafar, J. (2021). Artificial Intelligence (AI) in Medicine, Current Applications and Future Role with Special Emphasis on Its Potential and Promise in Pathology: Present and Future Impact, Obstacles Including Costs and Acceptance Among Pathologists, Practical and Philosophical Considerations. A Comprehensive Review. *Diagnostic Pathology*, *16*(1), 24. https://doi.org/10.1186/s13000-021-01085-4

Ahuja, A. S. (2019). The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician. *PeerJ*, *7*, e7702. https://doi.org/10.7717/peerj.7702

Asan, O., & Choudhury, A. (2021). Research Trends in Artificial Intelligence Applications in Human Factors Health Care: Mapping Review. *JMIR Human Factors*, *8*(2), e28236. https://doi.org/10.2196/28236

Beyer, F., Zierott, L., Fallenberg, E. M., Juergens, K. U., Stoeckel, J., Heindel, W., & Wormanns, D. (2007). Comparison of Sensitivity and Reading Time for the Use of Computer-Aided Detection (CAD) of Pulmonary Nodules at MDCT as Concurrent or Second Reader. *European Radiology*, *17*(11), 2941–2947. https://doi.org/10.1007/s00330-007-0667-1

Campbell, T. S., Johnson, J. A., Zernicke, K. A., Shaw, C., Hara, K., Hara, K., Folkman, S., Inada, S., Monaco, J., Weiss, S. M., Pressman, S. D., Kraft, T., Bowlin, S., Hamer, M., Williams, V. P., Williams, R. B., Williams, R. B., O'Hara, M., Newman, J., … Engeland, C. G. (2013). Workload. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 2068–2069). Springer New York. https://doi.org/10.1007/978-1-4419-1005-9_934

Carayon, P. (2006). Human Factors of Complex Sociotechnical Systems. *Applied Ergonomics*, *37*(4), 525–535. https://doi.org/10.1016/j.apergo.2006.04.011

Carayon, P., Cartmill, R., Hoonakker, P., Hundt, A. S., Karsh, B.-T., Krueger, D., Snellman, M., Thuemling, T., & Wetterneck, T. (2012). Human Factors Analysis of Workflow in Health Information Technology Implementation. In *Handbook of human factors and ergonomics in health care and patient safety* (Vol. 2). CRC Press.

Carayon, P., & Perry, S. (2021). Human Factors and Ergonomics Systems Approach to the Covid-19 Healthcare Crisis. *International Journal for Quality in Health Care*, *33*(Supplement_1), 1–3. https://doi.org/10.1093/intqhc/mzaa109

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2006). Work System Design for Patient Safety: The SEIPS Model. *BMJ Quality & Safety*, *15*, i50–i58. https://doi.org/10.1136/qshc.2005.015842

Carayon, P., Wetterneck, T. B., Rivera-Rodriguez, A. J., Hundt, A. S., Hoonakker, P., Holden, R., & Gurses, A. P. (2014). Human Factors Systems Approach to Healthcare Quality and Patient Safety. *Applied Ergonomics*, *45*(1), 14–25. https://doi.org/10.1016/j.apergo.2013.04.023

Carayon, P., Wooldridge, A., Hoonakker, P., Hundt, A. S., & Kelly, M. M. (2020). SEIPS 3.0: Human-Centered Design of the Patient Journey for Patient Safety. *Applied Ergonomics*, *84*, 103033. https://doi.org/10.1016/j.apergo.2019.103033

Choudhury, A. (2022). Toward an Ecologically Valid Conceptual Framework for the Use of Artificial Intelligence in Clinical Settings: Need for Systems Thinking, Accountability, Decision-making, Trust, and Patient Safety Considerations in Safeguarding the Technology and Clinicians. *JMIR Human Factors*, *9*(2), e35421. https://doi.org/10.2196/35421

Choudhury, A., & Asan, O. (2022). Impact of Accountability, Training, and Human Factors on the Use of Artificial Intelligence in Healthcare: Exploring the Perceptions of Healthcare Practitioners in the Us. *Human Factors in Healthcare*, *2*, 100021. https://doi.org/10.1016/j.hfh.2022.100021

Cruz Rivera, S., Liu, X., Chan, A.-W., Denniston, A. K., Calvert, M. J., The SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, Darzi, A., Holmes, C., Yau, C., Moher, D., Ashrafian, H., Deeks, J. J., Ferrante Di Ruffano, L., Faes, L., Keane, P. A., Vollmer, S. J., SPIRIT-AI and CONSORT-AI Consensus Group, Lee, A. Y., … Rowley, S. (2020). Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nature Medicine*, *26*(9), 1351–1363. https://doi.org/10.1038/s41591-020-1037-7

Dahlblom, V., Dustler, M., Tingberg, A., & Zackrisson, S. (2023). Breast Cancer Screening with Digital Breast Tomosynthesis: Comparison of Different Reading Strategies Implementing Artificial Intelligence. *European Radiology*, *33*(5), 3754–3765. https://doi.org/10.1007/s00330-022-09316-y

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Diao, K., Chen, Y., Liu, Y., Chen, B., Li, W., Zhang, L., YL, Q., Zhang, T., Zhang, Y., Wu, M., Li, K., & Song, B. (2022). Diagnostic Study on Clinical Feasibility of an AI-Based Diagnostic System as a Second Reader on Mobile CT Images: A Preliminary Result. *Annals of Translational Medicine*, *10*(12), 668. https://doi.org/10.21037/atm-22-2157

Donabedian, A. (1978). The Quality of Medical Care: Methods for Assessing and Monitoring the Quality of Care for Research and for Quality Assurance Programs. *Science*, *200*(4344), 856–864. https://doi.org/10.1126/science.417400

Federal Institue for Drugs and Medical Devices, Germany. (n.d.). *Medical devices*. Retrieved January 25, 2025, from https://www.bfarm.de/EN/Medical-devices/_artikel.html?nn=708878

Graafsma, J., Murphy, R. M., Van De Garde, E. M. W., Karapinar-Çarkit, F., Derijks, H. J., Hoge, R. H. L., Klopotowska, J. E., & Van Den Bemt, P. M. L. A. (2024). The Use of Artificial Intelligence to Optimize Medication Alerts Generated by Clinical Decision Support Systems: A Scoping Review. *Journal of the American Medical Informatics Association*, *31*(6), 1411–1422. https://doi.org/10.1093/jamia/ocae076

Han, R., Acosta, J. N., Shakeri, Z., Ioannidis, J. P. A., Topol, E. J., & Rajpurkar, P. (2024). Randomised Controlled Trials Evaluating Artificial Intelligence in Clinical Practice: A Scoping Review. *The Lancet Digital Health*, *6*(5), e367–e373. https://doi.org/10.1016/S2589-7500(24)00047-5

Hassan, A. E., Ringheanu, V. M., & Tekle, W. G. (2023). The Implementation of Artificial Intelligence Significantly Reduces Door-in-Door-Out Times in a Primary Care Center Prior to Transfer. *Interventional Neuroradiology*, *29*(6), 631–636. https://doi.org/10.1177/15910199221122848

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The Practical Implementation of Artificial Intelligence Technologies in Medicine. *Nature Medicine*, *25*(1), 30–36. https://doi.org/10.1038/s41591-018-0307-0

Hoonakker, P. L. T., Carayon, P., & Cartmill, R. S. (2017). The Impact of Secure Messaging on Workflow in Primary Care: Results of a Multiple-Case, Multiple-Method Study. *International Journal of Medical Informatics*, *100*, 63–76. https://doi.org/10.1016/j.ijmedinf.2017.01.004

Hoonakker, P. L. T., Carayon, P., McGuire, K., Khunlertkit, A., Wiegmann, D. A., Alyousef, B., Xie, A., & Wood, K. E. (2013). Motivation and Job Satisfaction of Tele-ICUD Nurses. *Journal of Critical Care*, *28*(3), 315.e13-315.e21. https://doi.org/10.1016/j.jcrc.2012.10.001

Hose, B.-Z., Carayon, P., Hoonakker, P. L. T., Brazelton, T. B., Dean, S. M., Eithun, B. L., Kelly, M. M., Kohler, J. E., Ross, J. C., & Rusy, D. A. (2023). Work System Barriers and Facilitators of a Team Health Information Technology. *Applied Ergonomics*, *113*, 104105. https://doi.org/10.1016/j.apergo.2023.104105

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial Intelligence in Radiology. *Nature Reviews Cancer*, *18*(8), Article 8. https://doi.org/10.1038/s41568-018-0016-5

Huang, L., Chen, Z., Yang, Z., & Huang, W. (2024). Advancing Healthcare Accessibility: Fusing Artificial Intelligence with Flexible Sensing to Forge Digital Health Innovations. *BME Frontiers*, *5*, 0062. https://doi.org/10.34133/bmef.0062

Institute of Medicine (with Olsen, L., Goolsby, W. A., & McGinnis, J. M.). (2009). *Leadership Commitments to Improve Value in Health Care: Finding Common Ground: Workshop Summary*. The National Academies Press. https://doi.org/10.17226/11982

International Ergonomics Association. (n.d.). *What Is Ergonomics (HFE)?* Retrieved January 3, 2025, from https://iea.cc/about/what-is-ergonomics/

International Organization for Standardization. (2018). *Ergonomics of Human-System Interaction (ISO Standard No. 9241-11:2018)* (Version 2).

Jasperson, Carter, & Zmud. (2005). A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems. *MIS Quarterly*, *29*(3), 525. https://doi.org/10.2307/25148694

Joshi, G., Jain, A., Araveeti, S. R., Adhikari, S., Garg, H., & Bhandari, M. (2024). FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. *Electronics*, *13*(3), 498. https://doi.org/10.3390/electronics13030498

Krupinski, E. A. (2016). Medical Imaging. In J. Chen, W. Cranton, & M. Fihn (Eds.), *Handbook of Visual Display Technology* (pp. 545–558). Springer International Publishing. https://doi.org/10.1007/978-3-319-14346-0_186

Langlotz, C. P. (2019). Will Artificial Intelligence Replace Radiologists? *Radiology: Artificial Intelligence*, *1*(3), e190058. https://doi.org/10.1148/ryai.2019190058

Lau, N., Hildebrandt, M., & Jeon, M. (2020). Ergonomics in AI: Designing and Interacting With Machine Learning and AI. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *28*(3), 3–3. https://doi.org/10.1177/1064804620915238

Lee, C., Britto, S., & Diwan, K. (2024). Evaluating the Impact of Artificial Intelligence (AI) on Clinical Documentation Efficiency and Accuracy Across Clinical Settings: A Scoping Review. *Cureus*, *16*(11), e73994. https://doi.org/10.7759/cureus.73994

Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., The SPIRIT-AI and CONSORT-AI Working Group, SPIRIT-AI and CONSORT-AI Steering Group, Chan, A.-W., Darzi, A., Holmes, C., Yau, C., Ashrafian, H., Deeks, J. J., Ferrante Di Ruffano, L., Faes, L., Keane, P. A., Vollmer, S. J., SPIRIT-AI and CONSORT-AI Consensus Group, Lee, A. Y., … Rowley, S. (2020). Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nature Medicine*, *26*(9), 1364–1374. https://doi.org/10.1038/s41591-020-1034-x

Marco-Ruiz, L., Hernández, M. Á. T., Ngo, P. D., Makhlysheva, A., Svenning, T. O., Dyb, K., Chomutare, T., Llatas, C. F., Muñoz-Gama, J., & Tayefi, M. (2024). A Multinational Study on Artificial Intelligence Adoption: Clinical Implementers' Perspectives. *International Journal of Medical Informatics*, *184*, 105377. https://doi.org/10.1016/j.ijmedinf.2024.105377

Matheny, M., Israni, S. T., Ahmed, M., & Whichers, D. (Eds.). (2019). *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. National Academy of Medicine.

Miyake, M., Iinuma, G., Taylor, S. A., Halligan, S., Morimoto, T., Ichikawa, T., Tomimatsu, H., Beddoe, G., Sugimura, K., & Arai, Y. (2013). Comparative Performance of a

Primary-Reader and Second-Reader Paradigm of Computer-Aided Detection for CT Colonography in a Low-Prevalence Screening Population. *Japanese Journal of Radiology*, *31*(5), 310–319. https://doi.org/10.1007/s11604-013-0187-7

Palmer, S., & Torgerson, D. J. (1999). Economics Notes: Definitions of Efficiency. *BMJ*, *318*(7191), 1136–1136. https://doi.org/10.1136/bmj.318.7191.1136

Park, Y., Jackson, G. P., Foreman, M. A., Gruen, D., Hu, J., & Das, A. K. (2020). Evaluating Artificial Intelligence in Medicine: Phases of Clinical Research. *JAMIA Open*, *3*(3), 326–331. https://doi.org/10.1093/jamiaopen/ooaa033

Pinto Dos Santos, D., Giese, D., Brodehl, S., Chon, S. H., Staab, W., Kleinert, R., Maintz, D., & Baeßler, B. (2019). Medical Students' Attitude Towards Artificial Intelligence: A Multicentre Survey. *European Radiology*, *29*(4), 1640–1646. https://doi.org/10.1007/s00330-018-5601-1

Salwei, M. E., & Carayon, P. (2022). A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery. *Journal of Cognitive Engineering and Decision Making*, *16*(4), 194–206. https://doi.org/10.1177/15553434221097357

Salwei, M. E., Carayon, P., Hoonakker, P. L. T., Hundt, A. S., Wiegmann, D., Pulia, M., & Patterson, B. W. (2021). Workflow Integration Analysis of a Human Factors-Based Clinical Decision Support in the Emergency Department. *Applied Ergonomics*, *97*, 103498. https://doi.org/10.1016/j.apergo.2021.103498

Strohm, L., Hehakaya, C., Ranschaert, E. R., Boon, W. P. C., & Moors, E. H. M. (2020). Implementation of Artificial Intelligence (AI) Applications in Radiology: Hindering and Facilitating Factors. *European Radiology*, *30*(10), 5525–5532. https://doi.org/10.1007/s00330-020-06946-y

Tang, Tam, R., Cadrin-Chênevert, A., Guest, W., Chong, J., Barfett, J., Chepelev, L., Cairns, R., Mitchell, J. R., Cicero, M. D., Poudrette, M. G., Jaremko, J. L., Reinhold, C., Gallix, B., Gray, B., & Geis, R. (2018). Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Canadian Association of Radiologists Journal = Journal l'Association Canadienne Des Radiologistes*, *69*(2), 120–135. https://doi.org/10.1016/j.carj.2018.02.002

The DECIDE-AI Steering Group. (2021). DECIDE-AI: New Reporting Guidelines to Bridge the Development-to-Implementation Gap in Clinical Artificial Intelligence. *Nature Medicine*, *27*(2), 186–187. https://doi.org/10.1038/s41591-021-01229-5

Topol, E. J. (2019). High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

U.S. Food and Drug Administration (Ed.). (2019). *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback*. https://www.fda.gov/media/122535/download

van Leeuwen, K. G., de Rooij, M., Schalekamp, S., van Ginneken, B., & Rutten, M. J. C. M. (2021). How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatric Radiology*. https://doi.org/10.1007/s00247-021-05114-8

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., Denniston, A. K., Faes, L., Geerts, B., Ibrahim, M., Liu, X., Mateen, B. A., Mathur, P., McCradden, M. D., Morgan, L., Ordish, J., Rogers, C., Saria, S., Ting, D. S. W., … Perkins, Z. B. (2022). Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI. *Nature Medicine*, *28*(5), 924–933. https://doi.org/10.1038/s41591-022-01772-9

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273–315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Wenderott, K., Franz, S., Friedrich, M. G., & Boos, M. (2023). Job Demands at the Patient's Bedside and Their Effects on Stress and Satisfaction of Nurses. *BMJ Open Quality*, *12*(1), e002025. https://doi.org/10.1136/bmjoq-2022-002025

Wenderott, K., Gambashidze, N., & Weigl, M. (2022). Integration of Artificial Intelligence into Sociotechnical Work Systems—Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review. *JMIR Research Protocols*, *11*(12), e40485. https://doi.org/10.2196/40485

Wenderott, K., Krups, J., Luetkens, J. A., & Weigl, M. (2024). Radiologists' Perspectives on the Workflow Integration of an Artificial Intelligence-Based Computer-Aided Detection System: A Qualitative Study. *Applied Ergonomics*, *117*, 104243. https://doi.org/10.1016/j.apergo.2024.104243

Wenderott, K., Krups, J., Zaruchas, F., & Weigl, M. (2024). Effects of Artificial Intelligence Implementation on Efficiency in Medical Imaging—A Systematic Literature Review and Meta-Analysis. *Npj Digital Medicine*, *7*(1), 265. https://doi.org/10.1038/s41746-024-01248-9

Whetton, S., & Georgiou, A. (2010). Conceptual Challenges for Advancing the Socio-Technical Underpinnings of Health Informatics. *The Open Medical Informatics Journal*, *4*(1), 221–224. https://doi.org/10.2174/1874431101004010221

Widner, K., Virmani, S., Krause, J., Nayar, J., Tiwari, R., Pedersen, E. R., Jeji, D., Hammel, N., Matias, Y., Corrado, G. S., Liu, Y., Peng, L., & Webster, D. R. (2023). Lessons Learned from Translating AI from Development to Deployment in Healthcare. *Nature Medicine*, *29*(6), 1304–1306. https://doi.org/10.1038/s41591-023-02293-9

Wolff, J., Pauling, J., Keck, A., & Baumbach, J. (2021). Success Factors of Artificial Intelligence Implementation in Healthcare. *Frontiers in Digital Health*, *3*. https://doi.org/10.3389/fdgth.2021.594971

Wong, S. H., Al-Hasani, H., Alam, Z., & Alam, A. (2019). Artificial Intelligence in Radiology: How Will We Be Affected? *European Radiology*, *29*(1), 141–143. https://doi.org/10.1007/s00330-018-5644-3

Wooldridge, A. R., Carayon, P., Hoonakker, P., Hose, B.-Z., Eithun, B., Brazelton, T., Ross, J., Kohler, J. E., Kelly, M. M., Dean, S. M., Rusy, D., & Gurses, A. P. (2020). Work System Barriers and Facilitators in Inpatient Care Transitions of Pediatric Trauma Patients. *Applied Ergonomics*, *85*, 103059. https://doi.org/10.1016/j.apergo.2020.103059

World Health Organization. (2013). *Transforming and Scaling up Health Professionals' Education and Training: World Health Organization Guidelines 2013*. World Health Organization. https://iris.who.int/handle/10665/93635

Xie, A., Carayon, P., Cartmill, R., Li, Y., Cox, E. D., & Kelly, M. M. (2014). Facilitators and Barriers to Multi-Stakeholder Collaboration in the Redesign of the Family-Centered

Rounds Process. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 175–179. https://doi.org/10.1177/1541931214581037

Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. *Journal of Medical Internet Research*, *23*(4), e25759. https://doi.org/10.2196/25759

3. Publications

This dissertation comprises six publications: Upon submission in February 2025 four were published and two manuscripts under review by the listed journals, these two publications were then accepted in May 2025 and published accordingly. Three of the publications (Chapters 3.1. to 3.3.) are based on a systematic review conducted as part of the work for this dissertation between 2021 and 2024 at the Institute of Patient Safety at the University Hospital Bonn. Two publications (Chapters 3.4. and 3.5.) result from an empirical research project carried out in collaboration with the Department of Radiology at the University Hospital Bonn between 2021 and 2023. The final publication (Chapter 3.6.), is based on a research project conducted from 2023 to 2024 in collaboration with the Institute of Human Genetics at the University Hospital Bonn and the Institute for Genomic Statistics and Bioinformatics at the University of Bonn.

3.1. Publication 1: Integration of Artificial Intelligence into Sociotechnical Work Systems – Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review

Wenderott K, Gambashidze N, Weigl M (2022)
JMIR Research Protocols, 11(12), e40485. doi: 10.2196/40485

Protocol

Integration of Artificial Intelligence Into Sociotechnical Work Systems—Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review

Katharina Wenderott, BSc, MSc; Nikoloz Gambashidze, MSc, MBA, PhD; Matthias Weigl, Prof Dr

Institute for Patient Safety, University Hospital Bonn, Bonn, Germany

Corresponding Author:
Katharina Wenderott, BSc, MSc
Institute for Patient Safety
University Hospital Bonn
Venusberg-Campus 1, Building A 02
Bonn, 53127
Germany
Phone: 49 228287 ext 13781
Email: katharina.wenderott@ukbonn.de

Abstract

Background: When introducing artificial intelligence (AI) into clinical care, one of the main objectives is to improve workflow efficiency because AI-based solutions are expected to take over or support routine tasks.

Objective: This study sought to synthesize the current knowledge base on how the use of AI technologies for medical imaging affects efficiency and what facilitators or barriers moderating the impact of AI implementation have been reported.

Methods: In this systematic literature review, comprehensive literature searches will be performed in relevant electronic databases, including PubMed/MEDLINE, Embase, PsycINFO, Web of Science, IEEE Xplore, and CENTRAL. Studies in English and German published from 2000 onwards will be included. The following inclusion criteria will be applied: empirical studies targeting the workflow integration or adoption of AI-based software in medical imaging used for diagnostic purposes in a health care setting. The efficiency outcomes of interest include workflow adaptation, time to complete tasks, and workload. Two reviewers will independently screen all retrieved records, full-text articles, and extract data. The study's methodological quality will be appraised using suitable tools. The findings will be described qualitatively, and a meta-analysis will be performed, if possible. Furthermore, a narrative synthesis approach that focuses on work system factors affecting the integration of AI technologies reported in eligible studies will be adopted.

Results: This review is anticipated to begin in September 2022 and will be completed in April 2023.

Conclusions: This systematic review and synthesis aims to summarize the existing knowledge on efficiency improvements in medical imaging through the integration of AI into clinical workflows. Moreover, it will extract the facilitators and barriers of the AI implementation process in clinical care settings. Therefore, our findings have implications for future clinical implementation processes of AI-based solutions, with a particular focus on diagnostic procedures. This review is additionally expected to identify research gaps regarding the focus on seamless workflow integration of novel technologies in clinical settings.

Trial Registration: PROSPERO CRD42022303439; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=303439

International Registered Report Identifier (IRRID): PRR1-10.2196/40485

(JMIR Res Protoc 2022;11(12):e40485) doi: 10.2196/40485

KEYWORDS

artificial intelligence; clinical care; clinical efficiency; sociotechnical work system; sociotechnical; review methodology; systematic review; facilitator; barrier; diagnostic; diagnosis; diagnoses; digital health; adoption; implementation; literature review; literature search; search strategy; library science; medical librarian; narrative review; narrative synthesis

Introduction

In medicine, vast changes in patient care because the development of artificial intelligence (AI) is foreseen and ongoing. AI is broadly defined as “the ability of computers to perform tasks that normally require human intelligence” [1]. The introduction of these technologies in medicine promises to improve the quality and safety in health care and accessibility of medical expertise [1]. In the future, AI-human collaboration can augment the ability of clinicians in health care delivery by extracting relevant information from big data sets or performing tasks with higher precision [2,3]. The areas where AI technologies can assist the health care professionals are manifold, for example, clinical diagnostics, decision-making, or health care administration [2,4,5]. These technologies “can be used as powerful tools and partners to enhance, extend, and expand human capabilities, delivering the types of care patients need, at the time and place they need them” [4].

When integrating AI applications into clinical practice, these technologies will become part of highly complex sociotechnical work systems. A model that considers the complexity and scope of the clinical care work environment is the systems engineering initiative for patient safety (SEIPS) 2.0 model [6]. On the basis of SEIPS 2.0, the conceptual model of workflow integration was developed to investigate the integration of a new technology into clinical work processes, which has also been applied to the integration of AI [7,8]. The model uses a sociotechnical system approach and proposes that the whole work system and workflow must be considered to evaluate the success of an AI technology implementation [8].

Some work systems in medicine are faster or more suitable in adopting AI-facilitated technologies. Especially, in specialties that are largely image-based or process big amounts of data, AI is expected to support physicians and improve patient care by leading to more effective and efficient diagnostics [9,10]. Health care providers in image-based medical disciplines handle a growing amount of imaging data that require thorough interpretation [11]. Moreover, the shortage of physicians in radiology and a limited time available per image to meet the current workload are common challenges [12]. The introduction of AI into clinical practice holds a significant potential for changes in clinicians’ duties and improvements such as advancing routine tasks and freeing clinicians’ time for other important tasks [1,2].

One of the main objectives in introducing AI into health care is efficiency improvement because AI is expected to take over not exceedingly complex but time-consuming tasks [1,13,14]. This goal can only be achieved if these technologies are seamlessly integrated into the existing clinical workflow [15]. Therefore, a correlation between workflow integration and usability outcomes, which include efficiency, effectiveness, and satisfaction, has been proposed [7,16]. Efficiency is defined as “resources used in relation to the results achieved. […] Typical resources include time, human effort, costs and materials” [16]. Drawing upon the conceptual model of workflow integration, efficiency-related clinician outcomes include the adaptation of workflow, time to complete tasks, and workload [7,13].

To our knowledge, there is currently no systematic literature review or structured synthesis available on whether the integration of AI into the clinical workflow is associated with improved efficiency. Therefore, comprehensive evidence is necessary, concerning the major promise of freeing physician time for other care activities, for example, direct patient care. As the potential fields of application for AI technologies in health care are diverse, we focus on AI used for medical imaging to enable comparability. In this review, efficiency-related clinician outcomes such as workflow adaptation, time to complete tasks, and workload will be considered. Moreover, reported facilitators or barriers for the successful integration of AI into the workflow will be reviewed as “workflow integration is crucial for making this kind of software [computer-aided detection based on AI] a success” [13].

Our systematic review addresses the following question: how do AI technologies influence the efficiency of workflows in medical imaging?

Specifically, it aims to synthesize the literature base concerning two specific objectives: (1) Identification and overall aggregation of the effects of AI technology implementation on efficiency-related clinician outcomes such as workflow adaptation, time to complete tasks, and clinicians’ workload; and (2) Description of the facilitators and barriers for the integration of AI into the workflow of medical imaging.

Methods

Protocol Registration and Reporting Information

A systematic literature review will be performed to assess the existing literature base and findings. The review’s protocol is registered in the PROSPERO database (registration: CRD42022303439). The protocol and subsequent systematic review follow the reporting guidelines of preferred reporting items for systematic review and meta-analysis protocols statement. The checklist is included in Multimedia Appendix 1.

Eligibility Criteria and Study Design

Only original studies retrieved in full-text and published in peer-reviewed journals will be included. The review will include prospective observational and interventional studies such as randomized controlled trials and nonrandomized studies of interventions, for example, before–after studies and those with an interrupted time series design.

Population

We will include studies conducted in health care facilities such as hospitals, clinics, or outpatient settings using medical imaging. All types of health care professionals, including all age groups, sexes, professions, and qualifications, will be included from the hospital and clinical care settings.

Exposure and Intervention

Studies targeting AI used for medical imaging and its effects on health care professionals interacting with the technology will be eligible for inclusion in this review, including a broad range of AI solutions and clinical work settings. Regarding clinical medical imaging and diagnostics, AI can be defined as “any computer system that can correctly interpret health data, especially in its native form as observed by humans” [17]. AI is often used in this context to identify or forecast a disease state [17]. This review will exclusively focus on AI used for image data interpretation for diagnostic purposes as well as medical imaging [2]. Therefore, our working definition for AI used for medical imaging activities as well as clinical diagnostics in this study will be as follows: any computer system used to interpret imaging data to make a diagnosis, support an image-based clinical (intervention) task, or screen for a disease, a task previously reserved for specialists.

Comparators

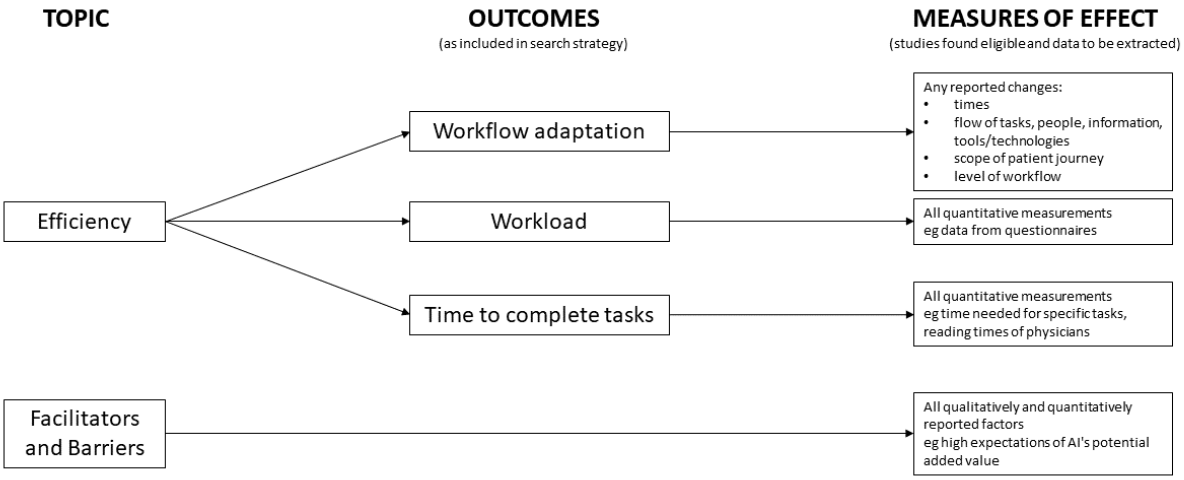
Studies comparing the use of AI in clinical diagnostics and medical imaging with only human specialists will be the comparison of interest; however, it is not a necessary condition for studies to be included in this review.

Outcomes

Overview

Our central study objective is to investigate the impact of AI solutions for clinical diagnostics on the workflow efficiency in clinical care settings. On the basis of our theoretical background, we will focus on three associated outcomes, namely, (1) workflow adaptation, (2) workload, and (3) time-to-complete tasks. In addition, we will systematically assess any facilitators and barriers of AI integration into practice that are mentioned in eligible studies (Figure 1).

Figure 1. Outcomes and measures of effect in this review. AI: artificial intelligence.



Workflow Adaptation

Workflow is defined as “the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules” [18]. This definition was given by the workflow management coalition for business processes but can be also used for clinical contexts [18]. Thus, we will systematically evaluate the adaptation of the workflow in form of any reported changes to the existing processes due to the introduction of an AI technology.

Workload

Workload is defined as “the task demand of accomplishing mission requirements for the human operator” [19,20]. Measuring and analyzing clinical workload is “dependent on the tasks performed, the total time needed to complete the tasks and other care delivery needs of patients” [20,21]. Workload can be measured using objective measures, for example, cases seen or physiological data, and subjective measures such as questionnaires [22]. We will include all forms of quantitative workload measurements that compare the use of an AI software to traditional or previous methods such as pre-existing IT solutions, tools, and technologies in the workplace.

Time to Complete Tasks

New technologies provide opportunities to reduce the time needed to complete tasks, such as the time needed to examine magnetic resonance (MR) or computed tomography images [7,13]. Therefore, we will consider all reported measures on the time-to-task completion or duration. Time to complete tasks will be included if time changes on tasks of interest, such as diagnostic reading of MR images or writing of patient reports, are reported quantitatively with a comparison between the use of AI and traditional methods.

Facilitators and Barriers

A facilitator is defined as any factor that promotes or expands the integration or use of the AI system in the workflow. A barrier is defined as any factor that limits or restricts the integration or use of the AI system. The definitions were developed based on a systematic review by Niezen and Mathijssen [23], and the reported results will be classified according to these definitions. We will extract and synthesize facilitators and barriers in a narrative form using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability framework for novel medical technologies in health care organizations [24].

Publication Types

We will include studies published from January 2000 onward because deep learning was developed in the early 2000s, which is thus marked as the beginning of a new area of AI use in medicine [25]. The article must be in English or German to be eligible for this review.

Owing to our rigorous scope, we limit our review to peer-reviewed journal articles and exclude dissertations, theses, and conference proceedings; as for the latter, the peer-review standards differ across conferences or disciplines. Furthermore, research on AI in medicine not used for medical imaging or diagnostics or research excluding the effects on the work system, such as studies on human interaction with the technology, will not be considered in this review.

Search Strategy

Literature will be retrieved through a structured literature search in several electronic databases: MEDLINE (PubMed), Embase, PsycINFO, Web of Science, IEEE Xplore, and Cochrane Central Register of Controlled Trials. We will use further the snowball method to identify literature not detected through electronic databases, thus screening through the references of identified studies and using Google Scholar. Table 1 outlines the search strategy, following the PICO framework. Because we have decided that comparator is not a necessary condition to be included in this review, we did not list it in the search strategy (see eligibility criteria above). To expand the list of search terms, a preliminary search will be performed before the main search.

Table 1. Search strategy according to the PICO framework.

Classification	Connector	Search term
Population	—[a]	"hospital" OR "clinic" OR "healthcare" OR "health care delivery" OR "clinical care" OR "medical" OR physician* OR clinician* OR doctor* OR nurse* OR "health care professional" OR "patient care" OR patient* OR surg* OR "oncology" OR "radiology" OR "health information"
Intervention	AND	"artificial intelligence" OR "machine intelligence" OR "machine learning" OR "deep learning" OR "neural network" OR "natural language processing" OR "AI " OR "automated image recognition" OR "decision-support" OR "AI application*"
Intervention	AND	"adoption" OR "deploy*" OR "implementation" OR "integration"
Intervention	AND	diagnos* OR "Magnetic Resonance Imaging" OR MRI OR "computer tomography" OR imag* OR detect* OR "data interpretation" OR "information system*" OR "health information technology*" OR "health IT*" OR "medical informatics" OR "electronic health record*" OR "medical record*" OR "patient data"
Outcomes	AND	"workload" OR "work reduction" OR load* OR "cognitive load" OR demand* OR time* OR stress* OR "satisfaction" OR "usability" OR "workflow" OR efficienc* OR "work system" OR "work adaptation" OR "turnaround" OR "clinician outcome" OR "performance"

[a]None.

Screening and Selection Procedure

All retrieved articles will be imported into the software *Zotero*, an open-source reference management software [26]. For title and abstract screening, *Rayyan*, a web application for an initial title and abstract screening, will be used [27,28]. In the first step, the titles and abstracts will be independently screened by 2 reviewers who will undergo training to increase interrater agreement. In case of disagreement, a third researcher from the team will be consulted to solve the conflict in a discussion. If the disagreement cannot be solved through obtaining consensus, the 3 researchers will solve the conflict democratically, that is, majority vote. In the second step, full texts of all eligible publications will be retrieved. These will also be screened by 2 reviewers, and potential conflicts on whether the articles should be included will be resolved in a discussion moderated by a third member of the study team. Studies that are excluded in the process will be recorded. A flow diagram presenting the study selection process will be prepared, following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram for new systematic reviews, which included searches of databases, registers, and other sources [29].

Data Collection Procedure

The study data will be extracted by 1 author and imported into MS Excel (Microsoft Corp). The study data contain details on study characteristics, sample, setting, type of intervention, type and assessment of outcomes, statistical analyses, reported results, moderators or control of confounders, and further information of interest (Textbox 1). The studies and extracted data will be checked at random by another reviewer from the study team. To obtain an agreement on relevant data to be extracted, data from the first 5 studies will be extracted by both reviewers, and a guideline for data extraction will be developed. The extracted data will be divided into several main categories. If any information is missing, the authors of that particular study will be contacted for further details. In case of multiple publications on 1 study, only the key publication will be included.

Textbox 1. Main categories for data to be extracted.

1. Study characteristics
- Authors
- Year of publication
- Location
- Study design

2. Sample
- Sample size
- Participants: demographics and professional characteristics

3. Setting
- Type of clinic
- Medical specialty
- Task

4. Type of intervention
- Artificial intelligence technology (category, reliability, and source)

5. Type and assessment of outcomes
- Workflow adaptation, workload, and times other reported outcome variables
- Facilitators and barriers (if reported)
- Sources of outcomes
- Assessment method (eg, interview, questionnaire, and observation)

6. Statistical analyses
- Types of statistical methods and analyses
- Means and variance metrics of outcomes (eg, standard deviations and confidence intervals)

7. Reported results
- Quantitative results
  - Coefficients (β, γ) and measures of strength of association between artificial intelligence and changes in outcome variables
  - Effect sizes (if reported or calculable)
  - *P* values
- Qualitative results
  - Named facilitators and barriers
  - Any reported analysis

8. Moderators or control of confounders
- Potential moderators or confounding variables (if reported)

9. Further information of potential interest
- Further information, for example, on limitations

Study Appraisal and Risk of Bias (Quality) Assessment

To assess the methodological quality of the included studies, a standardized risk of bias assessment will be performed. Three established tools to assess the risk of bias, applied by two independent reviewers, will be used. Cochrane Risk of Bias Tool (Rob2) [30] will be used for randomized controlled trials. For nonrandomized studies, the risk of bias in nonrandomized studies of interventions tool [31] will be used. These tools address different sources of bias, including the steps from selection to reporting. For observational studies, a checklist of quality of reporting of observational longitudinal research [32] will be used. In case of disagreement, a third reviewer will be consulted until consensus is achieved.

Strategy for Data Synthesis

First, we will qualitatively describe the overall sample and summarize the information extracted from each study. We will then provide an overview concerning the classification in our main categories (Textbox 1). The results of the risk of bias assessment will be provided in a narrative and tabular format. If an adequate set of studies of 5 or more studies is found eligible and the homogeneity level allows, we will perform a meta-analysis that reviews the effects of the introduction of AI on efficiency-associated outcomes. We will quantitatively synthesize data from the retrieved studies using the metafor package in R (R Core Team, R Foundation for Statistical Computing), which contains a set of functions for calculating meta-analyses such as effect-size calculation or model fitting to the data [33]. As we expect a level of heterogeneity of effects in the included studies, a random effects model will be used to estimate the average effect across studies. The heterogeneity across the included studies will be assessed using the Cochran Q test [34] and I^2 statistic [35]. If the number of studies (at least 5 studies per group) and heterogeneity among them allow, subgroup analyses concerning specific characteristics within our eligibility criteria (ie, participants' demographics, particular work settings, outcomes, study designs, and quality) will be performed.

If a meta-analysis is not possible, the results will be summarized in a narrative form and will also be presented in a tabular format. Regardless of the possibility of a meta-analysis, the results will be presented graphically to summarize the retrieved information in a user-friendly manner. We will also adopt a narrative synthesis approach for our additional outcomes, namely, facilitators and barriers. The narrative synthesis will be consistent with that of Strohm et al [36] who conducted an interview study on the factors facilitating and hindering the implementation of AI in radiology. They used the nonadoption, abandonment, scale-up, spread, and sustainability framework for new medical technologies in health care organizations, which will be also used in our data analysis [24,36].

Meta-biases

Regarding the potential sources of meta-bias (eg, publication bias across studies and selective reporting) in the results of the review and meta-analysis, we plan to create a funnel plot, which plots study size against the reported effect size. If a publication bias occurs, the resulting scatterplot is asymmetric with more studies showing a positive than a negative result [37]. We will include at least 10 studies (if possible) to check for small-study effects [38-40]. Additionally, we will use the critical appraisal tool for systematic reviews on randomized or nonrandomized studies of health care interventions AMSTAR-2, which consists of 16 items assessing the quality of conduct of our systematic review [41].

Confidence in Cumulative Evidence

The strength of the body of evidence will be assessed by using the Grading of Recommendations Assessment, Development and Evaluation, a system for rating the quality of evidence and strength of recommendations [42,43]. This rating system has been successfully used in clinical medicine, public health, and policy making, and more recently, in occupational and environmental health [44]. It supports the authors in rating their confidence whether the estimate of an effect is correct. In systematic reviews, the quality of evidence is rated separately for each outcome on a scale from high to very low [45].

Results

The search and screening for the systematic literature review are anticipated to be finished in October 2022. Data extraction, quality appraisal, and subsequent data synthesis will begin in November 2022. The review is expected to be completed by April 2023, and the study results will be published in 2023.

Discussion

Principal Findings

We propose a protocol for a systematic review on the influence of AI technologies on workflow efficiency in clinical care settings. Our review will summarize the existing literature and provide a comprehensive overview on the work system effects of AI technologies in clinical care. This will focus on efficiency outcomes as these are promising factors in the integration of AI into clinical practice. To our knowledge, no systematic overview has been yet conducted on this subject.

The focus in our review will be on workflow and clinician outcomes in imaged-based disciplines as in these fields AI technologies are predominantly and continuously integrated into clinical care practice. Presumably, in the future, almost every medical specialty will interact with AI-based technologies because of a broad range of potential AI application fields in this domain [46]. Contrary to the popular belief that AI will replace radiologists or other health care staff, the future of medicine will rather depend on optimized interactions between AI and humans, enabling AI systems to augment the physician's performance [12,47]. AI is foreseen to change clinicians' work environments and affect their work processes such as task flow and workload [3,46]. Notwithstanding the various promises being proposed with the introduction of AI in real-world care environments, current evidence concerning its effects on clinicians' workflow and practices is missing. Our review will therefore provide valuable insights into the existing evidence base on the immediate effects of AI implementation on work systems and clinician outcomes. Thus, our research synthesis will facilitate understanding if the current AI technologies live up to the expectation of significantly supporting clinicians in their work [48].

Comparison to Previous Research

Notably, in light of the current gap between the broad utilization of AI for research purposes and few AI applications being applied in routine patient care, facilitating AI implementation and adoption into clinical care has become essential. Although academic publications on AI solutions for medical imaging, diagnostic, and therapeutic contexts are numerous, only a few real-world solutions have been yet officially approved and implemented in the health care sector [49]. Furthermore, we expect that only a fraction of these solutions has been systematically evaluated regarding their impact on clinician outcomes or workflow integration. This expectation is supported by the review of Asan and Choudhury [3] who demanded systematic research that addresses AI's impact on clinical workflow and usability with emphasis on the importance of human factor research.

Because the seamless integration of AI is crucial for unfolding its potential in clinical practice, our review will specifically address the facilitators and barriers of implementation practices elicited from the retrieved studies [13]. The consideration of facilitating and hindering factors of AI adoption is an essential step in gaining a more detailed understanding of how AI implementation can be optimized in hospital and other clinical care settings. A study suggests that various process factors affect seamless AI adoption into hospital practices, such as a perceived high added value or hospital-wide innovation strategies, technical performance, and well-structured implementation processes [36]. We acknowledge that we will only extract the process characteristics from studies found eligible regarding clinician outcomes. Nevertheless, our synthesis approach, which draws upon a previously established framework, allows for a comprehensive understanding of AI implementation experiences and will expand the existing preliminary findings [36].

Limitations

Our review will focus on AI used for medical imaging used for diagnostic purposes. AI applications offer a great potential for image-based specialties and address a pressing issue, namely, the vastly growing amount of imaging data that need thorough interpretation [15,47]. Significant technological advancements have been made recently through the development of AI solutions and their application into clinical practice [1,50,51]. We solely focus on this clinical domain and a specific clinical task (eg, image-based tasks and diagnostics) to strengthen internal and external validity as well as to allow comparability across the work settings included. Nonetheless, we capture a medical field with the most extensive availability of AI technologies already integrated into clinical routine practices.

The algorithms or features used in the AI technologies included in this review might be different; however, this is not of central interest for answering our research question. We will not assess the quality or clinical effectiveness of the AI systems because this is covered by numerous systematic reviews with regard to the specific task for which comparable AI solutions were developed, such as in the reviews by Kunze et al [52] or Chidambaram et al [53]. Therefore, no specific conclusions regarding the technologies or characteristics of AI will be drawn as we will focus solely on the work system effects.

To achieve our goal of summarizing the existing literature on the impact of AI implementation on clinician outcomes, we will establish a rigorous list of exclusion criteria regarding study design, setting, and population. Therefore, conclusions will only be drawn for the specific setting of work environments where AI is used for image-based and diagnostic purposes. We acknowledge that this may result in limited generalizability of our results. In future research, it would be valuable to compare the workflow integration of AI across different health care settings such as ambulant care settings or nursing facilities. Our review approach may be an exemplary approach on how to systematically aggregate research findings on AI workflow integration, which can be transferred to other health care sectors and clinical domains.

Our outcome variables of interest draw upon the conceptual model of workflow integration [7]. Our key focus will be on clinician outcomes, workflow, and efficiency—the key issues for AI introduction. Notably, we will only address clinician outcomes named in the model, namely, those related to workload and efficiency. For future research, it would be valuable to include further outcomes such as perceived use and acceptance. Furthermore, it would be interesting to augment research with concepts such as trust and technology characteristics as these are important determinants of AI adoption [36,54,55].

Regarding our key concepts extracted from the conceptual model of workflow integration [7], there is substantial heterogeneity of applicable terms in the literature; for example, time to complete tasks is a collective term for measures such as physician's reading times [13] or time undertaken to review an image [56]. Moreover, some concepts used in this literature review, such as the use of AI in clinical diagnostics or facilitators and barriers for AI implementation, do not have a consistent definition in the literature. Therefore, we propose working definitions on the background of existing research [2,19,23]. Nonetheless, we acknowledge that key terms might be conceived differently in other contexts or publications. Thus, we limited the deviation from previous studies by conducting a pilot search and expanding our search terms to include common variants of key concepts.

Conclusions

Our review and meta-analysis or systematic narrative data analysis will allow first systematic conclusions on how AI for medical and diagnostic imaging affects clinician efficiency outcomes. We expect to provide a structured overview and systematic synthesis of the current literature. Thus, the findings of our review are expected to expand the existing knowledge on how AI affects clinical efficiency in medical imaging. Particularly, by providing a quality appraisal of the included studies, we will identify shortcomings of the current research. Moreover, our review will help to recognize research gaps regarding the seamless workflow integration of novel technologies into clinical settings. Our findings will eventually also provide guidance on provider-centered design and application of AI-based solutions in clinical settings, with potential improvements in clinical safety and performance. Furthermore, our consideration of the facilitators and barriers of AI implementation will provide an evidence-based foundation for hospital leadership and practitioners to successfully manage AI implementation in patient care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-P Checklist.
[PDF File (Adobe PDF File), 411 KB-Multimedia Appendix 1]

References

1. Ahmad Z, Rahim S, Zubair M, Abdul-Ghafar J. Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. Diagn Pathol 2021 Mar 17;16(1):24 [FREE Full text] [doi: 10.1186/s13000-021-01085-4] [Medline: 33731170]
2. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019 Jan;25(1):30-36 [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
3. Asan O, Choudhury A. Research trends in artificial intelligence applications in human factors health care: mapping review. JMIR Hum Factors 2021 Jun 18;8(2):e28236 [FREE Full text] [doi: 10.2196/28236] [Medline: 34142968]
4. Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc Manage Forum 2020 Jan 24;33(1):10-18. [doi: 10.1177/0840470419873123] [Medline: 31550922]
5. Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. J R Soc Med 2019 Jan;112(1):22-28 [FREE Full text] [doi: 10.1177/0141076818815510] [Medline: 30507284]
6. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
7. Salwei ME, Carayon P, Hoonakker PL, Hundt AS, Wiegmann D, Pulia M, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. Appl Ergon 2021 Nov;97:103498 [FREE Full text] [doi: 10.1016/j.apergo.2021.103498] [Medline: 34182430]
8. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak 2022 May 11;16(4):194-206. [doi: 10.1177/15553434221097357]
9. Maassen O, Fritsch S, Palm J, Deffge S, Kunze J, Marx G, et al. Future medical artificial intelligence application requirements and expectations of physicians in German University Hospitals: web-based survey. J Med Internet Res 2021 Mar 05;23(3):e26646 [FREE Full text] [doi: 10.2196/26646] [Medline: 33666563]
10. Pumplun L, Fecho M, Wahl N, Peters F, Buxmann P. Adoption of machine learning systems for medical diagnostics in clinics: qualitative interview study. J Med Internet Res 2021 Oct 15;23(10):e29301 [FREE Full text] [doi: 10.2196/29301] [Medline: 34652275]
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016 Feb;278(2):563-577 [FREE Full text] [doi: 10.1148/radiol.2015151169] [Medline: 26579733]
12. Hainc N, Federau C, Stieltjes B, Blatow M, Bink A, Stippich C. The bright, artificial intelligence-augmented future of neuroimaging reading. Front Neurol 2017 Sep 21;8:489 [FREE Full text] [doi: 10.3389/fneur.2017.00489] [Medline: 28983278]
13. van Leeuwen KG, de Rooij M, Schalekamp S, van Ginneken B, Rutten MJCM. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr Radiol 2022 Oct 12;52(11):2087-2093 [FREE Full text] [doi: 10.1007/s00247-021-05114-8] [Medline: 34117522]
14. Hazarika I. Artificial intelligence: opportunities and implications for the health workforce. Int Health 2020 Jul 01;12(4):241-245 [FREE Full text] [doi: 10.1093/inthealth/ihaa007] [Medline: 32300794]
15. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
16. ISO 9241-11 ergonomics of human-system interaction: part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2022-11-11]
17. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med 2019 Nov 19;11(1):70 [FREE Full text] [doi: 10.1186/s13073-019-0689-8] [Medline: 31744524]
18. Mueller ML, Ganslandt T, Frankewitsch T, Krieglstein CF, Senninger N, Prokosch HU. Workflow analysis and evidence-based medicine: towards integration of knowledge-based functions in hospital information systems. Proc AMIA Symp 1999:330-334 [FREE Full text] [Medline: 10566375]
19. Hooey BL, Kaber DB, Adams JA, Fong TW, Gore BF. The underpinnings of workload in unmanned vehicle systems. IEEE Trans. Human-Mach Syst 2018 Oct;48(5):452-467. [doi: 10.1109/thms.2017.2759758]
20. Fishbein D, Nambiar S, McKenzie K, Mayorga M, Miller K, Tran K, et al. Objective measures of workload in healthcare: a narrative review. IJHCQA 2019 Dec 20;33(1):1-17. [doi: 10.1108/ijhcqa-12-2018-0288]

21. Swiger PA, Vance DE, Patrician PA. Nursing workload in the acute-care setting: a concept analysis of nursing workload. Nurs Outlook 2016;64(3):244-254. [doi: 10.1016/j.outlook.2016.01.003] [Medline: 26944266]
22. Byrne A. Measurement of mental workload in clinical medicine: a review study. Anesth Pain 2011 Sep 21;1(2):90-94. [doi: 10.5812/aapm.2045]
23. Niezen MG, Mathijssen JJ. Reframing professional boundaries in healthcare: a systematic review of facilitators and barriers to task reallocation from the domain of medicine to the nursing domain. Health Policy 2014 Aug;117(2):151-169 [FREE Full text] [doi: 10.1016/j.healthpol.2014.04.016] [Medline: 24857559]
24. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
25. Kaul V, Enslin S, Gross SA. History of artificial intelligence in medicine. Gastrointest Endosc 2020 Oct;92(4):807-812. [doi: 10.1016/j.gie.2020.06.040] [Medline: 32565184]
26. Zotero. Vienna, Virginia: Corporation for Digital Scholarship URL: https://www.zotero.org/ [accessed 2022-11-11]
27. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
28. Rayyan. USA: Rayyan Systems Inc URL: https://www.rayyan.ai/ [accessed 2022-11-11]
29. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
32. Tooth L, Ware R, Bain C, Purdie DM, Dobson A. Quality of reporting of observational longitudinal research. Am J Epidemiol 2005 Feb 01;161(3):280-288. [doi: 10.1093/aje/kwi042] [Medline: 15671260]
33. Viechtbauer W. Conducting meta-analyses in with the metafor package. J Stat Soft 2010;36(3):1-48. [doi: 10.18637/jss.v036.i03]
34. Cochran WG. The combination of estimates from different experiments. Biometrics 1954 Mar;10(1):101. [doi: 10.2307/3001666]
35. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
36. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol 2020 Oct 26;30(10):5525-5532 [FREE Full text] [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
37. Lee W, Hotopf M. Critical appraisal: reviewing scientific evidence and reading academic papers. In: Wright P, Stern J, Phelan M, editors. Core Psychiatry. Edinburgh: Elsevier; 2012:131-142.
38. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis. J Clin Epidemiol 2000 Nov;53(11):1119-1129. [doi: 10.1016/s0895-4356(00)00242-0]
39. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
40. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011 Jul 22;343(jul22 1):d4002-d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]
41. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
42. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
43. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
44. Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghersi D, et al. GRADE: assessing the quality of evidence in environmental and occupational health. Environ Int 2016 Jul;92-93:611-616 [FREE Full text] [doi: 10.1016/j.envint.2016.01.004] [Medline: 26827182]
45. Schumemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. XX: XX; 2013.

46. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
47. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7:e7702 [FREE Full text] [doi: 10.7717/peerj.7702] [Medline: 31592346]
48. Coppola F, Faggioni L, Regge D, Giovagnoni A, Golfieri R, Bibbolino C, et al. Artificial intelligence: radiologists' expectations and opinions gleaned from a nationwide online survey. Radiol Med 2021 Jan 29;126(1):63-71. [doi: 10.1007/s11547-020-01205-y] [Medline: 32350797]
49. Wolff J, Pauling J, Keck A, Baumbach J. Success factors of artificial intelligence implementation in healthcare. Front Digit Health 2021 Jun 16;3:594971 [FREE Full text] [doi: 10.3389/fdgth.2021.594971] [Medline: 34713083]
50. Wong ST. Is pathology prepared for the adoption of artificial intelligence? Cancer Cytopathol 2018 Jun 17;126(6):373-375 [FREE Full text] [doi: 10.1002/cncy.21994] [Medline: 29663732]
51. Liu X, Gao K, Liu B, Pan C, Liang K, Yan L, et al. Advances in deep learning-based medical image analysis. Health Data Sci 2021 Jun 16;2021:1-14. [doi: 10.34133/2021/8786793]
52. Kunze KN, Rossi DM, White GM, Karhade AV, Deng J, Williams BT, et al. Diagnostic performance of artificial intelligence for detection of anterior cruciate ligament and meniscus tears: a systematic review. Arthroscopy 2021 Feb;37(2):771-781. [doi: 10.1016/j.arthro.2020.09.012] [Medline: 32956803]
53. Chidambaram S, Sounderajah V, Maynard N, Markar SR. Diagnostic performance of artificial intelligence-centred systems in the diagnosis and postoperative surveillance of upper gastrointestinal malignancies using computed tomography imaging: a systematic review and meta-analysis of diagnostic accuracy. Ann Surg Oncol 2022 Mar 11;29(3):1977-1990 [FREE Full text] [doi: 10.1245/s10434-021-10882-6] [Medline: 34762214]
54. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
55. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2018 Mar 19;294(1-2):567-592. [doi: 10.1007/s10479-018-2818-y]
56. Hains I, Georgiou A, Westbrook J. The impact of PACS on clinician work practices in the intensive care unit: a systematic review of the literature. J Am Med Inform Assoc 2012;19(4):506-513 [FREE Full text] [doi: 10.1136/amiajnl-2011-000422] [Medline: 22323392]

Abbreviations

AI: artificial intelligence
SEIPS: systems engineering initiative for patient safety
MR: magnetic resonance
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Leung; submitted 23.06.22; peer-reviewed by P Sockolow, L Weinert; comments to author 26.08.22; revised version received 16.09.22; accepted 20.10.22; published 01.12.22

Please cite as:
Wenderott K, Gambashidze N, Weigl M
Integration of Artificial Intelligence Into Sociotechnical Work Systems—Effects of Artificial Intelligence Solutions in Medical Imaging on Clinical Efficiency: Protocol for a Systematic Literature Review
JMIR Res Protoc 2022;11(12):e40485
URL: https://www.researchprotocols.org/2022/12/e40485
doi: 10.2196/40485
PMID:

3.2. Publication 2: Effects of Artificial Intelligence Implementation on Efficiency in Medical Imaging – A Systematic Literature Review and Meta-Analysis

Wenderott K, Krups J, Zaruchas F, Weigl M (2024)
npj Digital Medicine, 7(1), 265. doi: 10.1038/s41746-024-01248-9

The supplementary material can be accessed via this link.

npj | digital medicine

Review article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01248-9

Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis

Check for updates

Katharina Wenderott, Jim Krups, Fiona Zaruchas & Matthias Weigl

In healthcare, integration of artificial intelligence (AI) holds strong promise for facilitating clinicians' work, especially in clinical imaging. We aimed to assess the impact of AI implementation for medical imaging on efficiency in real-world clinical workflows and conducted a systematic review searching six medical databases. Two reviewers double-screened all records. Eligible records were evaluated for methodological quality. The outcomes of interest were workflow adaptation due to AI implementation, changes in time for tasks, and clinician workload. After screening 13,756 records, we identified 48 original studies to be incuded in the review. Thirty-three studies measured time for tasks, with 67% reporting reductions. Yet, three separate meta-analyses of 12 studies did not show significant effects after AI implementation. We identified five different workflows adapting to AI use. Most commonly, AI served as a secondary reader for detection tasks. Alternatively, AI was used as the primary reader for identifying positive cases, resulting in reorganizing worklists or issuing alerts. Only three studies scrutinized workload calculations based on the time saved through AI use. This systematic review and meta-analysis represents an assessment of the efficiency improvements offered by AI applications in real-world clinical imaging, predominantly revealing enhancements across the studies. However, considerable heterogeneity in available studies renders robust inferences regarding overall effectiveness in imaging tasks. Further work is needed on standardized reporting, evaluation of system integration, and real-world data collection to better understand the technological advances of AI in real-world healthcare workflows. Systematic review registration: Prospero ID CRD42022303439, International Registered Report Identifier (IRRID): RR2-10.2196/40485.

With a rising number of patients and limited staff available, the need for changes in healthcare is a pressing issue[1]. Artificial intelligence (AI) technologies promise to alleviate the current burden by taking over routine tasks, such as monitoring patients, documenting care tasks, providing decision support, and prioritizing patients by analyzing clinical data[2,3]. AI-facilitated innovations are claimed to significantly reduce the workload of healthcare professionals[4,5].

Several medical specialties have already introduced AI into their routine work, particularly in data-intensive domains, such as genomics, pathology, and radiology[4]. In particular, image-based disciplines have seen substantial benefits from the pattern recognition abilities of AI, positioning them at the forefront of AI integration in clinical care[3,6]. AI technologies expedite the processing of an increasing number of medical images, being used for detecting artifacts, malignant cells or other suspicious structures, and optionally for the succeeding prioritization of patients[7–9].

To successfully adopt AI in everyday clinical practice, different ways for effective workflow integration can be conceived, largely depending on the specific aim, that is, enhancing the quality of diagnosis, providing reinsurance, or reducing human workload[10,11]. Efficiency outcomes related to AI implementation include shorter reading times or a reduced workload of clinicians to meet the growing demand for interpreting an increasing number of images[12–14]. Thus, whether AI fulfills these aims and enables higher efficiency in everyday clinical work remains largely unknown.

Institute for Patient Safety, University Hospital Bonn, Bonn, Germany. e-mail: katharina.wenderott@ukbonn.de

Healthcare systems are complex, combining various components and stakeholders that interact with each other[15]. While the success of AI technology implementation highly depends on the setting, processes, and users, current studies largely focus on the technical features and capabilities of AI, not on its actual implementation and consequences in the clinical landscape[2,3,6,16,17]. Therefore, this systematic review aimed to examine the influence of AI technologies on workflow efficiency in medical imaging tasks within real-world clinical care settings to account for effects that stem from the complex and everyday demands in real-world clinical care, all not being existent in experimental and laboratory settings[18].

Results

Study selection

We identified 22,684 records in databases and an additional 295 articles through backward search. After the removal of duplicates, the 13,756 remaining records were included in the title/abstract screening. Then, 207 full texts were screened, of which 159 were excluded primarily because of inadequate study designs or not focusing on AI for interpreting imaging data (Supplementary Table 1). Finally, 48 studies were included in the review and data extraction. Twelve studies underwent additional meta-analyses. A PRISMA flow chart is presented in Fig. 1.

Study characteristics

Of the 48 extracted studies, 30 (62.5%) were performed in a single institution, whereas the 18 (37.5%) remaining studies were multicenter studies. One study was published in 2010, another in 2012, and all other included studies were published from 2018 onward. Research was mainly conducted in North America ($n = 21$), Europe ($n = 12$), Asia ($n = 11$), and Australia ($n = 3$). Furthermore, one study was conducted across continents. The included studies were stemming from the medical departments of radiology ($n = 26$), gastroenterology ($n = 6$), oncology ($n = 4$), emergency medicine ($n = 4$), ophthalmology ($n = 4$), human genetics ($n = 1$), nephrology ($n = 1$), neurology ($n = 1$), and pathology ($n = 1$). Most studies used computed tomography (CT) for imaging, followed by X-ray and colonoscopes. The most prominent indications were intracranial hemorrhage, followed by pulmonary embolism, and cancer screening. Table 1 presents the key characteristics of all included studies.

Concerning the purpose of using AI tools in clinical work, we classified the studies into three main categories. First, five studies (10.4%) described an AI tool used for segmentation tasks (e.g., determining the boundaries or volume of an organ). Second, 25 studies (52.1%) used AI tools to examine detection tasks to identify suspicious cancer nodules or fractures. Third, 18 studies (37.5%) investigated the prioritization of patients according to AI-detected critical features (e.g., reprioritizing the worklist or notifying the treating clinician via an alert).

Regarding the AI tools described in the studies, 34 studies (70.8%) focused on commercially available solutions (Table 2). Only Pierce et al. did not specify which commercially available algorithm was used[19]. Thirteen studies (27.1%) used non-commercially available algorithms, detailed information on these algorithms is provided in Table 3. Different measures were used to evaluate the accuracy of these AI tools, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the curve (AUC). Sensitivity and specificity were the most commonly reported measures (see Tables 2 and 3).

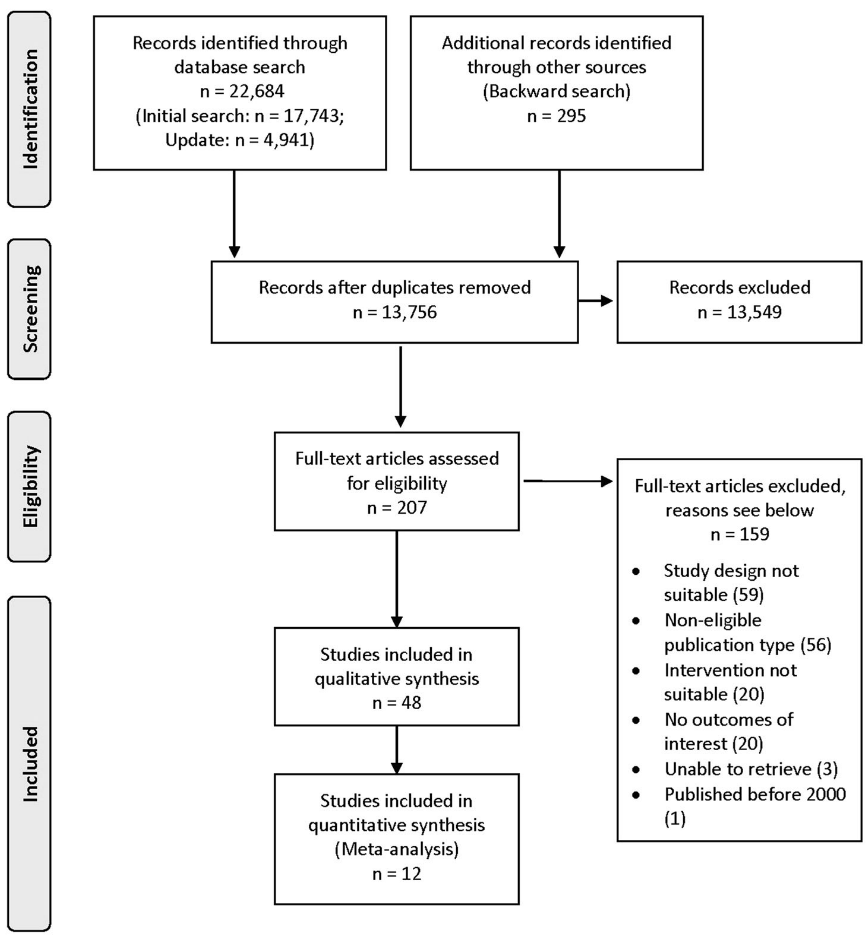


Fig. 1 | PRISMA flowchart. Visual representation of the search strategy, data screening and selection process of this systematic review.

https://doi.org/10.1038/s41746-024-01248-9

Table 1 | Key characteristics of included studies

Study	Year	Country	Setting	Medical specialty	Number of professionals	Imaging modality	Number of cases/ patients/ scans
Arbabshirani et al.[7]	2018	USA	Single-Center	Radiology	1	CT	347 patients
Batra et al.[34]	2023	USA	Single-Center	Radiology	32	CT	2501 examinations of 2197 patients
Carlile et al.[80]	2020	USA	Multi-Center	ED	112	X-Ray	1855 scans, survey on 202 scans
Cha et al.[38]	2021	USA	Single-Center	Oncology	18	CT	173 patients
Cheikh et al.[81]	2022	France	Multi-Center	ED	79	CT	7323 examinations
Chen et al.[53]	2022	China	Multi-Center	Radiology	4	CT	85 patients
Conant et al.[28]	2019	USA	Single-Center	Radiology	24	DBT	260 cases
Davis et al.[39]	2022	USA	Multi-Center	Radiology / ED	NI	CT	50,658 cases
Diao et al.[20]	2022	China	Multi-Center	Radiology	7	CT	251 patients
Duron et al.[21]	2021	France	Multi-Center	Radiology / ED	12	X-Ray	600 cases
Elijovich et al.[82]	2022	USA	Multi-Center	Neurology	NI	CT	680 patients
Ginat[83]	2021	USA	Single-Center	Radiology	5	CT	8723 scans
Hassan et al.[40]	2022	USA	Single-Center	Radiology / Neurology	NI	CT	63 patients
Hong et al.[84]	2022	South Korea	Single-Center	Radiology	60	X-Ray	1352 chest radiographs of 1319 patients
Jones et al.[85]	2021	Australia	Multi-Center	Radiology	11	X-Ray	2972 scans of 2665 patients
Kanagasingam et al.[22]	2018	Australia	Single-Center	Ophthalmology	4	Photographs	386 images of 216 patients
Kiljunen et al.[86]	2020	Finland/ Estonia/ Singapore	Multi-Center	Oncology	13	CT	45 scans of 30 patients
Ladabaum et al.[41]	2023	USA	Multi-Center	Gastroenterology	52	Colonoscopy	2329 patients
Levy et al.[87]	2022	Israel	Single-Center	Gastroenterology	30	Colonoscopy	4414 patients
Liu et al.[35]	2022	China	Multi-Center	Ophthalmology	2	OCT	1257 patients
Marwaha et al.[88]	2021	Canada	Single-Center	Human Genetics	15	Photographs	72 patients
Mueller et al.[8]	2022	Denmark	Single-Center	Radiology	2	CT	90 scans
Nehme et al.[29]	2023	USA	Single-Center	Gastroenterology	39	Colonoscopy	1041 patients
O'Neill et al.[89]	2021	USA	Single-Center	Radiology	NI	CT	6696 cases
Oppenheimer et al.[90]	2023	Germany	Single-Center	Radiology	2	X-Ray	1163 exams of 735 patients
Pierce et al.[19]	2021	USA	Single-Center	Radiology	NI	X-Ray	30,847 examinations
Potrezke et al.[54]	2023	USA	Single-Center	Nephrology	12	MRI	170 cases of 161 patients
Quan et al.[91]	2022	USA	Multi-Center	Gastroenterology	6	Colonoscopy	600 patients
Raya-Povedano et al.[36]	2021	Spain	Single-Center	Radiology	5	DM/DBT	15,986 patients
Repici et al.[24]	2020	Italy	Multi-Center	Gastroenterology	6	Colonoscopy	685 patients
Ruamviboonsuk et al.[92]	2022	Thailand	Multi-Center	Ophthalmology	12	Photographs	7651 patients
Sandbank et al.[93]	2022	Israel	Single-Center	Pathology	NI	Microscope	5954 cases
Schmuelling et al.[94]	2021	Switzerland	Single-Center	Radiology	3	CT	1808 scans of 1770 patients
Seyam et al.[95]	2022	Switzerland	Single-Center	Radiology	NI	CT	4450 patients
Sim et al.[96]	2022	Singapore	Single-Center	Radiology	NI	X-Ray	9431 datasets
Strolin et al.[97]	2023	Italy	Single-Center	Oncology	NI	CT	111 patients
Sun et al.[55]	2022	USA	Multi-Center	Radiology	NI	X-Ray	5335 images
Tchou et al.[31]	2010	USA	Single-Center	Radiology	5	DM	267 cases
Tricarico et al.[56]	2022	Italy	Single-Center	Radiology	NI	X-Ray	2942 scans
Vassallo et al.[32]	2019	Italy	Single-Center	Radiology	3	CT	225 patients
Wang et al.[26]	2019	China	Single-Center	Gastroenterology	8	Colonoscopy	1058 patients
Wang et al.[98]	2020	China	Multi-Center	Radiology	2	CT	2120 patients

https://doi.org/10.1038/s41746-024-01248-9

Table 1 (continued) | Key characteristics of included studies

Study	Year	Country	Setting	Medical specialty	Number of professionals	Imaging modality	Number of cases/ patients/ scans
Wittenberg et al.[33]	2012	Netherlands	Single-Center	Radiology	6	CT	209 patients
Wong et al.[99]	2021	Canada	Multi-Center	Oncology	39	CT	606 radiotherapy plans
Wong et al.[100]	2023	USA	Single-Center	Radiology	17	X-Ray	214 scans
Yacoub et al.[37]	2022	USA	Single-Center	Radiology	3	CT	390 scans
Yang et al.[101]	2022	China	Multi-Center	Ophthalmology	NI	Photographs	1001 patients
Zia et al.[30]	2022	Australia	Single-Center	Radiology	49	CT	1446 scans

ED Emergency Department, *CT* Computed Tomography, *DBT* Digital Breast Tomosynthesis, *DM* Digital Mammography, *MRI* Magnetic Resonance Imaging, *OCT* Optical Coherence Tomography.

In total only four studies followed a reporting guideline, three studies[20–22] used Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline[23] and Repici et al.[24] followed the CONSORT guidelines for randomized controlled trials[25]. Only two studies[24,26] pre-registered their protocol and none of the included studies provided or used an open-source available algorithm.

Appraisal of methodological quality

When assessing the methodological quality of the 45 non-randomized studies only one (2.2%) was rated with an overall "low" risk of bias. Four studies (8.9%) were rated "moderate", 28 studies (62.2%) were rated "serious", and 12 studies (26.7%) were rated "critical". All three randomized studies were appraised with an overall high risk of bias. Summary plots of the risk of bias assessments are shown in Fig. 2, full assessments can be found in Supplementary Figs. 1 and 2. The assessment of the quality of reporting using the *Methodological Index for Non-randomized Studies* (MINORS) is included in Supplementary Figs. 3 and 4. Higher scores indicate higher quality of reporting, with the maximum score being 24 for comparative studies and 16 for non-comparative studies[27]. Comparative studies reported a Median of 9 of 12 criteria with a median overall score of 15 (range: 9–23) and noncomparative studies reported a Median of 7 of 8 checklist items, with a median overall score of 7 (range: 6–14).

Outcomes

Of all included studies, 33 (68.8%) surveyed the effects of AI implementation on clinicians' time for task execution. The most frequently reported outcomes included (1) reading time (i.e., time the clinicians required to interpret an image); (2) report turnaround time (i.e., the time from completing the scan until the report is finalized); and (3) total procedure time (i.e., the time needed for colonoscopy)[28–30]. Times were assessed via surveys, recorded by researchers or staff, retrieved via time stamps, or self-recorded. Seventeen studies did not describe how they obtained the reported times.

Regarding our research question, whether AI use improves efficiency, 22 studies (66.6%) reported a reduction in time for task completion due to AI use, with 13 of these studies proving the difference to be statistically significant (see Table 4). Eight studies (24.2%) reported that AI did not reduce the time required for tasks. The remaining three studies (9.1%) chose a design or implementation protocol in which the AI was used after the normal reading, increasing the task time measured by study design[31–33].

For our meta-analyses, we established clusters with studies deploying similar methods, outcomes, and specific purposes. Concerning studies on detection tasks, we identified two main subgroups: studies using AI for interpreting CT scans ($n = 7$) and those using AI for colonoscopy ($n = 6$). Among studies using AI for interpreting CT images, a meta-analysis was performed for four studies reporting clinicians' reading times. As shown in Fig. 3a, the reading times for interpreting CT images did not differ between the groups: standardized mean error (SMD): −0.60 (95% confidence interval, −2.02 to 0.82; $p = 0.30$). Furthermore, the studies showed significantly high heterogeneity: $Q = 109.72$, $p < 0.01$, $I^2 = 96.35\%$. This heterogeneity may be associated with the different study designs included or the risk of bias ratings, with only one study being rated having a low risk of bias. Furthermore, Mueller et al.[8] reported no overall reading time but separated it for resident and attending physician, which we included separately in our meta-analysis. Concerning the use of AI for colonoscopy, five studies reported comparable measures. Our random effects meta-analysis showed no significant difference between the groups: SMD: −0.04 (95% CI, −0.76 to 0.67; $p = 0.87$), with significant heterogeneity: $Q = 733.51$, $p < 0.01$, $I^2 = 99.45\%$ (Fig. 3b). Four of the included studies had a serious risk of bias, whereas one randomized study included was rated with a high risk of bias. Among 11 studies that reported AI use for the prioritization of patients' scans, four measured the turnaround time. The study by Batra et al.[34] did not report variance measures and was therefore excluded from the meta-analysis. The remaining three studies used the AI tool Aidoc (Tables 2 and 4) to detect intracranial hemorrhage and reported the turnaround time for cases flagged positive. The meta-analysis showed no significant difference in turnaround time between cases with and without AI use: SMD: 0.03 (95% CI, −0.50 to 0.56; $p = 0.84$), with a significant heterogeneity across studies: $Q = 12.31$, $p < 0.01$, $I^2 = 83.75\%$ (Fig. 3c). All included studies were non-randomized studies, with two studies being rated with a serious risk of bias and one with a moderate risk of bias.

In total, 37 studies reported details on the actual workflow adaptations due to AI implementation, which we classified into four main variants (depicted exemplarily in Fig. 4). 16 studies (43.2%) used an AI tool as a triage system, i.e., the AI tool reprioritized the worklist or the AI tool sent an alert to the clinician or referred the patient to a specialist for further examination (Fig. 4a: AI triage). In two studies (5.4%), the AI tool acted as a gatekeeper, only referring cases labeled as suspicious to the clinician for further review, while excluding the remaining cases (Fig. 4a: AI gatekeeper). In 13 studies (35.1%), AI tools were used as a second reader for detection tasks in two variants (Fig. 4b: AI second reader). Eight studies reported that the AI tool functioned as a second reader in a concurrent mode, presenting additional information during the task to clinicians (e.g., in colonoscopy studies, where the workflow remained the same as before displaying additional information during the procedure). Five studies described a workflow in which the AI tool was used additionally after the normal detection task, resulting in a sequential second reader workflow. In five segmentation studies (13.5%), the AI tool served as a first reader with the clinician reviewing and then correcting the AI-provided contours (Fig. 4c: AI first reader).

In a single study (2.7%), the type of actual workflow implementation was at the radiologist's choice. Three studies used a study design with the AI tool as a second reader in a pre-specified reading sequence; therefore, we did not classify them as workflow adaptations. The remaining studies did not provide sufficient information on workflow implementation.

In our initial review protocol, we also aimed to include investigations on clinician workload[14]. Apart from three studies, Liu et al.[35], Raya-Povedano et al.[36], and Yacoub et al.[37], which calculated the saved workload in scans or patients because of AI use, no other study reported AI implementation effects on clinicians' workload (besides the time for tasks effects, see above). Other reported outcomes included evaluations of the AI

https://doi.org/10.1038/s41746-024-01248-9

performing the task (i.e., satisfaction)[8,38]; frequency of AI use[29,30]; patient outcomes, such as length of stay or in-hospital complications[39,40]; and sensitivity or specificity changes[8,21,24,28,41].

Risk of bias across studies

Funnel plots for the studies included in the meta-analyses were created (Supplementary Figs. 5–7). 19 studies declared a relevant conflict of interest and six other studies had potential conflicts of interest, which sum up to more than 50% of the included studies.

Additionally, we ran several sensitivity analyses to evaluate for potential selection bias. We first searched the dblp computer science bibliography, yielding 1159 studies for title and abstract screening. Therein, we achieved perfect interrater reliability (100%). Subsequently, only thirteen studies proceeded to full-text screening, with just one meeting our review criteria. This study by Wismueller & Stockmaster[42] was also part of our original search. Notably, this study was the only conference publication providing a full paper (refer to Supplementary Table 2).

Moreover, to ensure comprehensive coverage and to detect potentially missed publications due to excluding conference proceedings, we screened 2614 records from IEEE Xplore, MICCAI, and HICSS. Once again, our title and abstract screening demonstrated perfect interrater reliability (100%). However, despite including 31 publications in full-text screening, none met our inclusion criteria upon thorough assessment. Altogether, this additionally searches showed no significant indication for a potential selection bias and potentially missing out key work in other major scientific publication outlets.

Using AMSTAR-2 (A MeaSurement Tool to Assess Systematic Reviews)[43], we rated the overall confidence in the results as low, mainly due to our decision to combine non-randomized and randomized studies within our meta-analysis (Supplementary Fig. 8).

Discussion

Given the widespread adoption of AI technologies in clinical work, our systematic review and meta-analysis assesses efficiency effects on routine clinical work in medical imaging. Although most studies reported positive effects, our three meta-analyses with subsets of comparable studies showed no evidence of AI tools reducing the time on imaging tasks. Studies varied substantially in design and measures. This high heterogeneity renders robust inferences. Although nearly 67% of time-related outcome studies have shown a decrease in time with AI use, a noteworthy portion of these studies revealed conflicts of interest, potentially influencing study design or outcome estimation[44]. Our findings emphasize the need for comparable and independent high-quality studies on AI implementation to determine its actual effect on clinical workflows.

Focusing on how AI tools were integrated into the clinical workflow, we discovered diverse adoptions of AI applications in clinical imaging. Some studies have provided brief descriptions that lack adequate details to comprehend the process. Despite predictions of AI potentially supplanting human readers or serving as gatekeepers, with humans primarily reviewing flagged cases to enhance efficiency[10,11], we noted a limited adoption of AI in this manner across studies. In contrast, most studies reported AI tools as supplementary readers, potentially extending the time taken for interpretation when radiologists must additionally incorporate AI-generated results[18,45]. Another practice involved concurrent reading, which seems beneficial because it guides clinicians' attention to crucial areas, which potentially improves reading quality and safety without lengthening reading times[45,46]. Regardless of how AI was used, a crucial factor is its alignment with the intended purpose and task[15].

Although efficiency stands out in the current literature, we were also interested in whether AI affects clinicians' workload, besides the time measurements, such as number of tasks or cognitive load. We only found three studies on AI's impact on clinicians' workload, but no study assessed workload separately (e.g., in terms of cognitive workload changes)[18,35–37]. This gap in research is remarkable since human–technology interaction and human factors assessment will be a success factor for the adoption of AI in healthcare[47,48].

Our study included a vast variety of AI solutions reported in the publications. The majority was a large number of commercially available AI solutions which mostly had acquired FDA or CE clearance, ensuring safety of use in a medical context[49]. Nevertheless, it is desirable that future studies provide more detailed information about the accuracy of the AI solutions in their use case or processing times, which both can be crucial to AI adoption[50]. Regarding included studies which used non-commercially available algorithms, some of the studies did not specify the origin or source of the algorithm (i.e., developer). Especially with the specific characteristics and potential bias being introduced through the specific algorithm (e.g., for example stemming from a training bias or gaps in the underlying data), it is essential to provide information about the origins and prior validation steps of the algorithm in clinical use[51,52]. Interestingly, only four included studies discussed the possibility of bias in the AI algorithm[53–56]. Open science principles, such as data or code sharing, aid to mitigate the impact of bias. Yet, none of the studies in our review used open-source solutions or provided their algorithm[52]. Additionally, guidelines such as CONSORT-AI or SPIRIT-AI provide recommendations for the reporting of clinical studies using AI solutions[57], as previous systematic reviews have also identified serious gaps in the reporting on clinical AI solutions[58,59]. Our results corroborate this shortcoming, as none of the studies reporting non-commercial algorithms and only four studies overall followed a reporting guideline. Notwithstanding, for some included studies, AI-specific reporting guidelines were published after their initial publication. Nevertheless, comprehensive and transparent reporting remains insufficient.

With our review, we were able to replicate some of the findings by Yin et al., who provided a first overview on AI solutions in clinical practice, e.g., insufficient reporting in included studies[60]. By providing time for tasks and meta-analyses as well as workflow descriptions our review substantially extends the scope of their review, providing a robust and detailed overview on the efficiency effects of AI solutions. In 2020, Nagendran et al. provided a review comparing AI algorithms for medical imaging and clinicians, concluding that only few prospective studies in clinical settings exist[59]. Our systematic review demonstrated an increase in real-world studies in previous years and provides an up-to-date and comprehensive overview on AI solutions currently used in medical imaging practice. Our study thereby addresses one of the previously mentioned shortcomings, that benefits of the AI algorithm in silico or in retrospective studies might not transfer into clinical benefit[59]. This is also recognized by Han et al.[61] who evaluated randomized controlled trials evaluating AI in clinical practice and who argued that efficiency outcomes will strongly depend on implementation processes in actual clinical practice.

The complexities of transferring AI solutions from research into practice were explored in a review by Hua et al.[62] who evaluated the acceptability AI for medical imaging by healthcare professionals. We believe that for AI to unfold its full potential, it is essential to pay thorough attention to the adoption challenges and work system integration in clinical workplaces. Notwithstanding the increasing number of studies on AI use in real-world settings during the last years, many questions on AI implementation and workflow integration remain unanswered. On the one hand, limited consideration prevails on acceptance of AI solutions by professionals[62]. Although studies even discuss the possibility of AI as a teammate in the future[63,64], most available studies rarely include perceptions of affected clinicians[60]. On the other hand, operational and technical challenges as well as system integration into clinical IT infrastructures are major challenges, as many of the described algorithms are cloud-based. Smooth interoperability between new AI technologies and local clinical information systems as well as existing IT infrastructure is key to efficient clinical workflows[50]. For example, the combination of multimodal data, such as imaging and EHR data, could be beneficial for future decision processes in healthcare[65].

Our review has several limitations. First, publication bias may have contributed to the high number of positive findings in our study. Second, despite searching multiple databases, selection bias may have occurred, particularly as some clinics implementing AI do not systematically assess or

https://doi.org/10.1038/s41746-024-01248-9

Table 2 | Overview of the commercial AI tools used in the included studies

Source	Clearance	Body part	Purpose	Technology	Study	Sensitivity	Specificity	Processing time
Aidoc Medical, Tel Aviv, Israel/New York, NY, USA	FDA	Head	Prioritization	Convolutional neural network	Davis et al.[39]	95.0%	99.0%	near real-time
					Ginat [83]	88.4%	96.1%	3 min
					O'Neill et al.[89]	95.0%	99.0%	30–45 sec
					Seyam et al.[95]	87.2%	93.9%	NI
					Zia et al.[30]	85.7%	96.8%	NI
Aidoc Medical, Tel Aviv, Israel	CE, FDA	Chest	Prioritization	Convolutional neural network	Batra et al.[34]	83.3%	97.1%	NI
					Cheikh et al.[81]	92.6%	95.8%	NI
					Schmuelling et al.[94]	79.6%	95.0%	12.6 min
AITEM Solutions, Turin, Italy	NI	Chest	Prioritization	Convolutional neural network	Tricarico et al.[56]	78.2%	64.2%	NI
Annalise AI, Sydney, Australia	Pre-existing regulatory approval	Chest	Detection	Convolutional neural network	Jones et al.[85]	NI	NI	NI
Digital Diagnostics, Coralville, IA, USA	FDA	Eye	Prioritization	Deep learning and rule-based models	Kanagasingam et al.[22]	NI	92.0%	<3 min
EndoVigilant Inc., MD, USA	NI	Colon	Detection	NI	Quan et al.[91]	90.0%	97.0%	30 frames per sec
FDNA Inc., Sunrise, FL, USA	NI	Face	Detection	NI	Marwaha et al.[88]	NI	NI	NI
Gleamer, Paris, France	NI	Whole body	Detection	Convolutional neural network	Duron et al.[21]	79.4% (reader + AI, patient-wise)	93.6% (reader + AI, patient-wise)	NI
					Oppenheimer et al.[90]	86.9%	84.7%	3 min
Hologic, Marlborough, MA, USA	NI	Breast	Detection	NI	Tchou et al.[31]	NI	NI	NI
iCAD, Nashua, NH, USA	NI	Breast	Detection	Convolutional neural network	Conant et al.[28]	85.0% (reader + AI)	69.6% (reader + AI)	NI
Infervision Technology Co., Ltd., Beijing, China	CE, FDA	Chest	Detection	Deep learning	Diao et al.[20]	NI	NI	NI
Limbus AI, Regina, Saskatchewan, Canada	NI	Whole body	Segmentation	Deep learning	Wong et al.[99]	NI	NI	NI
Lunit, Seoul, South Korea	NI	Chest	Detection	Deep learning	Hong et al.[84]	74.8%	99.8%	NI
Medtronic, Minneapolis, MN, USA	FDA	Colon	Detection	NI	Ladabaum et al.[41]	NI	NI	NI
					Levy et al.[87]	NI	NI	NI
					Nehme et al.[29]	NI	NI	NI
					Repici et al.[24]	99.7%	NI	real-time
MVision AI Oy, Helsinki, Finland	CE, FDA	Whole body	Segmentation	Convolutional neural network	Kiljunen et al.[86]	NI	NI	NI
					Strolin et al.[97]	NI	NI	2.3 min
Philipps Healthcare, Best, The Netherlands	NI	Chest	Detection	NI	Wittenberg et al.[33]	96.0%	22.0%	NI
ScreenPoint Medical, Nijmegen, The Netherlands	CE, FDA	Breast	Prioritization	Deep learning	Raya-Povedano et al.[36]	84.1% (reader + AI)	NI	NI
Shanghai Wision AI Co., Ltd., Shanghai, China	NI	Colon	Detection	Deep learning	Wang et al.[26]	94.4% per image	95.9% per image	real-time
Shenzhen SiBright CO. Ltd., Shenzen, China	NIFDC	Eye	Detection	Ensemble of 3 convolutional neural networks	Yang et al.[101]	86.7%	96.1%	24 sec per eye

https://doi.org/10.1038/s41746-024-01248-9

Table 2 (continued) | Non-commercially available AI algorithms

Source	Clearance	Body part	Purpose	Technology	Study	Sensitivity	Specificity	Processing time
Siemens Healthcare, Erlangen, Germany	FDA	Chest	Detection	NI	Mueller et al.[8]	NI	NI	NI
					Yacoub et al.[37]	NI	NI	NI
Viz.ai, San Francisco, CA, USA	FDA	Head	Prioritization	NI	Elijovic et al.[82]	81.0%	NI	NI
					Hassan et al.[40]	87.6%	88.5%	NI

NI No information, *CE* Conformité Européenne, *FDA* Food and Drug Administration, *NIFDC* National Institutes for Food and Drug Control.

Table 3 | Non-commercially available AI algorithms

Study	Developers	Body part	Purpose	Technology	Sensitivity	Specificity	Processing time	Notes
Arbabshirani et al.[7]	SD	Head	Prioritization	Convolutional neural network	70.0%	87.0%	2.3 sec	
Carlile et al.[80]	SD	Lung	Detection	Convolutional neural network	82.8%	72.6%	real-time	
Cha et al.[38]	Elguindi et al.[102]	Multiple	Segmentation	Deep learning	NI	NI	NI	
Chen et al.[53]	SD	Head	Detection	Deep learning	33.3% (reader + AI)	91.5% (reader + AI)	NI	
Liu et al.[35]	Wang et al.[103]	Eye	Detection	Deep learning and rule decision models	98.5%	96.2%	21.4 hours	
Potretzke et al.[54]	SD	Kidney	Segmentation	NI	NI	NI	NI	
Ruamviboonsuk et al.[92]	Gulshan et al.[104]	Eye	Detection	Deep learning	91.4%	95.4%	real-time	Validated in Krause et al.[105], Ruamviboonsuk et al.[106]
Sandbank et al.[93]	SD	Breast	Detection	Multilayered convolutional neural networks	98.1%	96.2%	real-time	
Sim et al.[96]	SD	Lung	Detection	Deep learning	78.8%	97.1%	NI	
Sun et al.[55]	SD	Lung	Detection	NI	67.0%	77.0%	real-time	
Vassallo et al.[32]	Retico et al.[107]	Lung	Detection	NI	85.0% for lesions >3 mm	NI	19 min	Validated in Torres et al.[108]
Wang et al.[98]	SD	Lung	Prioritization	U-Net-based deep learning model	92.3%	85.1%	0.55 min	
Wong et al.[100]	Brown et al.[109]	Chest	Detection	Open-Source framework (SimpleMind)	88.0%	NI	3–4 min	

NI No information, *SD* Self-developed by authors, as described in respective publication.

https://doi.org/10.1038/s41746-024-01248-9

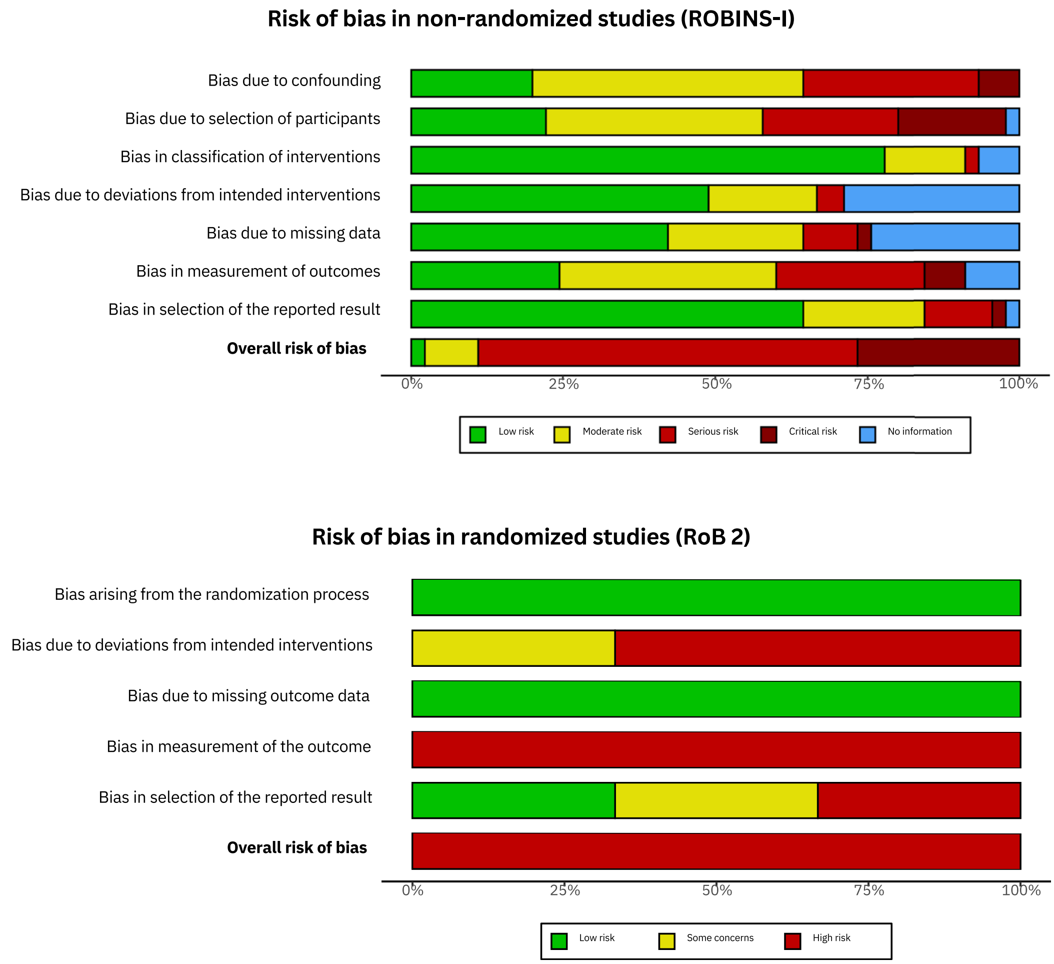


Fig. 2 | Quality assessment of included articles. Summary plots of the risk of bias assessments via *Risk of Bias in Non-randomized Studies of Interventions tool* (ROBINS-I) for non-randomized studies and the *Cochrane Risk of Bias tool* (Rob 2) for randomized studies.

publish their processes in scientific formats[60]. Moreover, we excluded conference publications which could be the source for potential biases. Nevertheless, we ran different sensitivity analyses for publication and selection bias, and did not find evidence for major bias introduced due to our search and identification strategy. Yet, aside from one conference paper, all other conference publications merely provided abstracts or posters, lacking a comprehensive base for the extraction of required details. Third, we focused exclusively on medical imaging tasks to enhance the internal validity of clinical tasks across diverse designs, AI solutions, and workflows. Fourth, the low quality rating of our review on the AMSTAR-2 checklist, which is due to the diverse study designs we included, calling for more comparable high quality studies in this field. Nevertheless, we believe that our review provides a thorough summary of the available studies matching our research question. Finally, our review concentrated solely on efficiency outcomes stemming from the integration of AI into clinical workflows. Yet, the actual impact of AI algorithms on efficiency gains in routine clinical work can be influenced by further, not here specified local factors, e.g., existent IT infrastructure, computational resources, processing times. Next to the testing of the AI solutions under standardized conditions or in randomized controlled trials, which can indicate whether AI solution are suitable for the transfer into routine medical care, careful evaluations of how AI solutions fit into everyday clinical workflow should be expanded, i.e., ideally before implementation. Exploring adoption procedures along with identifying key implementation facilitators and barriers provides valuable insights into successful AI technology use in clinical routines. However, it is important to note that AI implementation can address a spectrum of outcomes, including but not limited to enhancing patient quality and safety, augmenting diagnostic confidence, and improving healthcare staff satisfaction[8].

In conclusion, our review showed a positive trend toward research on actual AI implementation in medical imaging, with most studies describing efficiency improvements in course of AI technology implementation. We derive important recommendations for future studies on the implementation of AI in clinical settings. The rigorous use of reporting guidelines should be encouraged, as many studies reporting time outcomes did not provide sufficient details on their methods. Providing a protocol or clear depiction of how AI tools modify clinical workflows allows comprehension and comparison between pre- and post-adoption processes while facilitating learning and future implementation practice. Considering the complexity of healthcare systems, understanding the factors contributing to successful AI implementation is invaluable. Our review corroborates the need for comparable evaluations to monitor and quantify efficiency effects of AI in clinical real-world settings. Finally, future research should therefore explore success and potential differences between different AI algorithms in controlled trials as well as in real-world clinical practice settings to inform and guide future implementation processes.

Methods

Registration and protocol

Before its initiation, our systematic literature review was registered in a database (PROSPERO, ID: CRD42022303439), and the review protocol was

Table 4 | Outcomes organized by time type measured

Study	Time Type	Assessment Method	Statistical measures	Pre/ Without AI	Post/ With AI	Absolute Difference (%)	Significance	Workflow Adaptation
Batra et al.[34]	Reading time[a]	Timestamps extracted from EMR and radiologist dictation system	Mean	00:26:30	00:26:18	−00:00:12 (−0.75%)	n.s.	Triage
Cheikh et al.[81]	Reading time[a]	Survey	Mean (SD)	00:14:33 (00:09:05)	00:15:36 (00:09:46)	+00:01:03 (+7.22%)	***	Triage
Chen et al.[53]	Reading time	NI	Mean (SD)	00:03:39 (00:00:24)	00:02:56 (00:00:01)	−00:00:43 (−19.77%)	*	NI
Conant et al.[28]	Reading time	NI	Mean (CI)	00:01:04 (00:00:25)	00:00:30 (00:00:12)	−00:00:34 (−52.57%)	**	Second reader, concurrent
Diao et al.[20]	Reading time	Automatically recorded	Mean (SD)	00:04:30 (00:02:24)	00:03:43 (00:02:26)	−00:00:47 (−17.41%)	***	Second reader, sequential
Duron et al.[21]	Reading time	Automatically recorded	Mean	00:01:07	00:00:57	−00:00:10 (−14.93%)	n.s.	Second reader, concurrent
Mueller et al.[8]	Reading time – resident	NI	Mean (SD)	00:06:10 (00:02:49)	00:07:17 (00:02:29)	+00:01:07 (+18.11%)	n.s.	Depending on radiologist's choice
Mueller et al.[8]	Reading time – consultant	NI	Mean (SD)	00:06:06 (00:01:50)	00:06:20 (00:02:01)	+00:00:14 (+3.83%)	n.s.	Depending on radiologist's choice
O'Neill et al.[89]	Reading time[b]	NI	Median (CI)	00:04:50 (00:00:27)	00:06:14 (00:05:28)	+00:01:23 (+28.73%)	n.s.	Triage
Schmuelling et al.[94]	Reading time[a]	Timestamps in the clinical information system	Mean (SD)	01:25:30 (04:42:00)	01:18:30 (04:33:00)	−00:07:00 (-8.19%)	n.s.	Triage
Vassallo et al.[32]	Reading time	Recorded by investigator	Mean (SD)	00:04:56 (00:01:20)	00:05:29 (00:01:23)	+00:00:33 (+11.15%)	*	Sequential due to study design
Yacoub et al.[37]	Reading time	Self-measured with digital stopwatch	Mean (SD)	00:07:01 (00:02:55)	00:05:28 (00:02:02)	−00:01:33 (−22.09%)	***	Second reader, concurrent
Cha et al.[38]	Contouring time	Self-report	Median (IQR)	00:40:00 (00:43:00)	00:28:00 (00:10:00)	−00:12:00 (−30.00%)	**	First reader
Kiljunen et al.[86]	Contouring time	NI	Mean	00:27:00	00:15:00	−00:12:00 (−44.44%)	NI	First reader
Strolin et al.[97]	Contouring time	NI	Median (Range)	00:25:00 (01:47:00)	00:12:18 (00:46:54)	−00:12:42 (−50.80%)	***	First reader
Potretzke et al.[54]	Segmentation time[c]	..	..	..	..	..	NI	First reader
Tchou et al.[31]	Time to review AI results	Timestamp macro in Excel/ recording by investigator	Mean (SE)	00:01:58 (00:00:04)	..	00:00:23[d] (00:00:02)	NI	Sequential due to study design
Wittenberg et al.[33]	Time to review AI results	NI	Mean (Range)	00:01:15 (00:01:02)	..	00:00:22[d] (00:00:18)	NI	Sequential due to study design
Arbabshirani et al.[7]	Time to interpretation[b]	NI	Median (IQR)	08:32:00 (01:51:00)	00:19:00 (00:22:00)	−08:13:00 (−96.29%)	***	Triage
Ginat[83]	Wait time (ED cases)[b]	Automatically recorded	Mean (SD)	01:25:00 (03:14:00)	01:12:00 (02:57:00)	−00:13:00 (−15.29%)	n.s.	Triage
Ginat[83]	Wait time (inpatient cases)[b]	Automatically recorded	Mean (SD)	06:30:00 (06:08:00)	05:52:00 (05:15:00)	−00:38:00 (−9.74%)	**	Triage
Ginat[83]	Wait time (outpatient cases)[b]	Automatically recorded	Mean (SD)	11:14:00 (13:45:00)	01:10:00 (02:21:00)	−10:04:00 (−89.61%)	***	Triage
O'Neill et al.[89]	Wait time[b]	NI	Median (CI)	00:15:45 (00:00:46)	00:12:01 (00:01:55)	−00:03:44 (−23.75%)	***	Triage
Elijovich et al.[82]	Time to notification	Retrospective documentation	Median (IQR)	00:26:00 (00:14:00)	00:07:00 (00:04:00)	−00:19:00 (−73.08%)	***	Triage
Hong et al.[84]	Time to treatment	Retrospectively through analysis of electronic medical records	Mean (SD)	02:30:00 (03:24:00)	01:12:00 (19:30:00)	−01:18:00 (−4.91%)	n.s.	Second reader, concurrent
Batra et al.[34]	Report turnaround time	Timestamps	Mean	00:59:54	00:47:36	−00:12:18 (−20.53%)	***	Triage
Davis et al.[39]	Report turnaround time[a]	NI	Mean (SD)	01:03:30 (01:02:36)	00:52:30 (00:53:55)	−00:11:00 (−17.32%)	**	Triage
Seyam et al.[95]	Report turnaround time[b]	Timestamps extracted from the electronic medical record and PACS	Mean (CI)	01:00:00 (00:17:00)	01:03:00 (00:11:00)	+00:03:00 (+5.00%)	NI	Triage
Sim et al.[96]	Report turnaround time	Extracted timestamps from the hospital's RIS	Mean	00:09:00	00:07:00	−00:02:00 (−22.22%)	NI	Triage
Zia et al.[30]	Report turnaround time[b]	NI	Mean (SD)	01:06:42 (00:41:30)	01:20:00 (01:04:24)	+00:13:18 (+19.94%)	*	Second reader, sequential
Schmuelling et al.[94]	ED turnaround time[a]	Timestamps in the clinical information system	Mean (SD)	02:06:00 (01:04:12)	01:59:00 (01:41:00)	−00:07:00 (−5.56%)	n.s.	Triage
Hassan et al.[40]	DIDO time at PSC	NI	Mean (SD)	03:46:42 (04:02:54)	02:04:24 (00:57:36)	−01:42:18 (−45.13%)	*	Triage

Table 4 (continued) | Outcomes organized by time type measured

Study	Time Type	Assessment Method	Statistical measures	Pre/ Without AI	Post/ With AI	Absolute Difference (%)	Significance	Workflow Adaptation
Yang et al.[101]	Time for diagnosis	NI	Mean (SD)	00:00:38 (00:00:32)	00:00:24 (00:00:08)	−00:00:14 (−36.84%)	NI	NI
Ladabaum et al.[41]	Withdrawal time	NI	Mean (CI)	00:17:30 (00:01:30)	00:18:00 (00:01:36)	+00:00:30 (+2.86%)	n.s.	NI
Nehme et al.[29]	Withdrawal time	NI	Median (IQR)	00:17:00 (00:15:00)	00:18:00 (00:16:00)	+00:01:00 (+5.88%)	n.s.	NI
Repici et al.[24]	Withdrawal time	Stopwatch	Mean (SD)	00:07:15 (00:02:29)	00:06:57 (00:01:41)	−00:00:18 (−4.14%)	n.s.	NI
Wang et al.[26]	Withdrawal time	NI	Mean (SD)	00:06:23 (00:01:13)	00:06:53 (00:01:47)	+00:00:30 (+7.82%)	***	Second reader, concurrent
Ladabaum et al.[41]	Total procedure time	NI	Mean (CI)	00:26:06 (00:01:36)	00:26:42 (00:01:48)	+00:00:36 (+2.30%)	n.s.	NI
Levy et al.[87]	Total procedure time	Recorded by endoscopy nurse	Median (IQR)	00:24:00 (00:17:00)	00:22:00 (00:12:00)	−00:02:00 (−8.33%)	***	NI
Nehme et al.[29]	Total procedure time	NI	Median (IQR)	00:23:00 (00:16:00)	00:24:00 (00:19:00)	+00:01:00 (+4.35%)	n.s.	NI
Quan et al.[91]	Total procedure time	NI	Mean (SD)	00:19:30 (00:07:12)	00:21:24 (00:09:06)	+00:01:54 (+9.74%)	**	NI
Wang et al.[26]	Total procedure time	NI	Mean (SD)	00:12:06 (00:04:05)	00:12:31 (00:04:23)	+00:00:25 (+3.47%)	n.s.	Second reader, concurrent
Carlile et al.[80]	..	..	..	..	..	..	..	Second reader, concurrent
Jones et al.[85]	..	..	..	..	..	..	..	Second reader, concurrent
Kanagasingam et al.[22]	..	..	..	..	..	..	..	Triage + notification
Liu et al.[35]	..	..	..	..	..	..	..	Triage + notification
Marwaha et al.[88]	..	..	..	..	..	..	..	Sequential
Oppenheimer et al.[90]	..	..	..	..	..	..	..	Second reader, sequential
Pierce et al.[19]	..	..	..	..	..	..	..	Triage
Raya-Povedano et al.[36]	..	..	..	..	..	..	..	Gatekeeper
Ruamviboonsuk et al.[92]	..	..	..	..	..	..	..	Gatekeeper
Sandbank et al.[93]	..	..	..	..	..	..	..	Second reader, sequential
Sun et al.[55]	..	..	..	..	..	..	..	Second reader, sequential
Tricarico et al.[56]	..	..	..	..	..	..	..	Triage
Wang et al.[98]	..	..	..	..	..	..	..	Triage + notification
Wong et al.[99]	..	..	..	..	..	..	..	First reader
Wong et al.[100]	..	..	..	..	..	..	..	Second reader, concurrent

n.s. Not significant, *AI* Artificial intelligence, *CI* 95% confidence interval, *DIDO* Door-in-door out time, *ED* Emergency department, *EMR* Electronic medical record, *IQR* Interquartile range, *NI* No information, *PACS* Picture archiving and communication system, *PSC* Primary stroke center, *RIS* Radiology information system, *SD* Standard deviation, *SE* Standard error.
Time formats are hh:mm:ss. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
[a] Time measurements for scans that have been classified positive for pulmonary embolism.
[b] Time measurements for scans that have been classified positive for intracranial hemorrhage.
[c] Potretzke et al. reported a reduction in segmentation time but no concrete numbers.
[d] Additional reading time for AI use.

peer-reviewed (International Registered Report Identifier RR2-10.2196/40485)[14]. Our reporting adheres to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) statement reporting guidelines (Supplementary Table 3). During the preparation of this work, we used ChatGPT (version GPT-3.5, OpenAI) to optimize the readability and wording of the manuscript. After using this tool, the authors reviewed and edited the content as required and take full responsibility for the content of the publication.

