Research Article

Sheree May Plinz-Saßmannshausen*, Nazmun Nisat Ontika, Kieu Tran, Krzysztof Ukleja, Aparecido Fabiano Pinatti de Carvalho and Markus Rohde

Designing a human-centered AI solution with radiologists to improve prostate cancer diagnosis

https://doi.org/10.1515/icom-2025-0043
Received September 30, 2025; accepted November 14, 2025; published online December 19, 2025

Abstract: Prostate cancer is a significant health concern with rising incidence rates in Germany. While recent research highlights the potential of artificial intelligence (AI) to support radiologists by increasing diagnostic accuracy and efficiency, the successful integration of such technology depends on carefully designed, practice-oriented approaches. This paper presents a design case study (DCS) on a human-centered artificial intelligence (HCAI) solution to assist radiologists in diagnosing prostate cancer. In our contribution, we emphasize that by combining human-centered design (HCD) and practice-centered computing, we can contribute towards a seamless and usable integration of AI into healthcare. Through interdisciplinary collaboration among radiologists, user experience (UX) designers, AI developers, and human-computer interaction (HCI) researchers, we employed Design Thinking (DT) to iteratively develop low-, mid-, and high-fidelity prototypes of an AI solution for improving diagnostic workflow, which were evaluated by radiologists. Our findings advance the state-of-the-art in computing support for prostate cancer diagnosis by introducing an artifact designed with both HCD and practice-centered computing in mind.

Keywords: practice-centered design; artificial intelligence; human-centered AI; design thinking; prototyping; healthcare

1 Introduction

Prostate cancer is a prevalent and significant health concern in Germany, accounting for a considerable number of new cases and associated mortality. In 2022 alone, 74,895 new cases of prostate cancer were registered, leading to 15,196 deaths in 2023.[1] This malignancy predominantly affects elderly men, with a median diagnosis age of 71, and rarely occurs before 50. [1,2] Early detection and risk stratification are key elements in effectively managing prostate cancer, as identifying the disease at an early stage allows for timely and appropriate interventions that can improve patient outcomes. In contrast, delayed or missed diagnosis can worsen prognosis and require aggressive treatments.[3] Conversely, some prostate tumors remain indolent, and overdiagnosis may cause unnecessary side effects.[4]

In the diagnosis process for prostate cancer, various clinicians are involved. Among these, the radiologist plays a crucial role in the process by utilizing imaging techniques such as Magnetic Resonance Imaging (MRI) to diagnose, stage, and monitor the disease, which guides further treatment planning. However, the increasing number of cases exposes the limitation of available radiologists in Germany, as only 10,139 doctors are radiologists (2.32 % of the total),[5] making this scarcity a crucial obstacle in prostate cancer diagnosis.

Given the ongoing advancements in artificial intelligence (AI) research, incorporating AI-driven solutions into

Sheree May Plinz-Saßmannshausen and Nazmun Nisat Ontika contributed equally to this research.

***Corresponding author: Sheree May Plinz-Saßmannshausen**, Institute of Information Systems, University of Siegen, Siegen, Germany, E-mail: sheree.sassmannshausen@uni-siegen.de. https://orcid.org/0000-0002-9006-3907
Nazmun Nisat Ontika, Kieu Tran, Krzysztof Ukleja and Markus Rohde, Institute of Information Systems, University of Siegen, Siegen, Germany, E-mail: nazmun.ontika@uni-siegen.de (N. N. Ontika), kieu.tran@uni-siegen.de (K. Tran), uklejak@yahoo.de (K. Ukleja), markus.rohde@uni-siegen.de (M. Rohde)
Aparecido Fabiano Pinatti de Carvalho, Department of Informatics, University of Oslo, Oslo, Norway, E-mail: fabiano.pinatti@ifi.uio.no

medical workflows is becoming increasingly relevant. AI has made significant advancements in various medical fields, including radiology, where its application in medical imaging has shown promising potential.[6] Its ability to assist radiologists in managing the exponential surge in medical data volume and complexity while enhancing diagnostic accuracy makes this technology highly appealing. This assistance to healthcare professionals becomes critical in countries with proven shortages of radiologists. In Germany, for instance, where 46 % of hospitals face radiology staffing problems,[7] the gap between the number of practicing radiologists and the growing demand for radiology services emphasizes the need for additional tools to bridge this divide.

