RESEARCH-ARTICLE

Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support

HUBERT DARIUSZ ZAJAÇ, University of Copenhagen, Copenhagen, Hovedstaden, Denmark

TARIQ OSMAN ANDERSEN, University of Copenhagen, Copenhagen, Hovedstaden, Denmark

ELIJAH KWASA

RUTH WANJOHI, Nairobi Hospital, Nairobi, Nairobi, Kenya

MARY K ONYINKWA, Nairobi Hospital, Nairobi, Nairobi, Kenya

EDWARD K MWANIKI

View all

Open Access Support provided by:

The Aga Khan University Kenya

Nairobi Hospital

University of Copenhagen

Rigshospitalet

University of California, Irvine

Zealand University Hospital

Published: 21 April 2025
Online AM: 05 February 2025
Accepted: 28 December 2024
Revised: 01 August 2024
Received: 29 February 2024

Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support

HUBERT D. ZAJĄC and TARIQ O. ANDERSEN, University of Copenhagen, Copenhagen, Denmark
ELIJAH KWASA, Stratus Medical Imaging Solutions, Nairobi, Kenya
RUTH WANJOHI and MARY K. ONYINKWA, The Nairobi Hospital, Nairobi, Kenya
EDWARD K. MWANIKI, Nairobi Adventist Hospital, Nairobi, Kenya
SAMUEL N. GITAU, The Aga Khan University Hospital, Nairobi, Kenya
SHAWNIM S. YASEEN, Zealand University Hospital, Roskilde, Denmark
JONATHAN F. CARLSEN, Rigshospitalet Copenhagen University Hospital, Copenhagen, Denmark
MARCO FRACCARO, Unumed, Copenhagen, Denmark
MICHAEL B. NIELSEN, Rigshospitalet Copenhagen University Hospital, Copenhagen, Denmark
YUNAN CHEN, University of California, Irvine, California, USA

Despite recent advancements, real-world use of Artificial Intelligence (AI) in radiology remains low, often due to the mismatch between AI offerings and the situated challenges faced by healthcare professionals. To bridge this gap, we conducted a field study at nine medical sites in Denmark and Kenya with two goals: (1) to understand the challenges faced by radiologists during chest X-ray practice and (2) to envision alternative AI futures that align with collaborative clinical work. This study uniquely grounds the AI design insights in the comprehensive characterisation of diagnostic work across multiple geographical and institutional contexts. Building on ideas articulated by interviewed radiologists (N = 18), we conceptualised five visions that transcend the traditional notions of AI support. These visions emphasise that the clinical usefulness of AI-based systems depends on their configurability and flexibility across three dimensions: type of clinical site, expertise of medical professionals, and situational and patient contexts. Addressing these dependencies requires expanding the clinical AI design space by envisioning futures rooted in the realities of practice rather than solely following the trajectory of AI development.

This work was supported by Innovationsfonden (Grant No. 0176-00013B).
Authors' Contact Information: Hubert D. Zając (corresponding author), University of Copenhagen, Copenhagen, Denmark; e-mail: hdz@di.ku.dk; Tariq O. Andersen, Department of Computer Science, University of Copenhagen, Copenhagen, Denmark; e-mail: tariq@di.ku.dk; Elijah Kwasa, Stratus Medical Imaging Solutions, Nairobi, Kenya; e-mail: ekwasa@outlook.com; Ruth Wanjohi, Department of Radiology, The Nairobi Hospital, Nairobi, Kenya; e-mail: ruthwanjohi@nbihosp.org; Mary K. Onyinkwa, Department of Radiology, The Nairobi Hospital, Nairobi, Kenya; e-mail: onyxmary@gmail.com; Edward K. Mwaniki, Department of Radiology, Nairobi Adventist Hospital, Nairobi, Kenya; e-mail: edwardmwaniki7@yahoo.com; Samuel N. Gitau, Department of Radiology, The Aga Khan University Hospital, Nairobi, Kenya; e-mail: samuel.gitau@aku.edu; Shawnim S. Yaseen, Department of Radiology, Zealand University Hospital, Roskilde, Denmark; e-mail: shawnim1@gmail.com; Jonathan F. Carlsen, Department of Radiology, Rigshospitalet Copenhagen University Hospital, Copenhagen, Denmark; e-mail: jonathan.frederik.carlsen@regionh.dk; Marco Fraccaro, Unumed, Copenhagen, Denmark; e-mail: mf@unumed.com; Michael B. Nielsen, Department of Radiology, Rigshospitalet Copenhagen University Hospital, Copenhagen, Denmark; e-mail: michael.bachmann.nielsen@regionh.dk; Yunan Chen, Donald Bren School of Information and Computer Sciences, University of California, Irvine, California, USA; e-mail: yunanc@uci.edu.

ACM 1557-7325/2025/4-ART20
https://doi.org/10.1145/3715115

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; **Human computer interaction (HCI)**; • **Computing methodologies** → **Artificial intelligence**;

Additional Key Words and Phrases: human-centred AI, healthcare, radiology, medical imaging, responsible AI, configurable AI, design

ACM Reference format:
Hubert D. Zając, Tariq O. Andersen, Elijah Kwasa, Ruth Wanjohi, Mary K. Onyinkwa, Edward K. Mwaniki, Samuel N. Gitau, Shawnim S. Yaseen, Jonathan F. Carlsen, Marco Fraccaro, Michael B. Nielsen, and Yunan Chen. 2025. Towards Clinically Useful AI: From Radiology Practices in Global South and North to Visions of AI Support. *ACM Trans. Comput.-Hum. Interact.* 32, 2, Article 20 (April 2025), 38 pages.
https://doi.org/10.1145/3715115

1 Introduction

Artificial Intelligence (AI), defined as a technology using any modern data-driven machine learning technique, e.g., deep learning, has been hoped to significantly transform healthcare by improving patient care, reducing cost and improving the work life of healthcare providers [21]. A plethora of studies have substantiated these hopes by demonstrating increased detection rates of pathologies in medical images in controlled environments [77], supporting cancer detection on mammograms [42], identifying brain tumours on magnetic resonance images [59], detecting arrhythmias on electrocardiograms [9] or supporting polyp detection during colonoscopy [136]. However, only a small percentage of these systems make it from the laboratory to the real world [31, 152]. And even if they do, their positive impact on practice and patient outcomes is not guaranteed [15, 75]. Consequently, despite the growing popularity and ongoing technological developments, the successful use of AI in clinical practice remains low [8, 31, 46, 64, 102, 112, 137, 152].

The low adoption has been particularly apparent in radiology—one of the first medical fields to confront the potential of increasingly capable AI. Several confounding factors make radiology a promising domain to benefit from AI support: the abundance of digital data, the relative use of medical imaging across specialities and recent progress in vision algorithms performance [121]; and the fact that most countries suffer from severe staff shortages results in less time available per examination, stress and overburdening of healthcare professionals [105]. However, the benefits of using AI in practice have been vague. For example, among 62 models aimed to help with the detection and prognostics of COVID-19 on chest radiographs (X-rays) and thoracic **Computed Tomography (CT)**, not one was identified as clinically useful [106]. Currently, only a handful of systems targeting chest radiography have been approved by authorities in the United States and European Union [3], and their clinical utility remains mostly unclear [132].

Research suggests that the uncertainty of added value may be related to the misplaced intended use of AI-based systems [122] and the dominant focus on detection performance in environments isolated from clinical practice [138]. Case in point, breast radiography (mammography) has seen a significant uptake in using **Computer-Aided Detection (CAD)** systems. Yet, a successful implementation into clinical practice is not synonymous with improved patient outcomes [75]. When reviewing AI's ability to improve efficiency and health outcomes, van Leeuwen et al. reported that the clinical value of such systems is still unknown [132]. Similarly, Recht et al. argued for more clinical evaluations in radiology to understand these systems in practice better [101]. Finally, Strohm et al. pointed out that one of the main reasons for hitherto AI failure in radiology is the 'uncertain added value for clinical practice of AI applications' [122].

These issues were recognised by the **Human–Computer Interaction (HCI)** and **Computer Supported Cooperative Work (CSCW)** communities, who turned towards ethnographic research to understand the organisational context and work and care practices within entities implementing new information systems [19, 35, 66, 87, 89, 100, 125]. Engaging in formative research within the healthcare domain [13, 22, 30, 36, 50, 51, 95, 103], e.g., through ethnographic inquiries, workplace studies and user studies, has empowered researchers to design systems addressing the sociotechnical context, human factors and aligned with end-users needs. For example, Chen [30] identified a functionality gap in **Electronic Health Record (EHR)** systems and suggested including support for transitional information, which at the time was crucial to the clinical workflow and unaccounted for by the studied EHR system. Similarly, Zhou et al. [155] discovered that a physician order entry system was not used as intended due to 'a change in physical location, sufficient convenience, visibility of the information, and permanency of information.' They suggested reframing the requirements for such systems to include both formal and informal work practices. These formative studies transformed the early health information systems from a technological novelty to an integral part of clinical practice [38].

In response to the prevailing technology focus in AI development, the HCI community calls for a turn towards **Human-Centred AI (HCAI)**, emphasising the facilitation of social participation among end-users and the creation of AI that supports self-efficacy, responsibility and oversight [28, 118, 119]. With AI facing similar challenges to the early clinical information systems, inspired by the rich history of HCI and CSCW engagements, researchers are advocating for studies investigating the sociotechnical context of end-users work in the wild [31, 56, 93, 107, 130, 152], particularly, when designing with such difficult medium as AI [144]. New studies emerge, investigating AI's place in the broader ecosystem of collaborative medical work practices [56, 110, 116, 145], origin of trust [56, 98], explainability [74], bias [11] or collaboration [17]. However, to develop useful HCAI systems, designers and developers are called to engage in 'more holistic and in-the-wild' methods [8].

These ongoing efforts by the HCI community motivated our work, which was part of a larger research project to innovate, i.e., design, develop and deploy an AI-based chest X-ray support system for radiologists in Denmark and Kenya. The project was defined as a collaboration between the University of Copenhagen, Rigshospitalet Copenhagen University Hospital and Unumed,[1] a business partner delivering hospital management systems tasked with the commercialisation of the future system. This setup determined the choice of chest X-rays as the modality of focus, focus on radiologists as the intended end-users, access to medical data, the technical scope of the project, as well as the choice of the countries focal to this study: Denmark—the primary location of the research and development team and Kenya—a country with the strongest business presence of Unumed.

To contribute to HCI's understanding of the connection between practice and AI opportunities and to support the innovation of a clinically useful system for radiologists in both countries relevant to the project, we posed the following research questions:

RQ1 What are the challenges experienced by radiologists in chest X-ray practice?
RQ2 How can AI support radiologists in chest X-ray practice across clinical settings?

We answered them by conducting a field study at nine medical sites in Denmark and Kenya with 18 radiologists and 4 radiographers. We asked our participants how AI could support them in the context of their practice to explore the connection between the practice of radiologists and opportunities for AI support. By envisioning futures during their work practice, we wanted to ensure that their visions were pragmatic responses to actual challenges rather than hypothetical exercises attempting to retrofit preconceived AI models into their workflows. We drew inspiration

[1] https://unumed.com

from prior future-oriented HCI research and encouraged our participants to be critical towards technology and to envision their ideal future [61, 123, 128].

This study contributes the following to a growing body of HCI and HCAI research that investigates practice to inform design opportunities for clinical AI-based support systems:

(1) Unpacking chest X-ray practice, as part of collaborative diagnostic work, into three stages present across clinical sites in Denmark and Kenya: selecting, interpreting and reporting;
(2) Showing how HCI may expand the technology-shaped design space of clinical AI and bring it closer to the ideals of HCAI by engaging end-users and envisioning AI futures at the point of practice;
(3) Reflecting on how clinical usefulness depends on the configurability and flexibility of AI across three dimensions: type of clinical site, expertise of healthcare professionals and situational and patient context.

2 Related Work

2.1 Shortcomings of Technology-Centred AI Development

To some extent, the current state of AI in healthcare resembles the reality of computer science on the eve of the formation of HCI as a research domain in the 1980s. The development of computer systems, in this case, AI-based systems, is claimed, defined and envisioned through predominantly technology-centric approaches [118, 119, 142, 154]. This state of affairs is partially understandable, as technological progress enabled this new type of system. However, the main drawback of technology-centred AI development is the foundational character of the early design decision [5]. Research has shown how the very early decisions taken during data work affect the capabilities of AI models [88, 109, 151]. Going further, the capabilities of AI models determine the design space of potential solutions [85]. Any misunderstandings and misconceptions before and during AI model development may propagate to the final system, afflicting its real-world utility. Unfortunately, it is often difficult to mitigate them later, as the early stages of AI innovation require the most effort and resources to conduct [5, 88, 144]. Consequently, the capabilities of an AI model often determine the final system's capabilities. This remains true even for **Large Language Models (LLMs)** applied to expert tasks, which healthcare arguably falls under [4].

As a result, AI-based systems perform inconsistently when deployed in clinical settings across the globe, and their added clinical value is often dubious [71, 82, 122, 127, 133]. For example, in the Global North, Petitgand et al. [97] investigated the implementation of an AI-based **Decision Support System (DSS)** in an emergency department in a large hospital in Canada. They discovered that the system did not integrate with other IT systems and poorly modelled clinical practice. This resulted in support that was technically correct but not useful in practice. Another system evaluated by Hollander et al. [49] intended to support decisions on admission to the emergency department due to heart problems in a US university hospital. The system's output relied on specific cardiac markers available in the hospital's system after a clinician had already decided the admission, which rendered the support irrelevant. When looking into AI implementations in the Global South, Wang et al. [135] investigated the use of an AI-based clinical DSS in rural clinics in China. They reported on the misalignment of the AI-designed workflow and the reality of local practice, which was further exacerbated by 'technical limitations and usability barriers'. Similarly, Beede et al. [15] evaluated in 11 clinics in Thailand an AI-based system for diabetic retinopathy detection developed in the USA. This study captured the difficulty of translating AI-based systems across borders. The authors discovered that the US-trained model did not perform well on local data due to socio-environmental factors and that the envisioned AI workflow did not match the reality of local providers and their patients.

This is an opportunity for HCI researchers and practitioners to engage in formative work with medical practitioners to explore alternatives to the technology-shaped design space of AI-based support systems and to inform their design and development focusing on clinical value.

2.2 HCI Push for HCAI

HCI researchers noticed the challenges of developing useful AI-based systems in healthcare and argued for a shift in AI development to achieve HCAI by focusing on human values, responsibility, participation and oversight [118, 119]. So far, there has been no definitive approach to developing HCAI [28]. Instead, researchers explored different avenues. For example, Yildirim et al. [146] developed design resources to support meaningful ideation of AI futures. In a related study, Yildirim et al. [148] demonstrated how multidisciplinary teams may use those resources to explore clinical AI's problem-solution space. Another strand of HCI work recognised trust as a necessary component of the successful adoption of AI. Among others, the HCI community informed its clinical and practical origin [11, 25], relationship with accepting and acting on AI output [120], and dependency on organisational accountability [98]. Linked to trust is the 'black box' problem—a problem particular to AI, where users cannot inspect and understand the inner workings of a system [29]. Explainable AI has been the most promising answer to 'opening that box', enhancing trust, supporting oversight, but also increasing the perceived usefulness of medical AI [26, 43, 74, 141]. These new ways of reasoning by and with AI models prompted research into modes and premises of human-AI collaboration [17, 23, 27, 34, 84], investigating reliance, bias and its potential mitigation techniques [11], problematising AI's authority in such arrangement [60] and its place in the broader ecosystem of collaborative medical work practices [56, 84, 110, 116, 145]. Particularly, the alignment with work practices and sociotechnical context and responding to the actual needs of clinical end-users has been considered crucial when transitioning from technology-centred to HCAI [56, 107, 130, 134, 152].

The importance of understanding the clinical context, work practices and end-users should come as no surprise. HCI researchers showed time and time again that insufficient and delayed attention to the social considerations of work leads to inadequacy of developed systems and inevitable failure in practice [52, 53, 126]. The response to these challenges is one of the most foundational contributions of the HCI and CSCW communities—the push for the ethnography-based user and workplace studies to inform system design [7, 19, 22, 51, 66, 83, 87, 94, 113, 125, 156].

The need to understand end-users' clinical context and the sociality of work is just as important now as it was at the beginning of personal computing. HCI researchers describe how formative user studies may serve as an essential foundation for designing and developing clinically useful AI-based systems for pathology detection [44], chronic conditions [123], mental health [131], and collaborative environments of intensive care units [58]. However, conducting meaningful formative studies for AI in a clinical context is afflicted by difficulty in obtaining access, engaging healthcare professionals, and using AI as a design medium [32, 69, 115, 144]. In the context of still technology-driven AI development, HCI is in need of investigating new ways of involving users in foundational work to shape future systems and make them useful to healthcare professionals in clinical practice.