Search strategy and eligibility criteria

Articles were retrieved through a structured literature search in the following electronic databases: MEDLINE (PubMed), Embase, PsycINFO, Web of Science, IEEE Xplore, and Cochrane Central Register of Controlled Trials. We included original studies on clinical imaging, written in German or English, retrieved in full-text, and published in peer-reviewed journals from the 1st of January 2000 onward, which marked a new area of AI in healthcare with the development of deep learning[14,66]. The first search was performed on July 21st, 2022, and was updated on May 19th, 2023. Furthermore, a snowball search screening of the references of the identified studies was performed to retrieve relevant studies. Dissertations, conference proceedings, and gray literature were excluded. This review encompassed observational and interventional studies, such as randomized controlled trials and nonrandomized studies on interventions (e.g., before–after studies). Only studies that

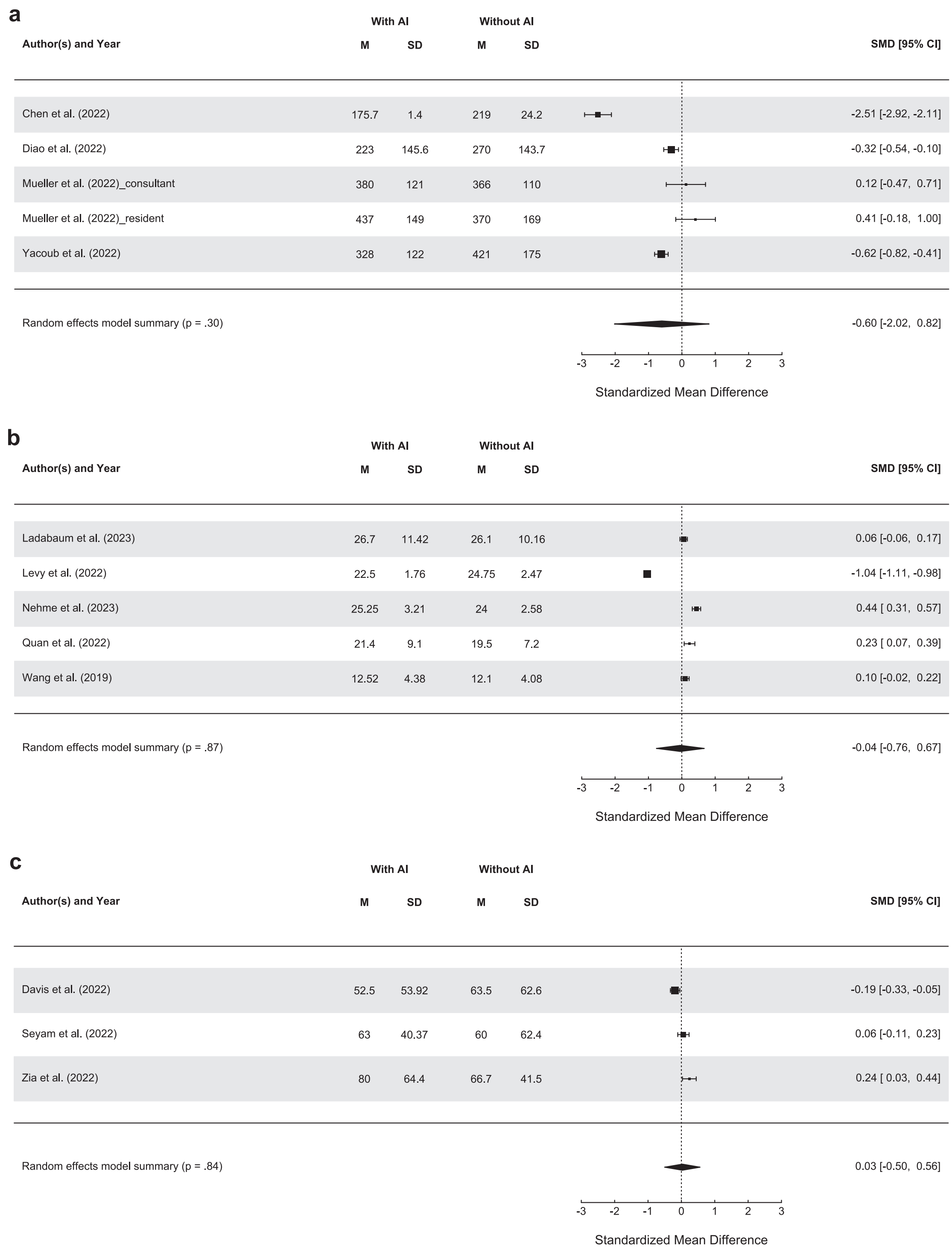


Fig. 3 | Results of meta-analyses. Graphical display and statistical results of the three meta-analyses: **a** Studies using AI for detection tasks in CT images and reported clinicians' reading time. **b** Studies using AI to detect polyps during colonoscopy and measured the total procedure time. **c** Studies that used AI for reprioritization and measured the turnaround times for cases flagged positive. All included studies used AIDOC for intracranial hemorrhage detection.

introduced AI to actual real-life clinical workflows were eligible, that is, those not conducted in an experimental setting or in a laboratory. The search strategy followed the PICO framework:

- Population: This review included studies conducted in real-world healthcare facilities, such as hospitals and clinics, using medical imaging and surveying healthcare professionals of varying expertise and qualifications.
- Exposure/interventions: This review encompassed studies that focused on various AI tools for diagnostics and their impact on healthcare professionals' interaction with the technology across various clinical

https://doi.org/10.1038/s41746-024-01248-9

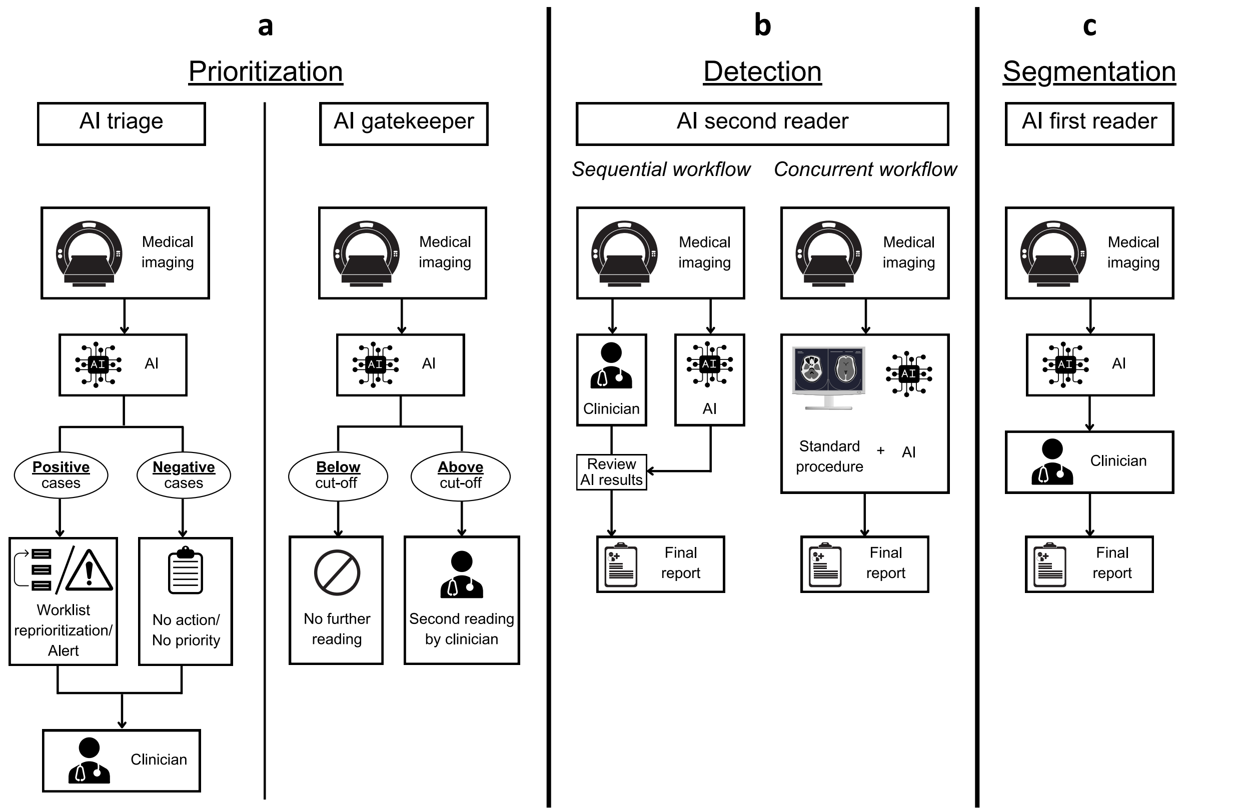


Fig. 4 | Prototypical workflows after AI implementation. Visual representation of the different workflows when using AI as reported in the included studies: **a** Workflows when using AI for prioritization tasks. **b** Workflow when using AI for detection. **c** Workflow when using AI for segmentation tasks. Figure created with Canva (Canva Pty Ltd, Sydney, Australia).

imaging tasks[67]. We exclusively focused on AI tools that interpret image data for disease diagnosis and screening[5]. For data extraction, we used the following working definition of AI used for clinical diagnostics: "any computer system used to interpret imaging data to make a diagnosis or screen for a disease, a task previously reserved for specialists"[14].

- Comparators: This review emphasized studies comparing the workflow before AI use with that after AI use or the workflow with AI use with that without AI use, although this was not a mandatory criterion to be included in the review.
- Outcomes: The primary aim of this study was to evaluate how AI solutions impact workflow efficiency in clinical care contexts. Thus, we focused on three outcomes of interest: (1) changes in time required for task completion, (2) workflow adaptation, and (3) workload.

(1) *Changes in time* for completion of imaging tasks were considered, focusing on reported quantitative changes attributed to AI usage (e.g., throughput times and review duration).
(2) *Workflow adaptation* encompasses changes in the workflow that result from the introduction of new technologies, particularly in the context of AI implementation (i.e., specifying the time and purpose of AI use).
(3) *Workload* refers to the demands of tasks on human operators and changes associated with AI implementation (e.g., cognitive demands or task load).

The detailed search strategy following the PICO framework can be found in Supplementary Table 4 and Supplementary Note 1.

Screening and selection procedure

All retrieved articles were imported into the *Rayyan tool*[68,69] for title and abstract screening. In the first step, after undergoing a training, two study team members (KW and JK/MW/NG) independently screened the titles and abstracts to establish interrater agreement. In the second step, the full texts of all eligible publications were screened by KW and JK. Any potential conflicts regarding the inclusion of articles were resolved through discussions with a third team member (MW). Reasons for exclusion were documented, as depicted in the flow diagram in Fig. 1[70].

Data extraction procedure

Two authors (JK and KW/FZ) extracted the study data and imported them into MS Excel which then went through random checks by a study team member (MW). To establish agreement all reviewers extracted data from the first five studies based on internal data extraction guidelines.

Study quality appraisal and risk of bias assessment

To evaluate the methodological quality of the included studies, two reviewers (KW and JK) used three established tools. The *Risk of Bias in Non-randomized Studies of Interventions tool* (ROBINS-I) for non-randomized studies and the *Cochrane Risk of Bias tool* (Rob 2) for randomized studies were used[71,72]. To assess the reporting quality of the included studies, the MINORS was used[27]. The MINORS was used instead of the Quality of Reporting of Observational Longitudinal Research checklist[73], as pre-specified in the review protocol, because this tool was more adaptable to all included studies. Appraisals were finally established through discussion until consensus was achieved.

Strategy for data synthesis
First, we describe the overall sample and the key information from each included study. Risk of bias assessment evaluations are presented in narrative and tabular formats. Next, where comparable studies were sufficient, a meta-analysis was performed to examine the effects of AI introduction. We used the method of Wan et al.[74] to estimate the sample mean and standard deviation from the sample size, median, and interquartile range because the reported measures varied across the included studies. Furthermore, we followed the Cochrane Handbook for calculating the standard deviation from the confidence interval (CI)[75]. The *metafor* package in R[76] was used to quantitatively synthesize data from the retrieved studies. Considering the anticipated heterogeneity of effects, a random-effects model was used to estimate the average effect across studies. Moreover, we used the DerSimonian and Laird method to determine cross-study variance and the Hartung–Knapp method to estimate the variance of the random effect[77,78]. Heterogeneity was assessed using Cochran's Q test[79] and the I^2 statistic[75]. In cases where a meta-analysis was not feasible, the results were summarized in narrative form and presented in tabular format.

Meta-biases
Potential sources of meta-bias, such as publication bias and selective reporting across studies, were considered. Funnel plots were created for the studies included in the meta-analyses.

To assess whether our review is subject to selection bias due to the choice of databases and publication types, we conducted an additional search in the dblp computer science bibliography (with our original search timeframe). As this database did not allow our original search string, the adapted version is found in Supplementary Note 2. Additionally, we performed searches on conference proceedings of the last three years, spanning publications from the January 1st 2020 until May 15th 2023. We surveyed IEEE Xplore and two major conferences not included in the database: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) and Hawaii International Conference on System Sciences (HICSS). We conducted an initial screening of titles and abstracts, with one reviewer (KW) screening all records and JK screening 10% to assess interrater reliability. Full-text assessments for eligibility were then performed by one of the reviewers, respectively (KW or JK). Furthermore, the AMSTAR-2 critical appraisal tool for systematic reviews of randomized and/or non-randomized healthcare intervention studies was used[43].

Data availability
All data generated or analyzed during this study is available from the corresponding author upon reasonable request.

Code availability
Code for meta-analyses available via https://github.com/katwend/metaanalyses.

Received: 3 April 2024; Accepted: 31 August 2024;
Published online: 30 September 2024

References
1. Yeganeh, H. An analysis of emerging trends and transformations in global healthcare. *IJHG* **24**, 169–180 (2019).
2. Asan, O., Bayrak, A. E. & Choudhury, A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J. Med. Internet Res.* **22**, e15154 (2020).
3. Park, C.-W. et al. Artificial intelligence in health care: current applications and issues. *J. Korean Med. Sci.* **35**, e379 (2020).
4. Ahmad, Z., Rahim, S., Zubair, M. & Abdul-Ghafar, J. Artificial Intelligence (ai) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. a comprehensive review. *Diagn. Pathol.* **16**, 24 (2021).
5. He, J. et al. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* **25**, 30–36 (2019).
6. Wong, S. H., Al-Hasani, H., Alam, Z. & Alam, A. Artificial intelligence in radiology: how will we be affected? *Eur. Radiol.* **29**, 141–143 (2019).
7. Arbabshirani, M. R. et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digit. Med.* **1**, 9 (2018).
8. Mueller, F. C. et al. Impact of concurrent use of artificial intelligence tools on radiologists reading time: a prospective feasibility study. *Acad. Radiol.* **29**, 1085–1090 (2022).
9. Pumplun, L., Fecho, M., Wahl, N., Peters, F. & Buxmann, P. Adoption of machine learning systems for medical diagnostics in clinics: qualitative interview study. *J. Med. Internet Res.* **23**, e29301 (2021).
10. Dahlblom, V., Dustler, M., Tingberg, A. & Zackrisson, S. Breast cancer screening with digital breast tomosynthesis: comparison of different reading strategies implementing artificial intelligence. *Eur. Radiol.* **33**, 3754–3765 (2023).
11. Miyake, M. et al. Comparative performance of a primary-reader and second-reader paradigm of computer-aided detection for CT colonography in a low-prevalence screening population. *Jpn J. Radio.* **31**, 310–319 (2013).
12. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. W. L. Artificial intelligence in radiology. *Nat. Rev. Cancer* **18**, 500–510 (2018).
13. van Leeuwen, K. G., de Rooij, M., Schalekamp, S., van Ginneken, B. & Rutten, M. J. C. M. How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatric Radiol.* **52**, 2087–2093 (2021).
14. Wenderott, K., Gambashidze, N. & Weigl, M. Integration of artificial intelligence into sociotechnical work systems—effects of artificial intelligence solutions in medical imaging on clinical efficiency: protocol for a systematic literature review. *JMIR Res. Protoc.* **11**, e40485 (2022).
15. Salwei, M. E. & Carayon, P. A Sociotechnical systems framework for the application of artificial intelligence in health care delivery. *J. Cogn. Eng. Decis. Making* **16**, 194–206 (2022).
16. Wolff, J., Pauling, J., Keck, A. & Baumbach, J. Success factors of artificial intelligence Implementation in Healthcare. *Front. Digit. Health* **3**, 594971 (2021).
17. Felmingham, C. M. et al. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. *Am. J. Clin. Dermatol.* **22**, 233–242 (2021).
18. Wenderott, K., Krups, J., Luetkens, J. A., Gambashidze, N. & Weigl, M. Prospective effects of an artificial intelligence-based computer-aided detection system for prostate imaging on routine workflow and radiologists' outcomes. *Eur. J. Radiol.* **170**, 111252 (2024).
19. Pierce, J. et al. Seamless integration of artificial intelligence into the clinical environment: our experience with a novel pneumothorax detection artificial intelligence algorithm. *J. Am. Coll. Radiol.* **18**, 1497–1505 (2021).
20. Diao, K. et al. Diagnostic study on clinical feasibility of an AI-based diagnostic system as a second reader on mobile CT images: a preliminary result. *Ann. Transl. Med.* **10**, 668 (2022).
21. Duron, L. et al. Assessment of an AI aid in detection of adult appendicular skeletal fractures by emergency physicians and radiologists: a multicenter cross-sectional diagnostic study. *Radiology* **300**, 120–129 (2021).
22. Kanagasingam, Y. et al. Evaluation of artificial intelligence–based grading of diabetic retinopathy in primary care. *JAMA Netw. Open* **1**, e182665 (2018).
23. Bossuyt, P. M. et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* **277**, 826–832 (2015).

24. Repici, A. et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. *Gastroenterology* **159**, 512–520.e7 (2020).
25. Schulz, K. F., Altman, D. G. & Moher, D. CONSORT Group CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c332–c332 (2010).
26. Wang, P. et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* **68**, 1813–1819 (2019).
27. Slim, K. et al. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument: methodological index for non-randomized studies. *ANZ J. Surg.* **73**, 712–716 (2003).
28. Conant, E. F. et al. Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. *Radiol. Artif. Intell.* **1**, e180096 (2019).
29. Nehme, F. et al. Performance and attitudes toward real-time computer-aided polyp detection during colonoscopy in a large tertiary referral center in the United States. *Gastrointest. Endosc.* **98**, 100–109.e6 (2023).
30. Zia, A. et al. Retrospective analysis and prospective validation of an Ai-based software for intracranial haemorrhage detection at a high-volume trauma centre. *Sci. Rep.* **12**, 19885 (2022).
31. Tchou, P. M. et al. Interpretation time of computer-aided detection at screening mammography. *Radiology* **257**, 40–46 (2010).
32. Vassallo, L. et al. A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies. *Eur. Radiol.* **29**, 144–152 (2019).
33. Wittenberg, R. et al. Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis—an observer study. *Radiology* **262**, 305–313 (2012).
34. Batra, K., Xi, Y., Bhagwat, S., Espino, A. & Peshock, R. Radiologist worklist reprioritization using artificial intelligence: impact on report turnaround times for CTPA examinations positive for acute pulmonary embolism. *Am. J. Roentgenol* **221**, 324–333 (2023).
35. Liu, X. et al. Evaluation of an OCT-AI-based telemedicine platform for retinal disease screening and referral in a primary care setting. *Transl. Vis. Sci. Technol.* **11**, 4 (2022).
36. Raya-Povedano, J. L. et al. AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. *Radiology* **300**, 57–65 (2021).
37. Yacoub, B. et al. Impact of artificial intelligence assistance on chest CT interpretation times: a prospective randomized study. *Am. J. Roentgenol.* **219**, 743–751 (2022).
38. Cha, E. et al. Clinical implementation of deep learning contour auto segmentation for prostate radiotherapy. *Radiother. Oncol.* **159**, 1–7 (2021).
39. Davis, M. A., Rao, B., Cedeno, P. A., Saha, A. & Zohrabian, V. M. Machine learning and improved quality metrics in acute intracranial hemorrhage by noncontrast computed tomography. *Curr. Probl. Diagn. Radiol.* **51**, 556–561 (2022).
40. Hassan, A., Ringheanu, V. & Tekle, W. The implementation of artificial intelligence significantly reduces door-in-door-out times in a primary care center prior to transfer. *Interv. Neuroradiol.* **29**, 631–636 (2022).
41. Ladabaum, U. et al. Computer-aided detection of polyps does not improve colonoscopist performance in a pragmatic implementation trial. *Gastroenterol*. **164**, 481–483 (2023).
42. Wismüller, A. & Stockmaster, L. A Prospective randomized clinical trial for measuring radiology study reporting time on artificial intelligence-based detection of intracranial hemorrhage in emergent Care Head CT (2020).
43. Shea, B. J. et al. Amstar 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* **358**, j4008 (2017).
44. Boutron, I. et al. Considering bias and conflicts of interest among the included studies. In *Cochrane Handbook for Systematic Reviews of Interventions* (eds Higgins, J. P. T. et al.) 177–204 (Wiley, 2019).
45. Beyer, F. et al. Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. *Eur. Radio.* **17**, 2941–2947 (2007).
46. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radio. Phys. Technol.* **13**, 6–19 (2020).
47. Asan, O. & Choudhury, A. Research trends in artificial intelligence applications in human factors health care: mapping review. *JMIR Hum. Factors* **8**, e28236 (2021).
48. Herrmann, T. & Pfeiffer, S. Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. *AI Soc.* **38**, 1523–1542 (2023).
49. Allen, B. The role of the FDA in ensuring the safety and efficacy of artificial intelligence software and devices. *J. Am. Coll. Radiol.* **16**, 208–210 (2019).
50. Wenderott, K., Krups, J., Luetkens, J. A. & Weigl, M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. *Appl. Ergon.* **117**, 104243 (2024).
51. Nazer, L. H. et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health* **2**, e0000278 (2023).
52. Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D. & Tzovara, A. Addressing bias in big data and AI for health care: a call for open science. *Patterns* **2**, 100347 (2021).
53. Chen, W. et al. Improving the diagnosis of acute ischemic stroke on non-contrast Ct using deep learning: a multicenter study. *Insights Imaging* **13**, 184 (2022).
54. Potretzke, T. et al. Clinical implementation of an artificial intelligence algorithm for magnetic resonance-derived measurement of total kidney volume. *Mayo Clin. Proc.* **98**, 689–700 (2023).
55. Sun, J. et al. Performance of a chest radiograph AI diagnostic tool for COVID-19: a prospective observational study. *Radiol. Artif. Intell.* **4**, e210217 (2022).
56. Tricarico, D. et al. Convolutional neural network-based automatic analysis of chest radiographs for the detection of COVID-19 pneumonia: a prioritizing tool in the emergency department, phase i study and preliminary 'real life' results. *Diagnostics* **12**, 570 (2022).
57. Ibrahim, H. et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. *Trials* **22**, 11 (2021).
58. Liu, X. et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit. Health* **1**, e271–e297 (2019).
59. Nagendran, M. et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* m689 (2020).
60. Yin, J., Ngiam, K. Y. & Teo, H. H. Role of artificial intelligence applications in real-life clinical practice: systematic review. *J. Med. Internet Res.* **23**, e25759 (2021).
61. Han, R. et al. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit. Health* **6**, e367–e373 (2024).
62. Hua, D., Petrina, N., Young, N., Cho, J.-G. & Poon, S. K. Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: a scoping review. *Artif. Intell. Med.* **147**, 102698 (2024).
63. Bruni, S., Freiman, M. & Riddle, K. Beyond the tool vs. teammate debate: exploring the sidekick metaphor in human-AI Dyads. In: Julia Wright and Daniel Barber (eds) Human Factors and Simulation. AHFE (2023) International Conference. AHFE Open Access, **83** (2023).

64. Flathmann, C. et al. Examining the impact of varying levels of AI teammate influence on human-AI teams. *Int. J. Hum.-Comput. Stud.* **177**, 103061 (2023).
65. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *npj Digit. Med.* **3**, 136 (2020).
66. Kaul, V., Enslin, S. & Gross, S. A. History of artificial intelligence in medicine. *Gastrointest. Endosc.* **92**, 807–812 (2020).
67. Dias, R. & Torkamani, A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med.* **11**, 70 (2019).
68. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Syst. Rev.* **5**, 210 (2016).
69. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev.* **5**, 210 (2016).
70. Page, M. J. et al. The Prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **372**, n71 (2021).
71. Sterne, J. A. et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* **355**, i4919 (2016).
72. Sterne, J. A. C. et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* **366**, l4898 (2019).
73. Tooth, L., Ware, R., Bain, C., Purdie, D. M. & Dobson, A. Quality of reporting of observational longitudinal research. *Am. J. Epidemiol.* **161**, 280–288 (2005).
74. Wan, X., Wang, W., Liu, J. & Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol.* **14**, 135 (2014).
75. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
76. Viechtbauer, W. Conducting meta-analyses in R with the metafor Package. *J Stat Softw.* **36**, 1–48 (2010).
77. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
78. Hartung, J. An alternative method for meta-analysis. *Biom. J. J. Math. Methods Biosci.* **41**, 901–916 (1999).
79. Cochran, W. G. The combination of estimates from different experiments. *Biometrics* **10**, 101 (1954).
80. Carlile, M. et al. Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department. *J. Am. Coll. Emerg. Phys. Open* **1**, 1459–1464 (2020).
81. Cheikh, A. B. et al. How artificial intelligence improves radiological interpretation in suspected pulmonary embolism. *Eur. Radiol.* **32**, 5831–5842 (2022).
82. Elijovich, L. et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. *J. NeuroIntervent Surg.* **14**, 704–708 (2022).
83. Ginat, D. Implementation of machine learning software on the radiology worklist decreases scan view delay for the detection of intracranial hemorrhage on CT. *Brain Sci.* **11**, 832 (2021).
84. Hong, W. et al. Deep learning for detecting pneumothorax on chest radiographs after needle biopsy: clinical implementation. *Radiology* **303**, 433–441 (2022).
85. Jones, C. M. et al. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. *BMJ Open* **11**, e052902 (2021).
86. Kiljunen, T. et al. A deep learning-based automated CT segmentation of prostate cancer anatomy for radiation therapy planning-A retrospective multicenter study. *Diagnostics* **10**, 959 (2020).
87. Levy, I., Bruckmayer, L., Klang, E., Ben-Horin, S. & Kopylov, U. Artificial intelligence-aided colonoscopy does not increase adenoma detection rate in routine clinical practice. *Am. J. Gastroenterol.* **117**, 1871–1873 (2022).
88. Marwaha, A., Chitayat, D., Meyn, M., Mendoza-Londono, R. & Chad, L. The point-of-care use of a facial phenotyping tool in the genetics clinic: enhancing diagnosis and education with machine learning. *Am. J. Med. Genet. A* **185**, 1151–1158 (2021).
89. O'Neill, T. J. et al. Active reprioritization of the reading worklist using artificial intelligence has a beneficial effect on the turnaround time for interpretation of head CT with intracranial hemorrhage. *Radiol. Artif. Intell.* **3**, e200024 (2021).
90. Oppenheimer, J., Lüken, S., Hamm, B. & Niehues, S. A prospective approach to integration of AI fracture detection software in radiographs into clinical workflow. *Life (Basel, Switzerland)* **13**, 223 (2023).
91. Quan, S. Y. et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. *Sci. Rep.* **12**, 6598 (2022).
92. Ruamviboonsuk, P. et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit. Health* **4**, e235–44 (2022).
93. Sandbank, J. et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. *npj Breast Cancer* **8**, 129 (2022).
94. Schmuelling, L. et al. Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: no significant effects on report communication times and patient turnaround in the emergency department nine months after technical implementation. *Eur. J. Radiol.* **141**, 109816 (2021).
95. Seyam, M. et al. Utilization of artificial intelligence-based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. *Radiol. Artif. Intell.* **4**, e210168 (2022).
96. Sim, J. Z. T. et al. Diagnostic performance of a deep learning model deployed at a National COVID-19 screening facility for detection of pneumonia on frontal chest radiographs. *Healthcare* **10**, 175 (2022).
97. Strolin, S. et al. How smart is artificial intelligence in organs delineation? Testing a CE and FDA-approved deep-learning tool using multiple expert contours delineated on planning CT images. *Front. Oncol.* **13**, 1089807 (2023).
98. Wang, M. et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *Lancet Digit. Health* **2**, e506–e515 (2020).
99. Wong, J. et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiat. Oncol.* **16**, 101 (2021).
100. Wong, K. et al. Integration and evaluation of chest X-ray artificial intelligence in clinical practice. *J. Med. Imaging* **10**, 051805 (2023).
101. Yang, Y. et al. Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study. *Ann. Transl. Med.* **10**, 1088 (2022).
102. Elguindi, S. et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Phys. Imaging Radiat. Oncol.* **12**, 80–86 (2019).
103. Wang, L. et al. An intelligent optical coherence tomography-based system for pathological retinal cases identification and urgent referrals. *Trans. Vis. Sci. Tech.* **9**, 46 (2020).
104. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **316**, 2402 (2016).
105. Krause, J. et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. *Ophthalmology* **125**, 1264–1272 (2018).

106. Ruamviboonsuk, P. et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *npj Digit. Med.* **2**, 25 (2019).
107. Retico, A., Delogu, P., Fantacci, M. E., Gori, I. & Preite Martinez, A. Lung nodule detection in low-dose and thin-slice computed tomography. *Comput. Biol. Med.* **38**, 525–534 (2008).
108. Lopez Torres, E. et al. Large scale validation of the M5L lung CAD on heterogeneous CT datasets. *Med. Phys.* **42**, 1477–1489 (2015).
109. Brown, M. S. et al. Automated endotracheal tube placement check using semantically embedded deep neural networks. *Acad. Radiol.* **30**, 412–420 (2023).

Acknowledgements
We sincerely thank Dr. Nikoloz Gambashidze (Institute for Patient Safety, University Hospital Bonn) for helping with the title and abstract screening. We thank Annika Strömer (Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn) for her statistical support. This research was financed through institutional budget, i.e., no external funding.

Author contributions
K.W.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing – original draft, writing – preparation, review, and editing; J.K.: data curation, investigation, visualization, writing – review and editing; F.Z.: investigation, writing – review and editing; M.W.: conceptualization, funding acquisition, supervision, validation. All authors have read and approved the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-024-01248-9.

Correspondence and requests for materials should be addressed to Katharina Wenderott.

3.3. Publication 3: Facilitators and Barriers to Implementing AI in Routine Medical Imaging: Systematic Review and Qualitative Analysis

Wenderott K, Krups J, Weigl M, Wooldridge AR (2025)
J Med Internet Res 2025;27:e63649. doi: 10.2196/63649

The supplementary material can be accessed via this link.

Review

Facilitators and Barriers to Implementing AI in Routine Medical Imaging: Systematic Review and Qualitative Analysis

Katharina Wenderott[1], BSc, MSc; Jim Krups[1], BA, BSc; Matthias Weigl[1], Dipl-Psych, Prof Dr; Abigail R Wooldridge[2], BSc, MSc, MEng, PhD

[1]Institute for Patient Safety, University Hospital Bonn, Bonn, Germany

[2]Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Katharina Wenderott, BSc, MSc
Institute for Patient Safety
University Hospital Bonn
Venusberg-Campus 1
Bonn, 53127
Germany
Phone: 49 228287 ext 13781
Email: katharina.wenderott@ukbonn.de

Abstract

Background: Artificial intelligence (AI) is rapidly advancing in health care, particularly in medical imaging, offering potential for improved efficiency and reduced workload. However, there is little systematic evidence on process factors for successful AI technology implementation into clinical workflows.

Objective: This study aimed to systematically assess and synthesize the facilitators and barriers to AI implementation reported in studies evaluating AI solutions in routine medical imaging.

Methods: We conducted a systematic review of 6 medical databases. Using a qualitative content analysis, we extracted the reported facilitators and barriers, outcomes, and moderators in the implementation process of AI. Two reviewers analyzed and categorized the data separately. We then used epistemic network analysis to explore their relationships across different stages of AI implementation.

Results: Our search yielded 13,756 records. After screening, we included 38 original studies in our final review. We identified 12 key dimensions and 37 subthemes that influence the implementation of AI in health care workflows. Key dimensions included evaluation of AI use and fit into workflow, with frequency depending considerably on the stage of the implementation process. In total, 20 themes were mentioned as both facilitators and barriers to AI implementation. Studies often focused predominantly on performance metrics over the experiences or outcomes of clinicians.

Conclusions: This systematic review provides a thorough synthesis of facilitators and barriers to successful AI implementation in medical imaging. Our study highlights the usefulness of AI technologies in clinical care and the fit of their integration into routine clinical workflows. Most studies did not directly report facilitators and barriers to AI implementation, underscoring the importance of comprehensive reporting to foster knowledge sharing. Our findings reveal a predominant focus on technological aspects of AI adoption in clinical work, highlighting the need for holistic, human-centric consideration to fully leverage the potential of AI in health care.

Trial Registration: PROSPERO CRD42022303439; https://www.crd.york.ac.uk/PROSPERO/view/CRD42022303439

International Registered Report Identifier (IRRID): RR2-10.2196/40485

(J Med Internet Res 2025;27:e63649) doi: 10.2196/63649

KEYWORDS

artificial intelligence; medical imaging; work system barriers and facilitators; implementation science; sociotechnical system; systems analysis; ergonomics; workflow; Systems Engineering Initiative for Patient Safety; SEIPS

Introduction

Background

Advancements in the development of artificial intelligence (AI) have increased the accessibility and awareness of AI solutions in health care [1,2]. AI in health care has numerous potential applications, which can be categorized into 4 areas of application: diagnostics, therapeutics, administration and regulation, and population health management [3]. AI is mostly applied to data-driven tasks due to its ability to adapt to input data. It can process and analyze large volumes of health care data more quickly [4,5].

In the United States and Europe, AI technologies in health care can be categorized as software as a medical device, referring to software designed for medical purposes without requiring hardware integration [6]. These purposes, as defined by the Food and Drug Administration, encompass treating, diagnosing, curing, mitigating, or preventing diseases or conditions [7]. The growing recognition of the potential of AI algorithms in health care is supported by the surge of Food and Drug Administration approvals since 2016 for AI-enabled devices [8]. Notably, >75% of approvals are related to radiology [8]. These numbers are consistent with reports that highlight image-based disciplines at the forefront of AI integration in clinical practice due to their data-driven nature and continuously increasing workload demands [3,5,9].

Despite the increasing availability of AI algorithms, there remains a limited understanding of their integration into clinical practice. A critical gap persists between broad research on algorithm development and limited evaluation of their actual use in clinical practice [10,11]. Most AI solutions are tested under controlled experimental conditions, which may underestimate the real-world impact of contextual factors on their utility and are therefore not necessarily transferable to clinical applications [12]. Depending on the users, the implementation process, and the clinical setting, the usefulness of AI solutions can significantly differ from previous evaluations or applications [13,14].

Complex sociotechnical systems, such as health care, "can be characterised by high uncertainty, multiple interacting elements and dynamic change" [15]. According to the sociotechnical systems theory, a sociotechnical system refers to the integration of humans, machines, environments, and organizational processes working together toward a shared objective. It consists of 2 interconnected subsystems: the technology subsystem, which encompasses tools and work organization, and the social subsystem, which involves individuals, teams, and coordination needs [15,16]. Sociotechnical frameworks of real-world clinical care offer a valuable approach to scrutinizing implementation complexities as well as the multiple intricacies of technology adoption [17,18].

A framework based on the sociotechnical systems theory that captures these complex demands and relations in health care settings is the Systems Engineering Initiative for Patient Safety (SEIPS) model [17]. The SEIPS model—most recently refined as SEIPS 3.0 [19]—proposes that sociotechnical systems consist of 5 interacting components: people, tasks, tools and technologies, organization, and environment. When one of the components changes, it affects the other components of the work system and subsequently the outcomes, that is, for patients, health care professionals, or organizations [17]. The model emphasizes the human as the center of the work system, which should be designed to support human performance and minimize negative impacts resulting from the work setting [17,19]. The SEIPS model can be applied to identify barriers and facilitators, which result from 1 element or the interaction between elements [20]. Hoonakker et al [21] introduced the concept of dimensions, which can function as either facilitators or barriers.

While the SEIPS model is useful for understanding work system dynamics, other frameworks also help analyze health care technology implementation. The Consolidated Framework for Implementation Research (CFIR) evaluates implementation processes in health services through 5 domains: intervention characteristics, outer setting, inner setting, individual characteristics, and the implementation process, overlapping with SEIPS in addressing the involved people and their environment [22,23]. The nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework examines factors influencing each of these outcomes and is specifically designed for technology implementation, while SEIPS covers broader work system design [24,25]. The integrate, design, assess, and share (IDEAS) framework, focusing on the full development cycle, is more suited for creating health technology solutions but less relevant to our study, which focuses on evaluating already implemented AI solutions [26]. The key distinction of SEIPS 3.0 is its human-centered approach, placing patients, clinicians, and caregivers at the core of the work system and emphasizing human-technology interaction and alignment in real-world clinical environments [19].

A thorough understanding of how professionals in real-world clinical settings use AI technologies and how these tools can support their performance seems imperative, given the increasing availability of AI in health care [27]. While current literature extensively addresses the potential of AI in overviews and opinion articles, limited empirical evidence stems from actual clinical care [11,28-30]. This leads to a critical lack of comprehensive understanding of AI implementation challenges and processes, potentially limiting the future development of evidence-based recommendations for successful AI technology implementation in clinical practice.

Objectives

Given the growing number of AI solutions in imaging-based disciplines, we aimed to explore and synthesize the existing literature on facilitators and barriers to AI implementation in routine medical imaging. We explored the relationships among AI implementation factors by drawing upon the SEIPS model. This approach allows for a concept-based and comprehensive synthesis of the available literature, generating a nuanced understanding of key process facilitators and barriers and their interactions in the implementation of AI technology into sociotechnical work systems in health care. Moreover, it contributes to a holistic picture of AI implementation in clinical work with consideration of important outcomes and moderating factors.

Methods

Registration and Protocol

Before starting, we registered our systematic literature review, which included qualitative analysis and synthesis, in the PROSPERO database (CRD42022303439) and published the review protocol (RR2-10.2196/40485) [28].

The primary aim of this study was to assess and synthesize facilitators and barriers to AI workflow integration in medical imaging. This study was part of a larger review project on the impact of AI solutions on workflow efficiency in medical imaging, with a separate publication on the effect of AI on efficiency outcomes [31]. Our report follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines (Multimedia Appendix 1).

Eligibility Criteria

We analyzed original clinical imaging studies in German or English published in peer-reviewed journals from January 2000 onward. Eligible studies implemented AI into real-world clinical workflows; therefore, we included observational and interventional studies (eg, randomized controlled trials) conducted in health care facilities using medical imaging. We focused on AI tools interpreting image data for disease diagnosis and screening.

We excluded dissertations, conference proceedings, and gray literature. In addition, due to our focus on real-world implementation of AI, we excluded studies conducted in experimental or laboratory settings.

Search Strategy

We searched the following electronic databases: MEDLINE (PubMed), Embase, PsycINFO, Web of Science, IEEE Xplore, and Cochrane CENTRAL. The databases were selected to reflect the interdisciplinary research on AI implementation in health care by including sources from medicine, psychology, and IT. Databases such as Cochrane, which only list systematic reviews or meta-analyses, were excluded in accordance with our eligibility criteria.

The detailed search strategy followed the PICO (population, intervention, comparison, and outcome) framework and can be found in the study by Wenderott et al [31]. The searches were performed on July 21, 2022, and on May 19, 2023. In a backward search, we identified additional relevant studies through screening the references of the included studies from the database search. Due to the time-consuming process of a systematic review with the in-depth qualitative analysis of the included studies, we performed an additional search on November 28, 2024, to identify relevant, recently published studies on facilitators and barriers to AI implementation in medical imaging [32]. This additional step ensured an update as well as the incorporation of interim published evidence on the topic. Further details are provided in Multimedia Appendix 2 [29,33-40].

Screening and Selection Procedure

All gathered articles were imported into the Rayyan tool (Rayyan) [41] for initial title and abstract screening. Two study team members (KW plus JK, MW, or Nikoloz Gambashidze), trained beforehand, individually assessed the titles and abstracts and reviewed their decisions in a consensus-oriented discussion. Subsequently, KW and JK screened the full texts of all eligible publications. Any disagreements regarding article inclusion were resolved through discussions with a third team member (MW). Exclusion reasons were documented and presented a flow diagram [42].

Data Extraction

For qualitative data extraction, full texts of all eligible articles were imported into MAXQDA 22 (VERBI Software GmbH) [43]. This program allows users to mark text segments with different semantic codes, in this case the key characteristics, and automatically creates Excel (Microsoft Corporation) files of all the marked segments. Two researchers (JK and Fiona Zaruchas) extracted key study characteristics, including country, sample size, and any reported conflicts of interest (for more details, refer to the study protocol [28]). Countries and authors were imported into RStudio (2025.05.1+513; Posit PBC) to create a map of the geographical distribution [44].

Regarding the reported stage and status of AI tool implementation in clinical practice, we used the studies by Bertram et al [45] and Pane and Sarno [46] to develop our classification of “level of implementation.” We defined 3 distinct levels: external validation, initial implementation, and full implementation (Textbox 1). We categorized all the included studies accordingly.

Textbox 1. Levels of artificial intelligence (AI) implementation in clinical practice.

External validation

- Evaluation of the AI solution using real-world data
- Participants (ie, clinicians) recruited for the study
- Participants potentially blinded to other patient data
- Approximate simulation of the routine workflow

Initial implementation

- Partial implementation into the usual workflow
- Participants recruited in their usual work
- Different study groups possible

Full implementation

- Used for all eligible patients
- Implemented into the routine workflow of clinicians

Data Analysis

We applied a multistep procedure for data analysis. We first used a structured qualitative content analysis in a stepwise process [47]. In the initial phase, JK and KW independently classified the following key content categories of AI technology process factors in all the retrieved study texts:

- Facilitators, defined as "any factor that promotes or expands the integration or use of the AI system in the workflow" [48].
- Barriers, defined as "any factor that limits or restricts the integration or use of the AI system" [48].
- Outcomes of AI use, defined as the impact the AI use has on clinicians, patients, organizations, or the workflow.
- Moderators, defined as external factors, independent of the AI tool, that influence its use, for example, the setting or user [33].

Subsequently, JK and KW engaged in a consensus-oriented discussion to reconcile all coded text segments [47,49]. In the following step, we defined subcategories following an inductive process. We noted a thematic overlap between topics being reported as a facilitator or barrier, depending on the study. Therefore, we decided to code categories that encompass facilitators as well as barriers, noting their valence (ie, positive or negative) separately. We organized the categories in a comprehensive codebook with corresponding definitions [47]. To establish consistency between raters throughout the coding process, the codebook underwent testing across 5 publications, where we discussed any coding issues and adjusted definitions as needed. Moving forward, both researchers (KW and JK) independently coded segments and subsequently discussed their codes to establish a consensus. Two researchers (KW and ARW) independently identified the proximally involved work system elements of the dimensions and then met to discuss their categorization and reached a consensus [20,50]. Using an inductive methodology, individual statements per dimension were clustered into themes that were mentioned frequently.

Epistemic Network Analysis

Epistemic network analysis (ENA) examines relationships between codes by modeling how frequently they co-occur in datasets. ENA was developed, validated, and widely applied in engineering education studies and has subsequently been used in research focused on human factors in health care [51-56]. ENA quantifies qualitative data by applying mathematics similar to social network analysis and principal component analysis to generate a weighted network of co-occurrences of codes. The matrix is then depicted graphically for each unit within the dataset. In each graph, the node size represents how frequently a code occurred in that unit; the thickness of the edges between the nodes corresponds to the weight, or frequency, at which a pair of codes co-occurred. The placement of each node is based on plotting vectors from the weighted co-occurrence matrix in a high-dimensional space, normalizing the vectors, reducing the dimensions using singular value decomposition (similar to principal component analysis), and then performing a rigid body rotation to preserve meaning. The x-axis is the dimension that accounts for the highest variation in the dataset, and the y-axis is a dimension orthogonal to the first that explains the next highest percentage of variance. Due to the preservation of meaning, these dimensions can be interpreted conceptually based on the qualitative data analysis. The fit of the resulting model can be evaluated both with Spearman and Pearson correlation coefficients. Importantly, ENA evaluates all networks concurrently, yielding a collection of networks that can be compared both visually and statistically. For more details on the method, including the mathematics and validation, please refer to the studies by Andrist et al [57], Bowman et al [58], Shaffer [59], Shaffer et al [56], and Shaffer and Ruis [60].

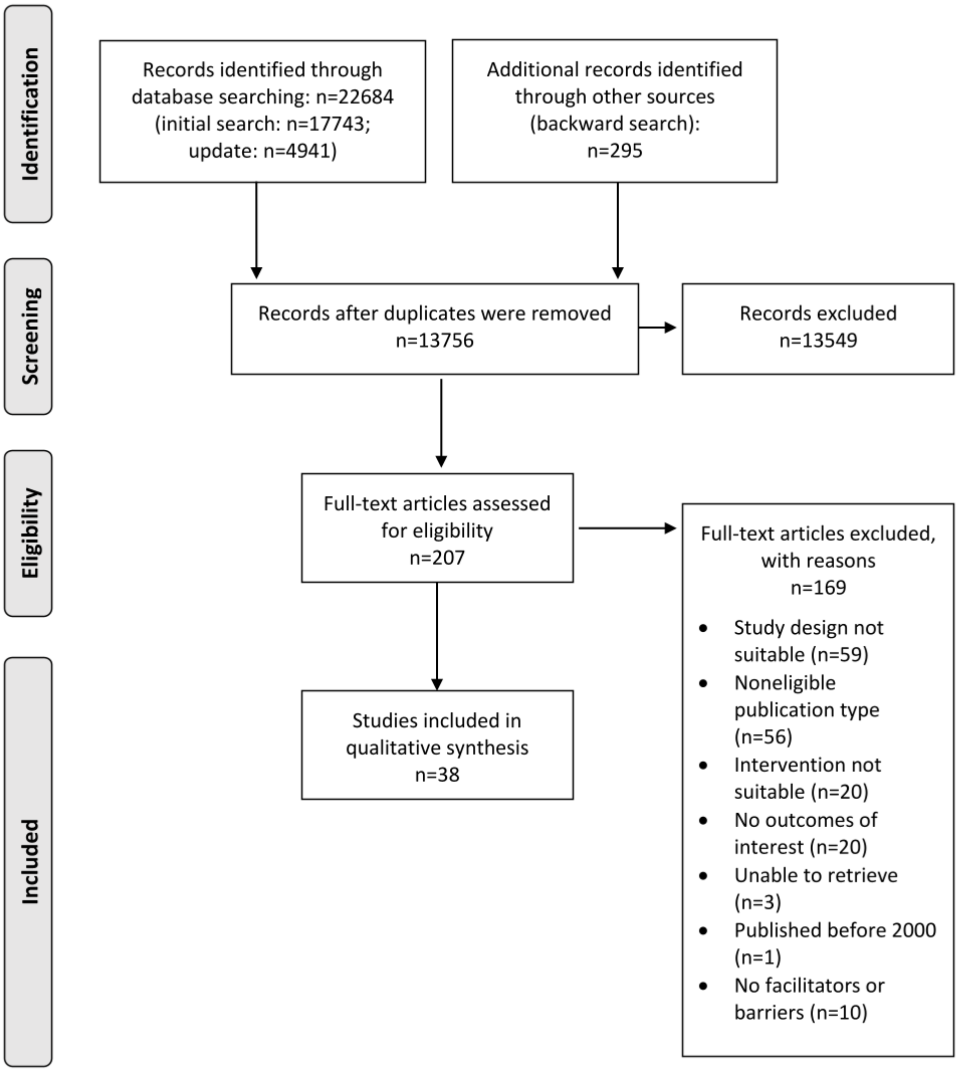
ENA serves as a valuable method to analyze and visualize the findings of our qualitative content analysis, that is, the co-occurrence of the dimensions of facilitators or barriers in the included studies [56,58-60]. In this study, we used the ENA web tool (version 1.7.0) [61]. The data were uploaded to the ENA web tool in a .csv file, with each row representing a barrier or facilitator identified through qualitative analysis; the columns included metadata such as the study, type of implementation, if that row contained a barrier or a facilitator, the dimension that specific barrier or facilitator was categorized as, and the coded excerpt from the study. ENA was used to generate 6 network graphs that depict the relationships between barriers or facilitators reported in each study, separated by the level of implementation. Thus, in each graph, the node size corresponds to the frequency that a barrier or facilitator occurred across all studies in that type of implementation; the thickness of the edges between nodes indicates how often a pair of barriers or facilitators co-occurred within the same study.