Radiology is expected to evolve in parallel with the capabilities and potential of AI.[8,9] At the same time, most radiologists anticipate the widespread adoption of AI in their field over the next decade, which is expected to significantly change it in the process.[10,11] Radiologists recognize AI as a valuable tool for improving diagnostic accuracy and efficiency, addressing staffing shortages, and ultimately enhancing patient care.[8,9,12–14]

AI applications for prostate cancer are rapidly growing to the point where AI may serve as decision support, decreasing inter-observer variability.[15,16] Though AI models have shown promising results, only a few AI-driven systems have transitioned from controlled laboratory environments to real-world clinical workflows.[17] Acknowledging the importance of the human experience within its contextual environment, a broader spectrum of factors relevant to the success of any technical solution has been emphasized over the years.[18–20] Nevertheless, the sociotechnical dimensions have been considerably underappreciated,[21] particularly at the intersection of AI and healthcare,[22,23] where the focus lies on technical performance and clinical outcomes. Not accounting for these aspects has been demonstrated to factor into the non-adoption of technology in healthcare.[24,25]

While AI excels in rapidly processing vast datasets and identifying nuances evading the human eye, the radiologist's strength lies in domain expertise and the ability to identify nearly optimal solutions.[6] Hence, the expected role of AI in radiology is not to replace the radiologist from the workflow but rather to serve as a second reader and optimize the workflow[10] by assisting the radiologists. Thus, involving radiologists in AI development to define clinical requirements and evaluate AI tools can help maintain expertise while avoiding over-reliance on technology.[26] Furthermore, engaging radiologists in the design phase and presenting them with transparent and comprehensible visual representations of the system's decisions aims to provide a "fair and responsible perception of the system-human decisions".[25,27]

Previous research has demonstrated how ethnographically informed approaches can be useful in understanding radiologist practices and uncovering user requirements for tools to support them with their duties.[28] However, little can be found in the literature about how this can be done during the design phase. Given the prominence of Design Thinking (DT) within Human-Centered Design (HCD),[29–31] especially for health interventions to improve outcomes, product quality, service adoption, and user satisfaction,[32–34] we have drawn on the approach for the design of an AI-based aid to support radiologists in prostate cancer diagnosis. In addition, we have used a practice-centered computing lens,[35,36] which acknowledges that only by understanding practices can we effectively design computer technologies to support people in their endeavors.

The aim of our research is, therefore, to introduce a Human-Centered AI (HCAI) solution for radiologists by embracing a human-centered design approach combined with practice-centered computing through involving radiologists and other stakeholders such as AI developers, UX designers, and HCI researchers to benefit from their various perspectives throughout the design process to ensure the AI solution was thoughtfully designed to fully harness its potential in the healthcare system to provide better improvements and user satisfaction.[13,37,38] In this paper, we present our Design Case Study (DCS) to contribute a human-centered AI solution to support radiologists in their current practice within the diagnosis of prostate cancer using MRI data.

From our previous ethnographically informed context study,[28] we gained insights into radiologists' practices and how these practices can inform the design of AI support, proposing a solution with the potential to improve their workflow by seamlessly integrating AI into radiology practices. Building on these insights, this paper contributes: 1) the design of low-, mid-, and high-fidelity prototypes through an interdisciplinary and iterative process spanning both a HCD and a practice-centered computing perspective; 2) the initial evaluation of the referred HCAI solution with radiologists, with reflections on the extent to which it helped us to achieve a solution that would respond to the sociotechnical aspects concerning the diagnosis of prostate cancer by radiologist; and finally, 3) deep insights and practical guidance on how the interdisciplinary capabilities of different stakeholders can be effectively utilized in collaborative settings to generate design ideas and how to address user