2.3 Performance Does Not Equal Clinical Usefulness

The natural consequence of centring AI development around algorithms, models and data is evaluation based on technical metrics, such as an **Area Under a Receiver Operating Characteristic Curve (AUC)**, accuracy, sensitivity and specificity. The performance-first evaluation approach permeates the general research of clinical AI. Li et al. [77] reviewed the added value of AI when diagnosing thoracic pathologies. The primary metrics reported were sensitivity, specificity, accuracy, AUC and time spent diagnosing. According to Wang et al. [137], such retrospective cohort studies constituted 98% of all the studies between 2015 and 2019 on AI in radiology.

Excelling at narrowly defined technical metrics may not be enough to bring meaningful change to medical practice [18, 62]. Keane and Topol [62] argued that none of these metrics ultimately relate to a change in patient outcomes. For example, breast radiography (mammography) has seen a significant uptake in the use of CAD systems. However, successful use in clinical practice did not improve patient outcomes [75]. This conundrum exemplifies that performance metrics are ultimately relevant during development but do not reflect the benefits of using an AI-based system in clinical practice [117].

This dichotomy has also been noticed by HCI researchers, who have been focusing on human-centred qualities of AI-based systems like trust [11, 25, 56] and usability [26, 68, 85], which support the actual use of a system in practice. However, similar to the technical metrics, on their own, they do not guarantee a positive impact on clinical practice and patient outcomes. Researchers investigated what other qualities make an AI-based system clinically useful, which was used as a general term describing positive contribution to clinical practice and patient outcomes [16, 39, 135]. Bossen and Pine found that flexible integration into the clinical workflow, support for sensemaking, awareness of unreliability, practitioners remaining in control and ability to experiment contributed to healthcare professionals considering the AI-based tool useful [23]. Awareness of those caveats of implemented AI-based systems may be raised through onboarding practices. Cai et al. [27] proposed calibration of AI through interaction with examples as a way to empower users and help them engage with new AI-based systems in a meaningful manner. When we look into the in-use perceptions, Wang et al. found that the perceived usefulness stemmed from the tool supporting the clinical diagnostic process, facilitating information search, offering training opportunities, and preventing adverse events [135]. These findings are crucial to our understanding of factors influencing the clinical usefulness of AI-based systems. However, to design for clinical usefulness, we need a better understanding of end-users' needs and expectations [131].

2.4 State of the Art in Radiology AI

For decades, radiology has been a space of innovation in medicine. Already in 1963, Lodwick et al. [80] presented the first DSS for chest radiographs that provided lung cancer prognosis. Since then, the broadly understood AI has progressed significantly to the point that prominent voices in the AI community heralded the imminent obsolescence of radiologists [86]. Despite the passing years, this vision has never materialised. While the number of FDA- or CE-approved AI-based systems in radiology (in the United States and European Union, respectively) has been steadily increasing, their clinical value is often uncertain [122, 132], and subsequently, their clinical use in radiology is not much more significant than in the times of Lodwick et al. [46, 112, 137].

Recent research from the frontier between health and HCI aims to address this lack of clinical value. For example, Leibig et al. investigated the effect of integration points on the final usefulness of AI [76]. They compared the same AI-based model trained on the same dataset in two different roles within a mammography cancer screening practice. In the first role, AI was used to substitute a radiologist and its performance was significantly worse than that of a trained healthcare professional. However, when AI was weaved into radiological practice—to pre-screen radiographs before being seen by a radiologist and post-screen to make sure no cases were missed—their joint performance significantly improved patient outcomes. In another study, AI was used to prioritise radiologists' work lists by screening for critical findings [12]. Baltruschat et al. showed that in a simulated environment, AI was able to reduce the average turnaround time for critical findings, indicating potential for improving patient safety. When looking at other points of support in radiology work, Xie et al. conducted a design study with ordering clinicians and radiologists. They explored how to support ordering clinicians at their point of practice, primarily focusing on useful explanations of AI prediction for chest X-rays [141]. These studies showcase that popular detection-focused AI-based systems have the potential to support radiology practice in other ways than decision support.

Similar grounded work was sparked by the advent of generative AI, e.g., LLMs, **Large Multimodal Models (LMMs)** and **Large Vision Language Models (LVLMs)**, which opened an entirely new design space linked to their novel capabilities. To support meaningful development and clinical usefulness of systems using LVLMs, Yildirim et al. [147] conducted an iterative design process with a strong foundational understanding of current work practices. They proposed new avenues of useful AI support that centre around ‘draft report generation, augmented report review, visual search and querying and patient imaging history highlights’. Importantly, the study by Yildirim et al. considered the radiology practice as a coherent part of diagnostic processes and involved clinicians using medical imaging, ensuring that their developed visions are relevant outside of the isolated laboratory contexts. Other researchers also investigated the opportunities for LLMs to support radiologists, in particular in drafting reports. For example, Yu et al. [149] proposed two new metrics (RadGraph F1 and RadCliQ) supporting automated evaluation of alignment between AI-generated radiological reports and radiologists. The latter was used by Hyland et al. [54] in the evaluation of their proof-of-concept LMM, which in the controlled laboratory environment showed promising results. These studies show how alternative approaches to thinking about AI in radiological practice may support its appropriation.

Radiology researchers second these findings and argue that the development of AI-based systems for radiology should search for new ways of supporting clinical practice rather than substitution [72, 129, 143]. This call is motivated not only by the sub-par performance of AI-based models in the wild but also by the complexity of radiologists’ practice that goes beyond the detection of pathologies on medical images [96]. For example, Saßmannshausen et al. [111] and Ontika et al. [91] conducted a practice-centred design study to support radiologists diagnosing prostate cancer. Their work adhered to the HCAI principles and was deeply rooted in understanding local work practices. As a result, they presented a co-designed vision for an AI-based system that embodied the hybrid intelligence approach, i.e., aims to augment users’ capabilities with AI while preserving their agency [57]. In the same tone, researchers argued for expanding the focus to ‘include quality of care and patient outcomes’ [64]; ensure transparency, trust [27] and understandability [91, 141]; and work towards the AI-based system integration within the clinical workflows [15, 108] and its appropriateness in local contexts [153].

In response to these challenges and to further the work conducted by the HCI community, we engage chest X-ray practice not only as an independent phenomenon detached from the clinical context but also as an integral part of diagnostic and treatment processes that involve a range of healthcare professionals. We use this holistic understanding to locate visions for AI support introduced by radiologists to support the development of clinically useful systems.

3 Methodology

The goal of this study is twofold: (1) to understand the challenges experienced by radiologists in chest X-ray practice and (2) to support the innovation of radiological AI by mapping out opportunities for AI support in chest X-ray practice across countries like Denmark and Kenya.

While the study is primarily focused on chest X-ray practice and supporting the innovation of systems that can be translated across borders, it is important to be aware of the socio-economic differences between the two countries. Denmark is characterised by its well-established welfare state and high-income economy. The Danish state boasts a vast, digitalised public healthcare system,

Table 1. Summary of Visited Medical Sites, Detailing the Types of Clinics, Radiologists Employed and the Country of Operation

Site	Type	Work lists based on	Radiologists	Country
D1	Specialised hospital	Speciality	100+	Denmark
D2	Imaging clinic	Single list	<5	Denmark
D3	General hospital	Modality	<20	Denmark
D4	General hospital	Modality	<5	Denmark
K1	Specialised hospital	Speciality	<20	Kenya
K2	Imaging clinic	Single list	1	Kenya
K3	General hospital	Single list	1	Kenya
K4	General hospital	Single list	<5	Kenya
K5	General hospital	Modality	10	Kenya

The table also outlines the organisation of radiologists' work lists within Picture Archiving and Communication System (Section 4.1.1), categorised as follows: (1) Speciality: work lists contain medical images from various modalities but focus on a single clinical domain, e.g., infectious diseases. (2) Modality: work lists include medical images exclusively from a single modality, e.g., X-rays. (3) Single list: all medical images are consolidated into one work list, regardless of modality or speciality.

where the number of people per radiologist totals 12,000 and up to 650,000 chest radiographs are captured yearly [48]. Kenya is an emerging economy with diverse natural resources, a growing population and a colonial history. Kenyan healthcare relies heavily on private medical centres to cater to the healthcare needs of its populace. There are approximately 265,000 people per radiologist in Kenya, and, to our best knowledge, there are no statistics about the annual number of captured images [1].

3.1 Study Design

This field study comprised in-situ participatory observations with radiologists and radiographers in nine medical sites in Denmark (4) and Kenya (5) (Table 1). To capture the complexity of work performed by radiologists and identify commonalities, we encompassed various medical sites providing medical imaging services. We recruited participants through email and professional contacts of our collaborators trying to reach a diverse group of professionals working at different medical sites. We grouped the visited sites by the catered population, specialisation level of medical staff, available resources and size. Two specialised hospitals (D1, K1) provided tertiary[2] and quaternary[3] care, handling the most complex medical procedures in their respective countries. Five general hospitals (D3, D4, K3, K4, K5) offered primary and secondary care and referred patients requiring more specialised care. Lastly, two imaging clinics (D2, K2) provided medical imaging services to patients referred by external physicians for chest radiographs, ultrasounds or CT scans (only K2). K2 also provided teleradiology services to clinics and hospitals in Kenya.

3.2 Participants

In total, 18 radiologists and 4 radiographers participated in this study (Table 2). Fifteen radiologists held senior (consultant) positions, while the remaining 3 were junior radiologists. In the context

[2]Tertiary care is highly specialised medical care usually over an extended period that involves advanced and complex procedures and treatments performed by medical specialists in state-of-the-art facilities.
[3]Quaternary care is an extension of tertiary care in reference to advanced levels of medicine which are highly specialised and not widely accessed, and usually only offered in a very limited number of national or international centres.

Table 2. Participants, Ordered by Sites

Participant	Position	Seniority	Works with Chest X-rays	Experience with AI in radiology	Site
P1	Radiographer	Senior	Daily	Yes	D1
P2	Resident Radiologist	Junior	Training dependent	Yes	D1
P3	Resident Radiologist	Junior	Training dependent	Yes	D1
P4	Radiologist	Senior	Daily	Yes	D1
P5	Radiologist	Senior	Daily	No	D1
P6	Resident Radiologist	Junior	Training dependent	Yes	D1
P7	Radiologist	Senior	Daily	Yes	D1
P8	Radiologist	Senior	Daily	Yes	D1
P9	Radiologist	Senior	Few days a week	Yes	D3
P10	Radiologist	Senior	Few days a week	No	D4
P11	Resident Radiologist	Junior	Training dependent	No	D4
P12	Radiologist	Senior	Daily	No	D2
P13	Radiographers	Senior	Daily	Yes	K2
P14	Radiologist	Senior	Daily	Yes	K2
P15	Radiographer	Junior	Training dependent	No	K3
P16	Radiologist	Senior	Daily	Yes	K3
P17	Radiologist	Senior	Daily	No	K4
P18	Radiographer	Senior	Daily	No	K4
P19	Radiologist	Senior	Few days a week	Yes	K5
P20	Radiologist	Senior	Few days a week	Yes	K5
P21	Radiologist	Senior	Few days a week	Yes	K5
P22	Radiologist	Senior	Daily	Yes	K1

of handling chest X-rays, senior radiologists had more experience, and their reports did not have to be approved by another radiologist. Conversely, junior radiologists focused on learning, and their reports had to be approved by a senior colleague before they could be shared with ordering clinicians. Radiographers, or radiologic technicians, were trained healthcare professionals who captured medical images, including chest radiographs but did not officially interpret them.

Thirteen radiologists had previous experiences with various AI-based systems, which covered a range of modalities. In particular, radiologists from D1 and D4 had a few DSSs for CT thorax at their disposal. These systems offered functionalities like generating a 3D model, correcting motion, segmenting lungs, and detecting findings. Participants working in K5 piloted a DSS detecting selected findings on chest X-rays. However, doctors from K1 and K2 had past experiences with systems detecting tumours in breast mammography, and the single doctor from K3 had past research experience with AI for chest X-rays. Despite the past exposure and the availability of AI in some clinics, only one of the participants used AI in their daily practice—a senior radiologist from D1 who self-described as a technology enthusiast.

3.3 Data Collection

To gain a broad understanding of how the work is truly conducted and subsequently use this knowledge to develop grounded-in-practice visions for future AI use, we conducted in-situ observations [40, 99]. This approach contrasts with learning from written or oral reports, which often distort the reality of the work, whether intentionally or not [65, 124]. As part of the observations, we conducted semi-structured interviews pertaining to informing the observation goals that were not observed or required clarifications [70].

Using these two methods, we wanted to learn about six knowledge areas: the clinic's profile, organisation and division of work at the clinic, patient characteristics, daily routines, work involved in handling chest radiographs and past experiences with AI-based systems. We aimed for this knowledge to encompass more than the work of interpreting chest X-rays. We wanted to be able to place specific tasks of radiologists within the broader and complex realities of healthcare delivery, which include variations in patient populations, clinical practices and technological infrastructures. Additionally, we gained insight into the overall context of clinical work, the collaboration among healthcare professionals and the rhythm of the clinics.

This knowledge was essential to meaningfully develop visions for future AI use in chest X-ray practice. Our approach was inspired by speculative design [10] and other non-formal methods of eliciting possible futures used in health and HCI studies focused on AI [61, 123, 128]. During the observations, we encouraged radiologists to envision AI support in relation to the work conducted by asking questions like, 'Can you imagine how AI could help you right now?' By doing this, we ensured that ideas generated during the observations were related to practice and practical challenges and were sensible to the current state of the art in AI but not limited by it. Moreover, we encouraged our participants to be critical in their visions and not limit themselves to improvements on already available systems by highlighting that we were at the beginning of the innovation process, not limited by preconceived ideas about support and not affiliated with any leading health-tech enterprise. Additionally, as researchers engaging in formative research, we positioned ourselves on the opposite side of the technology-first approach to innovating AI-based systems, which our participants reacted positively to by acknowledging that they were rarely asked for qualitative opinions on the evaluated and piloted systems.

3.3.1 Practical Execution. Each observation started with informing about the study and observation goals and, if any gaps in knowledge remained, ended by asking clarifying questions. During the observations, we took place behind the participants' side to observe and ask questions about their work. We halted observations and waited in common areas during meetings or other activities involving patients or unrelated to the focus of the observations. Since we aimed to introduce as little disruption as possible, we engaged in discussion only when our participants indicated readiness by opening up for conversation. Usually, these exchanges occurred during their breaks between examinations or when they encountered something they considered worth sharing, e.g., describing their routine or an interesting case. Often, participants themselves initiated this conversation by saying, e.g., 'I think artificial intelligence could be very helpful here to ...' [P9], which we followed up on with questions to better understand their visions.

Data collection took place from April 2021 to February 2023. We visited D1 in April 2021 (HDZ), D3 and D4 in February and March 2022 (HDZ), K1–K5 in January and February 2023 (HDZ, TOA) and D2 in February 2023 (HDZ, TOA). No personal information was recorded at any time. In cases where particular X-rays were discussed, they were pseudonymised, leaving out personal information like name and identification number.

In total, we conducted 67 hours of in-situ observations—35 hours in Denmark and 32 hours in Kenya. The length of observations varied based on the subject and situation in the clinic. This means that our participants decided the times and length of observations. On average, we observed each participant's work for 3 hours and 25 minutes (min. 30 minutes—max. 16 hours). Observations of P1, P9–P22 were audio recorded, transcribed and supplied with handwritten notes; during observations of P2–P8, only handwritten notes were taken.

Participants were not compensated. We collected written consent from all participants. Our study was considered a non-interventional, observational study, thus exempt from a formal ethical review according to the authors' institutions' institutional review boards.

3.4 Data Analysis

Based on the lived experience, information gathered during the interviews and observation notes, HDZ, TOA and YC mapped the practice of handling chest X-rays across the clinical settings. When conducting the observations, we realised the broad similarities between medical settings of the same type. This became apparent when we started to depict the workflows for the different settings and situations based on observation experiences and notes. Iteratively, we clarified the dimensions that determined the similarities, i.e., the catered population, specialisation level of medical staff, available resources and size, and converged on three workflows of chest X-ray practice (Figure 1). Importantly, the radiologist part of those workflows was shared across the settings and encompassed three general steps: selecting, interpreting and reporting. This work was done through April and May 2023 in a digital whiteboard—Miro.[4] The workflow models were validated by domain experts—co-authors of this article (EK, RW, MKO, EKM, SNG, SSY, JFC, MBN).

We used thematic analysis [24] to analyse the data collected during the in-situ participatory observations: recordings and handwritten notes. We conducted this analysis in Dovetail,[5] a web application for qualitative data analysis. Firstly, the first author familiarised themselves with the data by manually correcting machine-transcribed audio recordings from the observations and interviews and transferring handwritten notes to a digital format. Secondly, guided by the gained experience from the empirical work, the first author manually coded any references to work practices, radiographs and AI (both past experiences and envisioned future use) made by our participants. Thirdly, employing Dovetail's digital canvas, three authors (HDZ, TOA and YC) explored the codes for themes, and no AI was used to support this process. Through weekly meetings through May and June 2023, they iterated the breadth and meaning of created themes in the context of observed practice. As the data included a range of visions and challenges, we searched for similarities in the contexts and our participants' intent to bring out the underlying meaning and provide generalisable yet concrete visions for the future. In this step particularly important was the empirical knowledge gathered during the observations that allowed us to better understand and contextualise collected data. As a result, we conceptualised five themes representing challenges encountered in radiologist practice and related visions of AI support.