Results

Study Selection

We identified 22,684 records in the databases and an additional 295 articles through a backward search. After the removal of duplicates, 13,756 remaining records were included in the title and abstract screening. Afterward, 207 full texts were screened, of which 169 were excluded primarily because they did not meet the inclusion criteria, that is, experimental studies or studies not focusing on AI tools for interpreting imaging data (for more details, refer to the study by Wenderott et al [28]). A total of 10 studies were excluded because they did not describe any facilitator or barrier in the course of clinical implementation. Finally, 38 studies were included in the review and data extraction. A PRISMA flowchart is presented in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Study Characteristics

Of the 38 included studies, 24 (63%) were performed in a single institution and 14 (37%) were multicenter studies. Only 5% (2/38) of the studies were published before 2012, whereas all others (36/38, 95%) were published from 2018 onward. The geographical distribution of the studies is depicted in Figure 2. On the basis of the heterogeneity in the regulatory frameworks of AI in health care, we included a comparison across dimensions between the 2 main geographical clusters, the

European Union and the United States (Multimedia Appendix 3 [62-64]). Most studies (25/38, 66%) were conducted in radiology, followed by gastroenterology (5/38, 13%; Table 1). A total of 47% (18/38) of the studies reported a potentially relevant conflict of interest. For the risk of bias assessment, we used the Risk of Bias in Nonrandomized Studies of Interventions tool and the Cochrane Risk of Bias version 2 tool for the 1 included randomized study [65,66]. From the included 37 nonrandomized studies, only 1 (3%) study was classified as having a low risk of bias. In total, 11% (4/37) of the studies were rated as having a moderate risk, 65% (24/37) of the studies had a serious risk, and 22% (8/37) of the studies were assessed as having a critical risk of bias. The included randomized study was determined to have a high overall risk of bias. For a detailed risk of bias and quality of reporting assessment, refer to the supplementary material of the study by Wenderott et al [31].

Figure 2. Geographical distribution of the included studies (created with RStudio).

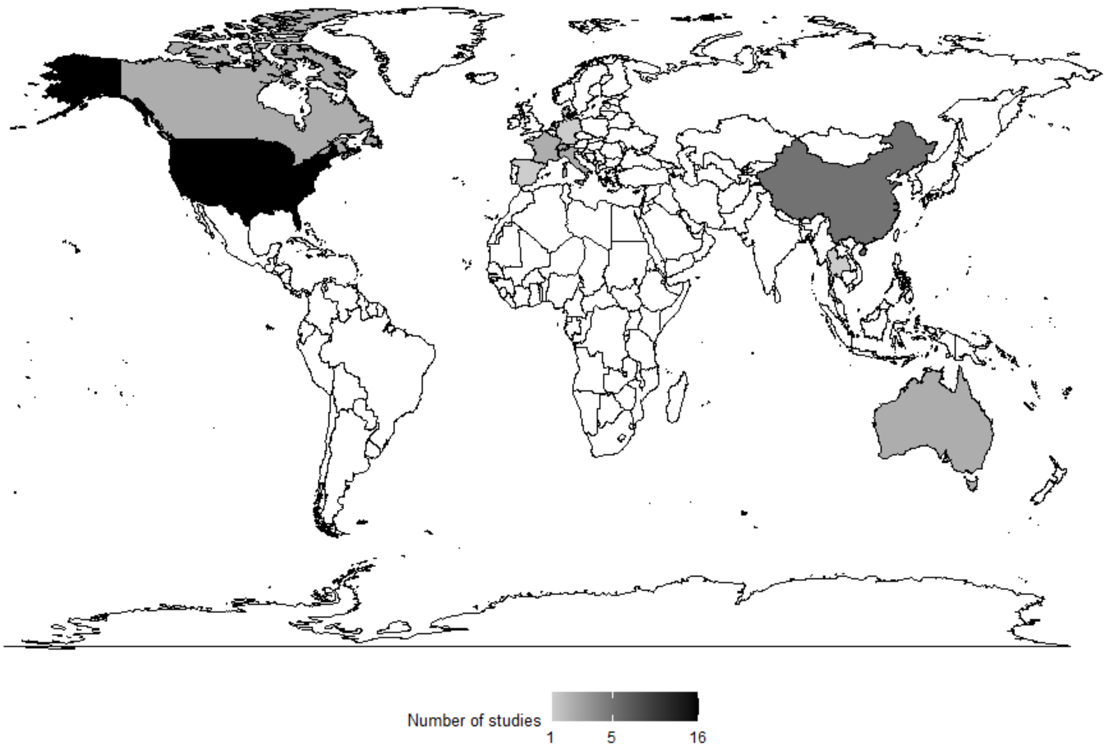


Table 1. Reported key characteristics of the included studies.

Study	Data collection	Source of data	Professionals, n	Cases, patients, or scans, n	Level of implementation
Arbabshirani et al [67]	Prospective	No information	Radiologists (not specified)	347 patients	Full
Batra et al [68]	Retrospective	Time stamps	32 radiologists	2501 examinations of 2197 patients	Full
Carlile et al [69]	Prospective	Survey	112 ED[a] physicians	1855 scans and a survey on 202 scans	Initial
Cha et al [70]	Prospective	Survey	18 physicians	173 patients	Full
Cheikh et al [71]	Retrospective	Performance metrics and survey	79 radiologists	7323 examinations	Initial
Chen et al [72]	Retrospective	Performance metrics and time measurement	4 radiologists	85 patients	External
Conant et al [73]	Retrospective	Performance metrics and time measurement	24 radiologists (including 13 breast subspecialists)	260 cases	External
Davis et al [74]	Prospective	Time stamps	Radiologists (not specified)	50,654 cases	Full
Diao et al [75]	Prospective	Time stamps and survey	7 radiologists	251 patients	Initial
Duron et al [76]	Retrospective	Performance metrics and time stamps	6 radiologists and 6 ED physicians	600 cases	External
Elijovich et al [77]	Retrospective	Chart review	Neurologists and neurointerventionalists (not specified)	680 patients	Full
Ginat [78]	Retrospective	Time stamps	5 radiologists	8723 scans	Initial
Hassan et al [79]	Retrospective	Chart review	Technologists, radiologists, ED physicians, neurologists, and interventionalists (not specified)	63 patients	Full
Jones et al [80]	Prospective	Survey	11 radiologists	2972 scans of 2665 patients	Initial
Ladabaum et al [81]	Retrospective	Chart review	52 endoscopists	2329 patients	Initial
Levy et al [82]	Retrospective	Performance metrics and time stamps	30 gastroenterologists	4414 patients	Full
Marwaha et al [83]	Retrospective	Survey	Genetic counselors and trainees (15 in total)	72 patients	Initial
Mueller et al [84]	Prospective	Observation, interview, and survey	2 radiologists	90 scans	Full
Nehme et al [85]	Prospective	Performance metrics, time stamps, and surveys	Endoscopists and staff members (45 in total)	1041 patients	Initial
Oppenheimer et al [86]	Prospective	Performance metrics	2 radiologists	1163 examinations of 735 patients	Full
Pierce et al [87]	Retrospective	Case review	Radiologists (not specified)	30,847 examinations	Full
Potrezke et al [88]	Prospective	Performance metrics	49 radiologists and 12 medical image analysts	170 cases of 161 patients	Initial
Quan et al [89]	Prospective	Performance metrics and time measurement	6 endoscopists	600 patients	Full
Raya-Povedano et al [90]	Retrospective	Performance metrics and workload	5 breast radiologists	15,986 patients	External
Ruamviboonsuk et al [91]	Prospective	Performance metrics and surveys	Staff members and nurses (12 in total)	7651 patients	Full
Sandbank et al [92]	Prospective	Performance metrics	Pathologists (not specified)	5954 cases	Full

Study	Data collection	Source of data	Professionals, n	Cases, patients, or scans, n	Level of implementation
Schmuelling et al [93]	Retrospective	Performance metrics and time stamps	Radiologists (not specified)	1808 scans of 1770 patients	Full
Seyam et al [94]	Retrospective	Performance metrics and time stamps	Radiologists (not specified)	4450 patients	Full
Tchou et al [95]	Prospective	Observation	5 radiologists	267 cases	External
Tricarico et al [96]	Prospective	Performance metrics	Radiologists (not specified)	2942 scans	Initial
Vassallo et al [97]	Retrospective	Observation and performance metrics	3 radiologists	225 patients	External
Wang et al [98]	Prospective	Performance metrics and time measurement	8 endoscopists	1058 patients	External
Wang et al [99]	Retrospective	Chart review	2 radiologists	2120 patients	External
Wittenberg et al [100]	Retrospective	Performance metrics and time measurement	6 radiologists	209 patients	External
Wong et al [101]	Prospective	Survey	Radiation therapists and oncologists (39 in total)	174 cases	Full
Wong et al [102]	Prospective	Performance metrics and survey	Radiologists and internists (17 in total)	214 scans	Initial
Yang et al [103]	Prospective	Performance metrics and time measurement	Ophthalmologists	1001 patients	Initial
Zia et al [104]	Prospective	Performance metrics, time stamps, and survey	49 radiologists	1446 scans	Initial

[a]ED: emergency department.

Regarding the level of AI implementation, we identified 24% (9/38) of the studies evaluating external validation, 34% (13/38) of the studies focusing on initial implementation, and 42% (16/38) of the studies focusing on an AI tool being fully integrated in the clinic. Table 1 presents the key characteristics of all the included studies. There was a substantial variety of AI technologies, with 42% (16/38) of the studies using commercial AI solutions and 55% (21/38) of the studies evaluating self-developed tools (1 study did not specify the source of the AI solution [87]). More details about the AI tools are provided in Multimedia Appendix 4 [67-104]. The methods that were most frequently used were the analysis of performance metrics (21/38, 55%) or time stamps (10/38, 26%). In total, 29% (11/38) of the studies used some form of survey or questionnaire to gather the opinions and experiences of clinicians. Most commonly, they used self-reports on the impact of AI use on the diagnosis and efficiency, followed by their attitude toward AI, their satisfaction or usefulness, as well as the usability of the AI tool. Notably, only the study by Jones et al [80] used an established tool, that is, the Systems Usability Scale. Further details on the surveys described in the studies are provided in Multimedia Appendix 5 [69,71,75,80,83-85,91,101,102,104].

Facilitators and Barriers to AI Implementation

Identification and Classification of Process Factors (Qualitative Content Analysis Results)

Overview

Drawing upon the qualitative analyses of the included studies, we identified 180 statements from the included publications that described the factors influencing AI implementation in clinical practice. These statements were systematically categorized into 12 overarching dimensions, as described in detail in Table 2. Within each dimension, we clustered recurring themes. This resulted in a total of 37 themes; the details and example quotations from the studies are listed in Multimedia Appendix 6 [67-104]. Many themes were stated simultaneously as facilitators and barriers, mostly depending on the presence or absence of the mentioned theme in the study (Figure 3). For example, the theme *impact on decision-making* was referenced positively in the study by Cheikh et al [71]:

> *Radiologists stressed the importance of AI to strengthen their conclusions, especially to confirm negative findings, or to ensure the absence of distal PE [pulmonary embolism] in poor-quality examinations.*

In contrast, Oppenheimer et al [86] stated the following:

> *In some edge cases, both residents reported feeling somewhat unsure of their diagnosis, in particular if*

they decided on a fracture and the AI result was negative.

With 64% (115/180) of the segments, we identified more facilitators in general than barriers (65/180, 36% segments). The dimensions *attitudes and values* and *stakeholder engagement* were mostly stated as facilitators, highlighting their positive impact on AI implementation. *Medicolegal concerns* was the only dimension that was exclusively mentioned as a barrier. In the subsequent sections, we describe the 3 dimensions with the most frequently coded segments in more detail.

Figure 3. Themes of reported facilitators and barriers to the implementation of artificial intelligence (AI) in medical imaging.

Table 2. Dimensions of facilitators and barriers to artificial intelligence (AI) implementation, including definitions and examples.

Dimensions	Definition	Codes, n	Work system elements					
			People	Tasks	TT[a]	Organization	PE[b]	EE[c]
Evaluation of AI use	Clinicians' or patients' evaluation of the usefulness of the AI tool impacting its integration.	37	✓	✓	✓			
Fit into the workflow	The AI is embedded into the workflow or processes of the local health care facility, including both clinical workflows and technical aspects such as data processing.	29		✓	✓	✓		
Implementation procedure	The AI implementation follows an implementation protocol or a prespecified plan, including users receiving training on the AI tool.	24	✓		✓	✓		
Explainability of AI	The capability of understanding and justifying the decisions made by the AI tool.	13			✓			
Attitudes and values	The beliefs, ethical principles, judgments, or priorities that might have been present before using AI influence clinicians' acceptance, adoption, and use of AI.	12	✓		✓			
Interoperability	Ensures that AI can seamlessly communicate and share data with other technologies used.	12			✓	✓		
Stakeholder involvement	In the course of implementing or using AI, important stakeholders are included in the process.	12	✓	✓	✓	✓		
Usability	Users can interact effectively and intuitively with the AI tool to accomplish their goals.	12		✓	✓			
Reliability	The reliability of the AI tool that impacts its use in the workflow.	11			✓			
Individual work organization	Fit of the AI tool with the individual preferences of the users' work organization.	7	✓	✓	✓			
Impact on the role of clinicians	AI use alters the role of clinicians, how they perceive autonomy, and whether they feel responsible for their diagnosis.	6	✓	✓	✓			
Medicolegal concerns	Intersection of medical practice and legal regulations, mitigation of legal risks, and safeguarding of patients and their rights when using the AI tool.	5			✓	✓		✓

[a]TT: tools and technologies.

[b]PE: physical environment.

[c]EE: external environment.

Evaluation of AI Use

The dimension *evaluation of AI use* reflected whether a positive or negative evaluation of the use of the AI solution aided the AI integration. This dimension was most frequently mentioned, reflecting the focus of the included studies on AI evaluation in clinical practice. We identified *people*, *tasks*, and *tools and technologies* as proximally involved work system elements. Two themes emerged in this dimension. Overall, the *usefulness* was the most frequently mentioned theme. This is supported by evidence that perceived usefulness or performance expectancy are strong determinants of the actual use of technologies [105,106], focusing on the behavior of users. The *impact on decision-making* emerged as a second theme in this dimension. Positively, clinicians valued the support provided by the AI tool, as AI use can increase the confidence of clinicians [107].

Negatively, the studies mentioned risks, such as alert fatigue [104], over trust [81,82], or insecurities due to diverging diagnostic decisions [86].

Fit Into the Workflow

The dimension *fit into the workflow* focused on how well AI technology fits into the workflow, which is an important factor to consider during the implementation of a novel technology [108,109]. The proximally involved work system elements were *tasks*, *tools and technologies*, and *organization*. In this dimension, 5 themes were identified. The most frequently and favorably mentioned theme was the *accessibility of results*, for example, by results being forwarded automatically to the clinicians [77] or providing a notification platform [78]. This also applied to the theme of *data processing*, where automatic and fast processing was a facilitating factor [67,68,77,78,97]. Regarding the themes distractions or disruptions due to AI, the facilitating factors were characterized by the absence of these, whereas the barriers reflected the negative influence of the AI tool on the workflow of the users, for example, through alarms that potentially distracted the clinicians. The theme *additional work steps* was only mentioned in the study by Batra et al [68].

Implementation Procedure

The dimension *implementation procedure* focused on the descriptions of the implementation process to install the AI system in the clinical workflow. The related work system elements were *people*, *tools and technologies*, and *organization*. In this dimension, the themes *internal testing* of the AI tool; *continuous maintenance*, that is, the ongoing monitoring of the AI tool with adaptations if necessary; and the *training of users* were exclusively mentioned as facilitators. Of the 38 studies, only 3 (8%) described a *deployment strategy* [81,87,88], with Ladabaum et al [81] describing that their minimalist approach was not sufficient to successfully implement the AI tool. In total, 13% (5/38) of the studies discussed the strategies or preconditions to the *technology readiness* of the organization, which can be defined as the willingness to “embrace and use new technologies to accomplish goals.... It is a combination of positive and negative technology-related beliefs” [110]. In the study by Ruamviboonsuk et al [91], the authors encountered the challenge that the hospital was still working with paper-based records, and the internet connectivity was slow, highlighting the role of the pre-existing digital infrastructure.

Comparison of Facilitators and Barriers Across the Levels of Implementation (Results of ENA)

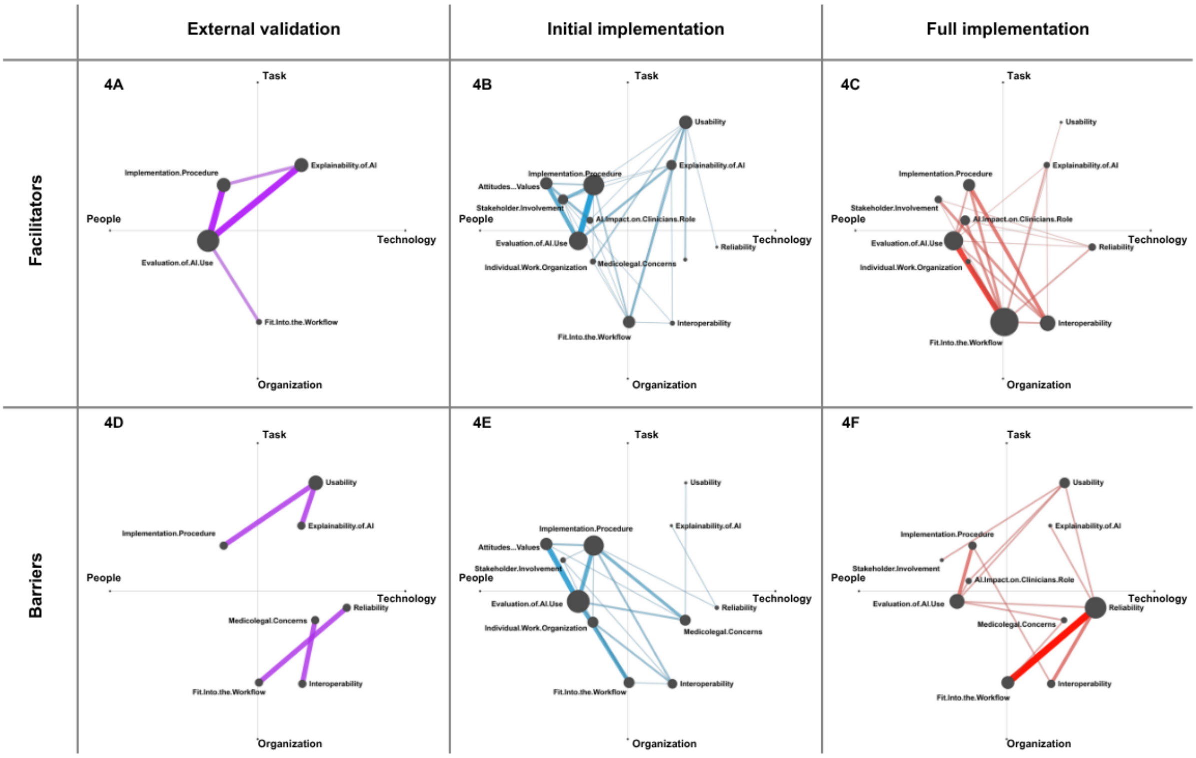
We used ENA to model the differences in facilitators and barriers across the level of implementation, resulting in 6 distinct network graphs (Figure 4). The axes identified in our ENA can be associated with work system elements of the SEIPS model [17]. The x-axis represents the work system element *people* in the negative direction, as indicated by the dimensions *attitudes and values* and *stakeholder involvement* being the farthest in this direction, and the work system element *technology* in the positive direction, which we concluded from the dimensions *reliability*, *interoperability*, and *usability* presented in this direction. For the x-axis and the y-axis, the coregistration correlations were 1 (both Pearson and Spearman), showing a strong goodness of fit [111]. The x-axis accounted for 37.2% of the variance. The y-axis accounted for 21% of the variance. The positive direction of the y-axis can be associated with the work system element *tasks*, with the ENA showing the dimension *usability* as the farthest node in this direction. In contrast, the negative side of the y-axis represents the work system element *organization*, which we inferred from the dimensions *fit into the workflow* and *interoperability* being the most distant nodes in this direction.

For the studies describing external validations of AI solutions, a total of 19 coded segments (segments per study: mean 2.11, SD 1.27; median 2, IQR 1-2) were included in the ENA. The resulting networks showed a small number of involved dimensions and connections, highlighting the dimensions *evaluation of AI use* and *explainability of AI* as facilitators and the dimension *usability* as a barrier (Figures 4A and 4D).

For the initial implementation studies, we analyzed 85 coded segments (segments per study: mean 6.54, SD 4.74; median 5, IQR 3-9). The facilitators showed an accumulation in the quadrant of the work system elements *tasks* and *people*, with the dimensions *implementation procedure* and *evaluation of AI use* being the largest nodes. The strongest connection for the facilitators was between the dimensions *evaluation of AI use* and *implementation procedure*, whereas the strongest connection for the barriers was between the dimensions *evaluation of AI use* and *attitudes and values*, with the dimension *implementation procedure* being also mentioned frequently (Figures 4B and 4E).

Regarding the publications reporting the full implementation of AI solutions, the network graphs were based on 76 coded segments (segments per study: mean 4.75, SD 4.11; median 3.5, IQR 2.5-7). The frequently mentioned facilitators were the dimensions *fit into the workflow* and *evaluation of AI use*, with a strong connection between these dimensions (Figure 4C). The barriers centered on the dimension *reliability*, with a strong connection to the dimension *fit into the workflow* (Figure 4F).

Figure 4. Facilitators and barriers to artificial technology (AI) technology implementation in medical imaging: network diagrams resulting from an epistemic network analyses separated by the level of implementation.



Reported Outcomes of AI Implementation

The included studies examined various outcomes stemming from the implementation of AI tools in medical imaging tasks. Of the 38 included studies, 31 (82%) reported efficiency outcomes, with 71% (22/31) of the studies showing enhanced efficiency, while 6% (2/31) of the studies reported a negative impact, and 23% (7/31) of the studies indicated no changes in efficiency. 13% (5/38) of the included studies assessed the impact of AI on workload or required work steps, with 80% (4/5) of the studies reporting reductions and 20% (1/5) of the studies indicating an increase. Of the 38 included studies, 16 (42%) reported on the performance of AI solutions in terms of changes in detection rates, need for human oversight, or quality of the AI-based results. In addition, 34% (13/38) discussed outcomes for patients, such as enhanced safety or quality control due to AI; a reduced time to diagnosis or treatment; prolonged stay in the emergency department; and increased detection rates, possibly leading to additional unnecessary treatments or increased workload [98]. The full details on the reported study outcomes are provided in Multimedia Appendix 7 [67-95,98,99,101-104].

Moderating Factors of AI Implementation

Of the 38 included studies, 18 (47%) identified moderators, which are defined as factors that influence AI use but are independent of the AI itself, such as the setting or the users. Details on the studies reporting moderators are provided in Multimedia Appendix 8 [69,70,75,77,78,80-82,84-86,91,93,95,98,100,102,103].

The setting, precisely the shifts, times of the day, or whether it was a weekday or a weekend, was mentioned by 5% (2/38) of the studies [78,86]. Schmuelling et al [93] and Wong et al [102] also highlighted the significant influence of the clinical environment or pre-existing clinical workflows on AI implementation [93,102].

In addition, 21% (8/38) of the studies described that the implementation and use of AI are impacted by how health care professionals use the AI system, such as through personal preferences concerning their workflow or change in behaviors when they are not being observed. In total, 11% (4/38) examined the impact of human behavior on the evaluation of AI solutions in terms of interobserver variability or the missing reporting of errors.

In total, 26% (10/38) of the studies listed task-related factors, for example, differences due to input image quality, task type, or criticality of the findings. Moreover, 18% (7/38) of the studies noted that job experience or familiarity with AI has an impact on AI use.

Of the 38 included studies, 5 (13%) investigated physician performance when using AI regarding their job experience, with 20% (1/5) of the studies reporting no association [80]. Furthermore, 40% (2/5) of the studies reported a more positive AI use evaluation [69,84] or an enhanced detection rate [85] for less experienced readers, while 20% (1/5) of the studies reported that "the time to review the CAD images increased with the experience of the reader" [95].

Additional Search to Include Recent Evidence

We searched 6 databases (PubMed, Web of Science, Embase, CENTRAL, Cochrane, and IEEE Xplore) to further identify recently published, relevant evidence, including review articles, in contrast to our original review process. While we retrieved and screened 1016 records, we identified 9 studies investigating facilitators and barriers of AI implementation in medical imaging. Among the 9 studies, 5 (56%) were scoping reviews, with 40% (2/5) of them focusing on AI implementation in health care in general [29,34], 40% (2/5) of the reviews studying AI for breast imaging [35,36], and 20% (1/5) of the reviews focusing on AI in radiology [37]. Only Chomutare et al [29] used a theoretical framework, the CFIR, to guide their analysis. All reviews provided a narrative synthesis of the results. In addition, of the 9 studies retrieved through the additional search, we identified 4 (44%) original studies, all using interviews as a qualitative methodology for studying facilitators and barriers of AI in medical imaging. Among those, 50% (2/4) of the studies did not study a specific AI implementation [38,39] and the other 50% (2/4) of the studies focused on specific AI solutions and were published after our second search [33,40]. Further details on these studies are provided in Multimedia Appendix 2.

Discussion

Principal Findings

Our systematic review provides, to the best of our knowledge, the first qualitative and quantitative synthesis that analyzes facilitators and barriers reported in studies on AI implementation in real-world clinical practice. Using our differentiation between the 3 levels of implementation, we were able to delve into the complexities of transferring AI technologies from model development and testing into the actual clinical environment [30]. To strengthen our conclusions, we used the SEIPS model, which is a strong asset for the system-based analysis of health care work environments [50]. In our analysis, we found that the frequency of various facilitators and barriers differed significantly across the stages of implementation. However, a consistently wide range of factors was identified, emphasizing the complex interplay of various elements when integrating AI into routine care processes. Consequently, our study offers a consolidated list of key factors that should be considered during AI implementation.

Focusing on categories across the implementation levels and matching them to work system elements can guide future implementation processes. In the conducted ENAs, the work system elements *tasks*, *tools and technology*, *organization*, and *people* were associated with the different axes, which provided a visualization of the importance of interactions between the work system elements. Missing in this categorization was the work system element *physical environment*, likely due to the diverse study settings and minimal impact of AI on work environments in the included studies. All studies focused on software as a medical device solutions that mostly did not alter their physical environment, and only 2 studies [89,104] reported physical changes because the AI solution was displayed on separate monitors. Referring to our resulting network graphs (Figure 4), it is noteworthy that the dimension *implementation procedure* was linked to work system elements *tasks* and *people*, while typically it is associated with organizational decisions [39,112]. Our classification showed that the included studies focused on evaluating AI on a microsystem level, that is, the individual health professionals and the tasks associated with AI use [113,114].

Studies describing external validations of AI solutions reported facilitators mostly related to the dimension *evaluation of AI use*, which was also the most prominent dimension overall. Barriers often stemmed from the AI technology itself, especially from the issues with *usability*. The focus of these networks highlights that external validation is still a part of the algorithm development process in which the clinical applicability of the AI solutions is being assessed. This is also supported by the outcomes reported in these studies, which were mostly time related, such as efficiency, treatment times, or workload. Moderating factors were not very prominent in these studies and were predominantly task related. These studies usually test the algorithm's interaction with various work system elements for the first time under realistic conditions, which is often not done during the AI development phase before clinical validation [115].

Studies focusing on the initial implementation tested how AI solutions can be fitted into the existing workflow, while not yet being applied to all patients or cases. Barriers and facilitators in these studies mainly focus on the work system elements *people* and *tasks*, with most connections in the ENA stemming from this quadrant. In addition, these studies presented a broader spectrum of outcomes, such as satisfaction or patient outcomes. Moderating factors to AI use in these studies were also diverse, including experience of clinicians and their behavior. This focus aligns with the SEIPS model, which prioritizes the people and a human-centered design [19]. This resonates well with the identified initial implementation studies that tested and studied AI integration into the work system, and determined the necessary optimizations. The rising recognition of the significance of human-centered design and stakeholder engagement in the adoption of AI in health care is supported by our findings [14,35,116-118].

In the network analysis of studies assessing AI solutions that have been fully integrated into routine care, the dimension *fit into the workflow* emerges as the largest node of facilitators, with also the most connections, supporting the literature that highlights the integration of AI into work processes as crucial for success [10,12,109]. The themes we observed as being most important were *accessibility of results* and *no disruptions due to AI*, with the latter being mentioned positively by the absence of AI-related disruptions to the workflow. As workflow disruptions can increase the procedure duration, this is highly relevant in medical imaging, as radiologists and other physicians face increasing workloads and time pressures due to the large amount of medical imaging data to be interpreted [119,120]. Interestingly, barriers in these studies showed a strong connection between the dimensions *reliability* and *fit into the workflow*. This aligns with our recent findings that technical issues can largely impact the workflow, contrasting with the literature that often emphasizes ethical debates, medicolegal concerns, or AI explainability, which were less prominent in our analysis [112,121]. Nevertheless, most outcomes reported in these studies were positive, such as increased efficiency, improved detection rates, or reduced treatment times, potentially reflecting that only the AI solutions that have overcome most barriers manage the transfer from the initial development stage to full implementation [29].

Comparison to Previous Work

Compared to previous research in the field, our results contribute important insights and show consistencies and discrepancies in AI implementation research. Few reviews have focused on the implementation of AI in clinical practice, and even fewer have specifically examined the facilitators and barriers to AI implementation. In our additional search, we only identified 5 scoping reviews targeting this topic in relation to AI for medical imaging. Hassan et al [34] provided a recent review on the facilitators and barriers to AI adoption, noting that most of the included studies focused on radiology and oncology. The authors identified 18 categories of facilitators and barriers, and similar to our findings, they observed that the same factor can be described as both a facilitator and a barrier [34]. However, because Hassan et al [34] do not offer a detailed overview of the included studies and only present a narrative synthesis, the comparison with our included studies, their settings, and designs is limited.

Lokaj et al [35] reviewed AI development and implementation for breast imaging diagnosis, identifying clinical workflow as a key facilitator. However, they emphasized technical aspects and algorithm development, with barriers such as data, evaluation, and validation issues. They noted the inclusion of very few prospective studies. In contrast, our review focuses on AI solutions evaluated after the development phase, in real-world clinical settings; therefore, technical aspects do not play a significant role in our developed set of facilitators and barriers.

Chomutare et al [29] also reviewed AI implementation in health care using the CFIR focusing on late-stage implementations. Despite including only 19 studies, they identified dimensions similar to ours, such as interoperability and transparency. Using ENAs based on implementation levels, our study provides a detailed overview of the facilitators and barriers at different implementation stages. Our findings further support the claim of Chomutare et al [29] that limited knowledge exists about the clinicians working with AI. Our review found that 29% (11/38) of the included studies incorporated user feedback, revealing a significant research gap. This underscores the need for research to adopt human-centered design, defined by the International Organization for Standardization standard 9241-210:2019 as follows: “an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance” [122]. Using human-centered design principles is crucial for developing AI systems that benefit clinicians and patients [116,118].

Factors influencing AI adoption in health care are similar to those for other health information technologies, for example, electronic health records or e-prescription systems [123-125]. Key success factors, such as stakeholder involvement and system usability, are comparable across these technologies [126,127]. Recommendations for AI implementation can be drawn from health information technology research, such as that by Yen et al [128], who emphasize the importance of the sociotechnical context and longitudinal studies over cross-sectional outcomes. Although few of our included studies reported on the implementation process over time, our network analyses by implementation level can help identify the criteria that must be met in the course of AI tool transitions from research to clinical practice. AI introduces unique considerations to health care workflows, such as shared decision-making and human oversight [129], and presents new challenges requiring a broader understanding of the technology [130].

Clinicians need to understand the data used to train AI tools, as biases and limitations can arise, a point highlighted by Pierce et al [87] through their educational campaign before AI implementation. As AI solutions present the possibility of algorithmic bias, which might not be detected by clinicians, it is noteworthy that we identified user training and transparency as facilitators of AI implementation. The diverse nature of algorithmic biases, for example, stemming from biased training data, data gaps on underrepresented groups, human bias of the developers, or a lack of data standards, is an important information to be considered by the users [131-133]. Algorithmic bias holds the potential for patient harm, especially for populations considered disadvantaged [132]. While we identified strategies that can limit the impact of bias, such as user training, continuous monitoring, or transparency, most of the included studies did not explicitly mention bias, as described in by Wenderott et al [31]. Beyond algorithmic bias, it is also essential to address the legal and ethical challenges surrounding AI-supported decisions in health care [134]. Although these topics are widely discussed in research and politics, only 13% (5/38) of the studies we reviewed discussed medicolegal concerns in terms of data privacy concerns and legal implications. Thus, although AI solutions have been successfully implemented into routine medical care, issues of liability remain unresolved [135,136]. As AI continues to evolve and becomes more integrated into clinical practice, it is crucial to carefully consider these factors to ensure its safe, effective, and responsible use in health care settings.

Limitations

Our study has a few limitations worth noting. First, we focused exclusively on AI tools in medical imaging, aiming to ensure the comparability of our findings. However, we encountered significant diversity in study settings, AI solutions, and purposes for decision-making or diagnostics. Because we only reviewed peer-reviewed original studies, some evaluations of AI implementation in health care might have been missed. Second, our findings showed more facilitators than barriers, which could be associated with a potential publication bias toward a more positive reporting of AI implementation, especially in combination with the high number of studies that reported a conflict of interest. In addition, we only searched for peer-reviewed literature, possibly missing reports on AI implementation from gray literature. AI implementation might also occur in clinical practice without scientific evaluation or reporting of results, which could also contribute to a publication bias. Third, the rapidly evolving nature of AI research indicates that certain processes or issues discussed in the studies may already be outdated by the time of publication, a challenge particularly relevant to the time-consuming process of systematic reviews, which often face delays from the literature search to final publication [32]. Therefore, while our review provides the first comprehensive, thorough, and methodologically rigorous overview of the facilitators and barriers to AI implementation in medical imaging, we recommend that future studies consider adopting shorter review cycles to ensure more timely publication and greater relevance in light of ongoing technical advancements. Fourth, facilitators and barriers were mainly extracted from study discussions, with separate reporting being rare, possibly introducing bias. In general, we noted that the descriptions of the implementation procedure and setting were sparse. Future research should provide details on their implementation strategy, processes, and subsequent adjustments to best integrate technology into the unique workflow [112]. This would enable comparisons across studies and facilitate learning in the scientific community. In addition, our established dimensions were formed inductively, requiring further validation. Fifth, while we used the SEIPS model for our analysis, we acknowledge that other frameworks exist such as the CFIR; the IDEAS framework; or the NASSS framework [22,24,26]. We planned to use the NASSS as specified in the review protocol but eventually chose the SEIPS model due to its human-centered and system-based approach [28]. Finally, our focus was on real-world investigations in clinical settings. Although our classification of "level of implementation" was useful for comparing different studies, its applicability to other clinical tasks, medical specialties, and work settings needs further examination. Furthermore, future studies should explore the impact of regulatory settings on research outcomes. While this was not feasible in our review due to the limited number of studies, the growing number of available AI algorithms and academic publications on AI in medicine will potentially provide sufficient data for these analyses [11,63].

Conclusions

In conclusion, the facilitators and barriers identified in medical imaging studies have produced a comprehensive list of dimensions and themes essential for AI implementation in clinical care. Our research underscores the pressing necessity for holistic investigations into AI implementation, encompassing not only the technical aspects but also their impact on users, teams, and work processes. Furthermore, our results corroborate the future need for transparent reporting of AI implementation procedures. This transparency fosters knowledge exchange within the scientific community, facilitating the translation of research findings into actionable strategies for clinical care. A deeper understanding of how AI solutions affect clinicians and their workflows can help reduce clinician workload and improve patient care.

Acknowledgments

This work was supported by a fellowship from the Deutscher Akademischer Austauschdienst (DAAD; German Academic Exchange Service) awarded to KW. The publication of this work was supported by the Open Access Publication Fund of the University of Bonn. The authors sincerely thank Dr Nikoloz Gambashidze and Fiona Zaruchas (Institute for Patient Safety, University Hospital Bonn) for helping with the title and abstract screening and data extraction. During the preparation of this paper, the authors used ChatGPT (version GPT-3.5, OpenAI) to optimize the readability and wording of the manuscript. This was done by asking ChatGPT for synonyms or the spelling of single words or for sentences using prompts such as "Can you check for spelling or grammar mistakes?" or "Can you enhance the readability of this sentence?" (Multimedia Appendix 9). After using this tool, the authors reviewed and edited the content as required and take full responsibility for the content of the paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KW was responsible for conceptualization. data curation, formal analysis, investigation, methodology, project administration, software development, and visualization. KW also led the writing of the original draft and contributed to the preparation, review, and editing of the manuscript. JK contributed to data curation, investigation, visualization, and the review and editing of the manuscript. MW was involved in conceptualization, funding acquisition, supervision, validation, and manuscript review and editing. ARW contributed to methodology, software development, supervision, validation, and the review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 53 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Additional search.
[DOCX File , 78 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Geographical comparison.
[DOCX File , 19 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Artificial intelligence solutions.
[DOCX File , 60 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Overview on surveys used in the included publications.
[DOCX File , 32 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Details on the extracted themes.
[DOCX File , 110 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Outcomes extracted from the included publications.
[DOCX File , 62 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Moderators extracted from the included publications.
[DOCX File , 55 KB-Multimedia Appendix 8]

Multimedia Appendix 9

ChatGPT transcript.
[DOCX File , 32 KB-Multimedia Appendix 9]

References

1. AlZaabi A, AlMaskari S, AalAbdulsalam A. Are physicians and medical students ready for artificial intelligence applications in healthcare? Digit Health. Jan 26, 2023;9:20552076231152167. [FREE Full text] [doi: 10.1177/20552076231152167] [Medline: 36762024]
2. Beets B, Newman TP, Howell EL, Bao L, Yang S. Surveying public perceptions of artificial intelligence in health care in the United States: systematic review. J Med Internet Res. Apr 04, 2023;25:e40337. [FREE Full text] [doi: 10.2196/40337] [Medline: 37014676]
3. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. Jan 2019;25(1):30-36. [FREE Full text] [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
4. Artificial intelligence and machine learning in software as a medical device. U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device [accessed 2022-04-29]
5. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. Oct 24, 2018;2(1):35. [FREE Full text] [doi: 10.1186/s41747-018-0061-6] [Medline: 30353365]
6. Software as a Medical Device (SaMD): key definitions. International Medical Device Regulators Forum. Dec 9, 2013. URL: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf [accessed 2025-06-15]
7. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) - discussion paper and request for feedback. U.S. Food & Drug Administration. 2019. URL: https://www.fda.gov/media/122535/download [accessed 2025-06-16]

8. Joshi G, Jain A, Araveeti SR, Adhikari S, Garg H, Bhandari M. FDA-approved artificial intelligence and machine learning (AI/ML)-enabled medical devices: an updated landscape. Electronics. Jan 24, 2024;13(3):498. [doi: 10.3390/electronics13030498]
9. Ahmad OF, Mori Y, Misawa M, Kudo SE, Anderson JT, Bernal J, et al. Establishing key research questions for the implementation of artificial intelligence in colonoscopy: a modified Delphi method. Endoscopy. Sep 2021;53(9):893-901. [FREE Full text] [doi: 10.1055/a-1306-7590] [Medline: 33167043]
10. Wolff J, Pauling J, Keck A, Baumbach J. Success factors of artificial intelligence implementation in healthcare. Front Digit Health. Jun 16, 2021;3:594971. [FREE Full text] [doi: 10.3389/fdgth.2021.594971] [Medline: 34713083]
11. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res. Apr 22, 2021;23(4):e25759. [FREE Full text] [doi: 10.2196/25759] [Medline: 33885365]
12. Wenderott K, Krups J, Luetkens JA, Gambashidze N, Weigl M. Prospective effects of an artificial intelligence-based computer-aided detection system for prostate imaging on routine workflow and radiologists' outcomes. Eur J Radiol. Jan 2024;170:111252. [FREE Full text] [doi: 10.1016/j.ejrad.2023.111252] [Medline: 38096741]
13. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. Jun 19, 2020;22(6):e15154. [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
14. Felmingham CM, Adler NR, Ge Z, Morton RL, Janda M, Mar VJ. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. Am J Clin Dermatol. Mar 22, 2021;22(2):233-242. [doi: 10.1007/s40257-020-00574-4] [Medline: 33354741]
15. Carayon P, Hancock P, Leveson N, Noy I, Sznelwar L, van Hootegem G. Advancing a sociotechnical systems approach to workplace safety--developing the conceptual framework. Ergonomics. Apr 02, 2015;58(4):548-564. [FREE Full text] [doi: 10.1080/00140139.2015.1015623] [Medline: 25831959]
16. Mumford E. The story of socio‐technical design: reflections on its successes, failures and potential. Inf Syst J. Sep 04, 2006;16(4):317-342. [doi: 10.1111/j.1365-2575.2006.00221.x]
17. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care. Dec 2006;15 Suppl 1(Suppl 1):i50-i58. [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
18. Hettinger LJ, Kirlik A, Goh YM, Buckle P. Modelling and simulation of complex sociotechnical systems: envisioning and analysing work environments. Ergonomics. Mar 11, 2015;58(4):600-614. [FREE Full text] [doi: 10.1080/00140139.2015.1008586] [Medline: 25761227]
19. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon. Apr 2020;84:103033. [FREE Full text] [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
20. Wooldridge AR, Carayon P, Hoonakker P, Hose BZ, Eithun B, Brazelton T3, et al. Work system barriers and facilitators in inpatient care transitions of pediatric trauma patients. Appl Ergon. May 2020;85:103059. [FREE Full text] [doi: 10.1016/j.apergo.2020.103059] [Medline: 32174347]
21. Hoonakker PL, Carayon P, Cartmill RS. The impact of secure messaging on workflow in primary care: results of a multiple-case, multiple-method study. Int J Med Inform. Apr 2017;100:63-76. [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.004] [Medline: 28241939]
22. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci. 2009;4:50. [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
23. Damschroder LJ, Reardon CM, Widerquist MA, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci. Oct 29, 2022;17(1):75. [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
24. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. Nov 01, 2017;19(11):e367. [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
25. Abell B, Naicker S, Rodwell D, Donovan T, Tariq A, Baysari M, et al. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement Sci. Jul 26, 2023;18(1):32. [FREE Full text] [doi: 10.1186/s13012-023-01287-y] [Medline: 37495997]
26. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res. Dec 16, 2016;18(12):e317. [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
27. Andersen TO, Nunes F, Wilcox L, Coiera E, Rogers Y. Introduction to the special issue on human-centred AI in healthcare: challenges appearing in the wild. ACM Trans Comput Hum Interact. Jun 30, 2023;30(2):1-12. [doi: 10.1145/3589961]
28. Wenderott K, Gambashidze N, Weigl M. Integration of artificial intelligence into sociotechnical work systems-effects of artificial intelligence solutions in medical imaging on clinical efficiency: protocol for a systematic literature review. JMIR Res Protoc. Dec 01, 2022;11(12):e40485. [FREE Full text] [doi: 10.2196/40485] [Medline: 36454624]

29. Chomutare T, Tejedor M, Svenning TO, Marco-Ruiz L, Tayefi M, Lind K, et al. Artificial intelligence implementation in healthcare: a theory-based scoping review of barriers and facilitators. Int J Environ Res Public Health. Dec 06, 2022;19(23):16359. [FREE Full text] [doi: 10.3390/ijerph192316359] [Medline: 36498432]
30. Han R, Acosta JN, Shakeri Z, Ioannidis JP, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. Lancet Digit Health. May 2024;6(5):e367-e373. [doi: 10.1016/s2589-7500(24)00047-5]
31. Wenderott K, Krups J, Zaruchas F, Weigl M. Effects of artificial intelligence implementation on efficiency in medical imaging-a systematic literature review and meta-analysis. NPJ Digit Med. Sep 30, 2024;7(1):265. [FREE Full text] [doi: 10.1038/s41746-024-01248-9] [Medline: 39349815]
32. Beller EM, Chen JK, Wang UL, Glasziou PP. Are systematic reviews up-to-date at the time of publication? Syst Rev. May 28, 2013;2:36. [FREE Full text] [doi: 10.1186/2046-4053-2-36] [Medline: 23714302]
33. Wenderott K, Krups J, Luetkens JA, Weigl M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. Appl Ergon. May 2024;117:104243. [FREE Full text] [doi: 10.1016/j.apergo.2024.104243] [Medline: 38306741]
34. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. JMIR Hum Factors. Aug 29, 2024;11:e48633. [FREE Full text] [doi: 10.2196/48633] [Medline: 39207831]
35. Lokaj B, Pugliese MT, Kinkel K, Lovis C, Schmid J. Barriers and facilitators of artificial intelligence conception and implementation for breast imaging diagnosis in clinical practice: a scoping review. Eur Radiol. Mar 02, 2024;34(3):2096-2109. [FREE Full text] [doi: 10.1007/s00330-023-10181-6] [Medline: 37658895]
36. Masud R, Al-Rei M, Lokker C. Computer-aided detection for breast cancer screening in clinical settings: scoping review. JMIR Med Inform. Jul 18, 2019;7(3):e12660. [FREE Full text] [doi: 10.2196/12660] [Medline: 31322128]
37. Eltawil FA, Atalla M, Boulos E, Amirabadi A, Tyrrell PN. Analyzing barriers and enablers for the acceptance of artificial intelligence innovations into radiology practice: a scoping review. Tomography. Jul 28, 2023;9(4):1443-1455. [doi: 10.3390/tomography9040115] [Medline: 37624108]
38. Swillens JE, Nagtegaal ID, Engels S, Lugli A, Hermens RP, van der Laak JA. Pathologists' first opinions on barriers and facilitators of computational pathology adoption in oncological pathology: an international study. Oncogene. Sep 16, 2023;42(38):2816-2827. [FREE Full text] [doi: 10.1038/s41388-023-02797-1] [Medline: 37587332]
39. Strohm L, Hehakaya C, Ranschaert ER, Boon WP, Moors EH. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur Radiol. Oct 26, 2020;30(10):5525-5532. [FREE Full text] [doi: 10.1007/s00330-020-06946-y] [Medline: 32458173]
40. Liao X, Yao C, Jin F, Zhang J, Liu L. Barriers and facilitators to implementing imaging-based diagnostic artificial intelligence-assisted decision-making software in hospitals in China: a qualitative study using the updated Consolidated Framework for Implementation Research. BMJ Open. Sep 10, 2024;14(9):e084398. [FREE Full text] [doi: 10.1136/bmjopen-2024-084398] [Medline: 39260855]
41. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. Dec 05, 2016;5(1):210. [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
42. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
43. VERBI software. MAXQDA. URL: https://www.maxqda.com/about [accessed 2023-04-21]
44. RStudio. 2015. URL: http://www.rstudio.com/ [accessed 2025-07-03]
45. Bertram RM, Blase KA, Fixsen DL. Improving programs and outcomes: implementation frameworks and organization change. Res Soc Work Pract. Jun 08, 2014;25(4):477-487. [doi: 10.1177/1049731514537687]
46. Pane ES, Sarno R. Capability maturity model integration (CMMI) for optimizing object-oriented analysis and design (OOAD). Procedia Comput Sci. 2015;72:40-48. [doi: 10.1016/j.procs.2015.12.103]
47. Kuckartz U, Radiker S. Qualitative Content Analysis: Methods, Practice and Software [Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung]. Weinheim, Germany. Beltz Juventa; 2022.
48. Niezen MG, Mathijssen JJ. Reframing professional boundaries in healthcare: a systematic review of facilitators and barriers to task reallocation from the domain of medicine to the nursing domain. Health Policy. Aug 2014;117(2):151-169. [FREE Full text] [doi: 10.1016/j.healthpol.2014.04.016] [Medline: 24857559]
49. Hopf C, Schmidt C. On the Relationship Between Intra-Familial Social Experiences, Personality Development and Political Orientations: Documentation and Discussion of the Methodological Procedure in a Study on This Topic. [Zum Verhältnis von innerfamilialen sozialen Erfahrungen, Persönlichkeitsentwicklung und politischen Orientierungen: Dokumentation und Erörterung des methodischen Vorgehens in einer Studie zu diesem Thema]. Hildesheim, Germany. Institut für Sozialwissenschaften der Universität Hildesheim; 1993.
50. Wooldridge AR, Carayon P, Hundt AS, Hoonakker PL. SEIPS-based process modeling in primary care. Appl Ergon. Apr 2017;60:240-254. [FREE Full text] [doi: 10.1016/j.apergo.2016.11.010] [Medline: 28166883]