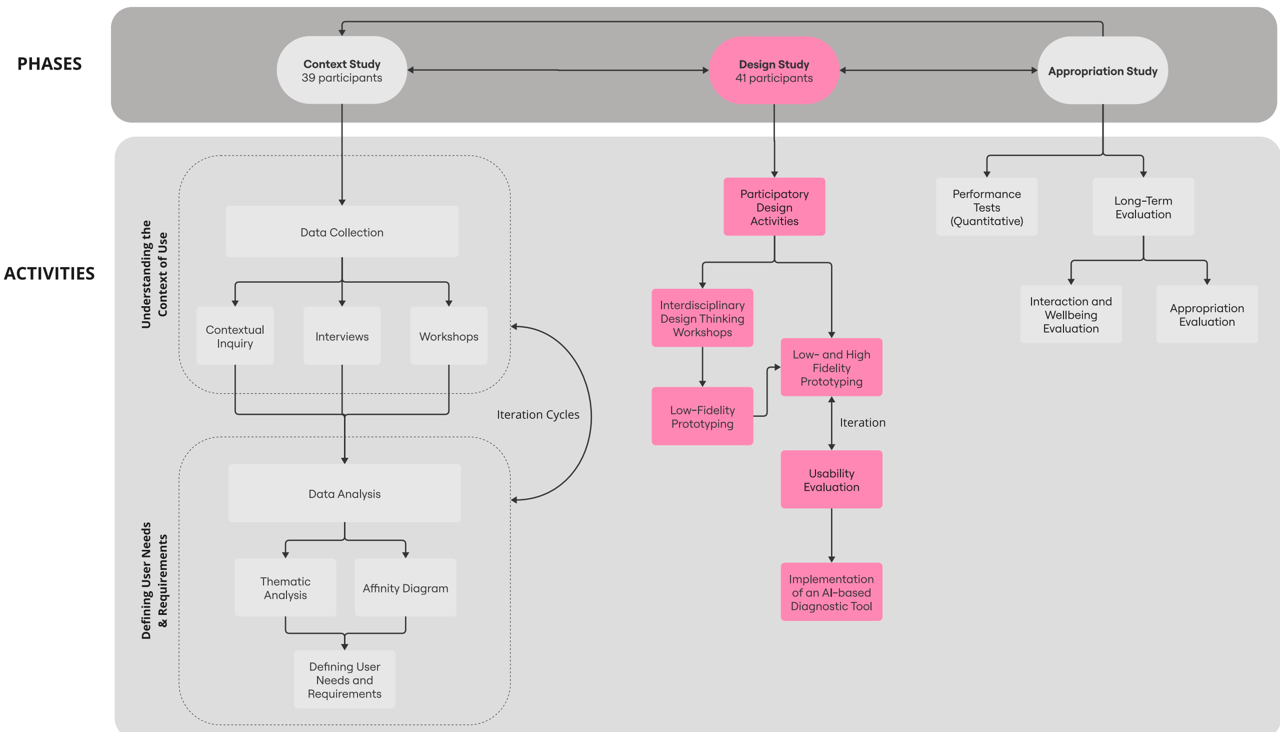


Figure 1: Our design case study framework, including context study, design study, and appropriation study.

needs, design preferences, and interactions with the AI system within the radiologists' diagnostic workflow.

2 Research background

The work presented in this paper was conducted within an interdisciplinary project aimed at supporting research on interactive technologies for health and quality of life. Between August 2021 and May 2024, we collaborated closely with a software development company in telemedicine and multiple associated partners in the field of radiology. Despite challenges arising from the COVID-19 pandemic, the involvement of the German Radiological Society[1] enabled us to generate considerable interest among radiologists, ultimately facilitating close collaboration with 13 professionals in the field.

We defined our work within the DCS framework introduced by Wulf and colleagues,[36,39] which conceptualizes practice-centered computing initiatives through three interrelated activities: context study, design study, and appropriation study.[35] According to Stevens et al., "good design case studies are characterized by an in-depth understanding and descriptions of social practices in the field of application, with a specific focus on how these empirical findings may contribute to the design of IT artifacts".[35]

To provide such an in-depth understanding of current practices, we also divided our work into three phases, depicted in Figure 1, which outlines the individual phases of our DCS, the associated activities, and their interrelations. The context study provides an in-depth understanding of the practice, which forms the foundation; building on that, the design study translates these insights into design concepts and prototypes, and the appropriation study examines how these artifacts are integrated into real-world practices. The three phases – context, design, and appropriation – were embedded in an iterative process, ensuring the ongoing emergence of insights and iterative refinements.