3.5 Methodological Limitations

Several methodological limitations in our study should be acknowledged. In Figure 1, we mapped out the chest X-ray practice starting from the X-ray order and finishing on the radiological report. While there are many healthcare workers involved in this process, we focused solely on supporting radiologists. The reason behind this choice was connected to the broader project that this study was a part of, and its applied aspect. Due to legal requirements in Denmark, only radiologists, as experts in their field, are allowed to interpret and report on chest X-rays. Thus, a system that supports the chest X-ray practice had to be centred on radiologists as the intended end-users. To this end, we conducted only a limited number of interviews with ordering clinicians in D1 and K1 (not included in this study) and radiographers, which offered contextual insights about the chest X-ray practice. For the same reason, we did not engage administrative workers who may play a role in determining the distribution of work within the departments.

Moreover, we acknowledge that the reporting stage was not explored with as much depth as other stages of the practice. This can be attributed to the fact that the participants' perspective on AI support was largely shaped by their practical experiences and challenges. Next, we acknowledge the recent developments and potential of LLMs. However, as they were not brought up in the

[4] https://miro.com
[5] https://dovetail.com

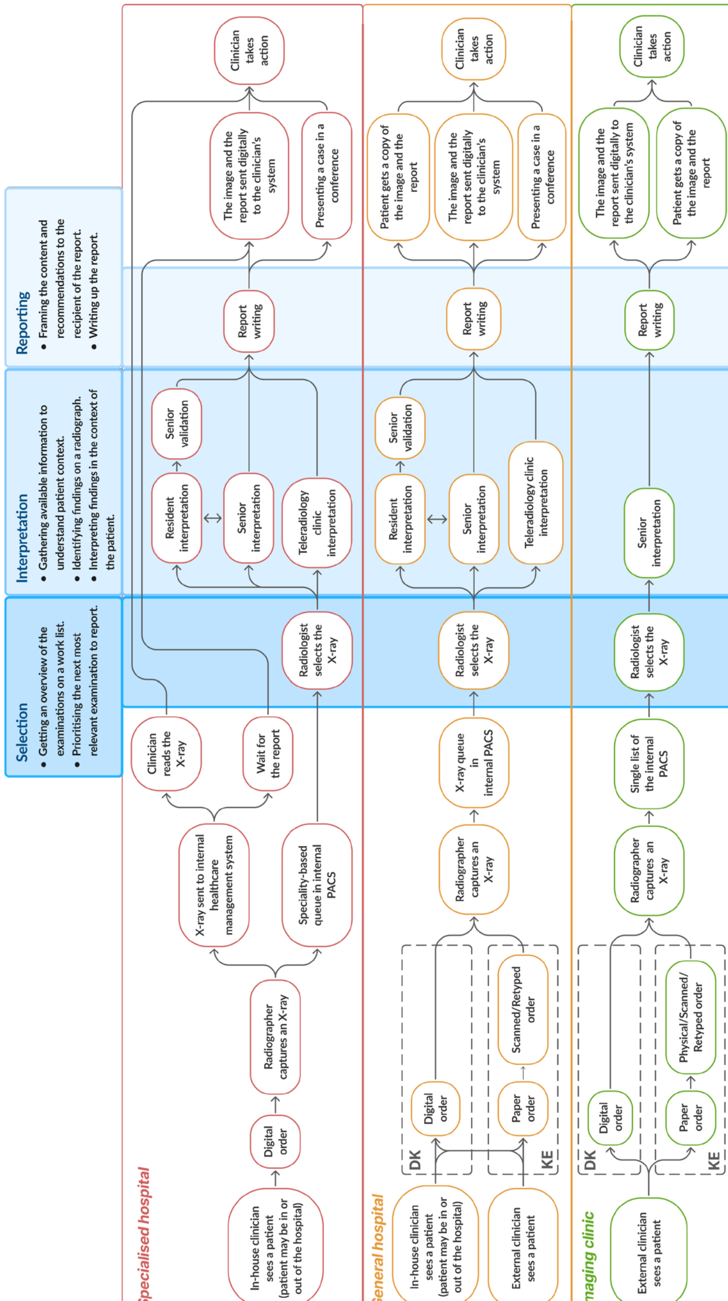

Fig. 1. A workflow of chest X-ray practice observed in the three types of clinical settings in Denmark and Kenya. The radiology part of the workflow is highlighted in blue, and three activities are distinguished. PACS, Picture Archiving and Communication System.

visions for future AI support, we do not have enough basis to address them. Moreover, the broader project plan was created prior to the LLM advancement, and due to that fact, they were not taken into consideration when specifying the scope of the project. Finally, while Denmark and Kenya are two countries from the Global North and the Global South, respectively, this study does not claim to address the radiology practice worldwide. Rather we offer a unique opportunity to see how we can innovate AI that translates across contexts. We caution that other practices may be observed in other countries, and visions for AI support may differ.

4 Vision of Clinically Useful AI Support Rooted in Chest X-Ray Practice

We expected stark differences between the radiologists' work practices from different countries, primed by the differences in Danish and Kenyan healthcare systems. However, the work practices concerning chest X-ray handling were remarkably similar. In fact, work practices in medical sites of similar type across the countries were more similar than practices within the same country but different types of medical sites, where the medical site types were defined as specialised hospitals (D1, K1), the imaging clinics (D2, K2) and the general hospitals (D3, D4, K3, K4, K5). Subtle differences linked to the digitalisation of the processes pertained to the medium of the referrals (digital or paper) and are depicted in Figure 1.

Across the settings, we observed three common activities that radiologists conducted when handling a single chest X-ray—selection, interpretation and reporting, present across all the visited medical sites in Denmark and Kenya (Figure 1). Importantly, we speak only to the similarities in chest X-ray practice and not socio-economic aspects of healthcare systems or other activities in radiologists work days like conferences with clinicians, supervising junior radiologists, preparing protocols for CT scans and performing ultrasounds or CT-guided biopsies.

Notably, the selection, interpretation and reporting were remarkably fast. The simplest of radiographs were said to be assessed within 30 seconds to 5 minutes. More complex ones reportedly took between 5 and 15 minutes. Generally, the more specialised setting (and thus the reader, not including residents), the shorter the assessment time. These estimates include all the work from selecting an examination from a list of radiographs captured in the system to disseminating a report.

The three outlined activities (selection, interpretation and reporting), while performed by radiologists, depended on communication and collaboration with other healthcare workers (ordering clinicians, radiographers and secretaries) involved in the chest X-ray practice (Figure 2). Ordering clinicians were radiologists' primary collaborators. They initiated the X-ray flow and were the final recipients of the reports. Radiographers and secretaries also played an important role in organising radiologist work throughout the day. Secretaries manually filled in the integration gaps between used IT systems and shared information to reduce the inconveniences of long waiting times for patients. However, radiographers captured X-rays. More experienced radiographers, although legally not allowed to report on X-rays, had extensive tacit knowledge gained during work and could often spot critical radiological findings in the X-rays. Additionally, they always personally interacted with patients. Due to these two points, sometimes they alerted radiologists about accidental medical emergencies spotted. Importantly, any of the actors could have initiated such collaboration, depending on their information needs and the patient's health. Moreover, communication often happened using different means and with varying support of technology, e.g., through referrals, phones, **Picture Archiving and Communication System (PACS)** or physical interaction. These findings paint a picture of chest X-ray practice as not solitary and independent of external factors but rather deeply embedded in clinical practice, mediated and collaborative.

To explore the connection between the situated chest X-ray practice and opportunities for clinically useful AI support, we enquired about radiologists' visions of how AI could help them at

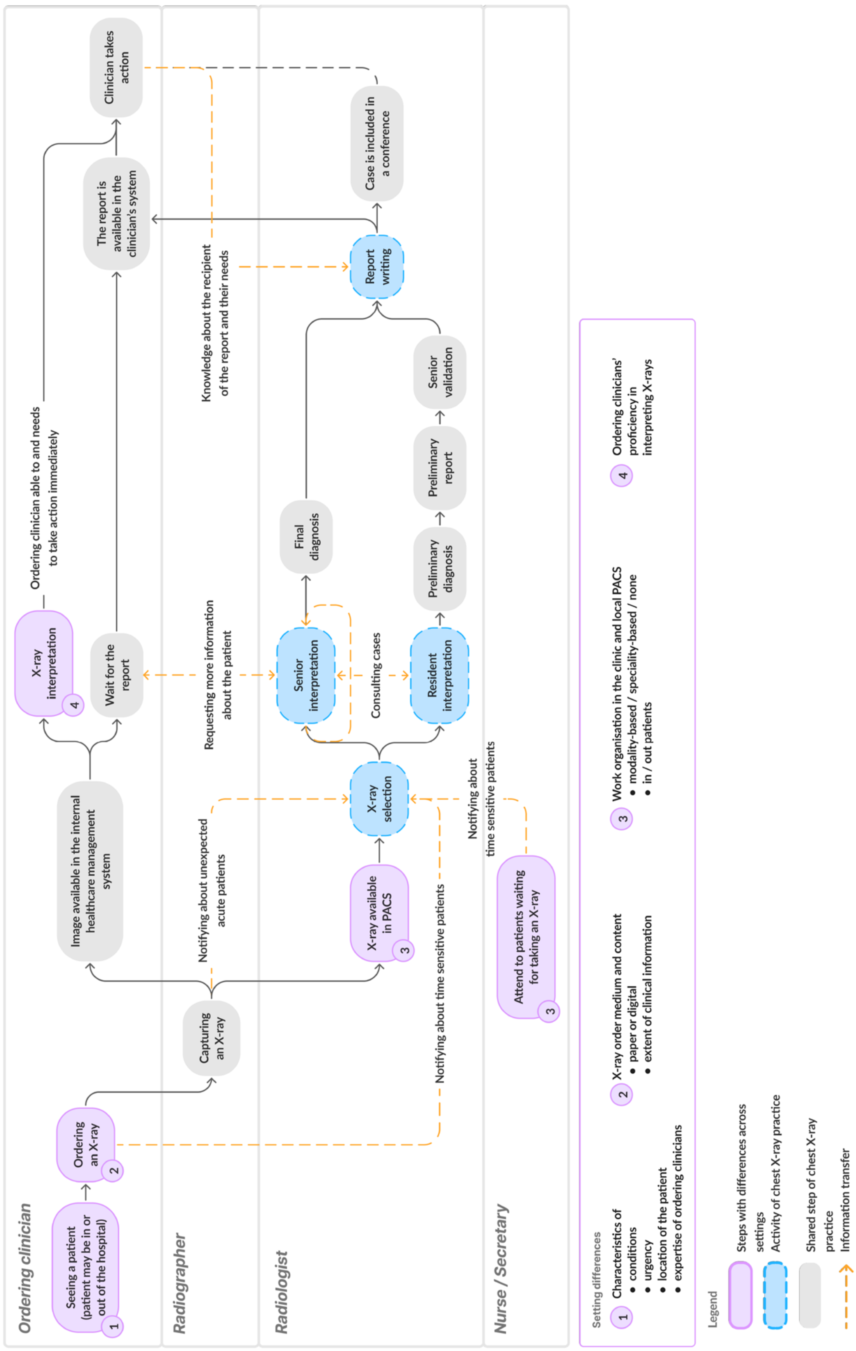

Fig. 2. A depiction of chest X-ray practice illustrating the collaboration and communication between involved healthcare professionals. PACS, Picture Archiving and Communication System.

Table 3. Challenges Encountered in Chest X-ray Practice and Envisioned AI Support

Selection	
	Challenge: Backlog of X-rays to report exceeding daily processing capacity
	Vision: AI distributing examinations by user's expertise
	Challenge: Selecting the next most relevant examination to report without an easy overview
	Vision: AI detecting medical emergencies
Interpretation	
	Challenge: Interpreting visually ambiguous findings
	Vision: AI providing decision support on subtle and difficult cases
	Challenge: Time-consuming process of obtaining additional clinical information
	Vision: AI measuring visual features and comparing changes across historical examinations
Reporting	
	Challenge: Conveying the right information in a report
	Vision: AI double-checking reports against radiographs for missed or misinterpreted findings

work. Their visions stemmed from practical challenges faced during everyday work and focused on providing actionable support rather than automating some of their tasks (Table 3). This was well captured by P4, a senior radiologist from a specialised hospital in Denmark (D1), *Make AI that helps me do my work better or faster*. Increasing efficiency and improving patient health outcomes were the basis for every envisioned functionality.

In this section, we will deconstruct the three activities of chest X-ray practice, highlight their collaborative aspects and relevant dependencies linked to the type of clinical sites. Finally, for each of the activities, we will locate the challenges and describe visions for alleviating them.

4.1 Selection: Practice of Prioritising

Radiologists in both countries started their work by logging in to the PACS, their primary working tool. This is where they selected, viewed, interpreted and reported examinations. The PACS also allowed for accessing historical images and referrals. A referral, also called a clinical history or an indication, typically includes a short description of a patient's current state and a clinical question justifying the X-ray order. In cases where the PACS was not integrated with a system used by ordering clinicians, paper-based referrals were delivered physically to the radiologist (K2), entered to the PACS by radiographers when capturing a radiograph (K3) or scanned and uploaded to the PACS when capturing a radiograph (K4, K5). The PACS was independent of the EHR system in every visited site, and there was little to no integration between them.

4.1.1 Challenge: Backlog of X-rays to Report Exceeding Daily Processing Capacity. In the PACS, radiologists accessed their work lists for a given day and gained an overview of the examinations to be reported. At that time, the radiographs captured in the evening of the previous day or at night were available for reporting. More X-rays were taken during the day. In most of the general hospitals in both countries (D3, D4, K4, K5) and the imaging clinic in Kenya (K2), the number of chest radiographs taken daily often reached or exceeded their reporting capabilities per day, which was voiced by P10, a senior radiologist from a general hospital in Denmark, *I just dream one day. I will open my list and just see ten examinations—always around or more than 50.* Usually, the number of patients fluctuated depending on the day of the week, season or proximity of important dates, e.g., beginning or end of school or holidays.

Typically, with several radiologists in a clinic, one radiologist was responsible for a single work list daily. A work list in the PACS held examinations of the same type that needed to be reported. The type was defined on the hospital or clinic level and could be based on, e.g., modality or the department of the ordering clinician. This means that radiologists were responsible for reporting examinations assigned to their specific list, which were usually similar. Sometimes, a radiologist was responsible for more than a single work list, especially in specialised hospitals and during increased patient visits. How radiologists and examinations were assigned to work lists varied depending on the medical site (Table 1).

— *Specialised Hospitals (D1, K1).* Work lists were created based on medical specialisations and departments in the hospitals, e.g., infectious diseases, oncology or heart medicine, and all types of the captured imaging were included as captures as part of their work, e.g., radiographs, CT or magnetic resonance imaging.
— *Big General Hospitals (D3, D4 and K5).* The work lists were created per imaging type (modality), i.e., all CTs were assigned to one list, and all radiographs were assigned to another, sometimes with few exceptions, e.g., oncological CTs. In these hospitals, radiologists worked on a weekly schedule, assigned to one of such lists a day.
— *Small General Hospitals and Medical Imaging Clinics (D2, K2, K3, K4).* Usually, there was only one list, which could have been filtered for a specific imaging type.

Vision: AI Distributing Examinations by User's Expertise. Radiologists envisioned an AI-based system that would 'filter the normal radiographs' in times of increased workload. It would screen all incoming chest radiographs for findings and assign them to one of two categories: those with abnormal findings and those without. P19, the head of the radiology department in a large Kenyan general hospital (K5), suggested that *the ones that are not flagged as abnormal are given to junior radiologists to clear the dispute.* This means that junior radiologists would ensure that a radiograph is truly normal and correct for errors. P19 continued, *And then the ones that have findings, I would assign to one of the senior radiologists. That would be useful.* This way, senior radiologists would use their time efficiently, focusing only on radiographs with findings. This suggestion sprouted when we asked about the backlog of 500 unreported radiographs. Through such an AI-supported distribution, P19 hoped for more radiographs to be reported within a given period to reduce the backlog to normal.

This suggestion was considered a useful tool to support the smooth work of a radiology department in times of increased workload. However, radiologists were aware of the potential negative effects on the education of junior radiologists should such a system be used daily, which was not their intention. Such a system was envisioned as a temporary means of support in moments of greater stress on the healthcare system *for the purpose of clearing the backlog. 500 [cases] is not too much. Once you get down to a reasonable number, then, of course, they can stop. But it'll be a good filtering tool to clear the queue* [P19]. After clearing this temporary backlog, the distribution of examinations would return to normal to re-engage junior radiologists in education.