51. Wooldridge AR, Carayon P, Shaffer DW, Eagan B. Quantifying the qualitative with epistemic network analysis: a human factors case study of task-allocation communication in a primary care team. IISE Trans Healthc Syst Eng. 2018;8(1):72-82. [FREE Full text] [doi: 10.1080/24725579.2017.1418769] [Medline: 30370395]
52. Weiler DT, Lingg AJ, Eagan BR, Shaffer DW, Werner NE. Quantifying the qualitative: exploring epistemic network analysis as a method to study work system interactions. Ergonomics. Oct 2022;65(10):1434-1449. [FREE Full text] [doi: 10.1080/00140139.2022.2051609] [Medline: 35258441]
53. Shaffer DW, Hatfield D, Svarovsky GN, Nash P, Nulty A, Bagley E, et al. Epistemic network analysis: a prototype for 21st-century assessment of learning. Int J Learn Media. May 2009;1(2):33-53. [FREE Full text] [doi: 10.1162/ijlm.2009.0013]
54. Arastoopour G, Chesler NC, Shaffer DW. Epistemic network analysis as a tool for engineering design assessment. In: Proceedings of the 122nd ASEE Annual Conference & Exposition. 2015. Presented at: ASEE 2015; June 14-17, 2015; Seattle, WA. URL: https://www.researchgate.net/publication/283441969_Epistemic_network_analysis_as_a_tool_for_engineering_design_assessment [doi: 10.18260/p.24016]
55. D'Angelo CM, Clark DB, Shaffer DW. Epistemic network analysis: an alternative analysis technique for complex STEM thinking. In: Proceedings of the National Association of Research on Science Teaching Conference. 2012. Presented at: NARST 2012; March 25-28, 2012; Indianapolis, Indiana. [doi: 10.4324/9781315617572-50]
56. Shaffer DW, Collier W, Ruis AR. A tutorial on epistemic network analysis: analyzing the structure of connections in cognitive, social, and interaction data. J Learn Anal. Dec 19, 2016;3(3):9-45. [doi: 10.18608/jla.2016.33.3]
57. Andrist S, Collier W, Gleicher M, Mutlu B, Shaffer D. Look together: analyzing gaze coordination with epistemic network analysis. Front Psychol. Jul 21, 2015;6:1016. [FREE Full text] [doi: 10.3389/fpsyg.2015.01016] [Medline: 26257677]
58. Bowman D, Swiecki Z, Cai Z, Wang Y, Eagan B, Linderoth J, et al. The mathematical foundations of epistemic network analysis. In: Proceedings of the Second International Conference on Advances in Quantitative Ethnography. 2021. Presented at: ICQE 2020; February 1-3, 2021; Malibu, CA. [doi: 10.1007/978-3-030-67788-6_7]
59. Shaffer DW. Quantitative Ethnography. Charlottesville, VA. Cathcart Press; 2017.
60. Shaffer DW, Ruis AR. Epistemic network analysis: a worked example of theory-based learning analytics. In: Lang C, Siemens G, Wise A, Grasevic D, editors. Handbook of Learning Analytics. Beaumont, AB. Society for Learning Analytics Research; 2017:175-187.
61. Marquart CL, Hinojosa C, Swiecki Z, Shaffer DW. Epistemic network analysis. Epistemic Network. 2018. URL: https://app.epistemicnetwork.org/login.html [accessed 2025-07-03]
62. Romagnoli A, Ferrara F, Langella R, Zovi A. Healthcare systems and artificial intelligence: focus on challenges and the international regulatory framework. Pharm Res. Apr 05, 2024;41(4):721-730. [doi: 10.1007/s11095-024-03685-3] [Medline: 38443632]
63. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digit Health. Mar 2021;3(3):e195-e203. [FREE Full text] [doi: 10.1016/S2589-7500(20)30292-2] [Medline: 33478929]
64. Vokinger KN, Gasser U. Regulating AI in medicine in the United States and Europe. Nat Mach Intell. Sep 2021;3(9):738-739. [FREE Full text] [doi: 10.1038/s42256-021-00386-z] [Medline: 34604702]
65. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. Oct 12, 2016;355:i4919. [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
66. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. Aug 28, 2019;366:l4898. [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
67. Arbabshirani MR, Fornwalt BK, Mongelluzzo GJ, Suever JD, Geise BD, Patel AA, et al. Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. NPJ Digit Med. Apr 4, 2018;1(1):9. [FREE Full text] [doi: 10.1038/s41746-017-0015-z] [Medline: 31304294]
68. Batra K, Xi Y, Bhagwat S, Espino A, Peshock RM. Radiologist worklist reprioritization using artificial intelligence: impact on report turnaround times for CTPA examinations positive for acute pulmonary embolism. Am J Roentgenol. Sep 2023;221(3):324-333. [doi: 10.2214/ajr.22.28949]
69. Carlile M, Hurt B, Hsiao A, Hogarth M, Longhurst CA, Dameff C. Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department. J Am Coll Emerg Physicians Open. Dec 05, 2020;1(6):1459-1464. [FREE Full text] [doi: 10.1002/emp2.12297] [Medline: 33392549]
70. Cha E, Elguindi S, Onochie I, Gorovets D, Deasy JO, Zelefsky M, et al. Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy. Radiother Oncol. Jun 2021;159:1-7. [FREE Full text] [doi: 10.1016/j.radonc.2021.02.040] [Medline: 33667591]
71. Cheikh AB, Gorincour G, Nivet H, May J, Seux M, Calame P, et al. How artificial intelligence improves radiological interpretation in suspected pulmonary embolism. Eur Radiol. Sep 22, 2022;32(9):5831-5842. [FREE Full text] [doi: 10.1007/s00330-022-08645-2] [Medline: 35316363]
72. Chen W, Wu J, Wei R, Wu S, Xia C, Wang D, et al. Improving the diagnosis of acute ischemic stroke on non-contrast CT using deep learning: a multicenter study. Insights Imaging. Dec 06, 2022;13(1):184. [FREE Full text] [doi: 10.1186/s13244-022-01331-3] [Medline: 36471022]

73. Conant EF, Toledano AY, Periaswamy S, Fotin SV, Go J, Boatsman JE, et al. Improving accuracy and efficiency with concurrent use of artificial intelligence for digital breast tomosynthesis. Radiol Artif Intell. Jul 31, 2019;1(4):e180096. [FREE Full text] [doi: 10.1148/ryai.2019180096] [Medline: 32076660]
74. Davis MA, Rao B, Cedeno PA, Saha A, Zohrabian VM. Machine learning and improved quality metrics in acute intracranial hemorrhage by noncontrast computed tomography. Curr Probl Diagn Radiol. Jul 2022;51(4):556-561. [doi: 10.1067/j.cpradiol.2020.10.007] [Medline: 33243455]
75. Diao K, Chen Y, Liu Y, Chen BJ, Li WJ, Zhang L, et al. Diagnostic study on clinical feasibility of an AI-based diagnostic system as a second reader on mobile CT images: a preliminary result. Ann Transl Med. Jun 2022;10(12):668. [FREE Full text] [doi: 10.21037/atm-22-2157] [Medline: 35845492]
76. Duron L, Ducarouge A, Gillibert A, Lainé J, Allouche C, Cherel N, et al. Assessment of an AI aid in detection of adult appendicular skeletal fractures by emergency physicians and radiologists: a multicenter cross-sectional diagnostic study. Radiology. Jul 2021;300(1):120-129. [doi: 10.1148/radiol.2021203886] [Medline: 33944629]
77. Elijovich L, Dornbos Iii D, Nickele C, Alexandrov A, Inoa-Acosta V, Arthur AS, et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. J Neurointerv Surg. Jul 20, 2022;14(7):704-708. [doi: 10.1136/neurintsurg-2021-017714] [Medline: 34417344]
78. Ginat D. Implementation of machine learning software on the radiology worklist decreases scan view delay for the detection of intracranial hemorrhage on CT. Brain Sci. Jun 23, 2021;11(7):832. [FREE Full text] [doi: 10.3390/brainsci11070832] [Medline: 34201775]
79. Hassan AE, Ringheanu VM, Tekle WG. The implementation of artificial intelligence significantly reduces door-in-door-out times in a primary care center prior to transfer. Interv Neuroradiol. Dec 2023;29(6):631-636. [doi: 10.1177/15910199221122848] [Medline: 36017543]
80. Jones CM, Danaher L, Milne MR, Tang C, Seah J, Oakden-Rayner L, et al. Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. BMJ Open. Dec 20, 2021;11(12):e052902. [FREE Full text] [doi: 10.1136/bmjopen-2021-052902] [Medline: 34930738]
81. Ladabaum U, Shepard J, Weng Y, Desai M, Singer S, Mannalithara A. Computer-aided detection of polyps does not improve colonoscopist performance in a pragmatic implementation trial. Gastroenterology. May 2023;164(6):S-152-S-153. [doi: 10.1016/s0016-5085(23)01329-x]
82. Levy I, Bruckmayer L, Klang E, Ben-Horin S, Kopylov U. Artificial intelligence-aided colonoscopy does not increase adenoma detection rate in routine clinical practice. Am J Gastroenterol. Nov 01, 2022;117(11):1871-1873. [doi: 10.14309/ajg.0000000000001970] [Medline: 36001408]
83. Marwaha A, Chitayat D, Meyn MS, Mendoza-Londono R, Chad L. The point-of-care use of a facial phenotyping tool in the genetics clinic: enhancing diagnosis and education with machine learning. Am J Med Genet A. Apr 08, 2021;185(4):1151-1158. [doi: 10.1002/ajmg.a.62092] [Medline: 33554457]
84. Müller FC, Raaschou H, Akhtar N, Brejnebøl M, Collatz L, Andersen MB. Impact of concurrent use of artificial intelligence tools on radiologists reading time: a prospective feasibility study. Acad Radiol. Jul 2022;29(7):1085-1090. [FREE Full text] [doi: 10.1016/j.acra.2021.10.008] [Medline: 34801345]
85. Nehme F, Coronel E, Barringer DA, Romero LG, Shafi MA, Ross WA, et al. Performance and attitudes toward real-time computer-aided polyp detection during colonoscopy in a large tertiary referral center in the United States. Gastrointest Endosc. Jul 2023;98(1):100-9.e6. [doi: 10.1016/j.gie.2023.02.016] [Medline: 36801459]
86. Oppenheimer J, Lüken S, Hamm B, Niehues SM. A prospective approach to integration of AI fracture detection software in radiographs into clinical workflow. Life (Basel). Jan 13, 2023;13(1):223. [FREE Full text] [doi: 10.3390/life13010223] [Medline: 36676172]
87. Pierce JD, Rosipko B, Youngblood L, Gilkeson RC, Gupta A, Bittencourt LK. Seamless integration of artificial intelligence into the clinical environment: our experience with a novel pneumothorax detection artificial intelligence algorithm. J Am Coll Radiol. Nov 2021;18(11):1497-1505. [doi: 10.1016/j.jacr.2021.08.023] [Medline: 34597622]
88. Potretzke TA, Korfiatis P, Blezek DJ, Edwards ME, Klug JR, Cook CJ, et al. Clinical implementation of an artificial intelligence algorithm for magnetic resonance-derived measurement of total kidney volume. Mayo Clin Proc. May 2023;98(5):689-700. [FREE Full text] [doi: 10.1016/j.mayocp.2022.12.019] [Medline: 36931980]
89. Quan SY, Wei MT, Lee J, Mohi-Ud-Din R, Mostaghim R, Sachdev R, et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. Sci Rep. Apr 21, 2022;12(1):6598. [FREE Full text] [doi: 10.1038/s41598-022-10597-y] [Medline: 35449442]
90. Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M. AI-based strategies to reduce workload in breast cancer screening with mammography and tomosynthesis: a retrospective evaluation. Radiology. Jul 2021;300(1):57-65. [FREE Full text] [doi: 10.1148/radiol.2021203555] [Medline: 33944627]
91. Ruamviboonsuk P, Tiwari R, Sayres R, Nganthavee V, Hemarat K, Kongprayoon A, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. Lancet Digit Health. Apr 2022;4(4):e235-e244. [FREE Full text] [doi: 10.1016/S2589-7500(22)00017-6] [Medline: 35272972]

92. Sandbank J, Bataillon G, Nudelman A, Krasnitsky I, Mikulinsky R, Bien L, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. NPJ Breast Cancer. Dec 06, 2022;8(1):129. [FREE Full text] [doi: 10.1038/s41523-022-00496-w] [Medline: 36473870]
93. Schmuelling L, Franzeck FC, Nickel CH, Mansella G, Bingisser R, Schmidt N, et al. Deep learning-based automated detection of pulmonary embolism on CT pulmonary angiograms: no significant effects on report communication times and patient turnaround in the emergency department nine months after technical implementation. Eur J Radiol. Aug 2021;141:109816. [FREE Full text] [doi: 10.1016/j.ejrad.2021.109816] [Medline: 34157638]
94. Seyam M, Weikert T, Sauter A, Brehm A, Psychogios MN, Blackham KA. Utilization of artificial intelligence-based intracranial hemorrhage detection on emergent noncontrast CT images in clinical workflow. Radiol Artif Intell. Mar 01, 2022;4(2):e210168. [FREE Full text] [doi: 10.1148/ryai.210168] [Medline: 35391777]
95. Tchou PM, Haygood TM, Atkinson EN, Stephens TW, Davis PL, Arribas EM, et al. Interpretation time of computer-aided detection at screening mammography. Radiology. Oct 2010;257(1):40-46. [doi: 10.1148/radiol.10092170] [Medline: 20679448]
96. Tricarico D, Calandri M, Barba M, Piatti C, Geninatti C, Basile D, et al. Convolutional neural network-based automatic analysis of chest radiographs for the detection of COVID-19 pneumonia: a prioritizing tool in the emergency department, phase I study and preliminary "real life" results. Diagnostics (Basel). Feb 23, 2022;12(3):570. [FREE Full text] [doi: 10.3390/diagnostics12030570] [Medline: 35328122]
97. Vassallo L, Traverso A, Agnello M, Bracco C, Campanella D, Chiara G, et al. A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies. Eur Radiol. Jan 15, 2019;29(1):144-152. [doi: 10.1007/s00330-018-5528-6] [Medline: 29948089]
98. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut. Oct 27, 2019;68(10):1813-1819. [FREE Full text] [doi: 10.1136/gutjnl-2018-317500] [Medline: 30814121]
99. Wang M, Xia C, Huang L, Xu S, Qin C, Liu J, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. Lancet Digit Health. Oct 2020;2(10):e506-e515. [FREE Full text] [doi: 10.1016/S2589-7500(20)30199-0] [Medline: 32984796]
100. Wittenberg R, Berger FH, Peters JF, Weber M, van Hoorn F, Beenen LF, et al. Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis--an observer study. Radiology. Jan 2012;262(1):305-313. [doi: 10.1148/radiol.11110372] [Medline: 22190659]
101. Wong J, Huang V, Wells D, Giambattista J, Giambattista J, Kolbeck C, et al. Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. Radiat Oncol. Jun 08, 2021;16(1):101. [FREE Full text] [doi: 10.1186/s13014-021-01831-4] [Medline: 34103062]
102. Wong KP, Homer SY, Wei SH, Yaghmai N, Estrada Paz OA, Young TJ, et al. Integration and evaluation of chest X-ray artificial intelligence in clinical practice. J Med Imaging (Bellingham). Sep 2023;10(5):051805. [FREE Full text] [doi: 10.1117/1.JMI.10.5.051805] [Medline: 37113505]
103. Yang Y, Pan J, Yuan M, Lai K, Xie H, Ma L, et al. Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study. Ann Transl Med. Oct 2022;10(20):1088. [FREE Full text] [doi: 10.21037/atm-22-350] [Medline: 36388839]
104. Zia A, Fletcher C, Bigwood S, Ratnakanthan P, Seah J, Lee R, et al. Retrospective analysis and prospective validation of an AI-based software for intracranial haemorrhage detection at a high-volume trauma centre. Sci Rep. Nov 18, 2022;12(1):19885. [FREE Full text] [doi: 10.1038/s41598-022-24504-y] [Medline: 36400834]
105. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform. Feb 2010;43(1):159-172. [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
106. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q. 2003;27(3):425-478. [doi: 10.2307/30036540]
107. Chanda T, Hauser K, Hobelsberger S, Bucher TC, Garcia CN, Wies C, Reader Study Consortium, et al. Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. Nat Commun. Jan 15, 2024;15(1):524. [FREE Full text] [doi: 10.1038/s41467-023-43095-4] [Medline: 38225244]
108. Salwei ME, Carayon P, Hoonakker PL, Hundt AS, Wiegmann D, Pulia M, et al. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. Appl Ergon. Nov 2021;97:103498. [FREE Full text] [doi: 10.1016/j.apergo.2021.103498] [Medline: 34182430]
109. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak. Dec 11, 2022;16(4):194-206. [FREE Full text] [doi: 10.1177/15553434221097357] [Medline: 36704421]
110. Godoe P, Johansen TS. Understanding adoption of new technologies: technology readiness and technology acceptance as an integrated concept. J Eur Psychol Students. May 06, 2012;3:38. [doi: 10.5334/jeps.aq]
111. Wooldridge AR, Morgan J, Ramadhani WA, Hanson K, Vazquez-Melendez E, Kendhari H, et al. Interactions in sociotechnical systems: achieving balance in the use of an augmented reality mobile application. Hum Factors. Mar 2024;66(3):658-682. [doi: 10.1177/00187208221093830] [Medline: 35549474]

112. Marco-Ruiz L, Hernández MÁ, Ngo PD, Makhlysheva A, Svenning TO, Dyb K, et al. A multinational study on artificial intelligence adoption: clinical implementers' perspectives. Int J Med Inform. Apr 2024;184:105377. [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105377] [Medline: 38377725]
113. Gunasekeran DV, Zheng F, Lim GY, Chong CC, Zhang S, Ng WY, et al. Acceptance and perception of artificial intelligence usability in eye care (APPRAISE) for ophthalmologists: a multinational perspective. Front Med (Lausanne). Oct 13, 2022;9:875242. [FREE Full text] [doi: 10.3389/fmed.2022.875242] [Medline: 36314006]
114. Lennon MR, Bouamrane MM, Devlin AM, O'Connor S, O'Donnell C, Chetty U, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res. Feb 16, 2017;19(2):e42. [FREE Full text] [doi: 10.2196/jmir.6900] [Medline: 28209558]
115. Widner K, Virmani S, Krause J, Nayar J, Tiwari R, Pedersen ER, et al. Lessons learned from translating AI from development to deployment in healthcare. Nat Med. Jun 29, 2023;29(6):1304-1306. [doi: 10.1038/s41591-023-02293-9] [Medline: 37248297]
116. Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. NPJ Digit Med. Oct 19, 2022;5(1):156. [FREE Full text] [doi: 10.1038/s41746-022-00699-2] [Medline: 36261476]
117. Herrmann T, Pfeiffer S. Keeping the organization in the loop: a socio-technical extension of human-centered artificial intelligence. AI Soc. Feb 18, 2022;38(4):1523-1542. [doi: 10.1007/s00146-022-01391-5]
118. Chen Y, Clayton EW, Novak LL, Anders S, Malin B. Human-centered design to address biases in artificial intelligence. J Med Internet Res. Mar 24, 2023;25:e43251. [FREE Full text] [doi: 10.2196/43251] [Medline: 36961506]
119. Koch A, Burns J, Catchpole K, Weigl M. Associations of workflow disruptions in the operating room with surgical outcomes: a systematic review and narrative synthesis. BMJ Qual Saf. Dec 23, 2020;29(12):1033-1045. [doi: 10.1136/bmjqs-2019-010639] [Medline: 32447319]
120. Gore JC. Artificial intelligence in medical imaging. Magn Reson Imaging. May 2020;68:A1-A4. [doi: 10.1016/j.mri.2019.12.006] [Medline: 31857130]
121. Mennella C, Maniscalco U, De Pietro G, Esposito M. Ethical and regulatory challenges of AI technologies in healthcare: a narrative review. Heliyon. Feb 29, 2024;10(4):e26297. [FREE Full text] [doi: 10.1016/j.heliyon.2024.e26297] [Medline: 38384518]
122. Ergonomics of human-system interaction: part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2025-01-14]
123. Kruse CS, Kothman K, Anerobi K, Abanaka L. Adoption factors of the electronic health record: a systematic review. JMIR Med Inform. Jun 01, 2016;4(2):e19. [FREE Full text] [doi: 10.2196/medinform.5525] [Medline: 27251559]
124. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst. Feb 2012;36(1):241-277. [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
125. Gagnon MP, Nsangou ER, Payne-Gagnon J, Grenier S, Sicotte C. Barriers and facilitators to implementing electronic prescription: a systematic review of user groups' perceptions. J Am Med Inform Assoc. 2014;21(3):535-541. [FREE Full text] [doi: 10.1136/amiajnl-2013-002203] [Medline: 24130232]
126. Sidek YH, Martins JT. Perceived critical success factors of electronic health record system implementation in a dental clinic context: an organisational management perspective. Int J Med Inform. Nov 2017;107:88-100. [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.08.007] [Medline: 29029696]
127. Fragidis LL, Chatzoglou PD. Implementation of a nationwide electronic health record (EHR): the international experience in 13 countries. Int J Health Care Qual Assur. Mar 12, 2018;31(2):116-130. [FREE Full text] [doi: 10.1108/IJHCQA-09-2016-0136] [Medline: 29504871]
128. Yen PY, McAlearney AS, Sieck CJ, Hefner JL, Huerta TR. Health information technology (HIT) adaptation: refocusing on the journey to successful HIT implementation. JMIR Med Inform. Sep 07, 2017;5(3):e28. [FREE Full text] [doi: 10.2196/medinform.7476] [Medline: 28882812]
129. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res. Jan 27, 2022;24(1):e32215. [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
130. Garvey KV, Thomas Craig KJ, Russell R, Novak LL, Moore D, Miller BM. Considering clinician competencies for the implementation of artificial intelligence-based tools in health care: findings from a scoping review. JMIR Med Inform. Nov 16, 2022;10(11):e37478. [FREE Full text] [doi: 10.2196/37478] [Medline: 36318697]
131. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. Patterns (N Y). Oct 08, 2021;2(10):100347. [FREE Full text] [doi: 10.1016/j.patter.2021.100347] [Medline: 34693373]
132. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med. Jun 14, 2023;6(1):113. [FREE Full text] [doi: 10.1038/s41746-023-00858-z] [Medline: 37311802]

133. Ratwani RM, Sutton K, Galarraga JE. Addressing AI algorithmic bias in health care. JAMA. Oct 01, 2024;332(13):1051-1052. [doi: 10.1001/jama.2024.13486] [Medline: 39230911]
134. Naik N, Hameed BM, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg. Mar 14, 2022;9:862322. [FREE Full text] [doi: 10.3389/fsurg.2022.862322] [Medline: 35360424]
135. Terranova C, Cestonaro C, Fava L, Cinquetti A. AI and professional liability assessment in healthcare. A revolution in legal medicine? Front Med (Lausanne). Jan 8, 2023;10:1337335. [FREE Full text] [doi: 10.3389/fmed.2023.1337335] [Medline: 38259835]
136. Eldakak A, Alremeithi A, Dahiyat E, El-Gheriani M, Mohamed H, Abdulrahim Abdulla MI. Civil liability for the actions of autonomous AI in healthcare: an invitation to further contemplation. Humanit Soc Sci Commun. Feb 23, 2024;11:305. [doi: 10.1057/s41599-024-02806-y]

Abbreviations

AI: artificial intelligence
CFIR: Consolidated Framework for Implementation Research
ENA: epistemic network analysis
IDEAS: integrate, design, assess, and share
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
PICO: population, intervention, comparison, and outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SEIPS: Systems Engineering Initiative for Patient Safety

Edited by Y Li; submitted 25.06.24; peer-reviewed by S Antani, B Mesko, T Donovan; comments to author 26.11.24; revised version received 15.01.25; accepted 15.05.25; published 21.07.25

Please cite as:
Wenderott K, Krups J, Weigl M, Wooldridge AR
Facilitators and Barriers to Implementing AI in Routine Medical Imaging: Systematic Review and Qualitative Analysis
J Med Internet Res 2025;27:e63649
URL: https://www.jmir.org/2025/1/e63649
doi: 10.2196/63649
PMID:

3.4. Publication 4: Prospective Effects of an Artificial Intelligence-Based Computer-Aided Detection System for Prostate Imaging on Routine Workflow and Radiologists' Outcomes

Wenderott K, Krups J, Luetkens J A, Gambashidze N, Weigl M (2024)
European Journal of Radiology, 170, 111252. doi: 10.1016/j.ejrad.2023.111252

The supplementary material can be accessed via this link.

European Journal of Radiology 170 (2024) 111252

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Prospective effects of an artificial intelligence-based computer-aided detection system for prostate imaging on routine workflow and radiologists' outcomes

Katharina Wenderott [a,*], Jim Krups [a], Julian A. Luetkens [b], Nikoloz Gambashidze [a], Matthias Weigl [a]

[a] Institute for Patient Safety, University Hospital Bonn, Germany
[b] Department of Radiology and Quantitative Imaging Lab Bonn (QILaB), University Hospital Bonn, Germany

ARTICLE INFO

Keywords:
Artificial intelligence
Magnetic resonance imaging
Workflow
Prostate

ABSTRACT

Objectives: Artificial intelligence (AI) is expected to alleviate the negative consequences of rising case numbers for radiologists. Currently, systematic evaluations of the impact of AI solutions in real-world radiological practice are missing. Our study addresses this gap by investigating the impact of the clinical implementation of an AI-based computer-aided detection system (CAD) for prostate MRI reading on clinicians' workflow, workflow throughput times, workload, and stress.
Materials and methods: CAD was newly implemented into radiology workflow and accompanied by a prospective pre-post study design. We assessed prostate MRI case readings using standardized work observations and questionnaires. The observation period was three months each in a single department. Workflow throughput times, PI-RADS score, CAD usage and radiologists' self-reported workload and stress were recorded. Linear mixed models were employed for effect identification.
Results: In data analyses, 91 observed case readings (pre: 50, post: 41) were included. Variation of routine workflow was observed following CAD implementation. A non-significant increase in overall workflow throughput time was associated with CAD implementation (mean 16.99 ± 6.21 vs 18.77 ± 9.69 *min*, $p = .51$), along with an increase in diagnostic reading time for high suspicion cases (mean 15.73 ± 4.99 vs 23.07 ± 8.75 *min*, $p = .02$). Changes in radiologists' self-reported workload or stress were not found.
Conclusion: Implementation of an AI-based detection aid was associated with lower standardization and no effects over time on radiologists' workload or stress. Expectations of AI decreasing the workload of radiologists were not confirmed by our real-world study.
Pre-Registration: German register for clinical trials https://drks.de/; DRKS00027391.

1. Introduction

Artificial intelligence (AI) is increasingly utilized and integrated into medical practice, especially in largely image-based disciplines such as radiology [1,2]. Despite the previous advancements in AI technologies, there is a significant knowledge gap concerning the actual implementation of AI technologies in clinical routines [3,4].

In radiology, there are many areas for applying AI, i.e. automating tasks that were previously reserved for humans such as medical image interpretation, quality evaluation or providing personalized reports [5,6]. AI is expected to improve efficiency, diagnostic accuracy, and standardization while decreasing radiologists' workload, costs or the number of errors [5,7,8]. Workload issues in radiology are pressing due to increased case volumes and enhanced image quality, coupled with limited trained radiologists, resulting in high workforce demands [7,9]. Next to image interpretation, radiologists have multiple other

Abbreviations: AI, Artificial intelligence; CAD, Computer-aided detection system; CE, Conformité Européenne; DWI, Diffusion weighted imaging; FDA, Food and Drug Administration; NASA-TLX, National Aeronautics and Space Administration-Task Load Index; PI-RADS, Prostate Imaging Reporting and Data System Guidelines; PSA, Prostate specific antigen; STAI, Spielberger State – Trait Anxiety Inventory; WTT, Workflow throughput time.

* Corresponding author at: Universitätsklinikum Bonn, Venusberg-Campus 1, Building A02, 53127 Bonn, Germany.
E-mail address: katharina.wenderott@ukbonn.de (K. Wenderott).

https://doi.org/10.1016/j.ejrad.2023.111252
Received 18 October 2023; Received in revised form 15 November 2023; Accepted 4 December 2023
Available online 6 December 2023

concurrent tasks such as reporting, consultations, or direct patient interaction [10,11].

Current research on AI applications in radiology is not taking the complexity of this work setting into account. Available studies focus mostly on model validation or diagnostic accuracy while being conducted retrospectively or in experimental settings [12,13]. When integrating AI in existent clinical workflows, it interacts with several components of the socio-technical work system, with important influences on radiologists [14]. To analyze how AI solutions improve healthcare delivery, it is imperative to evaluate not only the technologies themselves but also their actual implementation into clinical work systems [6,15].

To address this current lack of studies targeting real-world AI applications in everyday clinical practice and to respond to the demand for prospective studies on AI integration in radiology displayed [6,8,16], we thus studied the implementation process of an AI software into radiologists' routine workflow. As a use case, we focused on an AI-based computer-aided detection system (CAD) for the reading of prostate MRI scans. Prostate cancer is the second most frequent cancer type in men and the amount of cases is growing, with first AI solutions being released to meet rising diagnostic demands [13,17]. Given the increasing introduction of AI-facilitated diagnostic and decision-support tools in everyday radiological practice, empirical insights into work practices and user experiences in actual clinical settings are necessary. We hypothesized that the CAD introduction alters the workflow and impacts workflow throughput times per case reading thus affecting clinicians' workload and stress.

2. Materials and methods

2.1. Ethics committee approval

The study was part of a larger project which was reviewed by the Ethics committee of the medical faculty at the University of Bonn (Nr. 449/21). All participants provided informed consent. The study protocol was pre-registered in the German Register for Clinical Trials (DRKS00027391).

2.2. Study design

This study's reporting follows Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [18]. A prospective pre-post-design utilizing an interrupted time series was established to accompany the CAD implementation in a radiology department with a mixed-methods assessment. The department has an annual throughput of around 650 prostate MRI scans. We assessed the workflow by combining standardized expert observation with radiologists' self-reports for workload and stress. Data collection before implementation was scheduled on 17 days (12/2021–02/2022) and post-implementation within 16 days (06–09/2022). After CAD implementation and two months to familiarize with the system, data collection was repeated. Patient cases included incoming patients, who were not preselected, aligning with the study's intention to avoid altering everyday work practices. Patient cases were only read either during pre- or post-implementation i.e. only in one study phase, as the diagnostic reading was part of the routine patient journey.

The CAD system Quantib® Prostate (Version 1.2.0, Quantib® BV), a web-based deep-learning prostate MRI reading and reporting platform, which is FDA-cleared and CE-marked, was implemented in a radiology department. The CAD system offered a semi-automated process support, and aided prostate segmentation, calculating PSA density, creating heatmaps highlighting regions of interest, and guiding the PI-RADS assessment through a structured questionnaire. Additionally, it generated automatic patient reports based on the marked lesions and assessed PI-RADS scores. A detailed description of Quantib® Prostate was presented by Forookhi et al. [19]. Radiology residents received a training on Quantib® Prostate in the course of the CAD implementation.

2.3. Participants

This single-center study was conducted in the Radiology Department of the University Hospital Bonn, Germany. Eligible participants (N = 10) were radiology residents conducting prostate MRI readings, chosen through a convenience sampling approach. The study focused on residents as they were responsible for prostate MRI readings and was conducted within their routine work practices. We ran a-priori power analysis using G*Power version 3.1.9.7 [20] to determine the minimum sample size. To achieve 80 % power for detecting a medium effect, at a significance criterion of $\alpha = 0.05$, a sample size of $N = 54$ case readings for linear multiple regression with random effects was calculated.

2.4. Materials

2.4.1. Procedure of expert observations and questionnaire application

On-site expert observations were used to assess the workflow during pre- and post-CAD implementation. A pilot observation was conducted prior and local senior radiologists were consulted for testing observation procedure and defining specific time markers for the routine workflow. Three task steps with observable time markers were defined: preparation, diagnostic reading, and review. The preparation and the diagnostic reading task step were observed in our study, as the review with the responsible senior radiologist did not happen consecutively. After every case reading, the observed radiology resident was asked to fill in a short questionnaire assessing workload and stress. An overview of the process is presented in Fig. 1. As part of the larger project, radiologists were also interviewed on their work experience.

The observers were two trained researchers with a background in human factors and work psychology from the study team (KW and JK). Before starting, both had three to five training sessions on using the observer guideline. Times were taken with a digital stopwatch (Renkforce RF-SW-110).

2.4.2. Outcomes

2.4.2.1. Observational workflow measures. Expert observations measured the workflow throughput time (WTT) for the prostate MRI case readings. To assess use and effect of the CAD in the different task steps, separate WTT for preparation and diagnostic reading were noted (Fig. 1). Duration of interruptions was recorded to adjust the WTT. The highest PI-RADS score of the respective patient was also noted. Post implementation, observers noted when and how the CAD was used.

2.4.2.2. Radiologists' outcome reports. Immediately after each finalized case reading, radiologists were surveyed with a short, standardized questionnaire consisting of 12 questions (Appendix 1). A short, self-generated identifier was used to track radiologists over time. The questionnaire contained the German version of the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) with six items assessing workload due to the performed task with a scale from 0 to 100 [21]. Radiologists' stress was measured using the short version of the Spielberger State-Trait Anxiety Inventory (STAI) with six items on a scale from 0 to 3 [22]. It indicates cognitive, emotional, and physical stress at work [23].

2.5. Statistical analysis

First, all WTT data was corrected for the duration of interruptions. Mean scores of NASA-TLX and STAI were calculated. Case readings with irregularities, such as technical problems or missing images, or interruptions of more than 45 min were excluded. Data was checked for outliers via the interquartile range rule [24]. Case readings with missing

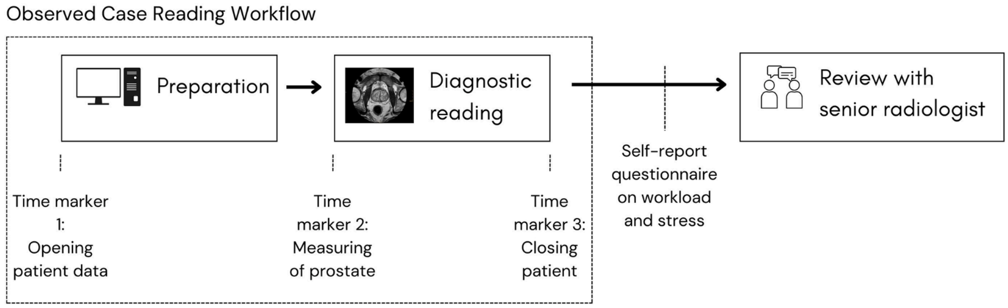


Fig. 1. Procedure of expert observations of prostate MRI case reading workflow: Radiologist's workflow assessment, task steps, and respective observable time.

values in the preparation task step were included in the overall WTT analysis using imputation with the mean of all case readings in this task step, to limit the loss of data. We performed pre-specified separate subgroup analyses for the two task steps (preparation and diagnostic reading).

For our main analyses, linear mixed models with random effects for radiologists were used. Predictors were CAD implementation and PI-RADS score. Additionally, the interaction between CAD implementation and PI-RADS score was included, to explore whether the effect of CAD implementation differed in relation to the case's PI-RADS score. Outcome variables were WTT corrected for interruptions, workload, and stress. The WTT corrected for interruptions and stress scale were transformed using logarithmic transformation to better fit the model assumptions. Statistical evaluation was conducted using RStudio (Version 2022.12.0, RStudio Team, 2020) [25]. Code for statistical analysis is available on GitHub (https://github.com/katwend/CAD_WorkflowRadiology).

3. Results

3.1. Participant and patient characteristics

Altogether, 106 case readings were observed, which were performed by nine radiology residents. A flow chart of the observations is presented in Fig. 2. Regarding the work experience of the radiologists, two were in the second year of their residency, three each in the third and fourth year, and one person in the fifth year. We did not record any details besides the work experience of the radiologists to ensure anonymity. For 33 days, observers were on-site in the department. On 9 days, no observations were conducted, due to no-show of patients, non-participation, or technical problems. Prior to CAD implementation, 50 case readings were conducted by five radiology residents, and post-CAD implementation, 56 performed by seven radiology residents were recorded. 16 observations post-CAD implementation were excluded due to technical issues, long interruptions or being identified as outliers via the interquartile range rule [24]. Finally, 41 observations post-CAD implementation were analyzed.

3.2. Descriptive data

Descriptive data of study variables are depicted in Table 1. We found no difference in mean PI-RADS scores between pre- and post-implementation case readings ($t(87.85) = -0.68$, $p = .50$), nor for interruption rates per hour ($t(81.22) = 1.85$, $p = .06$), therefore assuming that difficulty levels were comparable across pre- and post-implementation. Six case readings were observed by both observers and interrater reliability was calculated. There was a very good agreement between the two raters in the double-coded case readings, using the two-way random effect models and "single rater" unit, kappa = 1, $p < .001$.

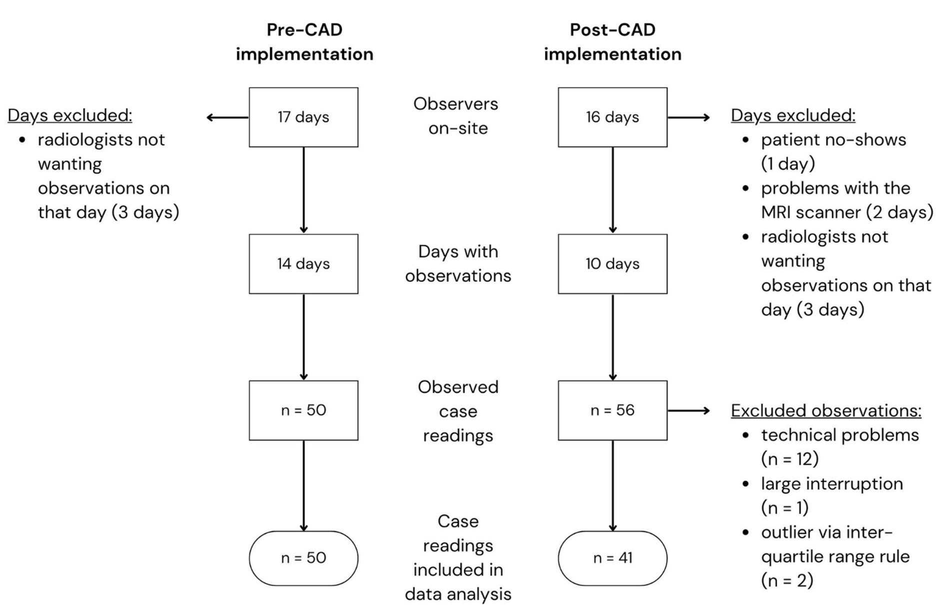


Fig. 2. Flowchart of observations.

Table 1
Descriptive data for observational and radiologists' self-reported outcomes.

	Pre-implementation			Post-implementation		
	n	M	SD	n	M	SD
Total workflow throughput time (min)[a]	50	16.99	6.21	41	18.77	9.69
WTT: Preparation (min)[a]	43	3.23	2.93	40	2.96	3.25
WTT: Diagnostic Reading (min)[a]	50	13.77	5.22	41	15.81	8.72
Workload (TLX Mean)	50	34.98	13.78	41	25.41	21.75
Stress (STAI Mean)	50	1.32	0.26	39	1.38	0.25
Interruption rate (per h)	50	10.17	8.43	41	13.67	9.41
PI-RADS score	50	3.02	1.74	41	2.81	1.27

n Number of analyzed cases, *M* Mean, *SD* Standard Deviation, *WTT* workflow throughput time, *TLX* National Aeronautics and Space Administration-Task Load Index, *STAI* Spielberger State-Trait Anxiety Inventory.
[a]corrected for duration of interruptions.

3.3. Workflow observation

The pre-implementation MRI case reading workflow consists of various steps including marking of possible lesions, followed by the calculation of prostate-specific antigen (PSA) density, analysis of diffusion-weighted imaging (DWI), and PI-RADS classification. After finishing the diagnostic reading, radiologists are drafting the report and upload it into the system (see Fig. 3A).

Since the transfer and upload of the MRI images into the CAD system took between 10 and 20 min, in most case readings (n = 36) radiologists used the CAD as a second opinion after going through the conventional diagnostic reading workflow during the upload time (see Fig. 3B). In five observed case readings, however, radiologists bridged the upload time by performing other clinical tasks and used the CAD during the diagnostic reading for lesion detection and segmentation as well as calculation of PSA density, DWI analysis, and PI-RADS classification (see Fig. 3C). All cases were reviewed with the responsible senior radiologist, which we did not observe (cf. Procedure and Fig. 1).

3.4. Mixed-linear models for effect identification

3.4.1. Workflow throughput times per case reading

Detailed results of the mixed linear regression with workflow throughput time per case reading as outcome variable can be found in Table 2. CAD implementation did not predict the WTT (p =.51). The PI-RADS score significantly predicted WTT (p <.01). We found no effect of the interaction between the PI-RADS score and the CAD implementation (p =.06).

Table 2
Results of linear mixed model with workflow throughput time per case reading as outcome variable (n = 91).

Predictor	b	*SE*(b)	*t*	*p*
(Intercept)	6.57	0.13	48.90	0.00
CAD implementation	0.09	0.13	0.67	0.51
PI-RADS	0.18	0.03	5.60	<0.01
CAD implementation x PI-RADS	−0.07	0.04	−1.89	0.06

$R^2_{GLMM(c)} = 0.64$, $R^2_{GLMM(m)} = 0.27$

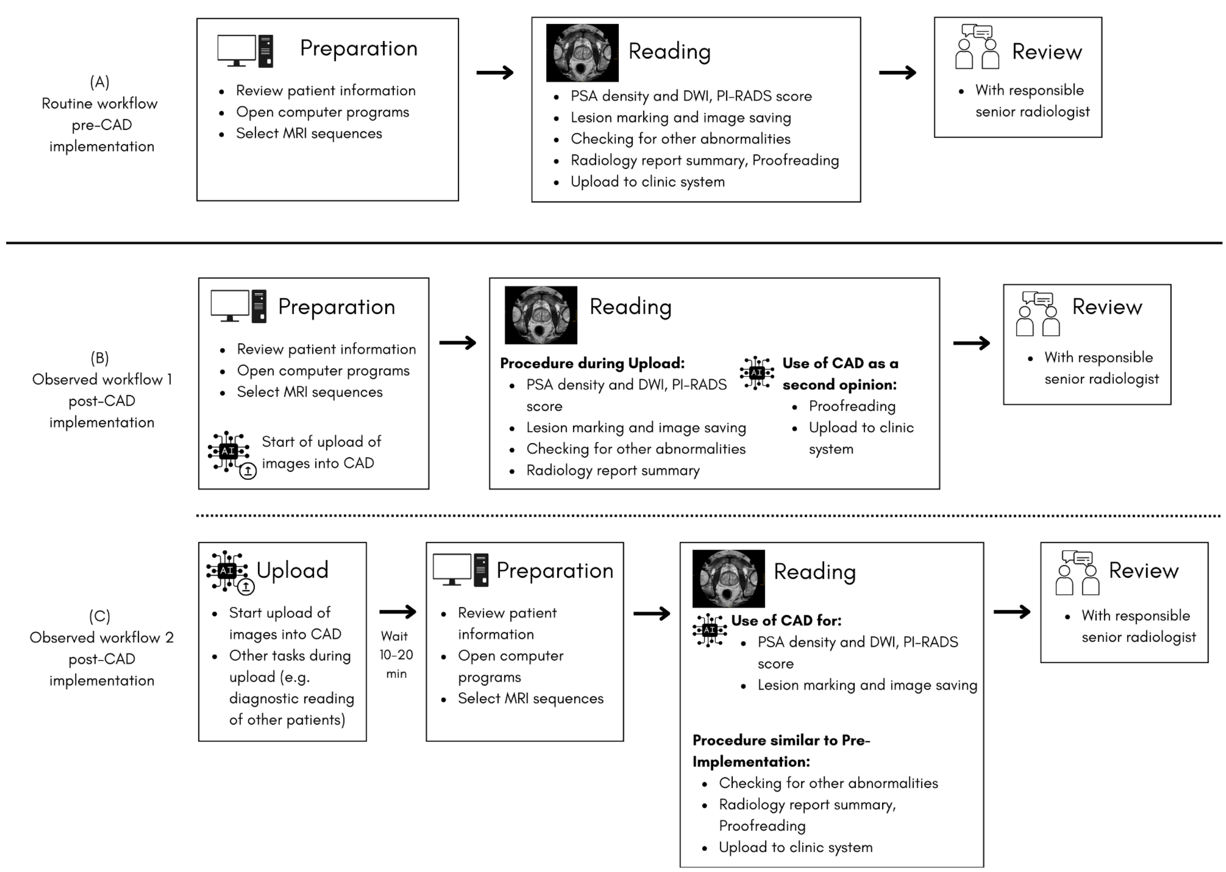


Fig. 3. Observed MRI prostate case reading workflows in the radiology department. (A) Workflow pre-CAD implementation. (B) Variant 1 of workflow post-CAD implementation. (C) Variant 2 of workflow post-CAD implementation.

3.4.2. Additional analyses per task step

For the preparation task step, we found neither an effect of CAD implementation for WTT ($b = -0.15$, $SD = 0.32$, $t(71) = -0.46$, $p = .65$) nor for the PI-RADS score ($b = 0.03$, $SD = 0.08$, $t(71) = 0.40$, $p = .70$). We also observed no effect for the interaction between PI-RADS score and CAD implementation ($b = 0.07$, $SD = 0.10$, $t(71) = 0.76$, $p = .45$).

For the diagnostic reading task step, we found no effect of the CAD implementation on WTT ($b = 0.13$, $SD = 0.16$, $t(79) = 0.79$, $p = .43$). Yet, the PI-RADS score predicted the WTT ($b = 0.20$, $SD = 0.04$, $t(79) = 5.25$, $p < .001$). Moreover, we identified a significant interaction between the PI-RADS score and CAD implementation for WTT ($b = -0.10$, $SD = 0.05$, $t(79) = -2.15$, $p = .03$; see Fig. 4). For descriptive data grouped by PI-RADS scores see Appendix 2, PI-RADS scores four and five were considered high and scores one, two and three as low.

This interaction was broken down by conducting separate mixed linear models for the high and low PI-RADS scores. The models included the CAD implementation as a predictor and random intercepts for radiologists. The analyses showed that CAD implementation predicted the WTT for PI-RADS scores of four or higher ($b = -0.33$, $t(25) = -2.9$, $p = .02$) and not for cases with a PI-RADS score lower than four ($b = -0.18$, $t(48) = -1.51$, $p = .14$).

3.4.3. Radiologists' outcome reports

The mixed linear regression analyses for stress and workload as outcome variables are summarized in Table 3. CAD implementation did not predict radiologists' workload ($p = .60$) or stress ($p = .66$). Higher PI-RADS scores were associated with increased workload reports ($p < .01$), but not with changes in radiologists' perceived stress ($p = .37$). We found neither an effect of the interaction between PI-RADS score and CAD implementation on the workload ($p = .14$) nor stress ($p = .90$).

4. Discussion

Our prospective observations revealed variations in how radiologists utilized the CAD, resulting in a decreased standardization of the workflow. CAD use did not result in shorter workflow throughput times for post-CAD implementation case readings. Yet, CAD use was associated with significantly longer workflow throughput times for diagnostic reading in cases with high PI-RADS scores. Interestingly, we did not observe any changes in radiologists' self-reported stress or workload following CAD implementation.

Our results do not support the broadly proposed expectation that AI leads to an increased efficiency in radiology in form of decreased workflow throughput times per case [7,26]. Our real-world observations thus contradict the findings of Cipollari et al. [27], who studied the same CAD in an experimental setting. A study which evaluated another AI tool for prostate MRI scan reading, also found an increase in radiologists' reading times when using the AI system with, yet, improved lesion detection performance [28]. We studied radiology residents i.e. readers with potentially less experience performing MRI scan readings, who in previous studies, benefitted more from using AI [29,30]. While we observed that the WTT increased for cases with high PI-RADS scores, subjective workload or stress assessments do not reflect this increase. This resonates well with previous studies: Rodriguez-Ruiz [31] found an increase in reading times for high suspicion cases in AI use on evaluating breast MRI scans; Shin et al. [32] had similar findings studying AI use for chest radiographs. A possible explanation for higher WTT for higher PI-RADS scores could be that additional work steps or editing had to be done for marking and evaluating lesions.

Our study corroborated empirically, that implementation of CAD in radiology does not necessarily lead to expected standardization [33]. There are several reasons for this discrepancy, mainly due to the complex socio-technical work system and environment of clinical workplace

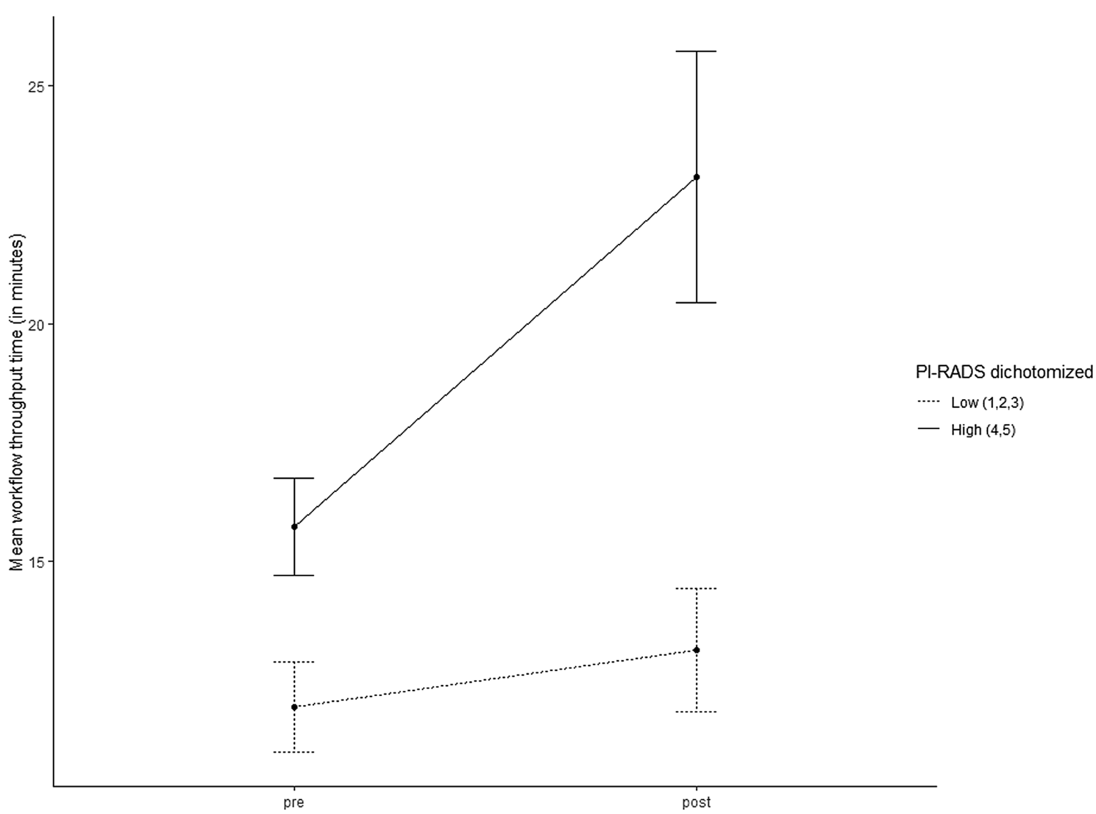


Fig. 4. Significant interaction of workflow throughput time with PI-RADS score for the diagnostic reading task step.

Table 3
Results of linear mixed models with radiologists' outcome reports (i.e. workload and stress) as outcome variables.