The design study, which is presented in this paper (highlighted in pink in Figure 1), is based on our context study, which we will briefly summarize in this section. The contextual empirical fieldwork and the analysis of the data fall out of the scope of this paper, as we are focusing here on the design study, and the main parts of the context study can already be read in our previous paper.[28] The appropriation study will be conducted as future work and therefore lies beyond the scope of this paper, too.

1 German: Die Deutsche Röntgengesellschaft, Website: https://www.drg.de.

2.1 Context study

We conducted qualitative research by working with different radiology centers across Germany for 2.5 years. We had 11 contextual inquiry sessions,[40] involving seven radiologists and other medical staff members, each lasting between three and five hours. Furthermore, we conducted 11 semi-structured interviews with radiologists, with a duration ranging from 40 to 80 minutes. Additionally, two workshops were organized: first, an ethics workshop with seven participants (excluding the first two authors) from the fields of ethics, radiology, HCI, and AI development, and second, an international workshop on Human-Centered AI in Healthcare held at the ECSCW conference in 2022, which involved 10 participants (excluding the first three authors) with backgrounds in HCI/CSCW, healthcare, and AI development.[41]

Through the context study, we gathered in-depth insights into the daily practices of radiologists.[42] We learned about individual steps within their workflows, contextual factors influencing the practice, usage of various artifacts and systems, internal and external communications, practitioners' needs, practitioners' challenges, and their perspectives towards AI as a potential assistant. This in-depth analysis aimed to establish a solid foundation of understanding and to define requirements, serving as the basis of the development of our HCAI solution. Part of our insights has been published in the proceedings of the CHI conference in 2024,[28] where we discussed a section of our iterative empirical study on radiologist practices in diagnosing prostate cancer and how AI can enhance decision-making in that context. While additional empirical insights emerged through our iterative work, which are not yet published, we will not discuss them in detail in this paper, as it focuses on the design study. But to comprehend our design decisions, we will briefly present key insights from the whole context study, including an overview of current prostate cancer diagnosis procedures using MRI, the user needs, and requirements for an AI tool to support radiologists in the diagnostic process.

2.1.1 Current practices in diagnosing prostate cancer

We categorized the complex diagnosis process into five stages: Patient inquiry, Image acquisition, Image interpretation, Reporting, and Verification, as shown in Figure 12 (in appendix).

Before the examination, radiologists gather relevant medical history, including Prostate Specific Antigen (PSA) value from the patients. Next, through an MRI exam, they acquire the Multiparametric MRI (mpMRI), which combines multiple imaging sequences including T2-Weighted (T2W), Diffusion-Weighted Imaging (DWI), and Dynamic Contrast-Enhanced (DCE) imaging to evaluate the prostate gland, with each sequence highlighting different areas of anatomy and prostate structure segmentation. An example of the most common sequences we have observed can be seen in Figure 11 (in the appendix).

Afterwards, the radiologist analyzes MRI imaging data to detect and assess lesions, considering that not every lesion is characterizable in every image, necessitating a comprehensive review of images from various perspectives. Prostate MRI interpretation is complex, with radiologists using the standardized scoring system Prostate Imaging Reporting and Data System (PI-RADS), which assigns scores from one to five based on the lesion's assessment from key mpMRI sequences to indicate the likelihood of clinical significance.[43] However, the sequence contributing to the score varies depending on the prostate zone in which the lesion is located. These zones encompass the Peripheral Zone (PZ), Transitional Zone (TZ), Central Zone (CZ), and Anterior Fibromuscular Stroma (AFS). After assigning scores to each lesion identified in the mpMRI images, the overall PI-RADS score is determined, equivalent to the score of the highest-rated lesion.