4.1.2 Challenge: Selecting the Next Most Relevant Examination to Report without an Easy Overview.

When selecting the next examination to report, radiologists in Denmark and Kenya constantly scanned for the next most relevant and urgent examination. In the observed sites, medical emergencies constituted urgency and were always prioritised by radiologists. A medical emergency was usually indicated by ordering clinicians on an X-ray order, as they were the ones who knew patients' health conditions (Figure 3). Importantly, these medical emergencies did not account for accidental findings, i.e., conditions that could constitute an emergency but were not expected by the ordering clinician.

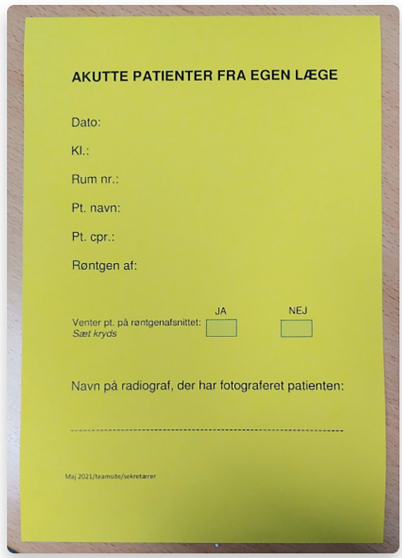

Fig. 3. A note from D4 indicating medical emergency. The emergency was indicated by an ordering clinician. A radiographer capturing the radiograph would take note of the emergency and bring this note to a reporting radiologist to prioritise it.

The meaning of urgency varied between the sites and situations. Not all consequential radiological findings, e.g., cancer, constituted urgency. This was explained by P22, a specialised radiologist from K1, *if you look at the medical emergency, things that need to be actioned, it's not everything. A lung nodule that has been sitting there for a few months, even if I report it on Monday [after a weekend], doesn't make a big difference, but a pneumothorax ... makes sense.* This quote emphasises the importance of timely treatment for selected conditions, e.g., pneumothorax, which almost always constitutes a medical emergency, as a delay of even a few days can negatively impact patient outcomes. Whereas prospects with other conditions, despite their seriousness, are less affected by a delay of a few days. Moreover, the conditions prevalent in different hospitals may depend on the type of the medical site and the population it serves. P17 highlighted the inherent complexity of deciding a priority among other cases that do not constitute a medical emergency. *I can't say this [a single radiological finding] is a high priority and ignore the other because it is one scenario, one package. I can't tell which one's more important.* In different contexts, different findings may be more or less critical, e.g., the air in the abdomen is an expected finding in a post-operative X-ray but an uncommon one in routine assessment, mandating immediate action in a routine X-ray.

While the utmost priority was given to urgent medical cases, healthcare professionals, whenever possible, tried to alleviate patient inconvenience. This prioritisation was observed when a patient had to wait for the report, often to bring it to the ordering clinician (K2, K4, K5). Such patients were noted by secretaries or radiographers, who informed radiologists about their situation. This way, the waiting patients' examinations were picked before other non-urgent examinations but after any constituting a medical urgency, as explained by P21: *You see some of these people are not waiting here. They got the exam and went home; others are sitting at the reception. So there's a receptionist there who comes to tell you: "This one is waiting here; please prioritise". So we will do that one and then move on with the rest of the list.*

Medical emergencies were communicated in several ways and, to varying degrees, supported by the existing IT systems. Radiologists typically received communication about the most critical cases through phone calls from the ordering clinicians. This was observed in both countries even when a priority attribute could be indicated on a radiograph order. Such calls were made primarily between doctors within the same hospital, as noted by P17: *[Even at night] they call us. If there is an urgent [case], they have to call me: "We have an urgent case. Please report it immediately." So I have to wake up and report.* However, maintaining the delicate balance between medical urgency and convenience was difficult. Some phone calls were motivated by the organisation of work at the requesting institution and not a patient emergency. *<After putting down the phone> Urgent thing ... or it's urgent because they're closing around five. It's always urgent. So around this time, everyone starts calling* [P14]. In such cases, the phone calls may not always achieve the desired prioritisation but instead introduce breakages to the workflow.

Instead, radiologists preferred the urgency conveyed through PACS as an attribute of a radiograph order. Using this attribute, they could order their work list to select the next most relevant examination to report. In K1, examinations were colour-coded based on the urgency attribute: *The red is an urgent, very urgent examination ... So this one you want to report within 1-2 hours, preferably before 1 ... The next colour will be semi-urgent, so within two hours, preferably before. Then there is this <colour>. Usually, the turnaround time is 24 hours. So the colour codes help us to know which exams we need to report first* [P22]. This way, radiologists could get an overview of their patients, plan their work accordingly and minimise interruptions caused by phone calls.

Without the prioritisation in the PACS, radiologists relied on secretaries and radiographers to inform them about patients requiring immediate attention. They also established other processes to gain an overview of their work lists and select the most relevant examination next, like quickly checking all referrals. Not every PACS in the visited sites supported that functionality (D2, D3, K2, K3, K4), and the system in K5, which did, was rendered unusable due to organisational challenges. Every examination in K5's PACS system was marked as urgent, while almost none of them were. As a result, radiologists ignored the urgency attribute and did not use it for prioritisation. P17, from a small general hospital in Kenya, usually quickly browsed through all the paper referrals, which were scanned in the PACS and linked to the examinations, to see which examination to report next.

Lastly, in all of the hospitals, the work and expertise of radiographers were crucial for radiologists' prioritisation of cases. They conveyed the urgency assigned by an ordering clinician in some of the general hospitals (D3, D4, K3, K4). Moreover, whenever assessing the quality of the captured X-rays, radiologists would gain experience in distinguishing certain findings. P9 explained why this was important, *Technicians [radiographers], they were wondering what this was in the breast, but she had both her breasts removed, and I think she had a hematoma, and that's what's remaining ... It's good for them to learn because then they can see if there's something urgent, and they will call us.* Based on this experience and patient interactions, radiographers could spot certain conditions on X-rays that necessitated urgency and inform radiologists about them.

Vision: AI Detecting Medical Emergencies. In all the visited sites, radiologists received examinations chronologically and constantly engaged in prioritisation work at the scene to report the most urgent and relevant examinations first. However, they had a little overview of the content of the examinations before opening them and relied primarily on clinician-indicated emergencies, which meant that inevitably, some of the urgent findings remained undetected until a radiologist saw the radiograph. Participants from Denmark and Kenya shared similar visions when asked whether they could envision AI supporting them in this prioritisation.

Radiologists envisioned an AI-based system that would screen all the incoming radiographs and detect findings constituting medical emergencies (K1, K2, K3, K5, D2, D4). While the desired findings to screen for varied between contexts and practitioners, pneumothorax—a lung collapse that may lead to respiratory or heart failure—almost always constituted a medical emergency. P14 envisioned AI that would detect it, *you can use it to triage trauma patients, ICU patients, and generally walk-ins … If there's a pneumothorax, it'll let me know. That's a very useful thing*. They would use that knowledge to prioritise relevant cases in their work lists and deliver crucial reports faster. Such screening would improve patient outcomes by faster diagnosis in time-sensitive cases. The main clinical value comes from providing faster diagnoses for patients with undetected medical emergencies.

4.2 Interpretation: Practice of Drawing Conclusions

After selecting an X-ray, radiologists proceeded to interpret it. The interpretation of chest X-rays involved understanding (1) a patient's situation, (2) visually detecting findings and (3) interpreting them in the patient context, i.e., why the X-ray was ordered, the patient's current state, their location (admitted to a hospital or at home) and their medical history. This results in a high degree of subjectivity. P20, a senior radiologist from a large general hospital in Kenya, warned us that *you can ask each of us [three senior radiologists in the room] and everybody will give a different opinion. That's the biggest problem with chest X-rays* [P20]. We observed two main reasons for this variability: the visual complexity of chest X-rays and the variation in the additional medical information required to interpret them.

4.2.1 Challenge: Interpreting Visually Ambiguous Findings. Chest radiographs are visually complex, making distinguishing and interpreting single findings difficult. P14 explained how X-rays are merely shadows of the complex structures in the human body cast on a plane. *On the surface, it [an X-ray] is extremely simple, but as you look at it, it becomes extremely complex … because of all the things that are in the chest and even the different densities within the chest from the lung, all the way to bones, mediastinum, and everything in between.* However, even when spotted, many abnormalities cannot be categorised unequivocally. Shadows of different conditions look the same; some pathologies may hide behind others. For example, during an observation P9 noted, *So I'm looking at this area and wonder if there is anything relevant. But there is something like this on the other side, so it probably is nothing. This is one of the very difficult areas because there are too many bones … and this patient is like this [wrongly positioned], and he's overweight.* P9 explained that they relied on their experience with similar cases to interpret such findings accurately.

One of the main ways radiologists resolve doubts about their interpretations is through collaborative interpretation with their colleagues (Figure 4). Such collaboration was well supported by a local PACS in D3 through an internal messaging system. As a result, such consultations became a norm. *I think this is one of the best things about our PAC system. We use it a lot … Instead of calling and saying, "Look at this CPR number [personal identification number]", and she is plotting it in [it takes extra time], now this is so easy [to just send it]* [P9]. Later, they encountered an X-ray that they were uncertain about. P9 decided to ask for a second opinion from their colleague: *So I just want to show her this. I don't know what happened, but it doesn't look good at all. Maybe it's nothing.* Then, they concurred at the colleague's desk and worked together to interpret the examination. This effort was deemed worthwhile even when faced with additional labour needed to consult with peers. As the only radiologist in K2, P14 had to rely on collaborating with other radiologists remotely, e.g., *I have a friend who's in the US … maybe two, three months ago, there was one case that … boggled my mind, he specialised in paediatric radiology, but he's also done musculoskeletal radiology so I would consult with him. There's another radiologist who works at (K5) whom I consult with.*

Junior radiologists relied even more on collaboration with their senior colleagues. As part of their training, they held regular meetings with their supervisor, which offered an opportunity to ask questions about difficult cases and their interpretations. However, when stuck on a particularly

Fig. 4. In K5, the reporting room was occupied by three radiologists at the time of the study. Radiologists often chatted about examinations to resolve doubts. They either turned and looked together at an examination on a single workstation or read out loud an identification number, which other radiologists used to find the examination in question and interpret within their own workstation.

difficult case, junior radiologists would reach out to their senior colleagues outside of the regimen of scheduled meetings, as explained by P11, *if I have any questions, I just need to find anyone who is at work [senior doctor qualified to approve x-rays] and I can just confer with them.* These scenarios exemplify the collaborative practices among radiologists from different medical sites and the critical role of such collaboration in chest radiograph interpretation.

Vision: AI Providing Decision Support on Subtle and Difficult Cases. Doctors from the majority of the clinical sites (D1, D2, K1, K2, K3, and K5) envisioned two AI-based functionalities that would help them during the interpretation of the radiographs—(1) providing decision support to increase certainty in difficult cases and (2) detecting subtle finding to reduce the risk of overlooking.

Firstly, our participants wished that AI could provide decision support when unsure about their interpretation. While our participants consulted their colleagues and friends when interpreting difficult cases, this support was often burdened with extra work to share the examinations, connect over the Internet or walk to another room to discuss them. P16, the only radiologist at a small general hospital in Kenya (K3), depicted the following intended use, *I think for me the utility will be more ... like a second opinion. I would want it to be like a second opinion to be more confident.* Similarly, P4 explained that in cases when they were not completely sure of a diagnosis, AI providing decision support could help decide on their interpretation, *If I'm 95% sure about the diagnosis, and find something else [an AI prediction] that suggests it, then I'm 100% sure.* While consulting colleagues may be more beneficial, it also takes more resources. Due to the high workload, radiologists often could not read through all the available medical history, let alone consult all the cases among each other. These constraints would not afflict an AI that could opinion a difficult radiograph in question.

Secondly, they envisioned a system that could help them not to miss subtle findings during the interpretation. *You are often in a hurry. It'll always help if a machine can ... show me something I haven't seen. Of course, I can't see everything. Of course, I'll make failures* [P12]. The awareness of the possibility of missing subtle findings was common among our participants. Similarly, P22 from K1 reflected, *As human beings, sometimes you miss something. You miss a small, tiny nodule in the chest x-ray, which after six months ... the patient comes back, and it's a big mass which you had* *missed, which happens. We're humans.* In these visions, radiologists wanted to ensure that they had not missed anything by having an AI double-check the radiograph for subtle findings.

These envisioned AI-based systems were contingent on the usefulness of the detected findings. Namely, the need for support was very low when interpreting simple cases with 'obvious' findings. This was emphasised by radiologists from Denmark and Kenya alike, e.g., P20 who previously evaluated an AI-based chest X-ray decision support tool mentioned, *when the findings are obvious, it wouldn't even be faster because then ... you'll see them quickly. It's with those subtle [findings] that AI could be useful.* The obviousness of the findings was also relative. While most radiologists found decision support detecting pneumothorax extremely useful, a specialised radiologist interpreting post-operative X-rays, where pneumothorax is a relatively frequent finding, held a contrary opinion: *Your tool will most likely detect pneumothorax. It's not that useful. We can easily see a pneumothorax* [P7]. These comments highlight the lack of clinical usefulness of AI-based systems providing decision support for obvious findings relative to the medical site and professional expertise.

4.2.2 Challenge: Time-Consuming Process of Obtaining Additional Clinical Information. Chest X-rays were only one of many diagnostic tools healthcare professionals in Denmark and Kenya employed to understand patient conditions. Insights from chest X-rays were only as good as radiologists' understanding of patients' context. Radiologists in both countries familiarised themselves with a radiograph referral to understand the clinical questions associated with a particular X-ray. Then, they gathered other available clinical information, e.g., historical images, to create a coherent story about a patient's condition. Only then, in the light of their understanding of a particular patient's situation, they interpreted the radiological findings seen on an X-ray. P9 pointed out the broader focus of radiologists' work: *It's all the big picture, and I can't do a proper description of the picture if I don't have all the information.*

A referral was typically available in digital (as a scanned document or a digital entry) or paper form. Radiologists used referrals to focus their efforts and guide their interpretation of visually similar findings. P17 stressed out that *if they [ordering clinician] didn't write it, I don't know what they look for, why this patient did the chest X-ray. Are we looking for an infection or pneumonia? But the patient had a trauma, so we are looking for a fracture. Maybe I would have missed it if they hadn't told me.* For this reason, all of our participants considered a referral necessary information and often would not report an X-ray without it.

Historical radiographs were another important source of information. Comparing patients' current X-rays against previous ones was useful to exclude potentially life-threatening diagnoses and discern between visually similar findings. P12, from an imaging clinic in Denmark (D2), captured this by referring to old X-rays as gold: *they are gold because when you can see he [a patient] was seen five years ago, you find the old picture and it looks exactly the same. Then you say, "Oh, it's not cancer".* In this scenario, P12 found the same finding of the same size on a historical X-ray. Hence, they could rule out cancer. Similarly, P21 from a large general hospital in Kenya (K5) used historical X-rays to guide their interpretation of ambiguous findings: *The images can look alike. But interpretation will differ depending on the history* [P21]. While providing invaluable insights, fetching historical images and comparing the development of findings across the radiographs were repetitive, time-consuming and poorly supported by the PACSs. *It takes a lot of time to open one [radiograph], look at it and see what was happening, then open the next one and see if it had resolved* [P22]. The need for extra time to use historical radiographs was a concern during times of heavy workload.

Medical records provided a wealth of information, e.g., test results or treatment progress, yet radiologists had to go out of their way to access them. Since radiologists worked in PACS and could access the referral and historical images directly, switching systems to search for patient information in the EHR was usually time-consuming. Because of this, it was used as a source of last resort for most complex cases. P10 explained: *Not all of us will do this because we do not have enough time. I can't check every patient in the SP [the local EHR system].* However, information obtained in such a way was considered very helpful, as expressed by P22: *I prefer to check the records cause you're able to get even more [information] than the clinician may give you.* The EHR was available in Danish hospitals, in the specialised hospital (K1) and the bigger general hospital (K5) in Kenya. Paper versions of the records were not used.

Finally, despite the overwhelming subjectivity of the interpretations, radiologists pointed out that they also use some objective metrics. Frequently, radiologists manually measured ratios or sizes of visual features on a radiograph to discern certain conditions. For example, they measured the width of the cardiac silhouette and the thoracic cavity. A ratio greater than 0.5 suggested cardiomegaly (enlarged heart). Similarly, P14 explained, *I am trying to discern whether this area is normal vascularity or whether it's an infection. An objective way to do that is to compare the intercostal spaces.* While measuring such features was common and not difficult for radiologists, it takes their valuable time away from using their expertise.

Vision: AI Measuring Visual Features and Comparing Changes Across Historical Examinations. To reduce the probability of errors and streamline their interpretative work, radiologists envisioned AI-based systems that could (1) pre-process radiographs by measuring visual features and (2) compare detected findings against historical radiographs.