Predictor	*n*	$R^2_{GLMM(c)}$	$R^2_{GLMM(m)}$	b	*SE*(b)	*t*	*p*
Workload	*91*	*0.56*	*0.06*				
(Intercept)				28.47	7.84	3.63	0.00
CAD implementation				3.75	7.18	0.52	0.60
PI-RADS				3.95	1.45	2.72	<0.01
CAD implementation x PI-RADS				−2.64	1.76	−1.50	0.14
Stress	*89*	*0.04*	*0.04*				
(Intercept)				0.4	0.08	2.98	0.00
CAD implementation				−0.04	0.10	−0.44	0.66
PI-RADS				0.02	0.03	0.90	0.37
CAD implementation x PI-RADS				−0.00	0.03	−0.12	0.90

in radiology, often being highly disruptive and loaded with multiple, simultaneous task demands. These real-world challenges and obstacles for clinical implementation are usually not reflected in experimental settings or preliminary tests of new clinical AI tools outside of the actual work environment. In fact, according to a study by Dhanoa et al. [10], image interpretation accounted for only 36.4 % of the total workload. On the days we observed the radiologists, prostate MRI readings accounted for a large proportion of their reading tasks. We assessed stress and workload after each MRI reading, using the NASA-TLX, a widely accepted measure in similar populations [34]. The STAI, chosen for its brevity, has been used previously in healthcare studies [35]. Questionnaires aimed for a balance between accuracy and brevity to minimize burden on clinicians, as well as showing a trade-off between specialized and established measures. Nevertheless, a bias due to the questionnaires cannot be ruled out, especially for the STAI as the answers were not normally distributed. Therefore, validating our findings with objective stress measures is warranted.

While technical aspects are often emphasized in AI development and user testing, there is a growing focus on challenges in the course of incorporating AI into routine care processes. Since the specific CAD tool in our study had already demonstrated its potential clinical efficacy in previous research, we did not specifically examine its diagnostic performance [27,36]. Nevertheless, CAD usage can have benefits in terms of increased diagnostic accuracy or as an educational tool for radiology residents [29,37,38]. As clinician acceptance is key to the implementation of innovative technology, we investigated rigorously the actual impact on workflow and radiologists. This is of particular importance, as radiologists who face constraints in workload or workflow will eventually not adopt AI tools, with, eventually, persistent shortcomings in patient care [39].

Our design was a prospective study assessing CAD implementation in a real-world clinical practice. To be emphasized is that we conducted the study in a pre-post design not relying on historic controls while using a combination of objective and subjective data. Beyond the benefits and valuable insights stemming from our approach, there are a few limitations inherent to our observational design (e.g., observer effects). In future studies, efforts could be made to automatically assess workflow-throughput times, e.g. through automatic recording on the workstation, limiting potential bias. A convenience sampling approach, with non-blinded radiologists in a single institutional setting, was established, and neither radiologists nor cases could be randomized. It is important to note that we focused on radiology residents, who might not be directly comparable to more experienced radiologists. However, this aligns with the department's standard practice, as they handle the initial reading of prostate MRI scans. Due to these factors, the generalizability of our findings may be limited. To address this limitation, a larger sample size could be considered. We did not track whether the reading results were confirmed by a second reader and whether the use of the CAD had an impact on clinical outcomes (i.e., diagnostic performance or patient morbidity). This might be an interesting approach for future study design also assessing the impact of AI technologies on quality of care and patient safety. Additionally, a comparison between experienced and inexperienced readers using the CAD in their routine would provide valuable insights into how work experience affects CAD use and acceptance.

From our study, we can further derive suggestions for workflow integration of AI tools in the future: First, prior to implementation, workflow analysis should be conducted to determine work system-level challenges in the course of appropriate AI utilization, taking into account previous studies on AI solutions [7]. Second, it will be crucial to involve radiologists in early stages of development and implementation processes, as they are key stakeholders with valuable insights into existent work practices, technology integration, and challenges in redesigning workflows [40]. Third, our findings may contribute to future research including prospective, multi-center studies in routine radiology settings. It may also innervate further investigations into AI implementation in everyday radiological work practices including liability, transparency, explainability, as well as workflow changes.

Our study evaluates the impact of AI implementation on radiologists and their routine workflow. We established a prospective study with very high ecological validity. By combining observations and self-reports we provide a thorough investigation into the effects of AI implementation on the workload of radiologists. As our findings do not support the widely proposed assumptions that AI reduces radiologists' workload, our study highlights the urgent need for high-quality research evaluating AI tools in routine clinical workflows. At the same time, carefully considering implementation issues around AI-facilitated technologies can improve the integration of AI solutions into routine workflow and help to solve the pressing issues of radiologists' workload and the rising number of cases while safeguarding patient care.

Declaration of generative AI in scientific writing

During the preparation of this work the authors used ChatGPT (Version GPT-3.5, OpenAI) in order to optimize readability and wording in the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Katharina Wenderott: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Jim Krups:** Investigation, Writing – review & editing, Visualization, Data curation. **Julian A. Luetkens:** Conceptualization, Supervision, Writing – review & editing. **Nikoloz Gambashidze:** Conceptualization, Writing – review & editing. **Matthias Weigl:** Conceptualization, Supervision, Validation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the radiology residents for participating in our study. We sincerely thank Hannah Klinkhammer (Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University of Bonn) for her statistical support.

Funding

This work was supported by the Open Access Publication Fund of the University of Bonn.

References

[1] Z. Ahmad, S. Rahim, M. Zubair, J. Abdul-Ghafar, Artificial Intelligence (ai) in Medicine, Current Applications and Future Role with Special Emphasis on Its Potential and Promise in Pathology: Present and Future Impact, Obstacles Including Costs and Acceptance Among Pathologists, Practical and Philosophical Considerations. a Comprehensive Review, Diagn. Pathol. 16 (2021) 24, https://doi.org/10.1186/s13000-021-01085-4.

[2] J. He, S.L. Baxter, J. Xu, J. Xu, X. Zhou, K. Zhang, The practical implementation of artificial intelligence technologies in medicine, Nat. Med. 25 (2019) 30–36, https://doi.org/10.1038/s41591-018-0307-0.

[3] J. Wolff, J. Pauling, A. Keck, J. Baumbach, Success factors of artificial intelligence Implementation in Healthcare, Front. Digit. Health. 3 (2021), https://doi.org/10.3389/fdgth.2021.594971.

[4] O.T. Jones, N. Calanzani, S. Saji, S.W. Duffy, J. Emery, W. Hamilton, H. Singh, N. J. de Wit, F.M. Walter, Artificial Intelligence Techniques That May Be Applied to Primary Care Data to Facilitate Earlier Diagnosis of Cancer: Systematic Review, J. Med. Internet Res. 23 (2021) e23483, https://doi.org/10.2196/23483.

[5] H. Kim, S.W. Kang, J.-H. Kim, H. Nagar, M. Sabuncu, D.J.A. Margolis, C.K. Kim, The role of AI in prostate MRI quality and interpretation: Opportunities and challenges, Eur. J. Radiol. 165 (2023), 110887, https://doi.org/10.1016/j.ejrad.2023.110887.

[6] B. Turkbey, M.A. Haider, Deep learning-based artificial intelligence applications in prostate MRI: brief summary, Br. J. Radiol. 95 (2022) 20210563, https://doi.org/10.1259/bjr.20210563.

[7] K.G. van Leeuwen, M. de Rooij, S. Schalekamp, B. van Ginneken, M.J.C.M. Rutten, How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr. Radiol. (2021) https://doi.org/10.1007/s00247-021-05114-8.

[8] M.R.S. Sunoqrot, A. Saha, M. Hosseinzadeh, M. Elschot, H. Huisman, Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges, Eur. Radiol. Exp. 6 (2022) 35, https://doi.org/10.1186/s41747-022-00288-8.

[9] A. Hosny, C. Parmar, J. Quackenbush, L.H. Schwartz, H.J.W.L. Aerts, Artificial Intelligence in Radiology, Nat. Rev. Cancer. 18 (2018) 500–510, https://doi.org/10.1038/s41568-018-0016-5.

[10] D. Dhanoa, T.S. Dhesi, K.R. Burton, S. Nicolaou, T. Liang, The Evolving Role of the Radiologist: The Vancouver Workload Utilization Evaluation Study, J. Am. Coll. Radiol. 10 (2013) 764–769, https://doi.org/10.1016/j.jacr.2013.04.001.

[11] A.P. Brady, Measuring Consultant Radiologist workload: method and results from a national survey, Insights Imaging 2 (2011) 247–260, https://doi.org/10.1007/s13244-011-0094-3.

[12] J. Yin, K.Y. Ngiam, H.H. Teo, Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review, J. Med. Internet Res. 23 (2021), e25759, https://doi.org/10.2196/25759.

[13] J.J. Twilt, K.G. Van Leeuwen, H.J. Huisman, J.J. Fütterer, M. De Rooij, Artificial Intelligence Based Algorithms for Prostate Cancer Classification and Detection on Magnetic Resonance Imaging: A Narrative Review, Diagnostics. 11 (2021) 959, https://doi.org/10.3390/diagnostics11060959.

[14] M.E. Salwei, P. Carayon, P.L.T. Hoonakker, A.S. Hundt, D. Wiegmann, M. Pulia, B. W. Patterson, Workflow integration analysis of a human factors-based clinical decision support in the emergency department, Appl. Ergon. 97 (2021), 103498, https://doi.org/10.1016/j.apergo.2021.103498.

[15] R.C. Li, S.M. Asch, N.H. Shah, Developing a delivery science for artificial intelligence in healthcare, NPJ Digit. Med. 3 (2020) 107, https://doi.org/10.1038/s41746-020-00318-y.

[16] S. Mehralivand, D. Yang, S.A. Harmon, D. Xu, Z. Xu, H. Roth, S. Masoudi, D. Kesani, N. Lay, M.J. Merino, B.J. Wood, P.A. Pinto, P.L. Choyke, B. Turkbey, Deep learning-based artificial intelligence for prostate cancer detection at biparametric MRI, Abdom. Radiol. 47 (2022) 1425–1434, https://doi.org/10.1007/s00261-022-03419-2.

[17] P. Rawla, Epidemiology of Prostate Cancer, World J Oncol. 10 (2019) 63–89, https://doi.org/10.14740/wjon1191.

[18] J.P. Vandenbroucke, E. Von Elm, D.G. Altman, P.C. Gøtzsche, C.D. Mulrow, S. J. Pocock, C. Poole, J.J. Schlesselman, M. Egger, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration, Int. J. Surg. 12 (2014) 1500–1524, https://doi.org/10.1016/j.ijsu.2014.07.014.

[19] A. Forookhi, L. Laschena, M. Pecoraro, A. Borrelli, M. Massaro, A. Dehghanpour, S. Cipollari, C. Catalano, V. Panebianco, Bridging the experience gap in prostate multiparametric magnetic resonance imaging using artificial intelligence: A prospective multi-reader comparison study on inter-reader agreement in PI-RADS v2.1, image quality and reporting time between novice and expert readers, Eur. J. Radiol. 161 (2023), 110749, https://doi.org/10.1016/j.ejrad.2023.110749.

[20] F. Faul, E. Erdfelder, A.-G. Lang, A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res Methods 39 (2007) 175–191.

[21] K. Flägel, B. Galler, J. Steinhäuser, K. Götz, The National Aeronautics and Space Administration-Task Load Index (NASA-TLX)-an instrument for assessing workload in family practice consultations: determination of psychometric properties. [Der National Aeronautics and Space Administration-Task Load Index (NASA-TLX) – ein Instrument zur Erfassung der Arbeitsbelastung in der hausärztlichen Sprechstunde: Bestimmung der psychometrischen Eigenschaften], Z. Evidenz Fortbild. Qual. Im, Gesundheitswesen. 147–148 (2019) 90–96, https://doi.org/10.1016/j.zefq.2019.10.003.

[22] T.M. Marteau, H. Bekker, The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI), Br. J. Clin. Psychol. 31 (1992) 301–306, https://doi.org/10.1111/j.2044-8260.1992.tb00997.x.

[23] S. Arora, T. Tierney, N. Sevdalis, R. Aggarwal, D. Nestel, M. Woloshynowych, A. Darzi, R. Kneebone, The Imperial Stress Assessment Tool (ISAT): A Feasible, Reliable and Valid Approach to Measuring Stress in the Operating Room, World J. Surg. 34 (2010) 1756–1763, https://doi.org/10.1007/s00268-010-0559-4.

[24] P.J. Rousseeuw, M. Hubert, Robust statistics for outlier detection, WIREs, Data Min. Knowl. Discov. 1 (2011) 73–79, https://doi.org/10.1002/widm.2.

[25] RStudio Team, RStudio: Integrated Development for R, (2020). http://www.rstudio.com/.

[26] O. Asan, A. Choudhury, Research trends in artificial intelligence applications in human factors health care: Mapping review, JMIR Hum. Factors. 8 (2021), e28236, https://doi.org/10.2196/28236.

[27] S. Cipollari, M. Pecoraro, A. Forookhi, L. Laschena, M. Bicchetti, E. Messina, S. Lucciola, C. Catalano, V. Panebianco, Biparametric prostate MRI: impact of a deep learning-based software and of quantitative ADC values on the inter-reader agreement of experienced and inexperienced readers, Radiol. Med. (torino) 127 (2022) 1245–1253, https://doi.org/10.1007/s11547-022-01555-9.

[28] S. Mehralivand, S.A. Harmon, J.H. Shih, C.P. Smith, N. Lay, B. Argun, S. Bednarova, R.H. Baroni, A.E. Canda, K. Ercan, R. Girometti, E. Karaarslan, A. R. Kural, A.S. Purysko, S. Rais-Bahrami, V.M. Tonso, C. Magi-Galluzzi, J. B. Gordetsky, R.S.E.S. Macarenco, M.J. Merino, B. Gumuskaya, Y. Saglican, S. Sioletic, A.Y. Warren, T. Barrett, L. Bittencourt, M. Coskun, C. Knauss, Y.M. Law, A.A. Malayeri, D.J. Margolis, J. Marko, D. Yakar, B.J. Wood, P.A. Pinto, P. L. Choyke, R.M. Summers, B. Turkbey, Multicenter Multireader Evaluation of an Artificial Intelligence-Based Attention Mapping System for the Detection of Prostate Cancer With Multiparametric MRI, Am. J. Roentgenol. 215 (2020) 903–912, https://doi.org/10.2214/AJR.19.22573.

[29] C. Balleyguier, K. Kinkel, J. Fermanian, S. Malan, G. Djen, P. Taourel, O. Helenon, Computer-aided detection (CAD) in mammography: Does it help the junior or the senior radiologist? Eur. J. Radiol. 54 (2005) 90–96, https://doi.org/10.1016/j.ejrad.2004.11.021.

[30] V. Giannini, S. Mazzetti, G. Cappello, V.M. Doronzio, L. Vassallo, F. Russo, A. Giacobbe, G. Muto, D. Regge, Computer-Aided Diagnosis Improves the Detection of Clinically Significant Prostate Cancer on Multiparametric-MRI: A Multi-Observer Performance Study Involving Inexperienced Readers, Diagnostics. 11 (2021) 973, https://doi.org/10.3390/diagnostics11060973.

[31] A. Rodriguez-Ruiz, K. Lång, A. Gubern-Merida, J. Teuwen, M. Broeders, G. Gennaro, P. Clauser, T.H. Helbich, M. Chevalier, T. Mertelmeier, M.G. Wallis, I. Andersson, S. Zackrisson, I. Sechopoulos, R.M. Mann, Can We Reduce the Workload of Mammographic Screening by Automatic Identification of Normal Exams with Artificial Intelligence? A Feasibility Study, Eur. Radiol. 29 (2019) 4825–4832, https://doi.org/10.1007/s00330-019-06186-9.

[32] H.J. Shin, K. Han, L. Ryu, E.-K. Kim, The impact of artificial intelligence on the reading times of radiologists for chest radiographs, Npj Digit. Med. 6 (2023) 82, https://doi.org/10.1038/s41746-023-00829-4.

[33] F. Pesapane, M. Codari, F. Sardanelli, Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine, Eur. Radiol. Exp. 2 (2018) 35, https://doi.org/10.1186/s41747-018-0061-6.

[34] A. Ramkumar, P.J. Stappers, W.J. Niessen, S. Adebahr, T. Schimek-Jasch, U. Nestle, Y. Song, Using GOMS and NASA-TLX to Evaluate Human-Computer Interaction Process in Interactive Segmentation, Int. J. Human-Computer Interact. 33 (2017) 123–134, https://doi.org/10.1080/10447318.2016.1220729.

[35] T. Augenstein, A. Schneider, M. Wehler, M. Weigl, Multitasking behaviors and provider outcomes in emergency department physicians: two consecutive, observational and multi-source studies, Scand. J. Trauma Resusc. Emerg. Med. 29 (2021) 14, https://doi.org/10.1186/s13049-020-00824-8.

[36] E. Faiella, D. Vertulli, F. Esperto, E. Cordelli, P. Soda, R.M. Muraca, L. P. Moramarco, R.F. Grasso, B. Beomonte Zobel, D. Santucci, Quantib® Prostate Compared to an Expert Radiologist for the Diagnosis of Prostate Cancer on mpMRI: A Single-Center Preliminary Study, Tomogr Ann Arbor Mich. 8 (2022) 2010–2019, https://doi.org/10.3390/tomography8040168.

[37] T. Way, H.-P. Chan, L. Hadjiiski, B. Sahiner, A. Chughtai, T.K. Song, C. Poopat, J. Stojanovska, L. Frank, A. Attili, N. Bogot, P.N. Cascade, E.A. Kazerooni, Computer-Aided Diagnosis of Lung Nodules on CT Scans: Acad, Radiol. 17 (2010) 323–332, https://doi.org/10.1016/j.acra.2009.10.016.

[38] D. Boone, S. Mallett, J. McQuillan, S.A. Taylor, D.G. Altman, S. Halligan, Assessment of the Incremental Benefit of Computer-Aided Detection (CAD) for Interpretation of CT Colonography by Experienced and Inexperienced Readers, PLoS One 10 (2015) e0136624.

[39] R.J. Holden, B.-T. Karsh, The technology acceptance model: its past and its future in health care, J. Biomed. Inform. 43 (2010) 159–172, https://doi.org/10.1016/j.jbi.2009.07.002.

[40] European Society of Radiology, What the radiologist should know about artificial intelligence - an ESR white paper, Insights Imaging 10 (2019) 44, https://doi.org/10.1186/s13244-019-0738-2.

3.5. Publication 5: Radiologists' Perspectives on the Workflow Integration of an Artificial Intelligence-Based Computer-Aided Detection System: A Qualitative Study

Wenderott K, Krups J, Luetkens J A, Weigl M (2024)
Applied Ergonomics, 117, 104243. doi: 10.1016/j.apergo.2024.104243.

The supplementary material can be accessed via this link.

Applied Ergonomics 117 (2024) 104243

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: A qualitative study

Katharina Wenderott [a,*], Jim Krups [a], Julian A. Luetkens [b,c], Matthias Weigl [a]

[a] Institute for Patient Safety, University Hospital Bonn, Venusberg-Campus 1, 53127, Bonn, Germany
[b] Department of Diagnostic and Interventional Radiology, University Hospital Bonn, Germany
[c] Quantitative Imaging Lab Bonn (QILaB), University Hospital Bonn, Germany

ARTICLE INFO

Handling Editor: Dr. P.R. Pennathur

Keywords:
Artificial intelligence
Workflow integration
Healthcare

ABSTRACT

In healthcare, artificial intelligence (AI) is expected to improve work processes, yet most research focuses on the technical features of AI rather than its real-world clinical implementation. To evaluate the implementation process of an AI-based computer-aided detection system (AI-CAD) for prostate MRI readings, we interviewed German radiologists in a pre-post design. We embedded our findings in the Model of Workflow Integration and the Technology Acceptance Model to analyze workflow effects, facilitators, and barriers. The most prominent barriers were: (i) a time delay in the work process, (ii) additional work steps to be taken, and (iii) an unstable performance of the AI-CAD. Most frequently named facilitators were (i) good self-organization, and (ii) good usability of the software. Our results underline the importance of a holistic approach to AI implementation considering the sociotechnical work system and provide valuable insights into key factors of the successful adoption of AI technologies in work systems.

1. Background

Artificial intelligence (AI) is increasingly used in various domains of our lives, such as in sensor devices, robots, and decision-support systems. These advancements in technology have the potential to automate tasks that were previously performed by human experts (Howard, 2019; Bruun and Duka, 2018). As a result, the way we work is changing, and there is even a possibility of jobs being replaced by technology. This shift due to large technological advancements, similar to the introduction of electricity or the internet, highlights the importance to work collaboratively, establishing a partnership between humans and AI, where AI technology ideally enhances and complements human capabilities (Howard, 2019; Jarrahi, 2018). As AI solutions might integrate differently depending on the setting and the users, it is indispensable to evaluate human factors integrating for example user experiences, workflow characteristics or user preferences (Asan et al., 2020; Felmingham et al., 2021; Asan and Choudhury, 2021). Next to known challenges, new problems arise that are unique to AI such as transparency, liability, or trust in human-machine interaction (Wang and Siau, 2019; Von Eschenbach, 2021; Esmaeilzadeh, 2020).

A discipline in the focus of AI-driven changes is healthcare. AI technologies hold immense potential to assist healthcare professionals in various tasks and processes, currently being implemented in largely image-based disciplines such as radiology (Ahmad et al., 2021; He et al., 2019). Potential tasks that can be augmented by AI are for example clinical diagnostics or decision-making (He et al., 2019; Chen and Decary, 2020; Reddy et al., 2019). In hindsight, the introduction of previous technological advancements in healthcare such as robotic surgery or electronic patient records showed the intricate challenges of the successful adoption of innovative technologies into the real-world complexities of clinical workplaces (Catchpoole et al., 2022). To analyze how AI solutions actually can improve healthcare delivery, it is necessary to evaluate not only the technologies themselves, but also their actual implementation into clinical practice (Li et al., 2020). Li et al. (2020) propose a delivery science for AI in healthcare with three basic principles: First, AI must adapt to the highly complex socio-technical work system. Second, AI should be perceived as part of a system-based solution and not a single product. Third, AI-driven solutions build a complex system including people, processes, and technologies. These three components offer a holistic approach to a seamless AI implementation into healthcare, highlighting that a successful implementation is not only dependent on the software, such as its technical features, but more importantly, on the work system it is implemented in (Salwei and Carayon, 2022).

Our study adds to the existing research by investigating the

* Corresponding author.
E-mail address: katharina.wenderott@ukbonn.de (K. Wenderott).

https://doi.org/10.1016/j.apergo.2024.104243
Received 2 August 2023; Received in revised form 18 December 2023; Accepted 23 January 2024
Available online 1 February 2024

Abbreviations

AI	Artificial Intelligence
AI-CAD	Artificial Intelligence-based Computer-aided Detection System
CE	Conformité Européenne
COREQ	Consolidated Criteria for Reporting Qualitative Research
FDA	Food and Drug Administration
MRI	Magnetic Resonance Imaging
SEIPS	Systems Engineering Initiative for Patient Safety
TAM	Technology Acceptance Model

implementation process of an AI system in a routine clinical workflow in a radiology department. It thus provides a thorough evaluation of process factors affecting AI adoption in healthcare. Current evaluations of AI solutions in healthcare are mostly focused on technical features, effectiveness, and performance in contrast to human performance, and to a much lesser extent, on their implementation (Wolff et al., 2021; The DECIDE-AI Steering Group, 2021). Therefore, several studies call for prospective studies to gather evidence from AI implementation in routine clinical care (Turkbey and Haider, 2022; Sunoqrot et al., 2022; Twilt et al., 2021; van Leeuwen et al., 2021; Yin et al., 2021). This urgent need has been emphasized by the World Health Organization in their recent report on big data adoption in healthcare and a systematic review by Abell et al. (2023), highlighting the frequent mention of a lack of integration as a significant barrier (Wolff et al., 2021). Previous research identified multiple barriers and facilitators to AI adoption in clinical settings, which also deserve attention as they can ease future implementation processes (Hemmer et al., 2022; Strohm et al., 2020).

The attitude towards AI and its future are especially in radiology a heated topic, as in 2009 an article stated that AI would take over much of radiologists' work (Langlotz, 2019; Obermeyer and Emanuel, 2016). This provoked discussions in publications and on conferences in the last years, with the current trend of striving for a human-AI collaboration and replacing radiologists who do not use AI (Strohm et al., 2020; Langlotz, 2019; Chockley and Emanuel, 2016; Meskó et al., 2018). Nevertheless, in diagnostic radiology AI technologies are already integrated into practice, as deep learning AI applications facilitate screening big amounts of data in shorter times (Ahmad et al., 2021; Wong, 2018). The hope of AI supporting clinicians is of particular importance in radiology, as the discipline has been facing a rising number of cases in the last years, while there is only a limited number of trained radiologists to interpret these images (van Leeuwen et al., 2021; Hosny et al., 2018; Litjens et al., 2017). Therefore, we chose as a use case for our study the implementation of an AI-based computer-aided detection (AI-CAD) system in the routine radiology workflow.

We used two models as a theoretical foundation for our study: Firstly, the Model of Workflow Integration by Salwei et al. (2021) is a conceptual framework designed to analyze the introduction of a novel technology in a highly complex socio-technical work system. It is based on the SEIPS (Systems Engineering Initiative for Patient Safety) model (Carayon et al., 2006), which provides a coherent base for understanding the role of human factors in healthcare work systems (Holden et al., 2013; Carayon et al., 2014). When integrating AI in the clinical workflow, it interacts with several components of the system, influences processes, and shapes provider as well as care outcomes (Salwei et al., 2021). The impact of technology on clinicians and patients can vary depending on how well it integrates with their workflow. Clinicians can experience positive effects such as acceptance, satisfaction, or a decrease in their workload. Conversely, they may also face negative consequences like stress, dissatisfaction, or an increase in their workload (Bates et al., 2021; Miles, 2020). Secondly, the Technology Acceptance Model (TAM) is commonly used to assess new technologies (Davis, 1989; Holden, 2011). TAM explains why people choose to use or not to use a technology, stating that perceived ease of use and usefulness influence attitudes towards it. Attitudes affect the intention to use the technology, which determines its acceptance and eventual usage (Holden and Karsh, 2010).

Drawing upon both models, our study aims to answer the following questions:

(a) What are radiologists' attitudes towards AI as well as their perceived benefits and risks before implementing the AI-CAD in their department?
(b) How is the AI-CAD implemented into the radiologists' workflow?
(c) What are facilitating and hindering factors in the course of the AI-CAD implementation process in radiology?
(d) How do radiologists evaluate the AI-CAD and its implementation process?

2. Material and methods

2.1. Study design

This prospective interview study with radiologists during the implementation of an AI-based computer-aided detection system was approved by the Ethics committee of the medical faculty, of Bonn University (no. 449/21). Participants gave oral and written consent directly before the interviews. This study was part of a larger research project, which was pre-registered in the German Register for clinical trials (DRKS00027391) (Wenderott et al., 2024). Study reporting follows the COREQ checklist (COnsolidated criteria for REporting Qualitative research; Appendix 1) (Tong et al., 2007).

Our study investigated the implementation of the AI-CAD Quantib® Prostate (Version 1.2.0, Quantib BV, Rotterdam, The Netherlands), an online platform for interpreting and documenting prostate MRIs. This software has obtained clearance from the FDA and has been marked with the CE certification. The AI-CAD system provided partial automation for prostate MRI reading, assisting in the segmentation of the prostate, generation of heat maps to emphasize areas of interest, and facilitation of the standardized prostate MRI assessment using a structured questionnaire. Furthermore, it automatically produced patient reports derived from identified lesions (Wenderott et al., 2024). Forookhi et al. (2023) provide an in-depth overview of Quantib Prostate. The workflow still comprised viewing of the cases with an attending radiologist and another second reader, ensuring quality and diagnostic safety.

Data collection was conducted in two waves: we interviewed participants before the AI-CAD was implemented, i.e., pre-AI-CAD implementation, when only the forthcoming implementation was announced. Pre-implementation interviews were conducted between January 2022 and March 2022. Prior to the implementation of the AI-CAD system, radiology residents underwent a training specifically focused on utilizing Quantib Prostate, where representatives of the company explained the AI-CAD and did some example cases using the software. Implementation into the radiology workflow started with a familiarization phase, where radiologists could test the software and initial problems were solved with the IT department. After this phase of approximately two months, the AI-CAD was used in the routine workflow, marking the start of our post-implementation evaluation. In the routine workflow radiologists were obliged by the departments' management to use the software. We conducted the post-AI-CAD implementation interviews between June 2022 and October 2022.

2.2. Setting and participants

Eligible participants were radiologists and radiology residents in training of the local radiology department where the AI-CAD system was implemented (i.e., convenience sampling). All had prior training in diagnostic imaging of prostate tumor evaluation and associated workflows. Clinicians work in shared offices located right next to the MRI scan rooms. Each workstation has three monitors equipped with standard technology and a speech-recognition dictation system. Their primary role does not involve direct interaction with patients, unless there is a shortage of staff or a privately insured patient requested a vis-à-vis explanation of their results. For additional information on the implementation process, a person from the hospital's IT department was interviewed (i.e., who was in charge of the AI-CAD deployment).

2.3. Interview procedure

The study was announced via e-mail, participants were approached in-person and prior to the interview informed about study goals, voluntary participation, and data protection measures. Interviews were semi-structured, conducted in German and audio recorded. To minimize disruptions, participants were interviewed in a separate room. All interviews were conducted in-person by KW (first author), a female researcher with a Master of Science in psychology. The interviewer introduced herself to the radiologists prior to data collection and presented the study as a part of her PhD project. Due to the routine rotations in the radiology department, we were not able to provide feedback to all participants after data analysis.

2.4. Measures and interview contents

During the study development, unstructured preliminary interviews were done with the head of the department and an unstructured workflow observation was made to identify relevant aspects for the interview guideline. After development, the interview guideline was tested beforehand with a team member to test for comprehensibility and clarity. The pre-implementation interviews aimed for a duration of 10–15 min and post-implementation of 15–20 min. Interview guidelines can be found in Appendix 2-3.

2.4.1. Work setting evaluation

Clinicians were asked to report on their average workload based on the assessment of prostate MRI scans. Additionally, clinicians were questioned about their satisfaction regarding the general procedure of prostate MRI assessment and technical equipment in use. We spoke with a radiologist and a urology expert in preliminary interviews and found that effective communication and sharing of information between different departments are crucial for ensuring high-quality care and patient safety.

2.4.2. Attitude towards AI

For the assessment of attitudes towards AI, we only asked one open question on general expectations regarding AI in medicine. Participants were asked to elaborate their answer. Post-implementation, participants were asked whether and how their attitude towards AI had changed. Additionally, participants were asked about the potential benefits and risks associated with the implementation of AI.

2.4.3. AI-CAD intention to use

Pre-implementation, participants were asked about their attitudes about an AI software for MRI prostate scan analysis, including its perceived usefulness, perceived ease of use, and their intention of using it. These dimensions were derived from the Technology Acceptance Model (TAM; Holden and Karsh, 2010).

2.4.4. Workflow integration

To assess the actual usability of the AI-CAD, workflow integration as well as facilitators and barriers after the implementation, we modified the interview guide for a semi-structured interview by Salwei et al. (2021) based on the SEIPS 2.0 and the conceptual Model of Workflow Integration. The original questionnaire was in English, and the adapted version was translated into German.

2.5. Analysis

All interviews were audio-recorded and transcribed verbatim. Only anonymized transcripts were then used for data analysis with MAXQDA Version 22.2.0 (Software, 2021), therefore we could not return the transcripts to the participants for comments or correction. Data analysis was done separately after both data-collection phases. During the transcription process, data saturation became clear on the basis of many repetitions across interviews and statements. We used structuring qualitative content analysis with a step-wise process to analyze the data (Kuckartz and Rädiker, 2022). In the first step, main content categories were derived deductively from the interview guide based on the Model of Workflow Integration (Salwei et al., 2021) and the TAM (Holden and Karsh, 2010), and a codebook was developed. To ensure interrater reliability across the coding, the researchers tested the codebook in two interviews and discussed problems with the codes and adjusted the definitions when necessary. Second, two researchers (KW, JK) coded separately all interviews and discussed all coded segments in a consensus-oriented approach (Kuckartz and Rädiker, 2022; Hopf and Schmidt, 1993). Third, subcategories were developed from the coded segments and also defined in a codebook. Fourth, both researchers coded independently all the interviews using the subcategories and then re-discussed all coded segments to reach a consensus.

3. Results

3.1. Participants

We conducted 19 interviews, 10 during pre- and 9 during post-implementation phase. Pre-implementation interviews took M = 08:27 min, SD = 01:48, and post-implementation M = 15:21 min, SD = 02:51. The study involved 9 radiology residents who looked at the initial prostate MRI scans and 3 attending radiologists who reviewed their interpretations. Seven radiologists were interviewed in both study phases. Among all potentially eligible participants, i.e. radiologists in the department involved with the prostate MRI reading process, we had in each phase one attending and one resident who chose not to participate. Therefore, we interviewed above 80 % of overall eligible radiologists. Due to routine rotations, five radiology residents were only interviewed in one study phase. To ensure anonymity, we did not collect any personal or demographical data except for interviewee's work experience. Eight radiologists had 2–4 years of work experience and four had over 5 years. In the department, five participants had an organizational tenure less than a year, four for 1–3 years, and three for 4–5 years. Since most participants were interviewed in both study phases, all study questions could be addressed by combining the interviews of both study phases, as they focused on different aspects and captured different experiences. A table on radiologists' work experience and the phase in which they were interviewed is included in the supplements (Appendix 4).

3.2. Work setting

Participating radiologists rated their average workload of prostate MRI readings being between 3 and 20 % of their weekly workload. Among radiology residents, seven rated their average workload for prostate MRI reading as 10 % or higher, and two answered their workload being around 5 %. This did not change from pre- to post-implementation. Supervising radiologists deemed their weekly workload for this task being between three and eight percent per week. Two of them reported an increase due to AI-CAD implementation from pre- to post-implementation. Generally, radiologists were quite satisfied with the prostate MRI reading workflow and the technical equipment in the department (each mentioned by ten participants). Residents highlighted the standardized procedure due to the reporting criteria and diagnostic reading process being very well-structured. Additionally, they mentioned a good personal connection to referring clinicians (10 out of 10 pre-implementation interviews), but noted that the transfer of written patient information could be improved (8 out of 10 pre-implementation interviews).

In the following, we present the resulting content categories of our interviews concerning AI workflow implementation with supplementary quotes to highlight meaning and key statements.

3.3. Attitudes towards AI, perceived benefits and risks

All but one radiologist expressed a positive attitude towards AI, while one radiologist remained undecided. After implementation, seven reported that their attitude remained unchanged after using the computer-aided detection (AI-CAD) system. One radiologist felt disappointed with their expectations, while another experienced a positive shift in attitude. Most radiologists (7 out of 12) considered AI as a form of backup or reinsurance. One participant viewed AI as relevant for training purposes. Five radiologists mentioned the possibility of autonomous AI, i.e. AI working without interaction with the radiologist, but only one favoured this idea: *"So the best thing would be a software that I have to interact with as little as possible. That just somehow gets the images and then spits out something on its own [...]. I just want so see an assessment or a report" (IV1 pre)*. Four radiologists believed human supervision was still necessary in such cases, an exemplary quote is: *"It [the AI] is not always right, either. [...] you have to sit behind it and see if it is right or not" (IV6 pre)*.

Two radiologists mentioned unclear liability and ethical considerations regarding AI usage. Nine radiologists (out of 12) mentioned that they do not believe that AI will substitute radiologists in the future.

Concerning the opportunities due to AI adoption in radiology, interviewees mentioned an increase in efficiency through prioritising tasks, speeding up the diagnostic process, and highlighting automatically detected lesions. They also mentioned higher standardization and improvement of quality, for example, by not being dependent on human concentration, and a high sensitivity (see Table 1). For example, one stated: *"[...] I see an opportunity in that we can perhaps guarantee consistent quality and, above all, simplify and accelerate the workflow and generate more volume overall [...]." (IV12 pre)*.

Risks associated with the use of AI were loss of competencies (i.e., concerns if future radiologists will be able to detect lesions themselves), mistakes or inaccuracies proposed by the AI software, relying and trusting completely on the AI system, and radiologists becoming dependent on using AI for diagnostics (see Table 2). An exemplary quote regarding the potential loss of competencies was: *"The main problem I see is that in the future, [radiologists] who grew up with many AI solutions [...] may no longer be able to check [MRI scans] because they simply lack the engrams regarding physiological or [...] pathological findings." (IV10 pre)*.

3.4. AI-CAD implementation

Before being implemented, nine out of ten radiologists expected that the AI-CAD would be useful, but highlighted the importance of the tool's usability in determining its actual adoption and effectiveness. Radiologists identified several conditions for utilizing the AI-CAD, namely the seamless integration into the workflow, smooth technical interfaces, user-friendly operation, and the absence of additional time requirements. Eight radiologists emphasized the tool's user surface and its ease of use. Seven radiologists were concerned about potential time constraints associated with the software. Seven out of ten said they intend to use the AI-CAD after implementation; the other three linked their decision to AI-CAD's usability.

After AI-CAD implementation, three of the radiology residents stated they were using the AI-CAD for all patients, and the other three merely for a part of their assigned patients. Among the attending radiologists who were reviewing the prostate MRI readings, two were using AI-CAD and one person was not. When using the system, six radiologists used the system as an add-on (i.e., as a supplement to their traditional workflow, mentioned 12 times) and seven also as a reinsurance (i.e., confirmation for preliminary diagnostic evaluations and decisions, mentioned 11 times). Post-implementation, residents described different workflows for current AI-CAD operation: three described a workflow where the AI-CAD was used parallel to their standard procedure, integrating the AI-CAD results when writing the radiology report (see workflow version 1, Fig. 1). One radiologist framed this as the desired workflow in the department. One person reported a workflow where the AI-CAD was only used after writing the radiology report as an add-on due to the time constraints (see workflow version 2, Fig. 1). The clinician stated: *"I always do my sessions in Impax [a digital working environment] in parallel and, yes, we are not actually supposed to do that at this time, but sometimes I also write the report if it all takes too long for me. And when the images can then be viewed in [AI-CAD], yes, then I do the evaluation there." (IV6 post)* Two radiologists used upload time of the images for the diagnostic reading of other patients, and after the completed uploads, proceeded with the AI-CAD for the initial diagnostic reading (see workflow version 3, Fig. 1).

The person from the IT department who was in charge of the AI-CAD system integration stated that an automatic upload into the system would have been beneficial for the workflow, minimizing the effort that radiologists need to invest to use the system.

When using the AI-CAD, seven radiologists reported AI-CAD-associated interruptions to their workflow, but only two radiologists reported them impacting the use of the AI-CAD. Other influential factors were patient characteristics, time pressure, workload, work organization, program default settings, such as that it always opened on a specific monitor and needed one specific browser, and study participation (see Table 3).

3.5. Facilitators and barriers for AI-CAD implementation

Main barriers in course of AI-CAD implementation into the workflow were the time constraints using the program, technical requirements not met, and an unstable performance of the AI-CAD (see Table 4): *"[...] that we simply have the problem that there are somehow technical, i.e. any problems with uploading images. This delays the diagnosis enormously." (IV2 post)*. Furthermore, radiologists described a delay in the workflow due to additional work steps to be taken when using the AI-CAD: *"The disadvantage is that at the moment we have to do some things twice, so to speak. So*

Table 1
Benefits through AI adoption (content categories and individual statements).

Benefit	Definition	Frequency (Times topic was mentioned)	Number of interviewees mentioning topic
Quality improvement	*Benefits for improving the quality of care through AI, e.g. by detecting more lesions.*	19	9
Increase in efficiency	*Acceleration or optimization of the work process and measurable outcome through AI e.g. increase of the amount of lesions detected.*	18	11
Standardization	*Contribution of AI to improve standardization and comparability of diagnostic tasks and the workflow.*	10	7

Table 2
Risks through AI adoption (content categories and individual statements).

Risk	Definition	Frequency (Times topic was mentioned)	Number of interviewees mentioning topic
Overreliance on AI	*Risk that radiologists do not critically appraise AI results and rely on AI results for their diagnosis, with the possible risk of false diagnosis when AI is incorrect.*	11	7
Loss of competencies	*Risk of losing their reading skills due to frequent AI use/radiologists in-training do not acquire manual reading skills due to learning MRI readings with AI assistance only.*	6	4
Mistakes through AI	*Risk of errors by AI such as incorrect classifications or biases in the training data.*	1	1
Dependence on AI	*Risk of being dependent on the AI for a diagnosis i.e. diagnostic decision making is not possible without AI, which is problematic when AI is not available or working.*	1	1

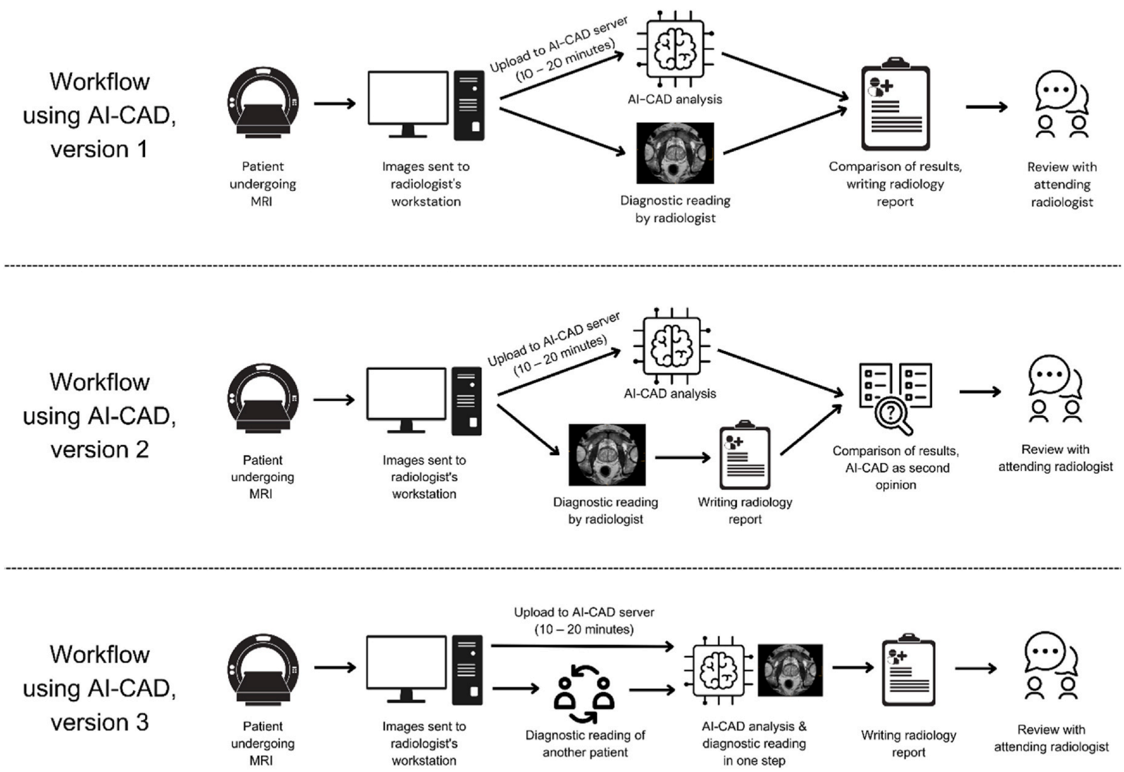


Fig. 1. Reported variety of workflows using the AI-CAD for prostate MRI readings after implementation.

Table 3
External factors influencing AI-CAD use.

Factor	Definition	Frequency (Times topic was mentioned)	Number of interviewees mentioning topic
Patient characteristics	*Patient characteristics influencing the workflow, e.g. impeding prostheses, previous prostate operations, private insurance.*	12	8
Time pressure	*Time pressure in the department influencing AI-CAD use.*	11	8
Workload	*Workload in the radiology department, e.g. amount of MRI scans.*	4	4
Work organization	*Way of organizing the work personally or in the department, e.g. radiologists using the AI-CAD simultaneously.*	3	2
Program default settings	*Default settings of the AI-CAD, e.g. need for a specific browser.*	2	2
Interruptions	*Interruptions of the workflow, e.g. colleagues asking questions.*	2	2
Participation in study	*Study participation leading to AI-CAD use.*	1	1

we evaluate visually and then again with the [AI-CAD] software." (IV8 post). Further barriers named were lack of integration in other programs or the interoperability with other software: *"The problem is that these are external and not integrated solutions, […] and they are only tailored to a specific problem." (IV12 post)*. This was also mentioned by the IT department.

The facilitator named most frequently was a good self-organization when using the software: *"If you use it, you have to be well organized, because you always have to send the images directly away and have to wait." (IV6 post)*. Moreover, radiologists mentioned the training and the familiarization phase for learning how to use the software and a good usability as facilitating factors (see Table 5). Also named were the

Table 4
Barriers for AI-CAD implementation into the workflow.

Barriers	Definition	Frequency (Times topic was mentioned)	Number of interviewees mentioning topic
Time delay in the work process	*Time delays in the workflow due to AI-CAD use, e.g. upload time.*	27	8
Additional work steps	*Additional work steps to be taken to use the AI-CAD, e.g. sending images to server.*	14	6
Unstable performance	*Unstable performance of AI-CAD, e.g. program crashes.*	13	6
Technical requirements not met	*Technical requirements needed for AI-CAD use (server, specific browser).*	9	7
Lack of flexibility	*Lack of flexibility in the AI-CAD's design, e.g. not being able to open two patients simultaneously.*	6	4
Lack of agreements on use	*Unstructured or parallel AI-CAD use of radiologists influencing the system's performance.*	6	2
Poor interoperability with other programs	*Poorly designed interoperability with standard programs influencing AI-CAD use.*	5	3
Poor usability	*AI-CAD not being intuitive or user-friendly.*	4	2
Lack of integration into standard programs	*AI-CAD not being integrated into standard programs like PACS etc.*		3
Lack of guidelines for handling errors	*No standardized guidelines for handling of errors.*	3	2
Summary of results	*Results being displayed in English (not in German) or having a different format than traditional radiology reports.*	2	1
Unclear purpose of use	*Purpose and benefit of AI-CAD use not being transparent.*	1	1

transparency of the AI-CAD (no black box system), clear agreements between colleagues when using the AI-CAD, as simultaneous use led to problems, and the motivation through colleagues.

3.6. Post-hoc evaluation of AI-CAD implementation

The evaluation of the AI-CAD was largely heterogeneous, which is also demonstrated by the comparison of the statements in Table 6. Six radiologists evaluated the usefulness of the AI-CAD negatively. In contrast, eight radiologists appraised it positively. Altogether this led to mixed assessment, i.e., with various, contradictory statements in the interviews. One clinician gave a negative evaluation of the workflow integration, while two clinicians negatively assessed the standardization. Radiologists appraised the usability, graphical representation of lesions, standardization, and interoperability with other programs. Two radiologists saw future potential in substituting the traditional workflow, using it for research, the visual presentation of diagnosis, and in developing further AI solutions. Limitations were the editing needed in the AI-CAD, no report interpretation by the program, critical assessment of findings, and that the software was still under development. Additionally, the interviewee from the IT department stressed the high technical, administrative, and managerial effort when setting up such a software, for example for setting up the servers, providing user accounts, and ensuring data security.

4. Discussion

In course of the successful adoption of AI-tools in work systems, various system-related challenges for workflow implementation need to be mastered. Key findings are that participating clinicians generally held a positive attitude towards AI in radiology, but were viewing it primarily as a backup tool. Benefits radiologists named were increased efficiency, standardization, and quality. However, they also acknowledged risks such as competency loss and overreliance. Workflow integration differed substantially among radiology residents which led to a lower standardization of the workflow. The findings contribute to the current literature base and theoretical concepts in various ways.

When embedding our results in the theoretical framework of the Technology Acceptance Model (Holden, 2011; Holden and Karsh, 2010), it becomes evident why our observations on the actual use were mixed (see Figs. 2 and 3). Prior to implementation, the perceived usefulness has been high, but radiologists highlighted the ease of use as a major precondition for their further use. Beyond an overall positive attitude, most radiologists intended to use the AI-CAD prior to the implementation. After the implementation process, the perceived usefulness post-implementation was characterized by contradictory statements within the interviews. Nonetheless, ease of use was evaluated mainly positively. The intention to use was influenced by the management's decision that residents were supposed to use the AI-CAD for all patients. Though, radiologists listed further influential factors for their decision to use the system for case readings, such as time pressure, patient characteristics or workload. Finally, the actual use was mixed and inconsistent, with half of the radiology residents using the system for all and the other half for parts of the patients. When referring to the TAM, as shown in Fig. 3, the evaluation and the actual use are heterogeneous, highlighting the multiple challenges to AI adoption that arise in a real-world clinical implementation.

Drawing upon the Model of Workflow Integration, we obtained comprehensive information concerning the fit of the AI-CAD software into the sociotechnical work system and its integration into the clinical workflow (see Fig. 4). This study thus contributes to previous work on this model and expands its application, to the best of our knowledge, for the first time to AI-facilitated tools in clinical practice. We discovered various factors within the work system that affect how the AI-CAD is

Table 5
Facilitators for AI-CAD implementation into the workflow.

Facilitators	Definition	Frequency (Times topic was mentioned)	Number of interviewees mentioning topic
Good self-organization	*Radiologists being well-organized in their workflow to optimally use the AI-CAD.*	8	5
Good usability	*AI-CAD system being user-friendly.*	4	4
Familiarization phase	*Familiarization phase to get used to the AI-CAD after its implementation.*	4	4
No black box	*AI-CAD system being transparent.*	2	2
Motivation by colleagues	*Other colleagues using the AI-CAD system led to motivation.*	2	1
Clear agreements on the work process	*No simultaneous AI-CAD use by more than one resident to improve the system's performance.*	1	1

Table 6
Examples of contradictory statements regarding AI-CAD implementation.

Category	Definition	Positive evaluation	Negative evaluation
Usefulness	*Clinical usefulness and benefits of the AI-CAD system.*	"[…] the reporting is very well possible with Quantib and [it] also gives you all the results, so that you can dictate it directly in the report." (IV9 post)	"So at the moment Quantib is not yet clinically useful, I would say, because it simply leads to longer time expenditure." (IV8 post)
Standardization	*The AI-CAD system's potential for standardization of diagnostics.*	"[…] you can also transfer the whole thing to a sheet in a standardized way, so [it] also contributes to the standardized handover to the clinical colleagues and so on." (IV4 post)	"And do you feel that this has changed anything in the standardization […] of your findings?" – "Standardization? No, not so far actually." (IV6 post)
Interoperability with other programs	*The AI-CAD system's interoperability with other (standard) programs.*	"So the good thing is that in Impax, with one mouse click, you can […] load it onto the [Quantib] server." (IV11 post)	"[…] but unfortunately we are still lacking a few basic requirements, including how interfaces can be […] integrated into our PACS, RIS, CAS." (IV11 post)
Usability	*The AI-CAD system's usability and user-friendliness.*	"[…] once you've done it once or twice, it's de facto self-explanatory and easy, yes." (IV11 post)	"It was unfortunately not intuitive, even though it is beautifully colorful." (IV10 post)
Presentation of results	*Graphic and linguistic presentation of the diagnostic results.*	"What it does totally well is visualize […] the results that it spits out. […] This volumetry and then the findings and then you get a result sheet directly." (IV2 post)	"Quantib also doesn't give a German report […]. A large part is actually English, so we can't copy and paste that directly into our findings." (IV9 post)

integrated, which are depicted in detail in Fig. 4. The system element "technologies" has been highlighted throughout the interviews mostly targeting migration issues and interoperability of the AI-CAD tool with available technologies such as the slow server connection and poor alignment with standard radiology programs. Another significant factor is the "organization" element, which has been discussed regarding an unclear purpose for using AI-CAD or a positive evaluation of enabling a training with a subsequent familiarization phase for the AI-CAD system. The "tasks" element was mentioned mainly in relation to barriers and conditions for use, such as workload or time pressure due to scheduled discussions with privately insured patients. On the other hand, the "person" element had facilitating attributes, such as good self-organization or motivation through colleagues. Our observations corroborate the multiple interactions of the work system components when adopting new AI tools. Most importantly our study supports the claim of Salwei et al. (Salwei and Carayon, 2022; Salwei et al., 2021) regarding the importance of temporal nature of workflow integration – which has also been a major barrier in our use case. Next to the extended upload time, also the when and how to use the AI software differed among radiologists, leading to workarounds in the workflow and a lack of sustained use.