Firstly, AI could assist with manual measurements of visual features. P22 envisioned an AI-based system that could support their interpretation by conducting such measurements: *AI could just measure for me the heart size and tell me if it's within normal. Because it's very manual, time-wasting … But you know, you can easily train a machine to do that.* This way, radiologists would only have to inspect the results, which would realistically reduce the time needed to interpret a radiograph.

Secondly, AI could automatically compare changes between historical radiographs. Similar to the manual measurements, fetching historical images and assessing their differences were a common and time-consuming task. Participants from K1, K5, D1, D3, and D4 envisioned an AI-based system that automatically fetched historical radiographs, detected relevant findings using image recognition and compared their development over time. Providing such support would expedite the manual search for findings across historical images and help radiologists assess whether changes currently observed are new developments or have been already present, resulting in improved clinical outcomes. Rather than spending time on manual fetching and comparing historical radiographs, P22, working in a specialised hospital in Kenya, envisioned: *It would help to have an AI that just compares … say the last five scans … highlights those changes for you so that you can just interpret what has happened. So that would be nice.* In this vision, the AI-based system would use image recognition to detect changes between historical chest radiographs and bring them to the radiologist's attention for interpretation. P9, from a general hospital in Denmark, envisioned a similar support tool to find the same finding across historical images and inform radiologists about its progression. *There are some small nodules in the right lung, and I have to compare … I think artificial intelligence could be very helpful here to spot the nodules and then compare them.* The envisioned functionality would allow radiologists to focus on interpreting the changes rather than fetching images and localising the changes.

4.3 Reporting: Practice of Contribution

The final stage of handling a chest radiograph encompassed writing a report. A report may seem to be solely a plain summary of an X-ray interpretation. However, this part of radiologists' work was also mediated by a patient's context and the expertise of the report recipient. In the visited sites,

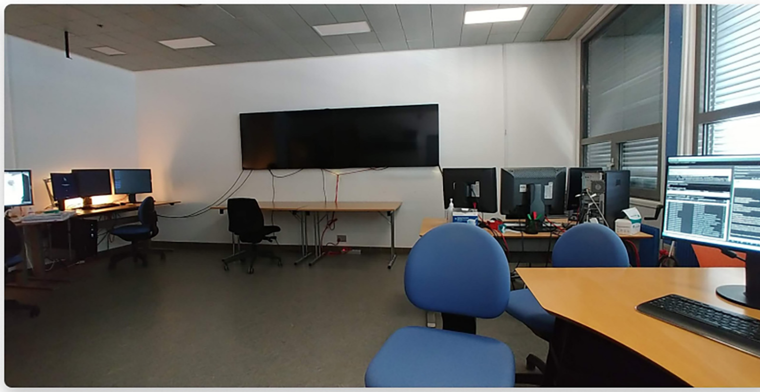

Fig. 5. In D1, conferences with clinicians were a part of the dissemination practices aimed at increasing the knowledge transfer between radiologists and ordering clinicians. In this room, radiologists responsible for one work list presented and explained the most important, difficult or interesting cases to clinicians from the departments that ordered them. This was also when they could ask questions about the reports of other patients.

radiologists usually communicated the three following elements in their reports: their impressions of what they saw on the radiograph, their interpretation of these findings in relation to the patient's condition and recommendations on the next steps for the ordering clinician (Figure 5).

4.3.1 Challenge: Conveying the Right Information in a Report. Radiologists framed their reports to provide maximum benefit to the patient based on the ordering clinic and the report recipient. The level of detail in their descriptions was finer if they knew that the ordering clinician could use this information. They also mentioned all the visual findings that may concern ordering clinicians to avoid confusion, even if they were not clinically relevant. *There are some [findings] that I can ignore mentioning. But the others not ... they [ordering clinicians] might think that you missed it because they're not aware what it is, they might worry that maybe it is something when it's not* [P16]. Similarly, the type of guidance radiologists provided varied between the recipients. P14 from a medical imaging and teleradiology clinic in Kenya mentioned, *My reports vary depending on who has sent it [an X-ray order]. So if it's a medical centre in Nairobi, I know they're going to interact with a healthcare practitioner who's fairly experienced... but the further out you go out, the less you have ...* P14 explained that clinics in the remote parts of Kenya lack medical doctors. As a result, nurses and clinical officers take care of patients. In some cases, they may lack the medical training necessary to navigate the complexities of diagnosing certain conditions. P14 continued, *I need to use my expertise to sort of guide them. Cause they'll put a lot of weight on what I say ... You want to give them some ideas: "This is what I think is going on. Please check for 1, 2, 3." And then they may not have thought of it and said: "Oh well, the child doesn't have a fever, and the child hasn't been coughing, so it's unlikely to be pneumonia." But if I write pneumonia, they're going to put the child on antibiotics whether or not the child has symptoms. So they say, "The doctor from Nairobi says pneumonia, so it's pneumonia." It's tricky ...* Overall, these quotes underscore the adaptability and responsibility of radiologists to provide tailored guidance in their reports based on the specific needs and capabilities of the recipients they report to.

Vision: AI Double-Checking Reports against Radiographs for Missed or Misinterpreted Findings. As a response to the problem originating at the interpretation stage—the complexity of reporting chest

X-rays and the high temporal cost of accessing support—and at the reporting stage—the fallibility of transcription software used—radiologists envisioned a system that would double-check whether all the relevant findings detected on the X-ray were mentioned in the report. This system would use two types of AI: **Natural Language Processing (NLP)** to read through a radiologist's report and image detection to detect relevant findings on an X-ray. Subsequently, the system would assess whether all the findings detected on an X-ray were also present in the accompanying report. In case of divergence or missing a finding, it would notify radiologists about potential errors.

Radiologists acknowledged their fallibility, which could be minimised with such AI use as envisioned by P17, *Even for us as consultants ... everyone is making mistakes, right? So I can report everything, and just with one click, I can ... see if I was wrong, if I missed something, if I have to think more about this, if I have to ignore this. It'll help everyone.* The report analysis is the main difference between this vision and the decision support described before. In this envisioned future, the AI's assessment of the radiograph is not available to radiologists before finishing the report. Thus, radiologists benefit from seamless quality assurance and interact with the AI only when an issue with the report is discovered, avoiding another layer of work for every X-ray.

5 Discussion

This study offers a unique perspective on radiology work practices and AI opportunities in two diverse countries. HCI researchers suggest that such joint focus on commonalities is necessary to balance the northern narrative in technology design, minimise biases and 'enable translation across geographies' of clinical AI-based systems [55, 73]. If not addressed, sociotechnical and political differences both within and between Global North and Global South countries impact the design and successful use of clinical IT systems [55, 90, 92, 139].

In this study, we contribute to the translational and bridging line of work. We investigated the feasibility of designing AI-based systems that effectively support radiologists across contexts, and not only respond to the practical challenges faced in the Global North. This is why we studied chest X-ray practices across nine clinical settings in Denmark and Kenya and focused on the relationship between practice and visions for AI support. Our findings point towards designing more useful AI-based systems that transcend national borders.

We found that radiologists' practices were similar across medical sites of the same type, regardless of the country, which may be due to the comparable medical education, shared biomedical model of medicine, scientific constructs and medical ethics in the undeniably technology-dependent field of radiology. Moreover, we show how these practices are more collaborative than expected and possess unique human contributions. We also detailed how local and contextual aspects of radiology work shape visions of AI support. These findings have implications not only for AI in radiology but also for HCI and the design of HCAI.

Taking a few steps back and looking beyond the practice, we did observe socio-economic differences between clinical sites across the two countries. For example, we noted differences between the types of healthcare systems, available IT infrastructure and prevalent medical conditions, which may cause diverse challenges to AI adoption across countries in the Global North and Global South [152] (see, e.g., the impact of socio-environmental factors in a study by Beede et al. [15]). However, their influence on concrete AI implementations should be further assessed, as this study focuses on uncovering valuable support for shared chest X-ray practices, motivated by the low clinical usefulness of hitherto radiological systems [122, 132]. Finally, we discuss these lessons and share reflections on what this means for designing clinically useful AI.

5.1 Practice-Grounded Envisioning of AI Futures: Searching for New Design Opportunities

Previous HCI studies revealed a disconnect between the functionality provided by AI-based systems and the functionality needed in clinical settings [56, 58, 98]. This suggests gaps in understanding the problems and needs of clinical end-users before considering AI-based solutions. To bridge this gap, researchers in the HCI community are increasingly voicing their concerns and advocating for an AI-paradigm change going from technology-centric to human-centric [8, 20, 107, 118, 119, 140, 150]. Results presented in this study adopt a human practice-centric perspective on AI support. We have demonstrated the potential to uncover new ways for AI support that align more closely with the needs and wishes of clinical end-users.

HCI researchers should envision AI functionality through engagement with real-world practices rather than merely following the trajectory of AI development. Engaging communities of practice before the development of AI systems is critical [58, 84, 123, 131], as once completed, AI model capabilities often dictate the final system's capabilities and, due to the high cost of the data and model work, they are unlikely to change [5, 152]. We advocate for a paradigm shift where clinical support opportunities arise from the practice, and not from retro-fitting already conceived AI models born out of access to data or external pressures to adopt AI. Practice-grounded envisioning of AI futures can reveal new ways these systems can support practitioners. Thanks to such engagements, in this study, we revealed opportunities for support in the complex, nuanced radiological workflows and collaborative dynamics inherent in clinical practices.

We also caution that enacting these opportunities in real-life projects may not be as straightforward. Critical aspects of such projects that shape the subsequent work and thus future systems, like regulatory approvals or data access, often need to be attended to prior to any development work. This was the case in our project, where the type and amount of data we could access were predetermined in the project description finalised before this study [151]. As a result, while we explored a broad range of visions addressing diverse challenges of chest X-ray practice, not all of the ideas can be incorporated into our future system. Addressing this challenge requires more profound changes to how we think about clinical AI innovation and joint efforts from legal, healthcare and computer science stakeholders.

However, as AI continues to increase its presence in healthcare, the first step is to shift the driving force behind AI development—from a technological opportunity to a desirable vision of the future rooted in the local sociotechnical context (local practices, hospital settings, expertise of clinical end-users).

5.2 New Design Opportunities for Clinically Useful AI in Radiology

The current understanding of handling chest X-rays has centred around an individual radiologist's interpretation work. This is reflected by hitherto AI-based support systems that have so far been aiming at supporting the interpretation and detection of findings on X-rays, also referred to as second opinion [33, 63, 67, 104, 138]. However, the clinical use of such systems in radiology remains low [46, 112, 137]. Research points out that one of the main reasons for the failure of AI in radiology is the 'uncertain added value for clinical practice of AI applications' [122].

The extent of visions for AI presented in this study brings attention to the importance of broadening the design space to include all three stages of chest X-ray practice, transcending the narrow potential of AI-based second opinion. This conceptual contribution is not merely speculative; it is grounded in the insights of radiologists who advocate for a pragmatic shift towards support mechanisms that enhance their professional expertise, as opposed to replacing them with models simplifying their practice [72, 76, 129, 143]. Our contribution lies in highlighting new directions for designing clinically useful AI in radiology and recommending improvements in how AI can support radiology work.

AI support has to account for the subtle collaborative practices in chest X-ray practice. We showed the collaboration between radiologists and other actors involved in the treatment process of patients. Fridell et al. previously argued that radiologists are active partners to clinicians in the diagnostic process [41]. We further expand the understanding of radiologist work by describing the subtle collaboration often mediated through various artefacts, such as referrals, urgency attributes and other data associated with X-rays. These findings contrast the dominant understanding of radiological practice, which, based on current AI-based systems offerings, focuses on the visual detection of findings on radiological examinations almost out of clinical context. We argue that to support radiologists in their practice, designers and developers of AI-based systems have to be aware of the, in fact, collaborative work practices. To truly contribute to radiologist work and have a positive effect on patient care, future AI-based systems must account for the already existing practices rather than ignore them, as these practices enable radiologists to work and support the best clinical outcomes for patients.

Meaningful AI support should not simplify the practice to one-dimensional finding detection. Our study foregrounded the human aspect and the range of tasks in radiologists' practice. Previously, Pesapane et al. highlighted, among others, educating students, communicating diagnoses and performing interventional procedures [96]. We add to this tasks related to handling medical images, e.g., drawing from various data sources to understand patient situations, anticipating ordering clinicians' questions and framing reports to contribute to clinicians' decision-making processes. These tasks are radiologists' contributions to the diagnostic process beyond the sole identification of findings on examinations. They require empathy, analysis of multimodal data and complex reasoning. To be able to contribute to radiological practice, designers and developers alike should be aware of the complexity of radiologists' work and take them into account when designing support functionalities.

One of the ways these two points could be achieved is deeper integration with adjacent systems and greater utilisation of multimodal medical data. Researchers have already taken steps to assess the feasibility and opportunities behind such systems [2, 79]. By combining data from various sources, AI could expand the provided support from one-dimensional task detection to, e.g., suggesting diagnoses most relevant in the context of a patient or drafting reports suitable to the recipient's expectations (inspired by the work by Yildirim et al. [147]). However, the amount of available clinical data grow constantly and knowledge of which data is relevant, and in which context is critical. We believe that using our detailed accounts of chest X-ray practice could help future projects determine which clinical data to focus on and guide its utilisation by multimodal AI-based systems.

To provide useful second opinions, AI needs to limit additional work needed to use it by adjusting to local conditions. Our findings do not negate the premise of supporting image interpretation. We argue that AI, as a second opinion, should be configurable to the local conditions, including the type of clinic, radiologist expertise, current workload and patient characteristics, to reduce the additional work required to benefit from the predictions. Our participants stressed the difficulty and subjectivity of interpreting chest X-rays, in line with previous research—the disagreement rates among radiologists on chest radiograph interpretations reach as high as 30% [37]. However, providing a second opinion on every examination based on an arbitrary set of pathologies, as experienced by some of our participants, resulted in a significant overhead. We know that the mental and temporal cost of discerning false-positive AI predictions may result in the failure of AI in clinical practice [11, 14, 81, 130]. This is critical in radiology where staff shortages and the increasing number of examinations result in reduced time available per examination, stress and overburdening [45, 105]. Thus, to provide a clinically useful second opinion, radiologists must be able to configure AI to reduce the additional work and spend their time on predictions relevant to their practice. We suggest that such configuration should be thoughtfully designed to address the social dependencies of radiology work [153].

AI support for radiology can transcend second-opinion. In this study, we explored the space of AI support thought clinically useful by radiologists. By contextualising chest X-ray practice within the broader diagnostic process, encompassing all its interdependencies, we uncovered several new opportunities for AI assistance. Particularly, we point to the selection and reporting stages as promising domains for AI designers and developers. Within these stages, radiologists envisioned AI to provide a much-needed overview of urgent cases, optimise case distribution based on the current clinic workload or double-check reports. Interestingly, none of the AI-based systems encountered by participating radiologists explored these practical tasks. However, promising work on providing alternative support has been conducted. For instance, Saßmannshausen et al. [111] conducted a human-centred design process and proposed new ways of supporting radiologists in prostate cancer diagnosis. Particularly, similarly to this study, they foregrounded the need to address the multidisciplinary collaboration and communication of radiologists and automation of the manual and repetitive tasks. This underscores the need for diverse support approaches, as the role of radiologists extends beyond pathology detection in medical images [96]. Our study offers a comprehensive account of radiological practice across diverse clinical settings that can serve as a starting point for future AI projects.

5.3 Clinically Useful AI Requires Configurable and Flexible AI

AI accuracy has historically been the main measure for evaluating the usefulness of AI systems in healthcare [77, 114, 133]. Achieving high AI accuracy has been the ethos of what entails a 'good' AI model, resulting in the pursuit of getting technology 'right' before anything else. However, technical excellence does not guarantee a positive impact on patient outcomes or the work of healthcare professionals [122, 138]. Researchers from both HCI and Health point out that technical excellence rarely translates directly to clinical usefulness [18, 27, 56].

5.3.1 Clinic Type and End-User Expertise Condition AI Support. Different types of clinical sites demand different types of AI support. This study revealed unexpected parallels in radiology work between Denmark and Kenya. However, while radiology work exhibited general similarities across the two countries, nuanced yet important differences appeared when comparing the work across types of medical sites. For example, in specialised hospitals in both Denmark and Kenya, X-ray work lists were created based on the hospital departments and according to medical specialisation, which stood in stark contrast to small general hospitals, where there was usually only one work list. These organisational differences affected the opportunities with AI for each medical site. For instance, the AI-based distribution of examinations by expertise was only envisioned in medical facilities with diversity among the employed radiologists. Similarly, differences in available resources in the radiology department were reflected in the envisioned AI support. In smaller clinical sites (D2, D4, D2, K3, K4) with a shortage of radiology staff, obtaining a second opinion was harder, which deemed visions for AI-based interpretation support more desirable. This suggests that the characteristics of the organisational context are important to consider when designing AI support in radiology. It highlights that different types of medical sites require different types of AI support.

Besides organisational differences, we also found that differences in radiologists' expertise may condition specific user requirements for AI. For example, radiologists relied on their experience with similar cases to accurately interpret the findings on visually complex X-rays where shadows of different conditions looked the same. On the other hand, junior radiologists undergoing training relied on collaboration with senior radiologists, especially when examinations were difficult to interpret. This suggests that end-users' level of expertise is a deciding factor with regard to which type of AI is considered useful, which has several implications for the design and development of AI in radiology.

Clinician-facing AI needs to be configurable to the clinic and end-user. It is common knowledge, both in academia and in the software industry, that favourable outcomes can be attributed to the configurability of clinical information systems [47, 78]. While designing for system configurability and tailorability is somewhat old news for HCI, it has yet to become part of the discourse on designing clinically useful AI. Emerging HCI studies investigating clinical AI indicate this direction. Researchers emphasise the significance of poor workflow integration, highlighting its equal importance alongside the well-known challenge of establishing trust by means of transparency and explainability [23, 56, 135]. Verma et al. [134] ascribe the limited impact of AI in medical practice to the poor connection between AI capabilities and clinical workflows. They argue for the importance of diverging from the existing 'one-size-fits-all' paradigm within HCAI discourses. Similarly, Wang et al. [135] report on tensions with the design of AI-based clinical decision-support systems in a rural clinical context and stress the reason being the misalignment with local context and workflow. Solutions to workflow integration problems with clinical AI have revolved around human-AI interaction, with extensive work on guidelines present (see, e.g., [6]) but also suggestions of creating multi-user systems and designing for time-constraint medical environments have been proposed [56].

When reflecting on the results of this study, we suggest that designers need to look beyond the interaction components of the interface. Specifically, there is a need to carefully consider how to make clinical AI systems configurable so that the set of AI features corresponds to the local needs of the medical site, the expertise of the end-users or the division of work and the number of medical staff. For example, this study showed how contextual medical emergencies are, i.e., different radiological findings may constitute emergencies in different contexts. We recommend that configuring the meaning of 'emergency' to fit local contexts would enhance the usefulness of AI-based prioritisation systems across various clinical sites, for example, the system evaluated by Baltruschat et al. [12]. As an immediate continuation of this recommendation, we conducted a follow-up study on the topic of AI configuration in relation to the social aspects of practice like clinic type, expertise of end-users, situation, patient context and medical knowledge [153]. We built on our understanding of chest X-ray practice and suggested how to configure four technical dimensions (AI functionality, AI medical focus, AI explainability and AI decision threshold) of radiological AI to ensure a better fit between human competencies and the capabilities of AI.

5.3.2 Situational Circumstances and Patient Context Influence Requirements for AI. Finally, radiologists rooted their visions for AI support in concrete situations and challenges from their local practice. For example, in time-limited situations where radiologists were 'in a hurry', the demand increased for AI that double-checks the radiographs for subtle findings or screens a list of examinations for medical emergencies that should be treated before the end of a radiologist's shift. Similarly, the type of welcomed support depended on the patient context, e.g., notifying radiologists about the detection of air in the abdomen may be a crucial safety feature when dealing with outpatients and a superfluous hindrance when applied to patients from a surgical ward. This means that radiologists' needs for AI are not constant; rather, they may change over time according to the situational circumstances and patient context.

Designing flexible AI is critical for ensuring clinical relevance. Bossen and Pine [23] have similarly found that AI being 'flexible' is an important factor for the successful use of AI in healthcare contexts. In their study of a NLP-based tool in the wild, they found that AI's utility was derived from its ability to act as an imperfect assistant that could be appropriated in use to fit particularly well with individual needs. While the timing of AI is important, i.e., it should be available when needed, HCI researchers suggest interactive machine learning as a way to make AI more useful for 'in-the-moment diagnostic needs' [26]. Cai et al. [26] provide important design recommendations for ways to improve AI's clinical relevance by enabling pathologists to interact with the AI tool in ways that fit with the particular cases. In both studies, it is clearly demonstrated how erroneous and imperfect AI algorithms can be made more clinically useful if designed for flexible use accommodating the needs of the situation. In light of these findings, we advise researchers and designers of clinician-facing AI to identify situational requirements and design for the appropriation of AI during use.

6 Conclusions

In this article, we present findings from a field study of chest X-ray practices in nine medical sites in Denmark and Kenya, aimed at informing innovation of clinically useful AI support for radiologists.

We presented how chest X-ray practice depends mostly on the type of clinical site and not the country in which the site is located. In other words, from the perspective of visions for AI support, the practice was more similar between two clinical sites of the same type that were located in Denmark and Kenya than between two clinical sites of different types located in the same country.

Next, we showed how radiology work was more collaborative than anticipated by being part of the overall diagnostic work. Moreover, the unique human contributions of radiologists surfaced as important yet often omitted when designing AI-based systems for chest X-rays.

These findings offer foundational insights into the design of AI support for radiologists. We use them to argue for expanding the space of AI in radiology, particularly to include AI-based distribution of examinations, measurements of visual features, assessments of historical changes, and report quality assurance. These visions transcend the traditional second-opinion systems and suggest that more opportunities for AI support that target other stages than interpretation should be explored. Rather than reacting to the trajectory of AI development, HCI researchers should shape AI innovation by facilitating future envisioning with domain experts in real-world practices.

Finally, we argued that the clinical usefulness of AI-based systems depends on their configurability and flexibility with regard to the type of clinical site, expertise of healthcare professionals, and situational and patient context. These reflections have implications for the design of clinician-facing AI and suggest new directions for future research on HCAI in health.

References

[1] [n. d.]. Digital X-ray On-The-Go in Kenya. Retrieved from https://radiology.ucsf.edu/blog/digital-x-ray-go-kenya

[2] Julián N. Acosta, Guido J. Falcone, Pranav Rajpurkar, and Eric J. Topol. 2022. Multimodal biomedical AI. *Nature Medicine* 28, 9 (Sep. 2022), 1773–1784. DOI: https://doi.org/10.1038/s41591-022-01981-2

[3] Scott J. Adams, Robert D. E. Henderson, Xin Yi, and Paul Babyn. 2021. Artificial intelligence solutions for analysis of X-ray images. *Canadian Association of Radiologists Journal* 72, 1 (Feb. 2021), 60–72. DOI: https://doi.org/10.1177/0846537120941671/ASSET/IMAGES/LARGE/10.1177\{_\}0846537120941671-FIG2.JPEG

[4] Amit Alfassy, Assaf Arbelle, Oshri Halimi, Sivan Harary, Roei Herzig, Eli Schwartz, Rameswar Panda, Michele Dolfi, Christoph Auer, Kate Saenko, et al. 2022. FETA: Towards specializing foundational models for expert task applications. *Advances in Neural Information Processing Systems* 35 (Dec. 2022), 29873–29888. Retrieved from https://ai-vision-public-datasets.s3.eu.cloud-object-storage.appdomain.cloud/

[5] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP '19)*, 291–300. DOI: https://doi.org/10.1109/ICSE-SEIP.2019.00042

[6] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. DOI: https://doi.org/10.1145/3290605.3300233

[7] Tariq Osman Andersen, Jørgen Peter Bansler, Finn Kensing, Jonas Moll, Troels Mønsted, Karen Dam Nielsen, Olav Wendelboe Nielsen, Helen Høgh Petersen, and Jesper Hastrup Svendsen. 2019. Aligning concerns in telecare: Three concepts to guide the design of patient-centred E-health. *Computer Supported Cooperative Work (CSCW)* 28 (2019), 1039–1072.
[8] Tariq Osman Andersen, Francisco Nunes, Lauren Wilcox, Enrico Coiera, and Yvonne Rogers. 2023. Introduction to the special issue on human-centred AI in healthcare: Challenges appearing in the wild. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–12. DOI: https://doi.org/10.1145/3589961
[9] Zachi I. Attia, Peter A. Noseworthy, Francisco Lopez-Jimenez, Samuel J. Asirvatham, Abhishek J. Deshmukh, Bernard J. Gersh, Rickey E. Carter, Xiaoxi Yao, Alejandro A. Rabinstein, Brad J. Erickson, et al. 2019. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *The Lancet* 394, 10201 (Sep. 2019), 861–867. DOI: https://doi.org/10.1016/S0140-6736(19)31721-0
[10] James Auger. 2013. Speculative design: Crafting the speculation. *Digital Creativity* 24, 1 (Mar. 2013), 11–35. DOI: https://doi.org/10.1080/14626268.2013.767276
[11] Anne Kathrine Petersen Bach, Trine Munch Nørgaard, Jens Christian Brok, and Niels van Berkel. 2023. "If I Had All the Time in the World": Ophthalmologists' perceptions of anchoring bias mitigation in clinical AI support. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. DOI: https://doi.org/10.1145/3544548.3581513
[12] Ivo Baltruschat, Leonhard Steinmeister, Hannes Nickisch, Axel Saalbach, Michael Grass, Gerhard Adam, Tobias Knopp, and Harald Ittrich. 2021. Smart chest X-ray worklist prioritization using artificial intelligence: A clinical workflow simulation. *European Radiology* 31, 6 (Jun. 2021), 3837–3845. DOI: https://doi.org/10.1007/s00330-020-07480-7
[13] Jakob E. Bardram and Claus Bossen. 2005. A web of coordinative artifacts: Collaborative work at a hospital ward. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '05)*. ACM Press, New York, NY, USA, 168. DOI: https://doi.org/10.1145/1099203.1099235
[14] Sally L. Baxter, Jeremy S. Bass, and Amy M. Sitapati. 2020. Barriers to implementing an artificial intelligence model for unplanned readmissions. *ACI Open* 4, 2 (Jul. 2020), e108–e113. DOI: https://doi.org/10.1055/s-0040-1716748
[15] Emma Beede, Elizabeth Baylor, Fred Hersch, Anna Iurchenko, Lauren Wilcox, Paisan Ruamviboonsuk, and Laura M. Vardoulakis. 2020. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–12. DOI: https://doi.org/10.1145/3313831.3376718
[16] Maura Bellio, Dominic Furniss, Neil P. Oxtoby, Sara Garbarino, Nicholas C. Firth, Annemie Ribbens, Daniel C. Alexander, and Ann Blandford. 2021. Opportunities and barriers for adoption of a decision-support tool for Alzheimer's disease. *ACM Transactions on Computing for Healthcare* 2, 4 (Oct. 2021), 1–19. DOI: https://doi.org/10.1145/3462764
[17] Arngeir Berge, Frode Guribye, Siri-Linn Schmidt Fotland, Gro Fonnes, Ingrid H. Johansen, and Christoph Trattner. 2023. Designing for control in nurse-AI collaboration during emergency medical calls. In *Proceedings of the 2023 23rd ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1339–1352. DOI: https://doi.org/10.1145/3563657.3596110
[18] Shlomo Berkovsky and Enrico Coiera. 2023. Moving beyond algorithmic accuracy to improving user interaction with clinical AI. *PLOS Digital Health* 2, 3 (Mar. 2023), e0000222. DOI: https://doi.org/10.1371/journal.pdig.0000222
[19] Pernille Bjørn and Lars Rune Christensen. 2011. Relation work: Creating socio-technical connections in global engineering. In *Proceedings of the 12th European Conference on Computer Supported Cooperative Work (ECSCW '11)*. Springer London, London, 133–152. DOI: https://doi.org/10.1007/978-0-85729-913-0\{_\}8
[20] Ann Blandford. 2019. HCI for health and wellbeing: Challenges and opportunities. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 41–51. DOI: https://doi.org/10.1016/j.ijhcs.2019.06.007
[21] T. Bodenheimer and C. Sinsky. 2014. From triple to quadruple aim: Care of the patient requires care of the provider. *The Annals of Family Medicine* 12, 6 (Nov. 2014), 573–576. DOI: https://doi.org/10.1370/afm.1713
[22] Claus Bossen and Randi Markussen. 2010. Infrastructuring and ordering devices in health care: Medication plans and practices on a hospital ward. *Computer Supported Cooperative Work (CSCW)* 19, 6 (Dec. 2010), 615–637. DOI: https://doi.org/10.1007/s10606-010-9131-x
[23] Claus Bossen and Kathleen H. Pine. 2023. Batman and Robin in healthcare knowledge work: Human-AI collaboration by clinical documentation integrity specialists. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–29. DOI: https://doi.org/10.1145/3569892
[24] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. DOI: https://doi.org/10.1191/1478088706qp063oa
[25] Eleanor R. Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuścińska, Suzanne Currie, J. Marc Overhage, Erika S. Poole, and Jofish Kaye. 2023. Healthcare AI treatment decision support: Design principles to enhance clinician

adoption and trust. In *Proceedings of the 2023 19th CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–19. DOI: https://doi.org/10.1145/3544548.3581251

[26] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, et al. 2019. Human-centered tools for coping with imperfect algorithms during medical decision-making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 1–14. DOI: https://doi.org/10.1145/3290605.3300234

[27] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. DOI: https://doi.org/10.1145/3359206

[28] Tara Capel and Margot Brereton. 2023. What is human-centered about human-centered AI? A map of the research landscape. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–23. DOI: https://doi.org/10.1145/3544548.3580959

[29] Davide Castelvecchi. 2016. Can we open the black box of AI? *Nature* 538, 7623 (Oct. 2016), 20–23. DOI: https://doi.org/10.1038/538020a

[30] Yunan Chen. 2010. Documenting transitional information in EMR. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1787–1796. DOI: https://doi.org/10.1145/1753326.1753594

[31] Enrico Coiera. 2019. The last mile: Where artificial intelligence meets reality. *Journal of Medical Internet Research* 21, 11 (2019), e16323. DOI: https://doi.org/10.2196/16323

[32] David Coyle and Gavin Doherty. 2009. Clinical evaluations and collaborative design: Developing new technologies for mental healthcare interventions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2051–2060. DOI: https://doi.org/10.1145/1518701.1519013

[33] Karin Dembrower, Alessio Crippa, Eugenia Colón, Martin Eklund, and Fredrik Strand. 2023. Artificial intelligence for breast cancer detection in screening mammography in Sweden: A prospective, population-based, paired-reader, non-inferiority study. *The Lancet Digital Health* 5, 10 (Sep. 2023), e703–e711. DOI: https://doi.org/10.1016/S2589-7500(23)00153-X

[34] Kate Donahue, Alexandra Chouldechova, and Krishnaram Kenthapadi. 2022. Human-algorithm collaboration: Achieving complementarity and avoiding unfairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 1639–1656. DOI: https://doi.org/10.1145/3531146.3533221

[35] Paul Dourish and Sara Bly. 1992. Portholes: Supporting awareness in a distributed work group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM Press, New York, New York, USA, 541–547. DOI: https://doi.org/10.1145/142750.142982

[36] Yrjo Engestrom, Ritva Engestrom, and Osmo Saarelma. 1988. Computerized medical records, production pressure and compartmentalization in the work activity of health center physicians. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work (CSCW '88)*. ACM Press, New York, NY, USA, 65–84. DOI: https://doi.org/10.1145/62266.62272

[37] Nicholas Fancourt, Maria Deloria Knoll, Breanna Barger-Kamate, John De Campo, Margaret De Campo, Mahamadou Diallo, Bernard E. Ebruke, Daniel R. Feikin, Fergus Gleeson, Wenfeng Gong, et al. 2017. Standardized interpretation of chest radiographs in cases of pediatric pneumonia from the PERCH study. *Clinical Infectious Diseases* 64, suppl_3 (Jun. 2017), S253–S261. DOI: https://doi.org/10.1093/CID/CIX082

[38] Geraldine Fitzpatrick and Gunnar Ellingsen. 2013. A review of 25 years of CSCW research in healthcare: Contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)* 22, 4-6 (2013), 609–665. DOI: https://doi.org/10.1007/s10606-012-9168-0

[39] Riccardo Fogliato, Shreya Chappidi, Matthew Lungren, Paul Fisher, Diane Wilson, Michael Fitzke, Mark Parkinson, Eric Horvitz, Kori Inkpen, and Besmira Nushi. 2022. Who goes first? Influences of human-AI workflow on decision making in clinical imaging. In *Proceedings of the 2022 22nd ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York, NY, USA, 1362–1374. DOI: https://doi.org/10.1145/3531146.3533193

[40] Diana E. Forsythe. 1999. "It's Just a Matter of Common Sense": Ethnography as invisible work. *Computer Supported Cooperative Work (CSCW)* 8, 1-2 (Mar. 1999), 127–145. DOI: https://doi.org/10.1023/A:1008692231284

[41] Kent Fridell, Lars Edgren, Lars Lindsköld, Peter Aspelin, and Nina Lundberg. 2007. The impact of PACS on radiologists' work practice. *Journal of Digital Imaging* 20, 4 (Dec. 2007), 411–421. DOI: https://doi.org/10.1007/s10278-006-1054-1

[42] Krzysztof J. Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S. Gene Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy, and Kyunghyun Cho. 2017. High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv:1703.07047. Retrieved from https://arxiv.org/abs/1703.07047

[43] Hongyan Gu, Jingbin Huang, Lauren Hung, and Xiang 'Anthony' Chen. 2021. Lessons learned from designing an AI-enabled diagnosis tool for pathologists. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (Apr. 2021), 1–25. DOI: https://doi.org/10.1145/3449084

[44] Hongyan Gu, Yuan Liang, Yifan Xu, Christopher Kazu Williams, Shino Magaki, Negar Khanlou, Harry Vinters, Zesheng Chen, Shuo Ni, Chunxu Yang, et al. 2023. Improving workflow integration with xPath: Design and evaluation of a human-AI diagnosis system in pathology. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–37. DOI: https://doi.org/10.1145/3577011

[45] Joseph Hawkins. 2001. Addressing the shortage of radiologists. *Radiology Management* 23, 4 (2001), 26–29. Retrieved from www.merritthawkins.com

[46] Jianxing He, Sally L. Baxter, Jie Xu, Jiming Xu, Xingtao Zhou, and Kang Zhang. 2019. The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine* 25, 1 (Jan. 2019), 30–36. DOI: https://doi.org/10.1038/s41591-018-0307-0

[47] Morten Hertzum and Jesper Simonsen. 2019. Configuring information systems and work practices for each other: What competences are needed locally? *International Journal of Human-Computer Studies* 122 (Feb. 2019), 242–255. DOI: https://doi.org/10.1016/j.ijhcs.2018.10.006

[48] Birthe Hojlund Bech. 2006. Danish society of radiology. *HealthManagement* (2006). Retrieved from https://healthmanagement.org/c/imaging/issuearticle/danish-society-of-radiology

[49] Judd E. Hollander, Keara L. Sease, Dina M. Sparano, Frank D. Sites, Frances S. Shofer, and William G. Baxt. 2004. Effects of neural network feedback to physicians on admit/discharge decision for emergency department patients with chest pain. *Annals of Emergency Medicine* 44, 3 (Sep. 2004), 199–205. DOI: https://doi.org/10.1016/j.annemergmed.2004.02.037

[50] Naja L. Holten Møller, Pernille Bjørn, Jonas Christoffer Villumsen, Tine C. Hansen Hancock, Toshimitsu Aritake, and Shigeyuki Tani. 2017. Data tracking in search of workflows. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2153–2165.