The Model of Workflow Integration also highlights four dimensions that help to understand how and to what extent the AI-CAD system is eventually adopted into clinical practice (Salwei et al., 2021). The first dimension is time, focusing on how well the technology fits into the workflow schedule. Radiologists used the AI-CAD either in parallel to their old workflow or sequentially after their conventional evaluation routine. Yet, this added extra tasks and time pressures, conceived as important barriers. Interviewees highlighted that there was a familiarization phase facilitating system integration. The second dimension – flow – consists of four parts: When it comes to task flow, barriers emerged due to additional work steps, inflexibility, and performance issues, causing interruptions. This also aligns with some radiologists reporting an increased workload, which is evidently a major barrier to routine use of the AI-CAD. Good self-organization was perceived helpful, which could also hint to a suboptimal fit, making adaptations and workarounds necessary. The second sub dimension, focusing on the flow of people, highlighted a problem: When multiple people tried to use the AI-CAD simultaneously, performance problems occurred, requiring mutual agreements when to use the software. This was a major barrier regarding multiple dimensions such as time, flow, and level of workflow implementation, and, eventually, impeding its actual adoption into the work process. Support from colleagues and clear process agreements aided adaptation. Flow of information barriers included lack of error-handling guidelines, results being presented in an incompatible format, and unclear usage purposes, impeding communication within the workflow. Analyzing the flow of technologies, issues like unmet

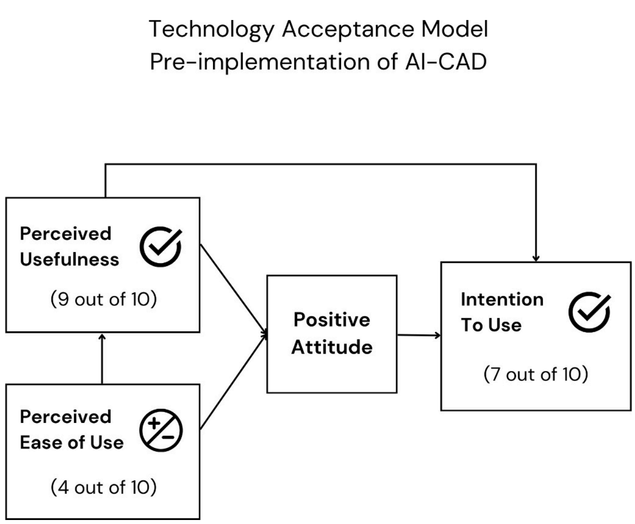


Fig. 2. Technology Acceptance Model (Holden and Karsh, 2010) with results from the pre-implementation interviews.

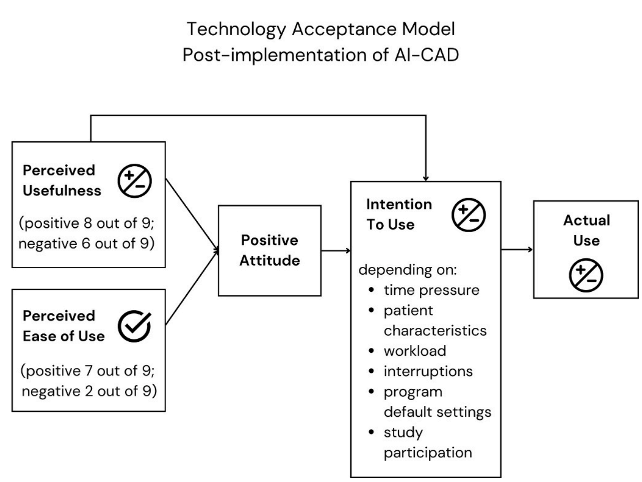


Fig. 3. Technology Acceptance Model (Holden and Karsh, 2010) with results from the post-implementation interviews.

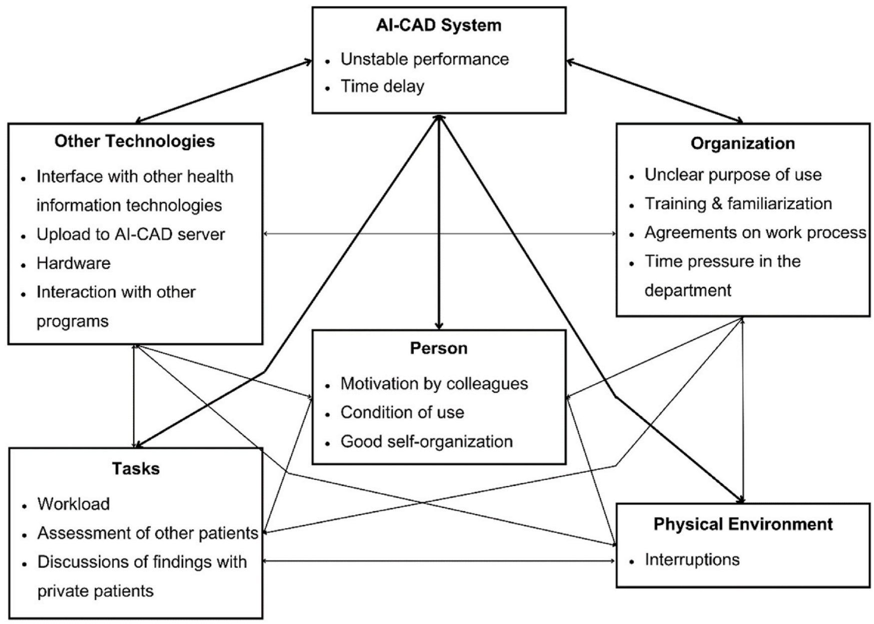


Fig. 4. Key factors for successful CAD implementation into the radiology work system based on the Model of Workflow Integration (Salwei and Carayon, 2022).

technical requirements, poor interoperability, and lack of integration with standard programs hindered AI-CAD's integration. The third dimension, patient journey scope, showed that radiologists used the AI-CAD mostly after the patients' visit but also appreciated the AI-CAD's graphical lesion presentations during the patient interaction. On the fourth dimension, workflow implementation level, we identified that the software was used by the radiologists in their individual workflow, but was implemented upon a decision by department management, limiting radiologists' autonomy in their individual workflow design. Usability was critical, seen as both a barrier and facilitator based on individual assessments. Interestingly, there were no reported differences in AI-CAD perception between radiology residents and attending physicians. We identified, however, differences in AI use that might be attributed to the organizational structure in the department. Normally, residents are responsible for MRI readings and attending radiologists are responsible for review, therefore usually not directly operating the AI system.

As demonstrated, the actual implementation of AI is impacted by a variety of facilitators and barriers (Wenderott et al., 2022). A study by Hemmer et al. (2022) conducted interviews with healthcare professionals on AI adoption factors, identifying user adaptiveness and time efficiency as most important factors for AI adoption. This is in line with the empirical findings in our study. Strohm et al. (2020) performed an interview study on another AI software in radiology, including radiologists but also hospital managers. Yet, our empirical results on facilitating process factors do not show much overlap with their findings. We assume that it could be attributed to the rather organizational-level perspective of their research and interviewees, whereas our sample only consisted of professionals who utilize and operate the AI-tools frequently. Our list of barriers for AI-CAD implementation, on the other hand, replicates and augments previous findings, such as inconsistent technical performance, absence of guidelines or best practices, and unstructured implementation in the workflow. Our results thus complement the research of Strohm et al. (2020), addressing the research gap on factors of AI workflow integration especially in real-world clinical work settings.

We found that technical features and disruptions to the standard workflow had an important impact on radiologists' use and the evaluation of the AI software. Usability aspects emerged as a key focus during the interviews, supporting previous research (Esmaeilzadeh, 2020). Nevertheless, these aspects are not unique to AI but are well known from previous software design and human-computer interaction as well as technology implementation processes. Specifically for healthcare, a variety of studies showed the adoption of electronic health records and health information systems in the hospital is driven by user acceptance and ease of use – especially in dynamic hospital environments (Schwappach and Ratwani, 2023). However, AI-specific aspects such as trust, black-box algorithms, regulatory concerns or liability issues were not in the primary focus of the reports, even though they are highlighted in current literature on AI implementation (Wang and Siau, 2019; Von Eschenbach, 2021; Esmaeilzadeh, 2020; Ferretti et al., 2018; Gille et al., 2020). Of greater concern was the potential of AI replacing radiologists which is also a topic fueled by discussions in the literature, with some radiologists stating their intention to only use the AI as back-up or reinsurance rather relying on their own competencies (Strohm et al., 2020; Langlotz, 2019; Obermeyer and Emanuel, 2016; Chockley and Emanuel, 2016; Meskó et al., 2018). Adding to this, it is worth noting that contrary to trending visions of the future current literature, our study did not find a prevalent view of human-AI collaboration (Asan et al., 2020; Sezgin, 2023; Lai et al., 2021).

4.1. Limitations of our study

Next to the strength of our study in evaluating a real-world implementation of AI, our study comes with several limitations inherent to this study design. Firstly, the study is limited to one department and one AI software. Therefore, its external validity may have limits due to its focus. At the same time, we were able to deeply analyze this implementation process over an extended period of time. Due to the organizational structure of the department, we were also not able to interview all radiologists pre- and post-implementation as routine rotations happened in the meantime. Additionally, as we relied on convenience sampling, we were only able to recruit from the department staff, resulting in a limited sample size. Furthermore, we had a relatively short interview duration, to not impose more time pressure on the participating radiologists and to ease compliance. Secondly, participants were aware of this study's objectives and socially desirable response behavior cannot be ruled out. This is particularly noteworthy, as study participation has been even mentioned by a participant as a factor for AI use or when describing the workflow as how it should be done vs. how it was actually done. Also, the researchers obtained the impression that the self-reported evaluation of the AI-CAD software has been more positive in the interviews than when observing the workflow in the department. Thirdly, we did not conduct a follow-up at a later time point, what limits inferences concerning long-lasting effects or sustainability of identified factors over time. Thus, we heard from the head of the radiology department that they stopped using the AI-CAD software not that far after our data collection. It is important to note that AI adoption can be influenced by multiple factors, some of them exceeding the scope of our study, such as trust in AI, explainability, regulatory concerns, or questions on liability. These AI-specific factors should be targeted in future studies determining their influence on actual AI adoption into practice.

4.2. Implications and conclusion

Our empirical findings revealed that AI technologies might be adopted differently into the workflow in real-world usage than in experimental studies. Since this approach of work-as-imagined is often in stark contrast to actual work practices (work-as-done), future research should focus on the transfer from experimental to actual clinical settings. Taking up the proposed delivery science for AI in healthcare by Li et al. (2020), our study also highlights the importance of a holistic approach to AI implementation with taking the users, work processes, and the technology itself into account. Additionally, our study showed that generic technological and user aspects were very relevant, thus they are not unique to AI technologies. In the future, the socio-technical work system should already be considered during developmental stages of AI systems, potentially mitigating problems arising from a deficient fit into the workflow. In these vendor-based stages of AI-tool development, radiologists as well as human factors or systems engineering experts should be included, since they are important stakeholders and can provide valuable insights to the workflow or AI implementation. In our use case, the usability has been evaluated very differently by the radiologists showing that extensive user testing prior to implementation with a variety of different skill levels is advised to optimize these systems.

When implementing AI tools in clinical care, the following implications of our study may be taken into account: it would be helpful to analyze and describe the everyday workflow and its potential variations before implementing an AI tool. This provides a reliable foundation and specifies when and how the new software should be used, as well as what kind of everyday variations need to be considered in the adoption process. The prior testing of interfaces with other programs used and exploring the possibility of integrating AI tools into standard programs should be taken carefully into account, i.e., through simulation on site or pilot runs. Before actual adoption in routine clinical work, a thorough training on the software and a following familiarization phase are necessary. Additionally, clinical users should know whom to contact when running into problems.

To conclude, our study provides valuable recommendations for the future implementation of AI technologies in healthcare. Our investigation addressed the existing research gap pertaining to the integration of AI solutions into routine workflows. By embedding our findings on a real-world AI implementation in a radiology department into the theoretical frameworks of the Model of Workflow Integration (Salwei et al., 2021) and the Technology Acceptance Model (Holden and Karsh, 2010) our research may inform future work in this area as well as serve as a resource for guiding the successful adoption of AI in healthcare.

Declaration of generative AI in scientific writing

During the preparation of this work the authors used ChatGPT (Version GPT-3.5, OpenAI) in order to optimize readability and wording in the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Open Access Publication Fund of the University of Bonn.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apergo.2024.104243.

References

Abell, B., Naicker, S., Rodwell, D., et al., 2023. Identifying barriers and facilitators to successful implementation of computerized clinical decision support systems in hospitals: a NASSS framework-informed scoping review. Implement. Sci. 18, 32. https://doi.org/10.1186/s13012-023-01287-y.

Ahmad, Z., Rahim, S., Zubair, M., Abdul-Ghafar, J., 2021. Artificial intelligence (ai) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. A comprehensive review. Diagn. Pathol. 16, 24. https://doi.org/10.1186/s13000-021-01085-4.

Asan, O., Choudhury, A., 2021. Research trends in artificial intelligence applications in human factors health care: mapping review. JMIR Human Factors 8, e28236. https://doi.org/10.2196/28236.

Asan, O., Bayrak, A.E., Choudhury, A., 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. J. Med. Internet Res. 22, e15154 https://doi.org/10.2196/15154.

Bates, D.W., Levine, D., Syrowatka, A., et al., 2021. The potential of artificial intelligence to improve patient safety: a scoping review. NPJ Digital Medicine 4, 54. https://doi.org/10.1038/s41746-021-00423-6.

Bruun, E.P.G., Duka, A., 2018. Artificial intelligence, jobs and the future of work: racing with the machines. Basic Income Stud. 13 https://doi.org/10.1515/bis-2018-0018.

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., et al., 2006. Work system design for patient safety: the SEIPS model. Qual. Saf. Health Care 15, i50–i58. https://doi.org/10.1136/qshc.2005.015842.

Carayon, P., Wetterneck, T.B., Rivera-Rodriguez, A.J., et al., 2014. Human factors systems approach to healthcare quality and patient safety. Appl. Ergon. 45, 14–25. https://doi.org/10.1016/j.apergo.2013.04.023.

Catchpoole, J., Nanda, G., Vallmuur, K., et al., 2022. Application of a machine learning-based decision support tool to improve an injury surveillance system workflow. Appl. Clin. Inf. 13, 700–710. https://doi.org/10.1055/a-1863-7176.

Chen, M., Decary, M., 2020. Artificial intelligence in healthcare: an essential guide for health leaders. Healthc. Manag. Forum 33, 10–18. https://doi.org/10.1177/0840470419873123.

Chockley, K., Emanuel, E., 2016. The end of radiology? Three threats to the future practice of radiology. J. Am. Coll. Radiol. 13, 1415–1420. https://doi.org/10.1016/j.jacr.2016.07.010.

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. 13, 319. https://doi.org/10.2307/249008.

Esmaeilzadeh, P., 2020. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med. Inf. Decis. Making 20, 170. https://doi.org/10.1186/s12911-020-01191-1.

Felmingham, C.M., Adler, N.R., Ge, Z., et al., 2021. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. Am. J. Clin. Dermatol. 22, 233–242. https://doi.org/10.1007/s40257-020-00574-4.

Ferretti, A., Schneider, M., Blasimme, A., 2018. Machine learning in medicine: opening the new data protection black box. European Data Protection Law Review 4, 320–332. https://doi.org/10.21552/edpl/2018/3/10.

Forookhi, A., Laschena, L., Pecoraro, M., et al., 2023. Bridging the experience gap in prostate multiparametric magnetic resonance imaging using artificial intelligence: a prospective multi-reader comparison study on inter-reader agreement in PI-RADS v2.1, image quality and reporting time between novice and expert readers. Eur. J. Radiol. 161, 110749 https://doi.org/10.1016/j.ejrad.2023.110749.

Gille, F., Jobin, A., Ienca, M., 2020. What we talk about when we talk about trust: theory of trust for AI in healthcare. Intelligence-Based Medicine 1–2, 100001. https://doi.org/10.1016/j.ibmed.2020.100001.

He, J., Baxter, S.L., Xu, J., et al., 2019. The practical implementation of artificial intelligence technologies in medicine. Nat. Med. 25, 30–36. https://doi.org/10.1038/s41591-018-0307-0.

Hemmer, P., Schemmer, M., Riefle, L., et al., 2022. Factors that influence the adoption of human-AI collaboration in clinical decision-making. https://doi.org/10.48550/ARXIV.2204.09082.

Holden, R.J., 2011. What stands in the way of technology-mediated patient safety improvements?: a study of facilitators and barriers to physicians' use of electronic health records. J. Patient Saf. 7, 193–203. https://doi.org/10.1097/PTS.0b013e3182388cfa.

Holden, R.J., Karsh, B.-T., 2010. The technology acceptance model: its past and its future in health care. J. Biomed. Inf. 43, 159–172. https://doi.org/10.1016/j.jbi.2009.07.002.

Holden, R.J., Carayon, P., Gurses, A.P., et al., 2013. Seips 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 56, 1669–1686. https://doi.org/10.1080/00140139.2013.838643.

Hopf, C., Schmidt, C., 1993. On the relationship between intra-familial social experiences, personality development and political orientations: Documentation and discussion of the methodological procedure in a study on this topic. [Zum Verhältnis von innerfamilialen sozialen Erfahrungen, Persönlichkeitsentwicklung und politischen Orientierungen: Dokumentation und Erörterung des methodischen Vorgehens in einer Studie zu diesem Thema. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-456148.

Hosny, A., Parmar, C., Quackenbush, J., et al., 2018. Artificial intelligence in radiology. Nat. Rev. Cancer 18, 500–510. https://doi.org/10.1038/s41568-018-0016-5.

Howard, J., 2019. Artificial intelligence: implications for the future of work. Am. J. Ind. Med. 62, 917–926. https://doi.org/10.1002/ajim.23037.

Jarrahi, M.H., 2018. Artificial intelligence and the future of work: human-AI symbiosis in organizational decision making. Bus. Horiz. 61, 577–586. https://doi.org/10.1016/j.bushor.2018.03.007.

Kuckartz, U., Rädiker, S., 2022. Qualitative content analysis: Methods, practice, computer support [Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung], fifth ed. Beltz Juventa, Weinheim, Germany https://www.beltz.de/produkt_detailansicht/47304-qualitative-inhaltsanalyse-methoden-praxis-computerunterstuetzung.html.

Lai, Y., Kankanhalli, A., Ong, D.C., 2021. Human-AI collaboration in healthcare: a review and research agenda. In: Hawaii International Conference on System Sciences 2021. University of Hawai'i at Manoa, Hamilton Library, Honolulu, HI. https://doi.org/10.24251/HICSS.2021.046.

Langlotz, C.P., 2019. Will artificial intelligence replace radiologists? Radiology: Artif. Intell. 1, e190058 https://doi.org/10.1148/ryai.2019190058.

Li, R.C., Asch, S.M., Shah, N.H., 2020. Developing a delivery science for artificial intelligence in healthcare. NPJ digital medicine 3, 107. https://doi.org/10.1038/s41746-020-00318-y.

Litjens, G., Kooi, T., Bejnordi, B.E., et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

Meskó, B., Hetényi, G., Győrffy, Z., 2018. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv. Res. 18, 545. https://doi.org/10.1186/s12913-018-3359-4.

Miles, K., 2020. Radiomics for personalised medicine: the long road ahead. Br. J. Cancer 122, 929–930. https://doi.org/10.1038/s41416-019-0699-8.

Obermeyer, Z., Emanuel, E.J., 2016. Predicting the future — big data, machine learning, and clinical medicine. N. Engl. J. Med. 375, 1216–1219. https://doi.org/10.1056/NEJMp1606181.

Reddy, S., Fox, J., Purohit, M.P., 2019. Artificial intelligence-enabled healthcare delivery. J. R. Soc. Med. 112, 22–28. https://doi.org/10.1177/0141076818815510.

Salwei, M.E., Carayon, P., 2022. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J. Cogn. Eng. Decis. Mak., 15553434221097357 https://doi.org/10.1177/15553434221097357.

Salwei, M.E., Carayon, P., Hoonakker, P.L.T., et al., 2021. Workflow integration analysis of a human factors-based clinical decision support in the emergency department. Appl. Ergon. 97, 103498 https://doi.org/10.1016/j.apergo.2021.103498.

Schwappach, D., Ratwani, R., 2023. Electronic health record usability contributions to patient safety and clinician burnout: a path forward. J Patient Saf Publish Ahead of Print. https://doi.org/10.1097/PTS.0000000000001130.

Sezgin, E., 2023. Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. DIGITAL HEALTH 9, 20552076231186520. https://doi.org/10.1177/20552076231186520.

Strohm, L., Hehakaya, C., Ranschaert, E.R., et al., 2020. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. Eur. Radiol. 30, 5525–5532. https://doi.org/10.1007/s00330-020-06946-y.

Sunoqrot, M.R.S., Saha, A., Hosseinzadeh, M., et al., 2022. Artificial intelligence for prostate MRI: open datasets, available applications, and grand challenges. Eur Radiol Exp 6, 35. https://doi.org/10.1186/s41747-022-00288-8.

The DECIDE-AI Steering Group, 2021. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nat. Med. 27, 186–187. https://doi.org/10.1038/s41591-021-01229-5.

Tong, A., Sainsbury, P., Craig, J., 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int. J. Qual. Health Care 19, 349–357. https://doi.org/10.1093/intqhc/mzm042.

Turkbey, B., Haider, M.A., 2022. Deep learning-based artificial intelligence applications in prostate MRI: brief summary. BJR 95, 20210563. https://doi.org/10.1259/bjr.20210563.

Twilt, J.J., Van Leeuwen, K.G., Huisman, H.J., et al., 2021. Artificial intelligence based algorithms for prostate cancer classification and detection on magnetic resonance imaging: a narrative review. Diagnostics 11, 959. https://doi.org/10.3390/diagnostics11060959.

van Leeuwen, K.G., de Rooij, M., Schalekamp, S., et al., 2021. How does artificial intelligence in radiology improve efficiency and health outcomes? Pediatr. Radiol. https://doi.org/10.1007/s00247-021-05114-8.

VERBI Software, 2021. MAXQDA 2022 [computer software]. Berlin, Germany: VERBI Software. Available from maxqda.com.

Von Eschenbach, W.J., 2021. Transparency and the black box problem: why we do not trust AI. Philos Technol 34, 1607–1622. https://doi.org/10.1007/s13347-021-00477-0.

Wang, W., Siau, K., 2019. Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. J. Database Manag. 30, 61–79. https://doi.org/10.4018/JDM.2019010104.

Wenderott, K., Gambashidze, N., Weigl, M., 2022. Integration of artificial intelligence into sociotechnical work systems—effects of artificial intelligence solutions in medical imaging on clinical efficiency: protocol for a systematic literature review. JMIR Res Protoc 11, e40485. https://doi.org/10.2196/40485.

Wenderott, K, Krups, J, Luetkens, JA, Weigl, M, 2024. Prospective effects of an artificial intelligence-based computer-aided detection system for prostate imaging on routine workflow and radiologists' outcomes. Eur. J. Radiol. 170, 111252. https://doi.org/10.1016/j.ejrad.2023.111252.

Wolff, J., Pauling, J., Keck, A., Baumbach, J., 2021. Success factors of artificial intelligence Implementation in Healthcare. Frontiers in Digital Health 3. https://doi.org/10.3389/fdgth.2021.594971.

Wong, S.T.C., 2018. Is pathology prepared for the adoption of artificial intelligence? Cancer cytopathology 126, 373–375. https://doi.org/10.1002/cncy.21994.

Yin, J., Ngiam, K.Y., Teo, H.H., 2021. Role of artificial intelligence applications in real-life clinical practice: systematic review. J. Med. Internet Res. 23, e25759 https://doi.org/10.2196/25759.

3.6. Publication 6: Workflow Analysis and Evaluation of a Next-Generation Phenotyping Tool: A Qualitative Study of Face2Gene

Wenderott K, Krups J, Zaruchas F, Krawitz PM, Weigl M, Lesmann H (2025)
Eur J Hum Genet (2025). doi: 10.1038/s41431-025-01875-0

The supplementary material can be accessed via this link.

ESHG

www.nature.com/ejhg

ARTICLE OPEN

Check for updates

Workflow analysis and evaluation of a next-generation phenotyping tool: A qualitative study of Face2Gene

Katharina Wenderott [1]✉, Jim Krups [1], Fiona Zaruchas [1], Peter Krawitz [2], Matthias Weigl [1,4] and Hellen Lesmann [2,3,4]

The diagnosis of rare genetic disorders often involves prolonged delays, with facial features serving as key diagnostic clues. Next-Generation Phenotyping (NGP) tools, such as Face2Gene, utilize Artificial Intelligence (AI)-driven algorithms to analyze patient photographs and list differential diagnoses based on facial dysmorphism. Despite their growing use and proven clinical value, their integration into clinical workflows remains poorly understood. This study evaluates Face2Gene's implementation into routine clinical care with barriers and facilitators to successful adoption. We conducted a literature review, followed by in-depth interviews with 15 geneticists across university hospitals in Germany. Results showed an overall positive appraisal of the tool among clinicians with emphasis on its usability. Key workflow barriers comprised IT integration and patient consent process. Despite being an additional step in the diagnostic pathway, Face2Gene has been effectively incorporated into geneticists' diagnostic routines, facilitating decision processes, and potentially expediting diagnoses for some patients. Our findings contribute to the existing literature on NGP technologies by demonstrating that effective integration of Face2Gene can enhance clinicians' efficiency and quality of work. To maximize impact of NGP technologies in genetic medicine, future implementation efforts should strive for clinicians' acceptance particularly through user-friendly design and sustained organizational support in course of workflow implementation. Study registration: German Register for Clinical Trials (DRKS) DRKS00032436

European Journal of Human Genetics; https://doi.org/10.1038/s41431-025-01875-0

BACKGROUND

Despite ongoing advances in sequencing technologies, diagnosing rare genetic disorders remains a lengthy and complex process, often leading to a prolonged diagnostic odyssey for affected individuals. On average, the time to diagnosis is 4–5 years, with some cases extending beyond a decade [1]. A key element in achieving a reliable diagnosis is often the phenotype of the patient. Since approximately 40% of rare diseases are associated with facial abnormalities, facial features can provide valuable diagnostic clues, particularly in disorders linked to facial dysmorphism [2, 3].

Alongside significant advancements in Next-Generation Sequencing, a growing array of Next-Generation Phenotyping (NGP) algorithms has emerged that are mostly based on recent advancements in computer vision and Artificial Intelligence (AI). The integration of such AI algorithms into more complex user interfaces results in NGP tools that are used in clinical practice to analyze phenotypes and suggest differential diagnoses based on facial images of patients [4–7]. The primary goal of these tools is to streamline the diagnostic process and reduce the time to diagnosis. Numerous studies have already demonstrated the technical validity and clinical utility of NGP by assessing its performance [8–10]. A recent large-scale study in Germany reported a significant improvement in the prioritization of exome data by using AI-based results of image analysis (PEDIA approach) [11, 12]. Face2Gene (F2G) is a software suite that provides access to NGP algorithms, namely DeepGestalt and GestaltMatcher, and is widely used by clinicians for analyzing patient photographs. F2G is also already implemented in clinical routine care in some clinics for human genetics [6, 7, 13].

Despite the growing use of NGP tools such as F2G in clinical practice, systematic investigations are lacking on how these tools are integrated into everyday clinical workflows or how efficiently they are utilized in routine care. Additionally, the factors influencing clinicians' acceptance of these technologies have not been thoroughly assessed. Understanding these challenges is crucial to evaluating the real-world impact of NGP tools, as their true value depends on their acceptance, effective integration, and consistent use in clinical settings [14]. The work system model developed by Carayon [15, 16] provides a useful framework to study the implementation of novel technologies such as NGP tools. This sociotechnical model conceives workflows as an interaction between five key elements: people, tasks, tools and technologies, physical environment, and organizational structures [15]. Introducing a new technology, such as an AI-driven or NGP tool, affects all elements of clinical work systems altering the relationships between these elements [17]. By adopting this systemic approach, the model captures significant changes in the work system as well as the impact on human interaction with the technology [14, 18, 19].

[1]Institute for Patient Safety, University Hospital Bonn, Bonn, Germany. [2]Institute for Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany. [3]Institute for Human Genetics, University Hospital Bonn, Bonn, Germany. [4]These authors contributed equally: Matthias Weigl, Hellen Lesmann. ✉email: katharina.wenderott@ukbonn.de

Received: 20 November 2024 Revised: 27 March 2025 Accepted: 12 May 2025
Published online: 23 May 2025

This study aims to conduct a provider-focused evaluation of the integration of NGP tools into clinical workflows, using F2G as a use case. Furthermore, it seeks to identify the barriers and facilitators to adopting NGP tools across different clinical settings, offering insights to support their broader implementation in the future.

MATERIALS AND METHODS

Study design

In a multi-stage study design, we first conducted a literature review assessing studies which applied F2G, which was followed by a cross-sectional interview study. The study was pre-registered (DRKS00032436) and approved by the Ethics Committee of the Medical Faculty, Bonn University (2023-161-BO). The reporting of this study adheres to the COREQ checklist (COnsolidated criteria for REporting Qualitative research) [20].

Use case: Face2Gene

Face2Gene (https://www.face2gene.com/; FDNA, Atlanta, GA) uses a deep convolutional neural network called DeepGestalt to detect syndromic phenotypes in facial photos [6]. DeepGestalt classifies 300 syndromes, with a 90% accuracy rate for including the correct diagnosis among the top 10 suggestions. F2G is offered as a web service accessible for healthcare providers and is currently used in 130 countries across 2.000 clinics according to the manufacturer FDNA [21].

Literature review

To provide a better foundation and contextualization of our results a literature review was conducted using 'Face2Gene' as a search term in the databases PubMed, Embase, Web of Science, and CENTRAL. The search was finalized on May 2nd 2024. Publications were de-duplicated in Rayyan and screened in Zotero. Inclusion decisions adhered to the PICOS framework and were conducted by one researcher (KW) (see Appendix 1). After screening the full-texts of studies matching the selection criteria, one researcher (KW) extracted the study design, condition, case/s, controls, population, F2G application used and outcomes assessed in regard to F2G into MS Excel. The extracted data were reviewed by a second researcher (JK).

Interview study with clinicians

Participants. A convenience sampling approach was applied. In Germany, F2G has already been widely adopted within clinical genetic departments and, to a growing extent, in pediatric settings. Clinicians from pediatric and genetics departments in German university hospitals were invited, with an expected sample size of 30. Licensed physicians in human genetics or pediatrics with B2-level German language proficiency were included. Excluded were professionals with no experience with F2G.

Measures and content. The semi-structured interview guide (Appendix 2) was adapted from a German study on AI integration [22]. During the study development, the research team met and discussed relevant aspects for the interview guideline and adapted it to the specific context. Following its development, the interview guide underwent preliminary testing with a team member to assess its comprehensibility and clarity. The interviews lasted approximately 30 min. After mutual introductions, participants were asked about their age, gender, clinic, department, job title, and their professional experience.

For the assessment of attitudes towards AI, participants were asked one open question on general expectations regarding AI in their medical discipline. Additional questions addressed chances, risks, and perceived effects of AI-based technology use on patient safety. If necessary, interviewees were asked to elaborate their answers. Additionally, they were asked if their appraisal on AI changed since they started working with an AI tool.

Regarding use and usability, participants were asked about their perceived usefulness, perceived ease of use, and actual use of F2G. In terms of workflow integration, the questions primarily focused on how F2G was incorporated into clinical routines. We explored factors that influenced the decision to use or not use the AI tool before its application, as well as conditions that facilitated or hindered its use after the decision was made, i.e., during its integration into the workflow. Facilitators were defined as "any factor that promotes or enhances the integration or use of the AI system in the workflow," while barriers were described as "any factor that limits or restricts the integration or use of the AI system" [23].

Procedure. Departments were contacted via email with a study description and Unipark survey link to schedule the interview and obtain informed consent. Interviews were conducted either in person or via telephone, for both settings it was ensured that the interviews took place in a separate room to minimize disturbances. For compensation, participants received a 60€ incentive. Semi-structured interviews in German were audio-recorded and conducted by graduate assistants with a background in medicine (FZ) and psychology (JK), after training by lead researcher KW. Demographic data was collected separately on paper (Appendix 3). Data saturation was indicated by no novel ideas emerging from the interviews and repetitions occurring across interviews and statements.

Analysis. Interviews were transcribed using audio.whisper in RStudio [24]. Data extraction on the use of F2G which were usage, frequency of use, purpose and medium was done by KW and checked by JK.

Next, we used qualitative content analysis following the method of Kuckartz and Rädiker [25] to identify barriers and facilitators clinicians experienced when using F2G. To guide this process, we created a codebook, which is a set of instructions defining the goals of the analysis, the structure, and the meaning of different codes. Two researchers (KW, JK) independently coded the transcripts using MAXQDA 24, resolving discrepancies through discussion or by consulting a third researcher. The list of main categories and their definitions can be found in Appendix 4. In the next step, we created subcategories using an inductive approach. We compiled these categories into a detailed codebook with clear definitions. To ensure consistency among raters, we tested the codebook in three interviews, discussed any coding discrepancies, and refined the definitions as necessary. Additionally, authors JK and KW individually identified the work system elements relating to the dimensions of facilitators and barriers, establishing a consensus through discussion; consistent to Wooldridge et al. [26, 27] and Wenderott et al. [14].

RESULTS

Literature review

The search yielded 192 publications, of which 40 were included after screening (see Fig. 1), a list of excluded studies is provided in Appendix 5. Appendix 6 lists key characteristics, populations, and outcomes of the included studies. Included studies comprised 11 case-control, 11 retrospective validations, 10 single-case, 5 multi-case, and 2 prospective studies. One study did not clearly describe the study setup [10].

Only one study, Marwaha et al. [13], assessed clinicians' experiences with F2G, while the vast majority of studies focused on case descriptions or accuracy. Marwaha et al. [13] used a six-question survey with ten respondents to evaluate F2G's use and usefulness one year after implementation. They also assessed its use as an educational tool for trainees. Overall, usability was rated positively, though participants expressed concerns about its usefulness, informed consent, and data security. To summarize, our literature review showed that the literature concerning systematic evaluations of clinicians' experiences is missing, while F2G has been integrated into routine use and demonstrated a good accuracy.

Clinicians' interviews

Sample. Seventeen interviews were conducted between October 2023 and April 2024, with two excluded due to no contact with F2G, leaving a final sample of 15 clinicians (see Table 1). Participants from six German cities were affiliated with human genetics departments at university hospitals. One was also a pediatric specialist. Mean age was 40.07 years (Standard deviation [SD] = 10.33), with 11 females and 4 males.

The interview duration was on average 22:08 min (SD = 06:16 min). Ten participants used F2G in their routine workflow, three used it occasionally, and two used it before but did not continue. These experience- and usage-levels were used to distinguish between the three user groups: (1) Non-Users, who had tried F2G but did not continue to use it; (2) Occasional Users, clinicians who

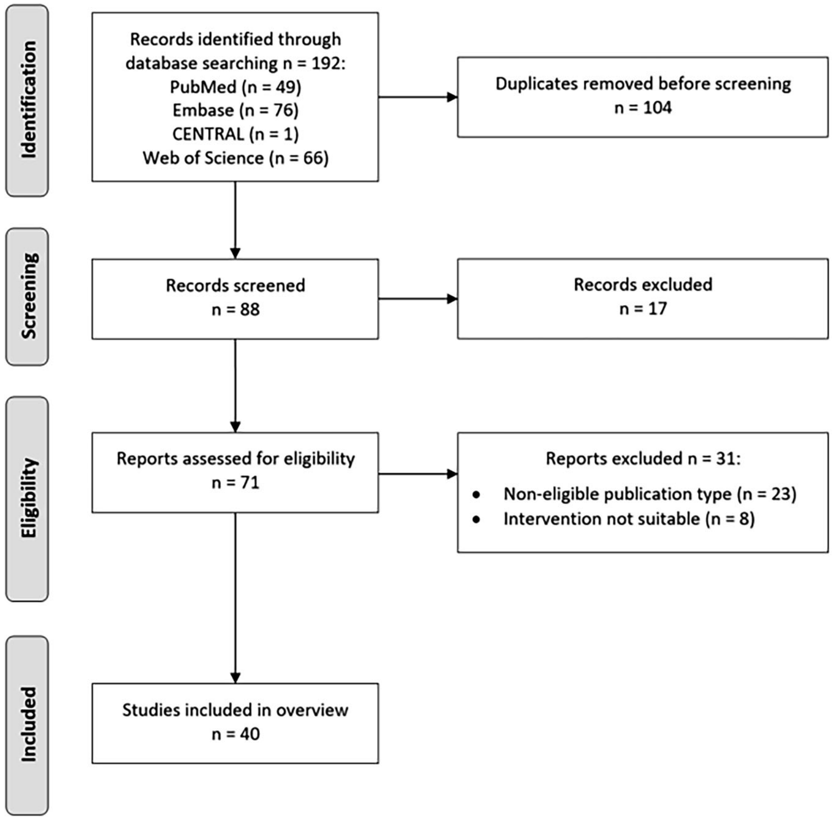


Fig. 1 Literature review flow chart.

Table 1. Details on interviewees.

No.	Position	Work experience (in years)	Use	User Group	Purpose for use	Medium	Reported frequency of tool use
2	Attending	14	No	Non-User	Patient care	Computer	10 times
4	Resident	1	Yes	Routine	Patient care	Computer	Almost daily, 1–2 times per day
5	Resident	5	Yes	Occasionally	Research, patient care	Smartphone	1–2 cases & a research project
6	Resident	2	Yes	Routine	Patient care, research	Computer	All patients with dysmorphic features
7	Attending	22	Yes	Routine	Patient care	Computer	Half of patients
8	Attending	19	Yes	Occasionally	Patient care	Computer	Two times a month
9	Resident	2	Yes	Routine	Patient care, research	Computer	All patients with dysmorphic features
10	Attending	25	No	Non-User	Patient care	No information	No use
11	Resident	2	Yes	Routine	Patient care	Computer	Once a week
12	Attending	23	Yes	Occasionally	Patient care	Computer	Every third patient (30–40 times a year)
13	Attending	14	Yes	Routine	Patient care	Computer	All patients with suspected genetic syndrome
14	Attending	21	Yes	Routine	Patient care	Computer	Twice a month
15	Attending	15	Yes	Routine	Patient care, research	Computer	All patients
16	Attending	15	Yes	Routine	Patient care	Computer	Almost all patients with dysmorphic syndrome
17	Resident	0.8	Yes	Routine	Patient care	Computer	All children

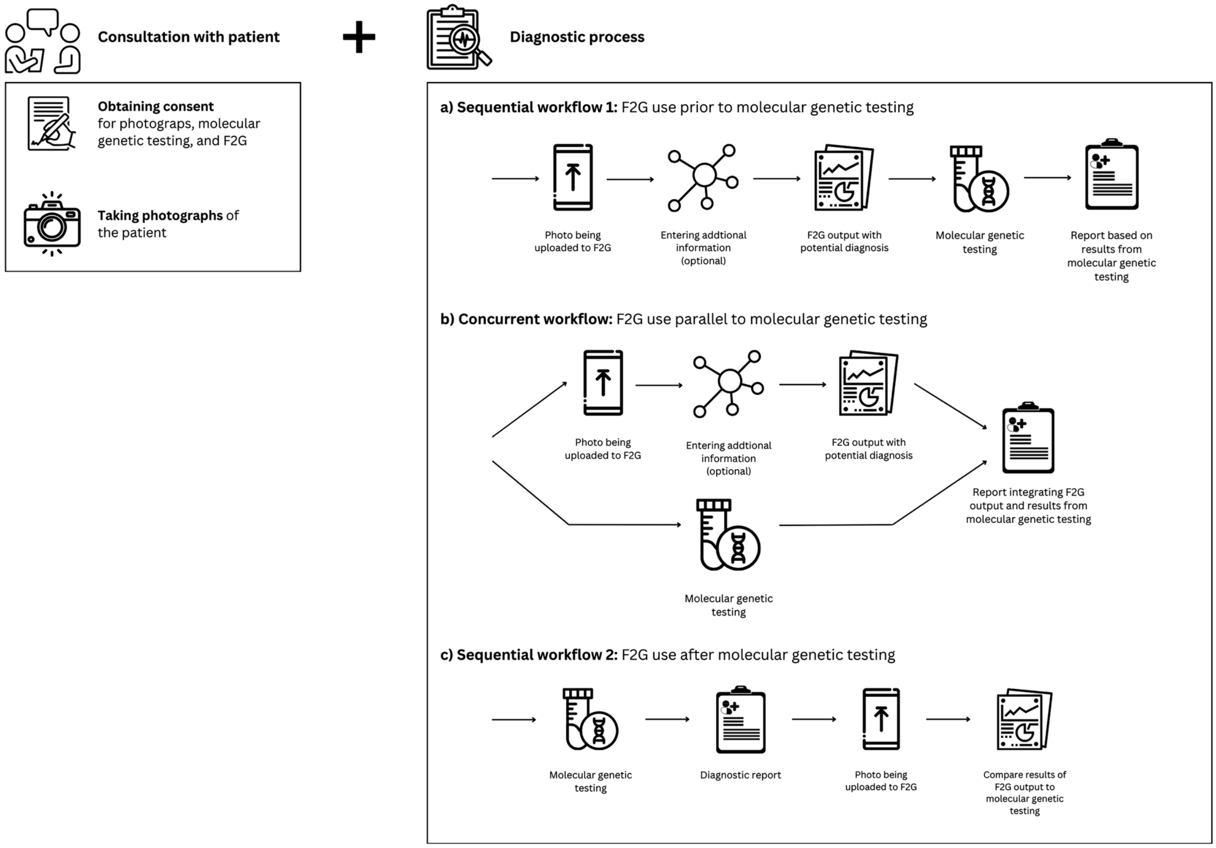


Fig. 2 Prototypical workflows for using F2G after the patient consultation.

used it only occasionally for specific cases; (3) Routine Users, who used it repeatedly for almost all their eligible patients.

Clinicians' reports on AI use in genetics. When participants were asked about their expectations regarding the use of AI in their field, they most frequently described several potential benefits: supporting clinical care processes (10/15; 10 out of 15 interviewees), enhancing efficiency (8/15), improving diagnostics (7/15), and easing the burden on healthcare professionals (4/15). However, they also expressed concerns about various risks associated with AI use, such as the potential for misuse (9/15), overtrust (8/15), bias in AI solutions (3/15), and the possible loss of professional skills (2/15). Details can be found in Appendix 7 and 8. Concerning factors that would encourage their acceptance of AI solutions, participants most often emphasized the importance of data protection (10/15), particularly when handling medical data. Transparency (9/15) was also a key factor, with one participant stating, "I definitely want to know how something works when I upload patient data there. Even if I can't always fully understand everything, I think such transparency greatly helps in the acceptance of these tools" (ID: IV08). Additionally, the source of the AI solution (6/15) and the data used to train the AI (6/15) were significant considerations. As one participant noted, "Those are definitely things I would question, just like whether a university hospital is behind it or if it's a commercial provider. And who is financing these commercial providers or what kind of organization is behind them" (IV09).

Evaluation of Face2Gene. Drawing upon the clinicians' reports of F2G use, we identified the following F2G workflow. It starts with patient consultation, where clinicians obtain consent from patients or guardians and take photographs. A separate informed consent is required for uploading these images to F2G for processing and providing diagnostic suggestions for report generation.

Only two participants used F2G during patient consultations. For using F2G after the patient contact, interviewees described three distinct workflows: nine clinicians used it sequentially before molecular genetic testing for initial diagnostic impressions; four applied it concurrently and combined the results from both methods when generating the diagnostic report; and three sometimes used it afterwards to describe cases or evaluate molecular results. These workflows are illustrated in Fig. 2.

Participants reported to switch between different workflows, as their actual decision whether and when they use F2G in routine patient care was influenced by various factors. Most interviewees (12/15) noted that time pressure significantly affected its use, while 11 mentioned that patient characteristics influenced their decision to use. Organizational factors, including workspace, digitalization level, and overall acceptance were also important, and three participants indicated that F2G's reputation in the professional community impacted their usage decisions. For detailed definitions and examples, refer to Table 2.

Twelve of the 15 participants reported positive experiences with F2G. One noted, "Yes, F2G is very useful, […] We would never make a diagnosis based on this alone without molecular correlation, but as a decision-making aid, it's definitely a good one" (IV07). In contrast, three participants shared negative experiences, including a non-user (IV02) who said, "I used it in its very early stages, and it didn't really help me. I was more frustrated. […]". Participants also indicated that the tool's usefulness varied based on factors like the specific patient (6/15), purpose of use (4/15), and the clinician's experience (3/15). One participant (IV06) remarked that "the usefulness is very high, especially at the beginning of one's career. […] Once you

Table 2. Reported factors influencing clinicians' decision to use Face2Gene.

Factor	Definition	Number of Statements	Number of Interviews	Example Statement
Time pressure	Workload and time pressures in the department influencing the use of F2G.	14	12	"I think if there is time pressure, I would use Face to Gene […] less than I normally would." (IV 09)
Patient characteristics	Patient characteristics which influence the decision to use, e.g. highly dysmorphic facial features of a patient.	12	11	"So, if I have a patient who looks noticeably dysmorphic for me, then I definitely want to use it." (IV 15)
Organizational factors	Hospital organization providing support, technical equipment or a positive attitude towards the use of F2G.	10	6	"[…] if I had a professional smartphone, I think I would use it more likely." (IV 05)
Popularity of F2G	The prominence and/or reputations of F2G having an impact on F2G usage.	4	3	"[…] you simply hear from other colleagues at congresses that they also use it and that it has also contributed to solving the disease in question." (IV 13)
Trustworthiness of the source of F2G	Trustworthiness attributed toward the source and/or developers of F2G impacting the decision to use F2G.	2	1	"[…] many colleagues […] don't get involved with projects like F2G for data protection reasons, because they say Face2Gene […], they have their servers in the USA […]" (IV 06)

gain experience, you might be able to figure things out on your own. I often see this with experienced colleagues—they already have their own ideas, which Face2Gene might then support".

Facilitators and barriers of using Face2Gene. A total of 87 statements was collected on facilitators and barriers for successfully integrating F2G into clinical work. Facilitators were mentioned more frequently (51/87 statements, 15/15 interviews) than barriers (36/87 statements, 12/15 interviews). These were categorized into eight dimensions, with definitions provided in Table 3. When examining the identified dimensions of workflow integration, usability, time investment, and accessibility emerged as the dimensions with the highest number of facilitators. These facilitators were largely related to the AI solution itself. Two dimensions—obtaining consent and additional work steps—were identified solely as barriers and related to tasks required of the users. Mixed evaluations were reported regarding local support and IT integration, which varied across organizations.

DISCUSSION

Artificial Intelligence in genetic medicine offers great potential for improving diagnosis and treatment. However, sustainable AI implementation in clinical settings presents challenges for providers and organizations. This study evaluates the use of Next-Generation Phenotyping tools in the diagnostic process, focusing on F2G. Our review revealed only one study addressing F2G user evaluations [13]. This lack of including healthcare provider viewpoints in AI integration is not unique to genetics, also appearing in other medical fields [28–30]. Our research contributes critical factors influencing the adoption and rejection of NGP tools and elucidates the conditions that either facilitate or hinder their utilization within clinical workflows in genetics.

Our review highlighted that F2G has successfully mastered the translation from usage in a research context into routine care, a critical step for AI tools in healthcare [29]. Although numerous studies utilizing F2G to report case descriptions, they frequently fail to include specific workflow details. The predominant focus on performance metrics, such as the accuracies of DeepGestalt or GestaltMatcher, often neglects clinicians' experiences, which is essential for the successful adoption of novel technologies [31, 32]. Marwaha et al. employed a brief questionnaire to capture user opinions one-year post-implementation at a single institution, providing an initial overview of user experiences [13]. Building on this, our study employed a semi-structured interview methodology with participants from diverse hospitals to enhance these findings. Although Marwaha et al. found F2G to be beneficial for diagnostic decision-making, they also identified significant reservations concerning accuracy, informed consent, confidentiality, and patient uptake [13]. In alignment with these findings, our participants reported favorable usability of F2G; however, they expressed concerns regarding the informed consent process.

Our interviewees highlighted that the use of F2G is just one component of the diagnostic process. Moreover, all users reported using it in addition to molecular genetic analyses. The extracted workflows align with research on integrating AI into work processes [28]. Interviewees described F2G as easy to use and evaluating its usefulness positively, which are key contributors for the intention to use of a novel technology [33]. Additionally, we identified several key factors impacting the decision to actually use F2G. Time pressure was the most frequently cited factor influencing participants' use of F2G, aligning with findings from other AI integration studies [18, 22]. However, time pressure was noted as less intense in genetics than in other specialties like pediatrics. Most users reported utilizing F2G for dysmorphic patients, while some did so out of curiosity or when a suspected condition was unclear. This is noteworthy, as the performance of

Table 3. Definitions, frequency, and exemplary statements of facilitators and barriers for Face2Gene implementation.

Dimension	Definition	Work System Elements						Facilitators			Barriers		
		P	T	T/T	O	PE	EE	No. of Statements	No. of IVs	Example	No. of Statements	No. of IVs	Example
Usability	Users can interact effectively and intuitively with F2G to accomplish their goals.	x		x				34	15	"So, I think it's a very clear program, you can very easily create new cases and simply upload the image by drag & drop […]" (IV 04)	8	5	"When you've uploaded the picture, where you then enter the features, so it's somehow not so clearly separated, so every time I stumble over it again […]" (IV 14)
Time investment	The amount of time needed to integrate and use F2G in the daily clinical workflow.		x	x				6	3	"So, it's also something you can use very quickly. It's not particularly time-consuming." (IV 07)	5	4	"You are dependent on the photos, which can take a bit of time, that you can't do it straight after the consultation [..]" (IV 11)
Obtaining consent	The process of getting the patients'/legal guardians' consent is affecting the use of F2G.	x	x		x			0	0	NA	9	4	"And all the paperwork, maybe you forgot to get the declaration of consent. You then have to obtain it afterwards." (IV 06)
Additional work step	Using F2G is an additional work step in the process.		x	x				0	0	NA	6	5	"Well, it's just another chunk of work, I say, that you don't necessarily have to do to get a diagnosis most of the time." (IV 02)
Local IT integration	Ensures that F2G can seamlessly communicate and share data with other technologies used.			x	x			1	1	"And the sending of patient photos via email works quite well for us, as we have a secure cloud for that." (IV 05)	6	6	"So that was not integrated and is not integrated yet. So, we have to download the images manually and then upload them back into Face2Gene. […]" (IV 02)
Accessibility	The accessibility of F2G, e.g. being accessible on multiple devices, influences its workflow integration.			x	x			5	5	"It fits in well, because you can do it whenever you want, even on your PC, and you can theoretically log in anywhere, no matter which device you're working on […]" (IV 04)	0	0	NA
Local support	At the institution clinicians have support when using F2G.	x	x		x			3	3	"There is support for it and some use it more, some use it less, but it is generally accepted […]." (IV 14)	1	1	"At the moment, it's more the case that only a few people even know about the app." (IV 05)
Influence by direct supervisor	The supervisor or head of the team is being supportive or unsupportive for using F2G.	x			x			2	2	"[…] my boss, who keeps suggesting that we should do it […]" (IV 02).	1	1	"And I definitely notice that not everyone is equally enthusiastic about these things and this influences people's use and sometimes also their interpretation." (IV 06)

No number, *IV* interview, *P* people, *T* tasks, *T/T* tools and technologies, *O* organization, *PE* physical environment, *EE* external environment, *F2G* Face2Gene, *IT* Information technology, *NA* not applicable.

these tools relies on the distinctiveness of facial dysmorphisms [7]. To optimize the cost-benefit ratio of time and AI performance, participants reported to select suitable patients for facial analysis, as indicated by their usage patterns. For most users, specific patient characteristics were central to their decision to use F2G, as reflected in their responses about how frequently they utilize the tool. Given the high proportion of diseases with facial dysmorphism [2, 3], NGP tools will continue to have ample use cases. In a scientific context, however, the analysis of less distinct disorders could also be useful, as AI can also recognize patterns that are not always apparent to humans [34].

An analysis of the workflow revealed that the consultation and patient photography are standard steps of genetic assessments. This may account for the observed absence of specific facilitators or barriers linked to F2G in this stage. However, from obtaining patient consent onward, additional implementation factors for successful F2G integration were identified. We observed more facilitators than barriers, potentially indicating a good workflow fit. The distribution of facilitators and barriers across different elements of the work system model highlights the complexity of adopting AI tools like F2G. Facilitators such as usability, time investment, and accessibility are closely tied to the AI solution itself namely the element of tools and technology, emphasizing the importance of good design when implementing AI in clinical settings. In contrast, barriers like obtaining consent and additional work steps are linked to the element of tasks, suggesting that administrative processes may hinder smooth integration. Additionally, the lack of work-issued smartphones or tablets for many participants presents a significant obstacle, forcing a reliance on external photographers and desktop computers, which complicates the workflow. These issues point to broader organizational challenges, such as the need for better local support and more seamless IT integration, which were met with varying levels of satisfaction among participants. The amount of barriers and facilitators associated with the element organization highlights the impact organizational support could have, especially as it was also among the factors the inform users' decision to use F2G. For example, the process of obtaining consent was a central barrier to use, which could be mitigated by the organization providing standardized forms for using F2G. It is crucial to clarify that patient consent is only required for research-related activities, such as the technical validation of the AI algorithm's performance. Similar to numerous Next-Generation Sequencing (NGS) platforms, NGP software is frequently designated 'for research purposes only', which may lead to user confusion. If these tools are approved as medical devices for decision support in the future and become part of accredited workflows, this usage would typically not necessitate additional patient consent. The work system element 'people' was identified mostly for task-related facilitators and barriers, but three users also identified their supervisor as having a strong impact on their workflow using F2G. We identified no facilitators or barriers related to the environment, likely because the introduction of F2G as a web-based AI solution did not lead to any changes.

Salwei and Carayon outlined three essential sociotechnical considerations for integrating AI into healthcare systems: the alignment with work systems, compatibility with existing workflows, and enhancement of clinical decision-making processes [17]. Through analyzing the work system barriers and facilitators, our study was able to demonstrate an overall good fit of F2G in the work system, supporting their first consideration. In examining how F2G fits into the workflow and the positive feedback regarding the time required to use it, we found that it integrates well with the existing workflow, which supports the second consideration. Although F2G is not a mandatory step in the diagnostic process, participants found it valuable for diagnostic purposes and case discussions, which aligns with findings from Marwaha et al. [13]. This supports the third consideration regarding workflow integration. Overall, we found that F2G is a well-integrated AI solution, yielding positive outcomes such as enhanced user satisfaction, shorter time to diagnosis, and greater acceptance of the technology. Clinicians' positive appraisal regarding the use of F2G corresponds with the primary advantages participants sought from AI in genetics, namely improved diagnostics and increased efficiency, which are relevant across healthcare setting [35, 36]. The information gained from this study can be used to successfully implement other NGP tools such as GestaltMatcher and Phenoscore [4, 7]. Even though the scientific benefit of these tools has been already proven by demonstrating their accuracy, their seamless transfer into clinical routine is essential to actually promote significant improvements in diagnoses of rare genetic disorders. This can only be achieved by identifying and addressing concerns of clinicians during the implementation ensuring a smooth fit into the workflow.

Participants generally viewed AI and NGP tools as beneficial for improving care processes, efficiency, diagnostics, and reducing healthcare workloads. However, they expressed concerns about potential misuse, over-reliance, bias, and loss of professional skills. This resonates with a study by Hallowell et al. which explored stakeholder perspectives on NGP technology [37]. They additionally highlighted that NGP tools could democratize access to diagnoses—an aspect not raised in our interviews, likely due to the widespread availability of genotyping methods in Germany. This shows a limitation of our study: all participants were recruited within Germany to ensure comparable workflows. Given that F2G is deployed globally, future studies should examine workflow integration across different countries and healthcare settings to identify universally applicable recommendations. Additionally, our literature review featured a rapid synthesis of available studies involving F2G, whereas a more systematic and rigorous approach might have been more effective in ensuring that all relevant studies were included (i.e., from gray literature). Although our response rate was lower than expected, sufficient data saturation was achieved, as later interviews did not reveal new themes. It must be noted, that we aimed to include both pediatricians and geneticists; however, only geneticists responded. Pediatricians, who may benefit most from NGP for patients with facial dysmorphism, often have shorter visit times, making their perspective on the cost-benefit ratio of NGP particularly valuable [38, 39]. Future studies should prioritize this user group to assess their specific needs and challenges. Finally, the limited number of interviews, particularly with non-users, may have introduced bias. Future research should explore detailed comparisons between various user groups, possibly examining correlations with diverse user characteristics [40].

Our study enhances the existing literature on NGP technologies by evaluating the integration of F2G as a specific use case across various institutions. Through clinician interviews, we mapped the workflow of using F2G, which has been effectively incorporated into routine practice, despite being an additional step in the diagnostic process without replacing molecular genetic testing. This indicates that NGP technology can significantly improve healthcare efficiency and quality, provided that clinicians' acceptance is further enhanced. Our analysis of facilitators and barriers using the work system model underscores essential considerations for future design and implementation, particularly the importance of user-friendly design and ensuring organizational support.

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are not publicly available due to the sensitive nature of the interview content, which may compromise participant anonymity, but are available from the corresponding author on reasonable request.

REFERENCES

1. Marwaha S, Knowles JW, Ashley EA. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. Genome Med. 2022;14:23. https://doi.org/10.1186/s13073-022-01026-w.
2. Lesmann H, Klinkhammer H, Krawitz PM. The future role of facial image analysis in ACMG classification guidelines. Medizinische Genetik. 2023;35:115–21. https://doi.org/10.1515/medgen-2023-2014.
3. Hart T, Hart P. Genetic studies of craniofacial anomalies: clinical implications and applications. Orthod Craniofacial Res. 2009;12:212–20. https://doi.org/10.1111/j.1601-6343.2009.01455.x.
4. Dingemans AJM, Hinne M, Truijen KMG, Goltstein L, Van Reeuwijk J, De Leeuw N, et al. PhenoScore quantifies phenotypic variation for rare genetic diseases by combining facial analysis with other clinical features using a machine-learning framework. Nat Genet. 2023;55:1598–607. https://doi.org/10.1038/s41588-023-01469-w.
5. Porras AR, Rosenbaum K, Tor-Diez C, Summar M, Linguraru MG. Development and evaluation of a machine learning-based point-of-care screening tool for genetic syndromes in children: a multinational retrospective study. Lancet Digital Health. 2021;3:e635–43. https://doi.org/10.1016/S2589-7500(21)00137-0.
6. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med. 2019;25:60–4. https://doi.org/10.1038/s41591-018-0279-0.
7. Hsieh T-C, Bar-Haim A, Moosa S, Ehmke N, Gripp KW, Pantel JT, et al. GestaltMatcher facilitates rare disease matching using facial phenotype descriptors. Nat Genet. 2022;54:349–57. https://doi.org/10.1038/s41588-021-01010-x.
8. Pantel JT, Hajjir N, Danyel M, Elsner J, Abad-Perez AT, Hansen P, et al. Efficiency of computer-aided facial phenotyping (DeepGestalt) in individuals with and without a genetic syndrome: diagnostic accuracy study. J Med Internet Res. 2020;22:e19263 https://doi.org/10.2196/19263.
9. Reiter AMV, Pantel JT, Danyel M, Horn D, Ott C-E, Mensah MA. Validation of 3 computer-aided facial phenotyping tools (DeepGestalt, GestaltMatcher, and D-Score): comparative diagnostic accuracy study. J Med Internet Res. 2024;26:e42904 https://doi.org/10.2196/42904.
10. Mishima H, Suzuki H, Doi M, Miyazaki M, Watanabe S, Matsumoto T, et al. Evaluation of Face2Gene using facial images of patients with congenital dysmorphic syndromes recruited in Japan. J Hum Genet. 2019;64:789–94. https://doi.org/10.1038/s10038-019-0619-z.
11. Schmidt A, Danyel M, Grundmann K, Brunet T, Klinkhammer H, Hsieh T-C, et al. Next-generation phenotyping integrated in a national framework for patients with ultrarare disorders improves genetic diagnostics and yields new molecular findings. Nat Genet. 2024;56:1644–53. https://doi.org/10.1038/s41588-024-01836-1.
12. Hsieh T-C, Mensah MA, Pantel JT, Aguilar D, Bar O, Bayat A, et al. PEDIA: prioritization of exome data by image analysis. Genet Med. 2019;21:2807–14. https://doi.org/10.1038/s41436-019-0566-2.
13. Marwaha A, Chitayat D, Meyn M, Mendoza-Londono R, Chad L. The point-of-care use of a facial phenotyping tool in the genetics clinic: Enhancing diagnosis and education with machine learning. Am J Med Genet Part A. 2021;185:1151–8.
14. Wenderott K, Krups J, Weigl M, Wooldridge AR. Reported Facilitators and Barriers to AI Implementation in Routine Medical Imaging: Systematic Review and Qualitative Analysis. J Med Internet Res (forthcoming). https://doi.org/10.2196/63649.
15. Carayon P. Human Factors of Complex Sociotechnical Systems. Appl Ergonomics. 2006;37:525–35. https://doi.org/10.1016/j.apergo.2006.04.011.
16. Carayon P. The Balance Theory and the Work System Model … Twenty Years Later. Int J Hum-Comput Interact. 2009;25:313–27. https://doi.org/10.1080/10447310902864928.
17. Salwei ME, Carayon P A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cognitive Eng Decision Making 2022:15553434221097357. https://doi.org/10.1177/15553434221097357.
18. Salwei ME, Hoonakker P, Carayon P, Wiegmann D, Pulia M, Patterson BW. Usability of a human factors-based clinical decision support in the emergency department: lessons learned for design and implementation. Human Factors 2022:187208221078625. https://doi.org/10.1177/00187208221078625 PT - Article.
19. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergonomics. 2020;84:103033. https://doi.org/10.1016/j.apergo.2019.103033.
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19:349–57. https://doi.org/10.1093/intqhc/mzm042.
21. FDNA INC. About FDNA n.d. https://fdna.com/about-us/?_ga=2.60898921.493362435.1668012522-103749772.1641811944. (accessed November 23, 2022).
22. Wenderott K, Krups J, Luetkens J.A, Weigl M. Radiologists' perspectives on the workflow integration of an artificial intelligence-based computer-aided detection system: a qualitative study. Appl Ergonomics. 2024;117:104243. https://doi.org/10.1016/j.apergo.2024.104243.
23. Niezen MGH, Mathijssen JJP. Reframing professional boundaries in healthcare: a systematic review of facilitators and barriers to task reallocation from the domain of medicine to the nursing domain. Health Policy. 2014;117:151–69. https://doi.org/10.1016/j.healthpol.2014.04.016.
24. Wijffels J audio.whisper: Transcribe Audio Files using the "Whisper" Automatic Speech Recognition Model. 2024.
25. Kuckartz U, Rädiker S Qualitative Content Analysis: Methods, Practice, Computer Support [Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung]. 5th ed. Weinheim, Germany: Beltz Juventa; 2022.
26. Wooldridge AR, Carayon P, Hundt AS, Hoonakker PLT. SEIPS-based process modeling in primary care. Appl Ergonomics. 2017;60:240–54. https://doi.org/10.1016/j.apergo.2016.11.010.
27. Wooldridge AR, Carayon P, Hoonakker P, Hose B-Z, Eithun B, Brazelton T, et al. Work system barriers and facilitators in inpatient care transitions of pediatric trauma patients. Appl Ergonomics. 2020;85:103059. https://doi.org/10.1016/j.apergo.2020.103059.
28. Wenderott K, Krups J, Zaruchas F, Weigl M. Effects of artificial intelligence implementation on efficiency in medical imaging—a systematic literature review and meta-analysis. Npj Digit Med. 2024;7:265 https://doi.org/10.1038/s41746-024-01248-9.
29. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res. 2021;23:e25759 https://doi.org/10.2196/25759.
30. Fujimori R, Liu K, Soeno S, Naraba H, Ogura K, Hara K, et al. Acceptance, barriers, and facilitators to implementing artificial intelligence–based decision support systems in emergency departments: quantitative and qualitative evaluation. JMIR Form Res. 2022;6:e36501 https://doi.org/10.2196/36501.
31. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res. 2020;22:e15154 https://doi.org/10.2196/15154.
32. Andersen TO, Nunes F, Wilcox L, Coiera E, Rogers Y. Introduction to the special issue on human-centred AI in healthcare: challenges appearing in the wild. ACM Trans Comput-Hum Interact. 2023;30:1–12. https://doi.org/10.1145/3589961.
33. Holden RJ, Karsh B-T. The technology acceptance model: its past and its future in health care. J Biomed Inf. 2010;43:159–72. https://doi.org/10.1016/j.jbi.2009.07.002.
34. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. Clin Transl Sci. 2023;16:2106–11. https://doi.org/10.1111/cts.13619.
35. Hoppe N, Härting R-C, Rahmel A Potential Benefits of Artificial Intelligence in Healthcare. In: Lim CP, Vaidya A, Chen Y-W, Jain V, Jain LC, editors. Artificial Intelligence and Machine Learning for Healthcare, vol. 229, Cham: Springer International Publishing; 2023, p. 225–49. https://doi.org/10.1007/978-3-031-11170-9_9.
36. Park C-W, Seo SW, Kang N, Ko B, Choi BW, Park CM, et al. Artificial intelligence in health care: current applications and issues. J Korean Med Sci. 2020;35. https://doi.org/10.3346/jkms.2020.35.e379.
37. Hallowell N, Badger S, McKay F, Kerasidou A, Nellåker C. Democratising or disrupting diagnosis? Ethical issues raised by the use of AI tools for rare disease diagnosis. SSM - Qualitative Res Health. 2023;3:100240. https://doi.org/10.1016/j.ssmqr.2023.100240.
38. Heinen F, Landgraf MN, Böhmer J, Borggräfe I, Bidlingmaier C, Krohn K, et al. Was vom Tage übrig bleibt – Arbeitsalltag einer Universitätsmedizin für Kinder und Jugendliche [What Remains of the Day – The Daily Work of a University Hospital for Children and Adolescents]. Z für Evidenz, Fortbild und Qualät im Gesundheitswesen [J Evid, Educ Qual Health]. 2016;117:20–6. https://doi.org/10.1016/j.zefq.2016.09.002.
39. Sukenik-Halevy R, Ludman MD, Ben-Shachar S, Raas-Rothschild A. The time-consuming demands of the practice of medical genetics in the era of advanced genomic testing. Genet Med. 2016;18:372–7. https://doi.org/10.1038/gim.2015.96.
40. Vedel I, Lapointe L, Lussier M-T, Richard C, Goudreau J, Lalonde L, et al. Healthcare professionals' adoption and use of a clinical information system (CIS) in primary care: Insights from the Da Vinci study. Int J Med Inf. 2012;81:73–87. https://doi.org/10.1016/j.ijmedinf.2011.11.002.

AUTHOR CONTRIBUTIONS

KW: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Visualization, Writing—Original Draft, Writing—Preparation, Review, and Editing. JK: Data Curation, Investigation, Visualization, Writing–Review and Editing. FZ: Data Curation, Investigation. PK: Supervision,

Validation, Writing—Review and Editing. MW: Conceptualization, Funding Acquisition, Supervision, Validation, Writing—Review and Editing. HL: Conceptualization, Investigation, Writing—Review and Editing, Supervision, Validation.

FUNDING
Institutional budget, no external funding. Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS
The authors declare no competing interests.

ETHICAL APPROVAL
The study was approved by the Ethics Committee of the Medical Faculty, University of Bonn (2023-161-BO). Informed consent was obtained from all participants prior to the commencement of the interviews.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING
During the preparation of this manuscript, the authors used ChatGPT (version GPT-4, OpenAI) to optimize the readability and wording. After using this tool, the authors reviewed and edited the content as required and take full responsibility for the content of the publication.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-025-01875-0.

Correspondence and requests for materials should be addressed to Katharina Wenderott.

4. Discussion with References

The work on this dissertation was influenced by major technological advances from 2021 to 2025, such as the widespread availability of generative AI, which have shaped both the research itself and the growing public interest in this topic (Grzybowski et al., 2024; Mishra et al., 2024). While AI is being increasingly adopted into healthcare, the aim of the work for this dissertation – studying the workflow integration of AI into the complex sociotechnical work system present in healthcare – remains as relevant as when this dissertation topic was proposed (Agarwal et al., 2024). While working on the topic of this dissertation, various reviews on factors influencing AI implementation in healthcare were published; however, few included only real-world clinical studies, potentially because only a limited number of prospective studies of AI implementation have been published (Ahmed et al., 2023; Chomutare et al., 2022; Hassan et al., 2024; O. Higgins et al., 2023; Hua et al., 2024; Lambert et al., 2023; Lokaj et al., 2023; Mennella et al., 2024; Tricco et al., 2023).

Therefore, a systematic review which only included studies that focused on AI implementation in clinical care was undertaken as the first major step of the work for the dissertation to synthesize the current literature base. This review (Chapters 3.1.-3.3.), including 48 studies, involved the systematic exploration of the impact of AI integration in routine clinical workflows in medical imaging on efficiency, clinician outcomes, and workflows (Wenderott, Krups, Zaruchas, et al., 2024). Overall a positive trend towards the impact of AI on the time clinicians needed for tasks was identified, which could help to accelerate the diagnostic process and reduce clinicians' workload (Khalifa & Albadawy, 2024; Marco-Ruiz et al., 2024). The identified AI-augmented workflows aligned with strategies from recent publications and matched the specific tasks for which AI was employed (M. Chen et al., 2024; Dahlblom et al., 2023; Ng et al., 2022). For example, in detection tasks, AI served as a second reader, either sequentially or concurrently, while in segmentation tasks, it typically was used as the first reader (Wenderott, Krups, Zaruchas, et al., 2024). These findings mirror those of M. Chen et al. (2024), who reviewed AI's impact on medical imaging, identifying reading paradigms (concurrent, second, or first reader) and noting significant time savings through the use of AI, particularly when AI served as a pre-screening tool. In contrast to their work, the systematic review presented in this dissertation was focused solely on studies conducted in real-world clinical settings.

To further enhance the understanding of AI implementation in clinical settings, in the third publication (Chapter 3.3.) a novel framework was introduced for comparing the extent to which AI tools are integrated into routine workflows (Wenderott, Krups, Weigl, et al., 2025). This new framework allowed for more meaningful comparisons across studies by differentiating the level of AI implementation. The added value has been also demonstrated by the epistemic network analysis (ENA) of facilitators and barriers identified in the included studies in relation to the different work system elements, which differed between the level of implementation (Wenderott, Krups, Weigl, et al., 2025).
The work for this dissertation also comprised two empirical investigations (Chapters 3.4.-3.6.) on use cases from real-world clinical care. The aim of both was to map the changes that occur through the introduction of an AI tool, to determine the effects on HCPs, and to identify barriers and facilitators of AI implementation. The first use case (Chapters 3.4. and 3.5.) involved the study of an AI tool which was an initial implementation, as it had not yet been fully integrated into clinical routines (Wenderott, Krups, Weigl, et al., 2025). In the second use case (Chapter 3.6.) most clinicians used the AI tool under study for nearly all eligible patients, therefore it was classified as a full implementation study.
The considered first use case (Chapters 3.4. and 3.5.) was the introduction of an AI tool designed to support prostate MRI scan interpretation. A unique feature of this empirical investigation was the opportunity to conduct work observations and interviews with radiologists both before and after the AI tool's introduction (Wenderott, Krups, Luetkens, Gambashidze, et al., 2024; Wenderott, Krups, Luetkens, & Weigl, 2024). Even though clinicians expressed a positive attitude towards the AI solution and their intentions to use it before the implementation, their actual use was largely dependent on existing time pressure and their workload. The tool was associated with an increase in reading time for more complex cases, which differed from previously published experimental studies on this AI solution (Cipollari et al., 2022; Faiella et al., 2022). This highlights the need for studies in actual clinical practice assessing the impact of AI under realistic conditions, including the complexities of the hospital work system (Choudhury & Asan, 2020).
Notably, on-site workflow observations revealed variations in AI-tool use among surveyed radiologists, including both concurrent and sequential reading paradigms, which reflected that there were no organizational guidelines for integrating the AI tool into their processes (Wenderott, Krups, Luetkens, & Weigl, 2024). In contrast to the findings of the systematic

literature review (Chapter 3.2.), the most common workflow involved radiologists using the AI tool as an add-on, incorporating its findings during the writing of the final report, after completing their traditional workflow. This was primarily due to long upload times, which was perceived as unhelpful and caused delays, ultimately leading to a deficient fit of the AI tool into the workflow (Wenderott, Krups, Luetkens, & Weigl, 2024). In agreement with the Conceptual Model of Workflow Integration this poor operability and fit caused negative clinician outcomes such as increased workload or workarounds as mentioned in the interviews, though these were not evident in the radiologist’s answers to the questionnaires on stress and workload (Salwei et al., 2021; Wenderott, Krups, Luetkens, Gambashidze, et al., 2024; Wenderott, Krups, Luetkens, & Weigl, 2024). In contrast to the negative evaluation of this AI tool and its workflow integration, the second use case (Chapter 3.6.) involved the study of a well-integrated AI tool in human genetics departments. Although the AI tool was also used as an add-on to the traditional procedure involving molecular genetic testing in different workflow variations, it was positively evaluated for its ease of use, its quick generation of results and its ability to provide an initial idea of the patient's diagnosis.

In addition to the analysis of the workflows and the AI tool adoption, in both use cases facilitators and barriers to AI integration were analyzed. When matching these to the different work system elements, it was evident that in the first use case (Chapter 3.5.) the identified facilitators were mostly linked to the task and the people, such as usability and the implementation process. Nevertheless, the primary barrier was the technology’s poor interoperability and poor fit with the workflow, resulting in a negative evaluation by the participants. As the technical problems stemmed from the interaction with the organization’s IT infrastructure, the barriers were linked to the people and the organization. These findings resonate well with the ENA results (Chapter 3.3.) for initial implementation studies, highlighting the key role of clinicians’ expectations and attitudes. This conclusion is also strengthened by the theoretical foundation in the TAM, which sheds light on the individuals’ decision to adopt a novel technology. In the second use case (Chapter 3.6.) a SEIPS-based examination of facilitators and barriers, comparable to the approach by Wooldridge et al. (2020), was included. In contrast to the ENA (Chapter 3.3.) for full implementation studies the results of this study revealed that facilitators were more attributable to task- and technology characteristics, instead of being linked to the organization and people. The identified barriers underscored the significant role of organizational factors in successful AI integration (Wenderott, Krups, Zaruchas, et al., 2025). This finding is particularly noteworthy as the multicenter nature of the study allowed for the inclusion of diverse organizational measures, enabling an assessment of their impact on AI workflow integration. Overall the results of the research underline the diversity of facilitators and barriers the can impact AI implementation into real-world clinical practice. By analyzing the associated work system elements, it was evident how complex the workflow integration is and how many different stakeholders on different levels are involved in the process (Choudhury, 2022; Salwei et al., 2021).

4.1. Strengths and Limitations of this Dissertation

To meet the diversity of AI workflow integration the work for this dissertation involved the employment of a variety of research methods, including workplace observations, time measurements, questionnaires, and interviews. Additionally, it incorporated three distinct study designs to offer a comprehensive as well as complementary analysis. While the systematic review established a baseline of existing research, the two use cases provided detailed analyses and enhanced the findings with profound insights and examples from clinical practice. However, it should be noted, that a systematic review is a time-intensive process and the rapid advancements in AI might lead to some methods, technologies, or conclusions being outdated upon publication (Borah et al., 2017; Mahuli et al., 2023).

A major strength of the work for this dissertation is the consistent use of theoretical frameworks (SEIPS and TAM), which provide a strong conceptual foundation, and emphasize the role of human factors in AI implementation (Carayon et al., 2006; Choudhury, 2022; Davis, 1989; Salwei & Carayon, 2022). To amplify this research perspective, the research teams for the empirical investigations in this dissertation always included a clinician or someone with a medical background from the relevant specialty. This interdisciplinary approach is noteworthy, as it ensured that the domain expertise needed for effective analysis and interpretation was an integral part of the research.

All research was conducted with a focus on real-world clinical settings to address the research-practice gap for AI in healthcare. However, this research focus and methodological approach also presents potential limitations. Since the studies were conducted in clinical environments, they were limited to specific use cases and rather small participant samples. Although working in real-world clinical settings allowed for a more valid and accurate depiction of the complexities of sociotechnical work systems and their impact on AI implementation, it limits the generalizability across different contexts and settings (Hettinger et al., 2015). Additionally, the studies did not involve any reporting on the technical properties of the AI systems (e.g., algorithm performance, training data), what may limit the scope of the holistic analysis and ability to draw conclusions concerning other AI tools or even updated versions of the studied AI tools. However, this research was focused on AI tools that were chosen by the organization or department to be used in clinical care, involving one commercial tool and one publicly available online tool, thereby presenting real-life organizational procedures.

Since AI implementation was and still is relatively new at the time of the work for this dissertation, standardized survey instruments were not yet available, resulting in a significant reliance on qualitative research. Despite adhering to quality standards, such as interviewer training, double coding, and use of reporting checklists, certain biases – including selection bias from convenience sampling or potential biases in data analysis – cannot be entirely ruled out (Williams et al., 2020).

A recurring question throughout this dissertation was what the novelty of AI is and what distinguishes it from other technologies that are used in healthcare. AI has unique capabilities that include processing large datasets, handling unstructured data, and learning from its input (Bajwa et al., 2021; Buch et al., 2018; Dipaola et al., 2024). In the past, the introduced algorithms were mostly frozen, which prevented the self-learning features (Mashar et al., 2023). Therefore, a majority of the extracted facilitators, barriers, and implementation strategies for AI that were observed are similar to those of other technologies used in healthcare such as electronic health records (Gagnon et al., 2010, 2014; Mennella et al., 2024). Nevertheless, the algorithms' opaque decision-making processes can create issues of trust, liability, and cognitive challenges for users, which will be even more important when algorithms change over time (Choudhury & Asan, 2020; Mennella et al., 2024; Pashkov et al., 2020). As a result, current research is focused on improving algorithm explainability and governments strive for regulatory frameworks such as the EU AI Act, where it is even more important to understand the users' and patients' needs (Aboy et al., 2024; H. Chen et al., 2022; Pierce et al., 2022). While these regulatory and policy challenges were not the focus of this dissertation, these questions should also be continuously considered with respect to future challenges of AI integration into sociotechnical systems in healthcare (Habli et al., 2020; Mennella et al., 2024).

4.2. Implications for Practice and Research

To optimize the integration of AI into healthcare settings, this dissertation includes several implications and recommendations for AI implementation for practice and research. Drawing upon the identified facilitators or barriers of AI implementation, best practices for future implementation processes were developed which should be considered when implementing AI. Nevertheless, since each clinical work setting is to some extent unique – as demonstrated by the human factors approaches which served as the theoretical foundation of the work for this dissertation – each implementation process should be adapted to the local context and the respective users. Thus, three key aspects evident in the included studies should be highlighted that can serve as best practice recommendations: First, successful AI implementation benefits greatly from the early inclusion of relevant stakeholders, such as clinicians, local supervisors, or the IT department. They should be involved in evaluating the need for an AI tool for a specific task, selecting a suitable solution, and designing the implementation process. Second, organizational support was frequently identified as a facilitator, or as a barrier when absent, therefore adequate change management procedures should be employed (Kim et al., 2023; Petersson et al., 2022). Examples from the conducted studies include user training, procedures for informing patients, and protocols for handling errors. Third, seamless integration with the local IT infrastructure is essential for implementing any new technology; without it, workarounds and frustration, as seen in the radiology use case, can occur. Additionally, designing an AI-augmented workflow that is iteratively tested and refined could be a beneficial strategy to ensure a good fit of the technology and that it is safe and reliable to use, as suggested by Agarwal et al. (2024).

The results of the work for this dissertation also highlight the importance of research on AI solutions in real-world conditions and emphasize that "work-as-imagined" often differs from "work-as-done" (Wenderott, Krups, Luetkens, & Weigl, 2024). Therefore, comparing results from experimental studies with real-world implementation settings is crucial for determining the impact of an AI tool in clinical care. Additionally, experimental studies can offer insights into factors like user training or varying levels of explainability that mostly influence user acceptance. These findings can then be used to inform real-world implementation processes. It should be noted that at every stage of technology development, appropriate study designs are necessary to determine the safe and effective use of the technology (D. Higgins & Madai, 2020; Marcus et al., 2024; Park et al., 2022). However, to draw a conclusion across different studies it was highlighted in the systematic review presented in this dissertation that there is a need for more comprehensive reporting and adhering to published reporting guidelines. Additionally, research independent from manufacturers and developers is necessary, as many studies identified in the systematic review were conducted by authors with potential conflicts of interest. Nevertheless, there are significant opportunities for researchers and practitioners to evaluate the introduction of AI in healthcare institutions, as many are implementing it without any thorough and systematic assessment concerning assumed effects and outcomes (Yin et al., 2021). The research presented in this dissertation also leads to the suggestion of the development of a checklist for AI implementation in clinical settings which could help to bridge the gap between research and practice. Owoyemi et al. (2024) are currently working on such a checklist and have published a pre-print of their work. Their work takes into account the complex sociotechnical work system in healthcare and involves the aim of facilitating the seamless integration of AI into clinical workflows.

4.3. Conclusion

To investigate the workflow integration of AI in clinical practice, the work for this dissertation covered a range of clinical workflows and AI solutions, by providing a systematic literature review and two in-depth analyses of clinical AI use cases. The conducted studies were focused mostly on the individual level, where a seamless fit into the workflow and a good usability of the AI tool were highlighted. Additionally, strong organizational influences, particularly regarding supportive IT infrastructure and change management procedures, were observed as relevant factors. By analyzing facilitators and barriers through the SEIPS model, recommendations for future AI implementation processes in healthcare were extracted and underline the practical relevance of the work for this dissertation. For research, the need for holistic assessments of AI implementation is emphasized to ensure that AI augments clinical workflows and reduces healthcare providers' workload. Thus, the work for this dissertation contributes to the safe and effective integration of AI in healthcare, hopefully in the long term leading to positive outcomes of AI integration for providers and patients such as improved care and reduced waiting times.

4.4. References

Aboy, M., Minssen, T., & Vayena, E. (2024). Navigating the EU AI Act: Implications for regulated digital medical products. *Npj Digital Medicine*, *7*(1), 237. https://doi.org/10.1038/s41746-024-01232-3

Agarwal, R., Dugas, M., & Gao, G. (Gordon). (2024). Augmenting Physicians with Artificial Intelligence to Transform Healthcare: Challenges and Opportunities. *Journal of Economics & Management Strategy*, *33*(2), 360–374. https://doi.org/10.1111/jems.12555

Ahmed, M. I., Spooner, B., Isherwood, J., Lane, M., Orrock, E., & Dennison, A. (2023). A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. *Cureus*. https://doi.org/10.7759/cureus.46454

Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: Transforming the practice of medicine. *Future Healthcare Journal*, *8*(2), e188–e194. https://doi.org/10.7861/fhj.2021-0095

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the Time and Workers Needed to Conduct Systematic Reviews of Medical Interventions Using Data from the PROSPERO Registry. *BMJ Open*, *7*(2), e012545. https://doi.org/10.1136/bmjopen-2016-012545

Buch, V. H., Ahmed, I., & Maruthappu, M. (2018). Artificial Intelligence in Medicine: Current Trends and Future Possibilities. *British Journal of General Practice*, *68*(668), 143–144. https://doi.org/10.3399/bjgp18X695213

Carayon, P., Schoofs Hundt, A., Karsh, B.-T., Gurses, A. P., Alvarado, C. J., Smith, M., & Flatley Brennan, P. (2006). Work System Design for Patient Safety: The SEIPS Model. *BMJ Quality & Safety*, *15*(suppl 1), i50–i58. https://doi.org/10.1136/qshc.2005.015842

Chen, H., Gomez, C., Huang, C.-M., & Unberath, M. (2022). Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence from a Systematic Review. *Npj Digital Medicine*, *5*(1), 156. https://doi.org/10.1038/s41746-022-00699-2

Chen, M., Wang, Y., Wang, Q., Shi, J., Wang, H., Ye, Z., Xue, P., & Qiao, Y. (2024). Impact of Human and Artificial Intelligence Collaboration on Workload Reduction in Medical Image Interpretation. *Npj Digital Medicine*, *7*(1), 349. https://doi.org/10.1038/s41746-024-01328-w

Chomutare, T., Tejedor, M., Svenning, T. O., Marco-Ruiz, L., Tayefi, M., Lind, K., Godtliebsen, F., Moen, A., Ismail, L., Makhlysheva, A., & Ngo, P. D. (2022). Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. *International Journal of Environmental Research and Public Health*, *19*(23), 16359. https://doi.org/10.3390/ijerph192316359

Choudhury, A. (2022). Toward an Ecologically Valid Conceptual Framework for the Use of Artificial Intelligence in Clinical Settings: Need for Systems Thinking, Accountability, Decision-making, Trust, and Patient Safety Considerations in Safeguarding the Technology and Clinicians. *JMIR Human Factors*, *9*(2), e35421. https://doi.org/10.2196/35421

Choudhury, A., & Asan, O. (2020). Human Factors: Bridging Artificial Intelligence and Patient Safety. *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, *9*(1), 211–215. https://doi.org/10.1177/2327857920091007

Cipollari, S., Pecoraro, M., Forookhi, A., Laschena, L., Bicchetti, M., Messina, E., Lucciola, S., Catalano, C., & Panebianco, V. (2022). Biparametric Prostate MRI: Impact of a Deep Learning-Based Software and of Quantitative ADC Values on the Inter-Reader Agreement of Experienced and Inexperienced Readers. *La Radiologia Medica*, *127*(11), 1245–1253. https://doi.org/10.1007/s11547-022-01555-9

Dahlblom, V., Dustler, M., Tingberg, A., & Zackrisson, S. (2023). Breast Cancer Screening with Digital Breast Tomosynthesis: Comparison of Different Reading Strategies Implementing Artificial Intelligence. *European Radiology*, *33*(5), 3754–3765. https://doi.org/10.1007/s00330-022-09316-y

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Dipaola, F., Gebska, M. A., Gatti, M., Levra, A. G., Parker, W. H., Menè, R., Lee, S., Costantino, G., Barsotti, E. J., Shiffer, D., Johnston, S. L., Sutton, R., Olshansky, B., & Furlan, R. (2024). Will Artificial Intelligence Be "Better" Than Humans in the Management of Syncope? *JACC: Advances*, *3*(9), 101072. https://doi.org/10.1016/j.jacadv.2024.101072

Faiella, E., Vertulli, D., Esperto, F., Cordelli, E., Soda, P., Muraca, R. M., Moramarco, L. P., Grasso, R. F., Beomonte Zobel, B., & Santucci, D. (2022). Quantib Prostate Compared to an Expert Radiologist for the Diagnosis of Prostate Cancer on mpMRI: A Single-Center Preliminary Study. *Tomography (Ann Arbor, Mich.)*, *8*(4), 2010–2019. https://doi.org/10.3390/tomography8040168

Gagnon, M.-P., Desmartis, M., Labrecque, M., Car, J., Pagliari, C., Pluye, P., Frémont, P., Gagnon, J., Tremblay, N., & Légaré, F. (2010). Systematic Review of Factors Influencing the Adoption of Information and Communication Technologies by Healthcare Professionals. *Journal of medical systems*, *36*(1), 241–277. https://doi.org/10.1007/s10916-010-9473-4

Gagnon, M.-P., Nsangou, É.-R., Payne-Gagnon, J., Grenier, S., & Sicotte, C. (2014). Barriers and Facilitators to Implementing Electronic Prescription: A Systematic Review of User Groups' Perceptions. *Journal of the American Medical Informatics Association: JAMIA*, *21*(3), 535–541. https://doi.org/10.1136/amiajnl-2013-002203

Grzybowski, A., Pawlikowska–Łagód, K., & Lambert, W. C. (2024). A History of Artificial Intelligence. *Clinics in Dermatology*, *42*(3), 221–229. https://doi.org/10.1016/j.clindermatol.2023.12.016

Habli, I., Lawton, T., & Porter, Z. (2020). Artificial Intelligence in Health Care: Accountability and Safety. *Bulletin of the World Health Organization*, *98*(4), 251–256. https://doi.org/10.2471/BLT.19.237487

Hassan, M., Kushniruk, A., & Borycki, E. (2024). Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review. *JMIR Human Factors*, *11*, e48633. https://doi.org/10.2196/48633

Hettinger, L. J., Kirlik, A., Goh, Y. M., & Buckle, P. (2015). Modelling and Simulation of Complex Sociotechnical Systems: Envisioning and Analysing Work Environments. *Ergonomics*, *58*(4), 600–614. https://doi.org/10.1080/00140139.2015.1008586

Higgins, D., & Madai, V. I. (2020). From Bit to Bedside: A Practical Framework for Artificial Intelligence Product Development in Healthcare. *Advanced Intelligent Systems*, *2*(10), 2000052. https://doi.org/10.1002/aisy.202000052

Higgins, O., Short, B. L., Chalup, S. K., & Wilson, R. L. (2023). Artificial intelligence (AI) and machine learning (ML) based decision support systems in mental health: An integrative review. *International Journal of Mental Health Nursing*, *32*(4), 966–978. https://doi.org/10.1111/inm.13114

Hua, D., Petrina, N., Young, N., Cho, J.-G., & Poon, S. K. (2024). Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. *Artificial Intelligence in Medicine*, *147*, 102698. https://doi.org/10.1016/j.artmed.2023.102698

Khalifa, M., & Albadawy, M. (2024). AI in Diagnostic Imaging: Revolutionising Accuracy and Efficiency. *Computer Methods and Programs in Biomedicine Update*, *5*, 100146. https://doi.org/10.1016/j.cmpbup.2024.100146

Kim, J. Y., Boag, W., Gulamali, F., Hasan, A., Hogg, H. D. J., Lifson, M., Mulligan, D., Patel, M., Raji, I. D., Sehgal, A., Shaw, K., Tobey, D., Valladares, A., Vidal, D., Balu, S., & Sendak, M. (2023). Organizational Governance of Emerging Technologies: AI Adoption in Healthcare. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1396–1417. https://doi.org/10.1145/3593013.3594089

Lambert, S. I., Madi, M., Sopka, S., Lenes, A., Stange, H., Buszello, C.-P., & Stephan, A. (2023). An Integrative Review on the Acceptance of Artificial Intelligence Among Healthcare Professionals in Hospitals. *Npj Digital Medicine*, *6*(1), 111. https://doi.org/10.1038/s41746-023-00852-5

Lokaj, B., Pugliese, M.-T., Kinkel, K., Lovis, C., & Schmid, J. (2023). Barriers and Facilitators of Artificial Intelligence Conception and Implementation for Breast Imaging Diagnosis in Clinical Practice: A Scoping Review. *European Radiology*, *34*(3), 2096–2109. https://doi.org/10.1007/s00330-023-10181-6

Mahuli, S. A., Rai, A., Mahuli, A. V., & Kumar, A. (2023). Application ChatGPT in Conducting Systematic Reviews and Meta-Analyses. *British Dental Journal*, *235*(2), 90–92. https://doi.org/10.1038/s41415-023-6132-y

Marco-Ruiz, L., Hernández, M. Á. T., Ngo, P. D., Makhlysheva, A., Svenning, T. O., Dyb, K., Chomutare, T., Llatas, C. F., Muñoz-Gama, J., & Tayefi, M. (2024). A Multinational Study on Artificial Intelligence Adoption: Clinical Implementers' Perspectives. *International Journal of Medical Informatics*, *184*, 105377. https://doi.org/10.1016/j.ijmedinf.2024.105377

Marcus, H. J., Ramirez, P. T., Khan, D. Z., Layard Horsfall, H., Hanrahan, J. G., Williams, S. C., Beard, D. J., Bhat, R., Catchpole, K., Cook, A., Hutchison, K., Martin, J., Melvin, T., Stoyanov, D., Rovers, M., Raison, N., Dasgupta, P., Noonan, D., Stocken, D., … Paez, A. (2024). The IDEAL Framework for Surgical Robotics: Development, Comparative Evaluation and Long-Term Monitoring. *Nature Medicine*, *30*(1), 61–75. https://doi.org/10.1038/s41591-023-02732-7

Mashar, M., Chawla, S., Chen, F., Lubwama, B., Patel, K., Kelshiker, M. A., Bachtiger, P., & Peters, N. S. (2023). Artificial Intelligence Algorithms in Health Care: Is the Current Food and Drug Administration Regulation Sufficient? *JMIR AI*, *2*, e42940. https://doi.org/10.2196/42940

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and Regulatory Challenges of AI Technologies in Healthcare: A Narrative Review. *Heliyon*, *10*(4), e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

Mishra, T., Sutanto, E., Rossanti, R., Pant, N., Ashraf, A., Raut, A., Uwabareze, G., Oluwatomiwa, A., & Zeeshan, B. (2024). Use of Large Language Models as Artificial Intelligence Tools in Academic Research and Publishing Among Global Clinical Researchers. *Scientific Reports*, *14*(1), 31672. https://doi.org/10.1038/s41598-024-81370-6

Ng, A., Glocker, B., Oberije, C., Fox, G., Gilbert, F. J., Nash, J., Karpati, E., Kerruish, S., & Kecskemethy, P. (2022). Safe and Effective Integration of AI as Supporting Reader in Double Reading Breast Cancer Screening. *Insights into Imaging*, *14*, 366. https://dx.doi.org/10.1186/s13244-022-01337-x PT - Conference Abstract

Owoyemi, A., Osuchukwu, J., Salwei, M., & Boyd, A. (2024). *Checklist to Support the Development and Implementation of AI in Clinical Settings*. https://doi.org/10.1101/2024.08.08.24311701

Park, S. H., Han, K., Jang, H. Y., Park, J. E., Lee, J.-G., Kim, D. W., & Choi, J. (2022). Methods for Clinical Evaluation of Artificial Intelligence Algorithms for Medical Diagnosis. *Radiology*, 220182. https://doi.org/10.1148/radiol.220182

Pashkov, V. M., Harkusha, A. O., & Harkusha, Y. O. (2020). Artificial Intelligence in Medical Practice: Regulative Issues and Perspectives. *Wiadomosci Lekarskie (Warsaw, Poland: 1960)*, *73*(12 cz 2), 2722–2727.

Petersson, L., Larsson, I., Nygren, J. M., Nilsen, P., Neher, M., Reed, J. E., Tyskbo, D., & Svedberg, P. (2022). Challenges to Implementing Artificial Intelligence in Healthcare: A Qualitative Interview Study with Healthcare Leaders in Sweden.

BMC Health Services Research, *22*(1), 850. https://doi.org/10.1186/s12913-022-08215-8

Pierce, R. L., Van Biesen, W., Van Cauwenberge, D., Decruyenaere, J., & Sterckx, S. (2022). Explainability in Medicine in an Era of AI-Based Clinical Decision Support Systems. *Frontiers in Genetics*, *13*, 903600. https://doi.org/10.3389/fgene.2022.903600

Salwei, M. E., & Carayon, P. (2022). A Sociotechnical Systems Framework for the Application of Artificial Intelligence in Health Care Delivery. *Journal of Cognitive Engineering and Decision Making*, *16*(4), 194–206. https://doi.org/10.1177/15553434221097357

Salwei, M. E., Carayon, P., Hoonakker, P. L. T., Hundt, A. S., Wiegmann, D., Pulia, M., & Patterson, B. W. (2021). Workflow Integration Analysis of a Human Factors-Based Clinical Decision Support in the Emergency Department. *Applied ergonomics*, *97*, 103498. https://doi.org/10.1016/j.apergo.2021.103498

Tricco, A. C., Hezam, A., Parker, A., Nincic, V., Harris, C., Fennelly, O., Thomas, S. M., Ghassemi, M., McGowan, J., Paprica, P. A., & Straus, S. E. (2023). Implemented Machine Learning Tools to Inform Decision-Making for Patient Care in Hospital Settings: A Scoping Review. *BMJ Open*, *13*(2), e065845. https://doi.org/10.1136/bmjopen-2022-065845

Wenderott, K., Krups, J., Luetkens, J. A., Gambashidze, N., & Weigl, M. (2024). Prospective Effects of an Artificial Intelligence-Based Computer-Aided Detection System for Prostate Imaging on Routine Workflow and Radiologists' Outcomes. *European Journal of Radiology*, *170*, 111252. https://doi.org/10.1016/j.ejrad.2023.111252

Wenderott, K., Krups, J., Luetkens, J. A., & Weigl, M. (2024). Radiologists' Perspectives on the Workflow Integration of an Artificial Intelligence-Based Computer-Aided Detection System: A Qualitative Study. *Applied Ergonomics*, *117*, 104243. https://doi.org/10.1016/j.apergo.2024.104243

Wenderott, K., Krups, J., Weigl, M., & Wooldridge, A. R. (2025). Facilitators and Barriers to Implementing AI in Routine Medical Imaging: Systematic Review and Qualitative Analysis. *J Med Internet Res*, *27*, e63649. https://doi.org/10.2196/63649

Wenderott, K., Krups, J., Zaruchas, F., Krawitz, P. M., Weigl, M., & Lesmann, H. (2025). Workflow Analysis and Evaluation of a Next-Generation Phenotyping Tool: A Qualitative Study of Face2Gene. *Eur J Hum Genet*. https://doi.org/10.1038/s41431-025-01875-0

Wenderott, K., Krups, J., Zaruchas, F., & Weigl, M. (2024). Effects of Artificial Intelligence Implementation on Efficiency in Medical Imaging—A Systematic Literature Review and Meta-Analysis. *Npj Digital Medicine*, *7*(1), 265. https://doi.org/10.1038/s41746-024-01248-9

Williams, V., Boylan, A.-M., & Nunan, D. (2020). Critical Appraisal of Qualitative Research: Necessity, Partialities and the Issue of Bias. *BMJ Evidence-Based Medicine*, *25*(1), 9–11. https://doi.org/10.1136/bmjebm-2018-111132

Wooldridge, A. R., Carayon, P., Hoonakker, P., Hose, B.-Z., Eithun, B., Brazelton, T., Ross, J., Kohler, J. E., Kelly, M. M., Dean, S. M., Rusy, D., & Gurses, A. P. (2020). Work System Barriers and Facilitators in Inpatient Care Transitions of Pediatric Trauma Patients. *Applied Ergonomics*, *85*, 103059. https://doi.org/10.1016/j.apergo.2020.103059

Yin, J., Ngiam, K. Y., & Teo, H. H. (2021). Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. *Journal of Medical Internet Research*, *23*(4), e25759. https://doi.org/10.2196/25759

5. Acknowledgements

First of all, I would like to thank my supervisor, Prof. Dr. Matthias Weigl. Without your constant support and encouragement, this PhD would not have been possible! I want to thank you for all the opportunities for learning and development, the constructive feedback, and for sharing your projects and professional network. I am very grateful for my wonderful four years at the Institute for Patient Safety and will always look back fondly on this time.

I would like to extend special thanks to Jim Krups, my colleague and student assistant during my PhD project. Thank you for all the hours you shared with me collecting and analyzing data.

Further, I want to express my gratitude to Prof. Dr. Abigail R. Wooldridge for hosting me at the Human Factors in Sociotechnical Systems (HFSS) Laboratory at the University of Illinois at Urbana-Champaign, IL, USA. Thank you for trusting me to join your lab, for all your support, the opportunities you provided, and the professional advice you shared. Thank you for being an inspiring female role model in the scientific community!

My sincere gratitude also goes to my committee members: Prof. Dr. Dr. Jens Kleesiek, for providing feedback and support throughout the four years of my PhD, and for reading and reviewing my thesis as my second supervisor. Prof. Dr. Ulrike Attenberger, for being part of my dissertation committee and enabling my collaboration with the Department of Radiology. Prof. Dr. Nicole Ernstmann, for creating a great support structure for PhD students at the IfPS and facilitating collaboration with her former work group, the Health Services and Health Communication Research unit at the University Hospital Bonn.

I would also like to thank our collaboration partners at the University Hospital Bonn: Dr. Julian Luetkens from the Department of Radiology, Prof. Dr. Peter Krawitz from the Institute for Genomic Statistics and Bioinformatics, and Hellen Lesmann from the Institute for Human Genetics. I am also deeply grateful to the clinicians who participated in my studies and shared their insights with me.

The support provided by the team at the IfPS was greatly appreciated, especially by my two colleagues, Judith Hammerschmidt and Kathrin Adamietz, who shared an office with me.

Lastly, my sincere gratitude goes to my friends and family for supporting me throughout my PhD.

6. Statement

When creating the content for this doctoral thesis, I used ChatGPT (versions GPT-3.5 and 4.0, OpenAI) to optimize the readability and wording of the dissertation. After using this tool, I reviewed and edited the content as required and take full responsibility for the content. For the creation of figures in this dissertation and the included publications Canva (Canva Pty Ltd) was used. Additionally, I utilized the professional editing services of “Mentorium” for proofreading and plagiarism check.