[51] Naja L. Holten Møller and Signe Vikkelsø. 2012. The clinical work of secretaries: Exploring the intersection of administrative and clinical work in the diagnosing process. In *From Research to Practice in the Design of Cooperative Systems: Results and Open Challenges*. Springer London, London, 33–47. DOI: https://doi.org/10.1007/978-1-4471-4093-1\{_\}3

[52] John Hughes, Val King, Tom Rodden, and Hans Andersen. 1994. Moving out from the control room: Ethnography in system design. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work (CSCW '94)*. ACM Press, New York, NY, USA, 429–439. DOI: https://doi.org/10.1145/192844.193065

[53] John A. Hughes and Dan Shapiro. 1992. Faltering from ethnography to design. *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, 115–122.

[54] Stephanie L. Hyland, Shruthi Bannur, Kenza Bouzid, Daniel C. Castro, Mercy Ranjit, Anton Schwaighofer, Fernando Pérez-García, Valentina Salvatelli, Shaury Srivastav, Anja Thieme, et al. 2023. MAIRA-1: A specialised large multimodal model for radiology report generation. arXiv:2311.13668. Retrieved from https://arxiv.org/abs/2311.13668

[55] Azra Ismail and Neha Kumar. 2021. AI in global health: The view from the front lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–21. DOI: https://doi.org/10.1145/3411764.3445130

[56] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: A sociotechnical lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. DOI: https://doi.org/10.1145/3411764.3445385

[57] Mohammad Hossein Jarrahi, Christoph Lutz, and Gemma Newlands. 2022. Artificial intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data and Society* 9, 2 (Jul. 2022), 205395172211428. DOI: https://doi.org/10.1177/20539517221142824

[58] Annika Kaltenhauser, Verena Rheinstädter, Andreas Butz, and Dieter P. Wallach. 2020. "You Have to Piece the Puzzle Together"—Implications for designing decision support in intensive care. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 1509–1522. DOI: https://doi.org/10.1145/3357236.3395436

[59] Konstantinos Kamnitsas, Enzo Ferrante, Sarah Parisot, Christian Ledig, Aditya V. Nori, Antonio Criminisi, Daniel Rueckert, and Ben Glocker. 2016. DeepMedic for brain tumor segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10154. Springer Verlag, 138–149. DOI: https://doi.org/10.1007/978-3-319-55524-9\{_\}14

[60] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. "Because AI is 100% right and safe": User attitudes and sources of AI authority in India. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. DOI: https://doi.org/10.1145/3491102.3517533

[61] Naveena Karusala, Shirley Yan, Nupoor Rajkumar, VictoriaG, and Richard Anderson. 2023. Speculating with care: Worker-centered perspectives on scale in a chat-based health information service. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sep. 2023), 1–26. DOI: https://doi.org/10.1145/3610210

[62] Pearse A. Keane and Eric J. Topol. 2018. With an eye to AI and autonomous diagnosis. *NPJ Digital Medicine* 1, 1 (Aug. 2018), 40. DOI: https://doi.org/10.1038/s41746-018-0048-y

[63] Brendan S. Kelly, Conor Judge, Stephanie M. Bollard, Simon M. Clifford, Gerard M. Healy, Awsam Aziz, Prateek Mathur, Shah Islam, Kristen W. Yeom, Aonghus Lawlor, et al. 2022. Radiology artificial intelligence: A systematic review and evaluation of methods (RAISE). *European Radiology* 32, 11 (Nov. 2022), 7998–8007. DOI: https://doi.org/10.1007/s00330-022-08784-6

[64] Christopher J. Kelly, Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. 2019. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine* 17, 1 (Oct. 2019), 195. DOI: https://doi.org/10.1186/s12916-019-1426-2

[65] Finn Kensing, Jesper Simonsen, and Keld Bodker. 1998. MUST: A method for participatory design. *Human–Computer Interaction* 13, 2 (Jun. 1998), 167–198. DOI: https://doi.org/10.1207/s15327051hci1302\{_\}3

[66] Finn Kensing, Jesper Simonsen, and Keld Bødker. 1998. Participatory design at a radio station 1. *Computer Supported Cooperative Work* 7 (1998), 243–271.

[67] David Killock. 2020. AI outperforms radiologists in mammographic screening. *Nature Reviews Clinical Oncology* 17, 3 (Mar. 2020), 134–134. DOI: https://doi.org/10.1038/s41571-020-0329-7

[68] Josua Krause, Adam Perer, and Kenney Ng. 2016. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 5686–5697. DOI: https://doi.org/10.1145/2858036.2858529

[69] Leah Kulp and Aleksandra Sarcevic. 2018. Design in the "medical" wild: Challenges of technology deployment. In *Proceedings of the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, Vol. 2018-April. ACM, New York, NY, USA, 1–6. DOI: https://doi.org/10.1145/3170427.3188571

[70] Steinar Kvale. 1996. *InterViews: An Introduction to Qualitive Research Interviewing*. Sage.

[71] Chantelle C. Lachance and Melissa Walter. 2020. *Artificial Intelligence for Classification of Lung Nodules: A Review of Clinical Utility, Diagnostic Accuracy, Cost-Effectiveness, and Guidelines*. 1–23. Retrieved from http://europepmc.org/books/NBK562929https://europepmc.org/article/nbk/nbk562929

[72] Curtis P. Langlotz. 2019. Will artificial intelligence replace radiologists? *Radiology: Artificial Intelligence* 1, 3 (May 2019), e190058. DOI: https://doi.org/10.1148/ryai.2019190058

[73] Shaimaa Lazem, Danilo Giglitto, Makuochi Samuel Nkwo, Hafeni Mthoko, Jessica Upani, and Anicia Peters. 2022. Challenges and paradoxes in decolonising HCI: A critical discussion. *Computer Supported Cooperative Work (CSCW)* 31, 2 (Jun. 2022), 159–196. DOI: https://doi.org/10.1007/s10606-021-09398-0

[74] Min Hun Lee and Chong Jun Chew. 2023. Understanding the effect of counterfactual explanations on trust and reliance on AI for human-AI collaborative clinical decision making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sep. 2023), 1–22. DOI: https://doi.org/10.1145/3610218

[75] Constance D. Lehman, Robert D. Wellman, Diana S. M. Buist, Karla Kerlikowske, Anna N. A. Tosteson, and Diana L. Miglioretti. 2015. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Internal Medicine* 175, 11 (Nov. 2015), 1828. DOI: https://doi.org/10.1001/jamainternmed.2015.5231

[76] Christian Leibig, Moritz Brehmer, Stefan Bunk, Danalyn Byng, Katja Pinker, and Lale Umutlu. 2022. Combining the strengths of radiologists and AI for breast cancer screening: A retrospective analysis. *The Lancet. Digital Health* 4, 7 (Jul. 2022), e507–e519. DOI: https://doi.org/10.1016/S2589-7500(22)00070-X

[77] Dana Li, Lea Marie Pehrson, Carsten Ammitzbøl Lauridsen, Lea Tøttrup, Marco Fraccaro, Desmond Elliott, Hubert Dariusz Zając, Sune Darkner, Jonathan Frederik Carlsen, and Michael Bachmann Nielsen. 2021. The added effect of artificial intelligence on physicians' performance in detecting thoracic pathologies on CT and chest X-ray: A systematic review. *Diagnostics* 11, 12 (Nov. 2021), 2206. DOI: https://doi.org/10.3390/diagnostics11122206

[78] Henry Lieberman, Fabio Paternò, Markus Klann, and Volker Wulf. 2006. End-user development: An emerging paradigm. In *End User Development*. Springer Netherlands, Dordrecht, 1–8. DOI: https://doi.org/10.1007/1-4020-5386-X\{_\}1

[79] Jana Lipkova, Richard J. Chen, Bowen Chen, Ming Y. Lu, Matteo Barbieri, Daniel Shao, Anurag J. Vaidya, Chengkuan Chen, Luoting Zhuang, Drew F. K. Williamson, et al. 2022. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* 40, 10 (Oct. 2022), 1095–1110. DOI: https://doi.org/10.1016/j.ccell.2022.09.012

[80] Gwilym S. Lodwick, Theodore E. Keats, and John P. Dorst. 1963. The coding of Roentgen images for computer analysis as applied to lung cancer. *Radiology* 81, 2 (Aug. 1963), 185–200. DOI: https://doi.org/10.1148/81.2.185

[81] Stina Matthiesen, Søren Zöga Diederichsen, Mikkel Klitzing Hartmann Hansen, Christina Villumsen, Mats Christian Højbjerg Lassen, Peter Karl Jacobsen, Niels Risum, Bo Gregers Winkel, Berit T Philbert, Jesper Hastrup Svendsen, et al. 2021. Clinician preimplementation perspectives of a decision-support tool for the prediction of cardiac arrhythmia based on machine learning: Near-live feasibility and qualitative study. *JMIR Human Factors* 8, 4 (Nov. 2021), e26964. DOI: https://doi.org/10.2196/26964

[82] Mohammad H. Rezazade Mehrizi, Simon H. Gerritsen, Wouter M. de Klerk, Chantal Houtschild, Silke M. H. Dinnessen, Luna Zhao, Rik van Sommeren, and Abby Zerfu. 2023. How do providers of artificial intelligence (AI) solutions propose and legitimize the values of their solutions for supporting diagnostic radiology workflow? A technography study in 2021. *European Radiology* 33, 2 (Feb. 2023), 915–924. DOI: https://doi.org/10.1007/s00330-022-09090-x

[83] Tina W Milbak, Jakob Grue Simonsen, Marco B. Hansen, and Naja L. Holten Møller. 2023. Designing with awareness* building an agenda for worker-and patient well-being. In *Proceedings of the 21st European Conference on Computer-Supported Cooperative Work*. Retrieved from https://dl.eusset.eu/items/aa962d38-1696-4caf-b405-9485861036aa/full

[84] Jesper Molin, Paweł W. Woźniak, Claes Lundström, Darren Treanor, and Morten Fjeld. 2016. Understanding design for automated image analysis in digital pathology. *In Proceedings of the 9th Nordic Conference on Human-Computer Interaction.* ACM, New York, NY, USA, 1–10. DOI: https://doi.org/10.1145/2971485.2971561

[85] Cecily Morrison, Kit Huckvale, Bob Corish, Richard Banks, Martin Grayson, Jonas Dorn, Abigail Sellen, and Sân Lindley. 2018. Visualizing ubiquitously sensed measures of motor ability in multiple sclerosis. *ACM Transactions on Interactive Intelligent Systems* 8, 2 (Jun. 2018), 1–28. DOI: https://doi.org/10.1145/3181670

[86] Siddhartha Mukherjee. 2017. A.I. versus M.D. What happens when diagnosis is automated? Retrieved from https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

[87] Claudia Müller, Cornelius Neufeldt, David Randall, and Volker Wulf. 2012. ICT-development in residential care settings: Sensitizing design to the life circumstances of the residents of a care home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2639–2648. DOI: https://doi.org/10.1145/2207676.2208655

[88] Michael Muller, Christine T. Wolf, Josh Andres, Michael Desmond, Narendra Nath Joshi, Zahra Ashktorab, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Evelyn Duesterwald, et al. 2021. Designing ground truth and the social life of labels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–16. DOI: https://doi.org/10.1145/3411764.3445402

[89] Trine Rask Nielsen, Thomas Gammeltoft-Hansen, and Naja Holten Møller. 2024. Mobile phone data transforming casework in asylum decision-making: Insights from the Danish case. *ACM Journal on Responsible Computing* 1, 4 (2024), 1–29.

[90] John Nolan, Peter McNair, and Jytte Brender. 1991. Factors influencing the transferability of medical decision support systems. *International Journal of Bio-Medical Computing* 27, 1 (1 1991), 7–26. DOI: https://doi.org/10.1016/0020-7101(91)90018-A

[91] Nazmun Nisat Ontika, Sheree May Sassmannshausen, Aparecido Fabiano Pinatti De Carvalho, and Volkmar Pipek. 2023. PAIRADS: Hybrid interaction between humans and AI in radiology. In *Frontiers in Artificial Intelligence and Applications*. Vol. 368. IOS Press, 395–397. DOI: https://doi.org/10.3233/FAIA230108

[92] Evan W. Orenstein, Katherine Yun, Clara Warden, Michael J. Westerhaus, Morgan G. Mirth, Dean Karavite, Blain Mamo, Kavya Sundar, and Jeremy J. Michel. 2019. Development and dissemination of clinical decision support across institutions: Standardization and sharing of refugee health screening modules. *Journal of the American Medical Informatics Association* 26, 12 (Dec. 2019), 1515–1524. DOI: https://doi.org/10.1093/jamia/ocz124

[93] Tariq Osman Andersen, Francisco Nunes, Lauren Wilcox, Elizabeth Kaziunas, Stina Matthiesen, and Farah Magrabi. 2021. Realizing AI in healthcare: Challenges appearing in the wild. In *Proceedings of the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–5. DOI: https://doi.org/10.1145/3411763.3441347

[94] Sun Young Park, Yunan Chen, and Scott Rudkin. 2015. Technological and organizational adaptation of EMR implementation in an emergency department. *ACM Transactions on Computer-Human Interaction* 22, 1 (Mar. 2015), 1–24. DOI: https://doi.org/10.1145/2656213

[95] Sun Young Park, So Young Lee, and Yunan Chen. 2012. The effects of EMR deployment on doctors' work practices: A qualitative study in the emergency department of a teaching hospital. 81, 3 (2011), 204–217. DOI: https://doi.org/10.1016/j.ijmedinf.2011.12.001

[96] Filippo Pesapane, Marina Codari, and Francesco Sardanelli. 2018. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *European Radiology Experimental* 2, 1 (10 2018), 35. DOI: https://doi.org/10.1186/s41747-018-0061-6

[97] Cécile Petitgand, Aude Motulsky, Jean-Louis Denis, and Catherine Régis. 2020. Investigating the barriers to physician adoption of an artificial intelligence-based decision support system in emergency care: An interpretative qualitative study. *Studies in Health Technology and Informatics* 270 (Jun. 2020), 1001–1005. DOI: https://doi.org/10.3233/SHTI200312

[98] Rob Procter, Peter Tolmie, and Mark Rouncefield. 2023. Holding AI to account: Challenges for the delivery of trustworthy AI in healthcare. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–34. DOI: https://doi.org/10.1145/3577009

[99] Dave Randall, Richard Harper, and Mark Rouncefield. 2007. *Fieldwork for Design: Theory and Practice*. Springer London, London. DOI: https://doi.org/10.1007/978-1-84628-768-8

[100] Rebecca Randell, Stephanie Wilson, and Peter Woodward. 2011. Variations and commonalities in processes of collaboration: The need for multi-site workplace studies. *Computer Supported Cooperative Work (CSCW)* 20, 1-2 (Apr. 2011), 37–59. DOI: https://doi.org/10.1007/s10606-010-9127-6

[101] Michael P. Recht, Marc Dewey, Keith Dreyer, Curtis Langlotz, Wiro Niessen, Barbara Prainsack, and John J. Smith. 2020. Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations. *European Radiology* 30, 6 (Jun. 2020), 3576–3584. DOI: https://doi.org/10.1007/s00330-020-06672-5

[102] Madhu Reddy, Lena Mamykina, and Andrea Grimes Parker. 2012. Designing interactive systems in healthcare. *Interactions* 19, 1 (Jan. 2012), 24–27. DOI: https://doi.org/10.1145/2065327.2065334

[103] Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2006. Temporality in medical work: Time also matters. *Computer Supported Cooperative Work (CSCW)* 15, 1 (Feb. 2006), 29–53. DOI: https://doi.org/10.1007/s10606-005-9010-z

[104] Mohammad Hosein Rezazade Mehrizi, Peter van Ooijen, and Milou Homan. 2021. Applications of artificial intelligence (AI) in diagnostic radiology: a technography study. *European Radiology* 31, 4 (Apr. 2021), 1805–1811. DOI: https://doi.org/10.1007/s00330-020-07230-9

[105] Abi Rimmer. 2017. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ* 359 (Oct. 2017), j4683. DOI: https://doi.org/10.1136/bmj.j4683

[106] Michael Roberts, Derek Driggs, Matthew Thorpe, Julian Gilbey, Michael Yeung, Stephan Ursprung, Angelica I. Aviles-Rivero, Christian Etmann, Cathal McCague, Lucian Beer, et al. 2021. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence* 3, 3 (Mar. 2021), 199–217. DOI: https://doi.org/10.1038/s42256-021-00307-0

[107] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. 2021. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. DOI: https://doi.org/10.1145/3411764.3445748

[108] Alejandro Rodriguez-Ruiz, Kristina Lång, Albert Gubern-Merida, Jonas Teuwen, Mireille Broeders, Gisella Gennaro, Paola Clauser, Thomas H. Helbich, Margarita Chevalier, Thomas Mertelmeier, et al. 2019. Can we reduce the workload of mammographic screening by automatic identification of normal exams with artificial intelligence? A feasibility study. *European Radiology* 29, 9 (Sep. 2019), 4825–4832. DOI: https://doi.org/10.1007/s00330-019-06186-9

[109] Nithya SambasivanShivani KapaniaHannah Highfll, Diana AkrongPraveen ParitoshLora Aroyo, Nithya Sambasivan, Shivani Kapania, Hannah Highfll, Diana Akrong, Praveen Paritosh, and Lora Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data cascades in high-stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15. DOI: https://doi.org/10.1145/3411764.3445518

[110] Sahil Sandhu, Anthony L. Lin, Nathan Brajer, Jessica Sperling, William Ratliff, Armando D. Bedoya, Suresh Balu, Cara O'Brien, and Mark P. Sendak. 2020. Integrating a machine learning system into clinical workflows: Qualitative study. *Journal of Medical Internet Research* 22, 11 (Nov. 2020), e22421. DOI: https://doi.org/10.2196/22421

[111] Sheree May Saßmannshausen, Nazmun Nisat Ontika, Aparecido Fabiano Pinatti De Carvalho, Mark Rouncefield, and Volkmar Pipek. 2024. Amplifying human capabilities in prostate cancer diagnosis: An empirical study of current practices and AI potentials in radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–20. DOI: https://doi.org/10.1145/3613904.3642362

[112] Shier Nee Saw and Kwan Hoong Ng. 2022. Current challenges of implementing artificial intelligence in medical imaging. *Physica Medica* 100 (Aug. 2022), 12–17. DOI: https://doi.org/10.1016/J.EJMP.2022.06.003

[113] Kjeld Schmidt. 2000. The critical role of workplace studies in CSCW. *Workplace Studies* (May 2000), 141–149. DOI: https://doi.org/10.1017/CBO9780511628122.007

[114] Jarrel C. Y. Seah, Cyril H. M. Tang, Quinlan D. Buchlak, Xavier G. Holt, Jeffrey B. Wardman, Anuar Aimoldin, Nazanin Esmaili, Hassan Ahmad, Hung Pham, John F. Lambert, et al. 2021. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: A retrospective, multireader multicase study. *The Lancet Digital Health* 3, 8 (Aug. 2021), e496–e506. DOI: https://doi.org/10.1016/S2589-7500(21)00106-0

[115] Kate Sellen, Dominic Furniss, Yunan Chen, Svetlena Taneva, Aisling Ann O'Kane, and Ann Blandford. 2014. Workshop abstract: HCI research in healthcare: Using theory from evidence to practice. In *Proceedings of the 2014 CHI Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 87–90. DOI: https://doi.org/10.1145/2559206.2559240

[116] Mark Sendak, William Ratliff, Dina Sarro, Elizabeth Alderton, Joseph Futoma, Michael Gao, Marshall Nichols, Mike Revoir, Faraz Yashar, Corinne Miller, et al. 2020. Real-world integration of a sepsis deep learning technology into routine clinical care: Implementation study. *JMIR Medical Informatics* 8, 7 (Jul. 2020), 1–16. DOI: https://doi.org/10.2196/15182

[117] Nigam H. Shah, Arnold Milstein, and Steven C. Bagley. 2019. Making machine learning models clinically useful. *JAMA* 322, 14 (Oct. 2019), 1351. DOI: https://doi.org/10.1001/jama.2019.10306

[118] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (Apr. 2020), 495–504. DOI: https://doi.org/10.1080/10447318.2020.1741118

[119] Ben Shneiderman. 2020. Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction* 12, 3 (Sep. 2020), 109–124. DOI: https://doi.org/10.17705/1thci.00131

[120] Venkatesh Sivaraman, Leigh A. Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. 2023. Ignore, trust, or negotiate: Understanding clinician acceptance of AI-based treatment recommendations in health care. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. DOI: https://doi.org/10.1145/3544548.3581075
[121] Jonathon E. Stewart, Frank J. Rybicki, and Girish Dwivedi. 2020. Medical specialties involved in artificial intelligence research: Is there a leader. *Tasman Medical Journal* 2, 1 (Feb. 2020), 20–27. Retrieved from https://tasmanmedicaljournal.com/2020/02/medical-specialties-involved-in-artificial-intelligence-research-is-there-a-leader/
[122] Lea Strohm, Charisma Hehakaya, Erik R. Ranschaert, Wouter P. C. Boon, and Ellen H. M. Moors. 2020. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors. *European Radiology* 30, 10 (Oct. 2020), 5525–5532. DOI: https://doi.org/10.1007/s00330-020-06946-y
[123] Zhaoyuan Su. 2022. "What is Your Envisioned Future?": Toward human-AI enrichment in data work of asthma care. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28. DOI: https://doi.org/10.1145/3555157
[124] Lucy Suchman. 1987. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press. DOI: https://doi.org/10.11225/JCSS.6.191
[125] Lucy A. Suchman. 1983. Office procedure as practical action: Models of work and system design. *ACM Transactions on Information Systems* 1, 4 (Oct. 1983), 320–328. DOI: https://doi.org/10.1145/357442.357445
[126] Lucille Alice Suchman. 2006. *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, 314. Retrieved from https://cds.cern.ch/record/1991545
[127] Amara Tariq, Saptarshi Purkayastha, Geetha Priya Padmanaban, Elizabeth Krupinski, Hari Trivedi, Imon Banerjee, and Judy Wawira Gichoya. 2020. Current clinical applications of artificial intelligence in radiology and their best supporting evidence. *Journal of the American College of Radiology* 17, 11 (Nov. 2020), 1371–1381. DOI: https://doi.org/10.1016/j.jacr.2020.08.018
[128] Angelique Taylor, Michele Murakami, Soyon Kim, Ryan Chu, and Laurel D. Riek. 2022. Hospitals of the future: Designing interactive robotic systems for resilient emergency departments. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–40. DOI: https://doi.org/10.1145/3555543
[129] Sian Taylor-Phillips and Karoline Freeman. 2022. Artificial intelligence to complement rather than replace radiologists in breast screening. *The Lancet Digital Health* 4, 7 (Jul. 2022), e478–e479. DOI: https://doi.org/10.1016/S2589-7500(22)00094-2
[130] Anja Thieme, Danielle Belgrave, and Gavin Doherty. 2020. Machine learning in mental health. *ACM Transactions on Computer-Human Interaction* 27, 5 (Oct. 2020), 1–53. DOI: https://doi.org/10.1145/3398069
[131] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–50. DOI: https://doi.org/10.1145/3564752
[132] Kicky G. van Leeuwen, Maarten de Rooij, Steven Schalekamp, Bram van Ginneken, and Matthieu J. C. M. Rutten. 2022. How does artificial intelligence in radiology improve efficiency and health outcomes? *Pediatric Radiology* 52, 11 (Oct. 2022), 2087–2093. DOI: https://doi.org/10.1007/s00247-021-05114-8
[133] Kicky G. van Leeuwen, Steven Schalekamp, Matthieu J. C. M. Rutten, Bram van Ginneken, and Maarten de Rooij. 2021. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology* 31, 6 (Jun. 2021), 3797–3804. DOI: https://doi.org/10.1007/s00330-021-07892-z
[134] Himanshu Verma, Jakub Mlynar, Roger Schaer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquoz, and Adrien Depeursinge. 2023. Rethinking the role of AI with physicians in oncology: Revealing perspectives from clinical and research workflows. In *Proceedings of the 2023 19th CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–19. DOI: https://doi.org/10.1145/3544548.3581506
[135] Dakuo Wang, Liuping Wang, and Zhan Zhang. 2021. Brilliant AI doctor in rural clinics: Challenges in AI-powered clinical decision support system deployment. In *Proceedings of the Conference on Human Factors in Computing Systems*. ACM, 1–18. DOI: https://doi.org/10.1145/3411764.3445432
[136] Pu Wang, Xiao Xiao, Jeremy R. Glissen Brown, Tyler M. Berzin, Mengtian Tu, Fei Xiong, Xiao Hu, Peixi Liu, Yan Song, Di Zhang, et al. 2018. Development and validation of a deep-learning algorithm for the detection of polyps during colonoscopy. *Nature Biomedical Engineering* 2, 10 (Oct. 2018), 741–748. DOI: https://doi.org/10.1038/s41551-018-0301-3
[137] Shanshan Wang, Guohua Cao, Yan Wang, Shu Liao, Qian Wang, Jun Shi, Cheng Li, and Dinggang Shen. 2021. Review and prospect: Artificial intelligence in advanced medical imaging. *Frontiers in Radiology* 1 (Dec. 2021), 781868. DOI: https://doi.org/10.3389/FRADI.2021.781868
[138] Thomas Weikert, Joshy Cyriac, Shan Yang, Ivan Nesic, Victor Parmar, and Bram Stieltjes. 2020. A practical guide to artificial intelligence-based image analysis in radiology. *Investigative Radiology* 55, 1 (Jan. 2020), 1–7. DOI: https://doi.org/10.1097/RLI.0000000000000600

[139] Batia Mishan Wiesenfeld, Yin Aphinyanaphongs, and Oded Nov. 2022. AI model transferability in healthcare: A sociotechnical perspective. *Nature Machine Intelligence* 4, 10 (Oct. 2022), 807–809. DOI: https://doi.org/10.1038/s42256-022-00544-x

[140] Christine Wolf and Jeanette Blomberg. 2019. Evaluating the promise of human-algorithm collaborations in everyday work practices. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–23. DOI: https://doi.org/10.1145/3359245

[141] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang 'Anthony' Chen. 2020. CheXplain: Enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In *Proceedings of the 2020 20th CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. DOI: https://doi.org/10.1145/3313831.3376807

[142] Wei Xu. 2019. Toward human-centered AI: A perspective from human-computer interaction. *Interactions* 26, 4 (Jun. 2019), 42–46. DOI: https://doi.org/10.1145/3328485

[143] Ling Yang, Ioana Cezara Ene, Reza Arabi Belaghi, David Koff, Nina Stein, and Pasqualina Lina Santaguida. 2022. Stakeholders' perspectives on the future of artificial intelligence in radiology: A scoping review. *European Radiology* 32, 3 (Mar. 2022), 1477–1495. DOI: https://doi.org/10.1007/s00330-021-08214-z

[144] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. *Proceedings of the Conference on Human Factors in Computing Systems*, 1–13. DOI: https://doi.org/10.1145/3313831.3376301

[145] Qian Yang, John Zimmerman, Aaron Steinfeld, Lisa Carey, and James F. Antaki. 2016. Investigating the heart pump implant decision process: Opportunities for decision support tools to help. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 4477–4488. DOI: https://doi.org/10.1145/2858036.2858373

[146] Nur Yildirim, Changhoon Oh, Deniz Sayar, Kayla Brand, Supritha Challa, Violet Turri, Nina Crosby Walton, Anna Elise Wong, Jodi Forlizzi, James McCann, et al. 2023. Creating design resources to scaffold the ideation of AI concepts. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. ACM, New York, NY, USA, 2326–2346. DOI: https://doi.org/10.1145/3563657.3596058

[147] Nur Yildirim, Hannah Richardson, Maria Teodora Wetscherek, Junaid Bajwa, Joseph Jacob, Mark Ames Pinnock, Stephen Harris, Daniel Coelho De Castro, Shruthi Bannur, Stephanie Hyland, et al. 2024. Multimodal healthcare AI: Identifying and designing clinically relevant vision-language applications for radiology. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–22. DOI: https://doi.org/10.1145/3613904.3642013

[148] Nur Yildirim, Susanna Zlotnikov, Deniz Sayar, Jeremy M. Kahn, Leigh A. Bukowski, Sher Shah Amin, Kathryn A. Riman, Billie S. Davis, John S. Minturn, Andrew J. King, et al. 2024. Sketching AI concepts with capabilities and examples: AI innovation in the intensive care unit. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18.

[149] Feiyang Yu, Mark Endo, Rayan Krishnan, Ian Pan, Andy Tsai, Eduardo Pontes Reis, Eduardo Kaiser Ururahy Nunes Fonseca, Henrique Min Ho Lee, Zahra Shakeri Hossein Abad, Andrew Y. Ng, et al. 2023. Evaluating progress in automatic chest X-ray radiology report generation. *Patterns* 4, 9 (Sep. 2023), 100802. DOI: https://doi.org/10.1016/j.patter.2023.100802

[150] Hubert D. Zając. 2022. Designing ground truth for machine learning—Conceptualisation of a collaborative design process between medical professionals and data scientists. In *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. DOI: https://doi.org/10.48340/ecscw2022\{_\}p04

[151] Hubert D. Zaja̧ c, Natalia R. Avlona, Finn Kensing, Tariq O. Andersen, and Irina Shklovski. 2023. Ground truth or dare: Factors affecting the creation of medical datasets for training AI. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 351–362. DOI: https://doi.org/10.1145/3600211.3604766

[152] Hubert D. Zaja̧ c, Dana Li, Xiang Dai, Jonathan F. Carlsen, Finn Kensing, and Tariq O. Andersen. 2023. Clinician-facing AI in the wild: Taking stock of the sociotechnical challenges and opportunities for HCI. *ACM Transactions on Computer-Human Interaction* 30, 2 (Apr. 2023), 1–39. DOI: https://doi.org/10.1145/3582430

[153] Hubert Dariusz Zaja̧ c, Jorge Miguel Neves Ribeiro, Silvia Ingala, Simona Gentile, Ruth Wanjohi, Samuel Nguku Gitau, Jonathan Frederik Carlsen, Michael Bachmann Nielsen, and Tariq Osman Andersen. 2024. "It depends": Configuring AI to improve clinical usefulness across contexts. In *Proceedings of the Designing Interactive Systems Conference*. ACM, New York, NY, USA, 874–889. DOI: https://doi.org/10.1145/3643834.3660707

[154] Nan-ning Zheng, Zi-yi Liu, Peng-ju Ren, Yong-qiang Ma, Shi-tao Chen, Si-yu Yu, Jian-ru Xue, Ba-dong Chen, and Fei-yue Wang. 2017. Hybrid-augmented intelligence: Collaboration and cognition. *Frontiers of Information Technology and Electronic Engineering* 18, 2 (2 2017), 153–179. DOI: https://doi.org/10.1631/FITEE.1700053

[155] Xiaomu Zhou, Mark S. Ackerman, and Kai Zheng. 2009. I just don't know why it's gone: maintaining informal information use in inpatient care. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2061–2070. DOI: https://doi.org/10.1145/1518701.1519014

[156] Xiaomu Zhou, Mark S. Ackerman, and Kai Zheng. 2010. Doctors and psychosocial information: Records and reuse in inpatient care. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1767–1776. DOI: https://doi.org/10.1145/1753326.1753592

Received 29 February 2024; revised 1 August 2024; accepted 28 December 2024