The findings from the diagnosis are documented in a report together with the patient's metadata and the previously collected data about the medical history in the fourth phase. The final diagnosis also entails the PI-RADS score with required diagnostic metrics, a standardized prostate sector map with identified lesions marked, and recommendations for subsequent actions.

The optional fifth phase involves occasional communication with other doctors, primarily for verification purposes in complex cases, but routine feedback is seldom received in daily radiological practice.[44]

2.1.2 Identified user needs

During our observations, we identified several manual tasks performed by radiologists that might be addressed through AI. For example, they manually measure the prostate's length, width, and height from MRI and calculate prostate volume using a private calculator, subsequently manually calculate PSA density by dividing the PSA level by the volume. Additionally, report writing involves copy-paste and a dictation system that requires frequent manual corrections due to limited accuracy. Radiologists also use pre-printed prostate sector maps to manually mark lesion locations, further adding to their workload.

Determining individual prostate zones is important for assessing lesion severity, but both zone segmentation and lesion identification can be challenging. Although PI-RADS is a standardized scoring system, radiologists criticized it as it gives the impression of being straightforward. The accuracy of prostate volume calculation can vary among radiologists due to manual measurement, which can potentially lead to issues, as PI-RADS imposes stringent limits for critical values. PI-RADS also suggests having a double analysis policy for prostate exams, improving anomaly detection and diagnostic accuracy by having two radiologists assess each case. However, we observed that most of the clinics don't have the resources, e.g., staff and time to follow that approach, and each case is diagnosed by one radiologist (except in complex cases).

2.2 Requirements of the AI-tool

The insights and challenges we have identified through our context study are transformed into user needs and requirements. According to the often-mentioned user needs, we have prioritized the most relevant requirements:

1. Identification of the prostate location
2. Segmentation of distinct prostate regions
3. Detection and localization of potential lesions
4. Classification of identified lesions following the PI-RADS scheme
5. Calculation of prostate volume and PSA density
6. Standardization of the reporting system
7. Giving feedback about AI results

Later, in explaining the prototyping process in Section 5, we will refer to the requirements numeration, which doesn't imply a priority.

3 Related work

In this section, we explore the existing literature that informs our research, engaging ourselves in the field of AI in healthcare, specifically in radiology. We look at how AI developments have improved diagnosis accuracy and efficiency. We also explore the significance of maintaining humans at the heart of AI development lifespans through HCD.[45,46]

3.1 AI and decision-support systems in radiology and prostate cancer diagnosis

AI is being used more and more for a variety of activities, but decision support is the primary application it serves.[47] AI-driven decision aids are increasingly adopted in hospitals and clinics to support medical practitioners in analyzing medical data and interpreting results more accurately.[48] Clinical Decision Support (CDS)/Clinical Decision Support Systems (CDSS) are developed to offer timely, evidence-based recommendations for diagnosis and treatment.[49–51] With advanced AI AI-based Medical Diagnosis Support Systems (AIMDSS) are growing into a significant component in pathology and radiology,[52] while Computer-Aided Detection (CAD)/Computer-Aided Intelligent Diagnosis (CAID) shifted from just detecting suspicious regions to also interpreting and diagnosing those regions.[49,53,54]

AI has the potential to revolutionize radiology practice by offering valuable support to radiologists in various aspects and steps of their workflow.[25,55,56] There exists a variety of studies on the use of AI for prostate cancer diagnosis. As the prostate exhibits high variability in shape and appearance, Convolutional Neural Network (CNN) can better cope with these issues, gaining popularity for segmentation.[57] Many examples[58–65] demonstrate CNN applications in medical imaging for prostate cancer detection and segmentation, with a primary focus on utilizing MRI images for image classification, analysis, and segmentation.

Litjens et al., developed an automated CAD for prostate cancer using MRI, comprising steps such as prostate segmentation, feature extraction, and candidate classification. Their evaluation showed superior performance, suggesting potential usefulness for radiologists in both first- and second-reader settings.[66] To understand physicians' views on AI, Buck et al.[67] interviewed German physicians and medical professionals closely tied to radiology. They discovered that while specialists acknowledged AI risks, they tended to double-check its recommendations, leading to extra effort. One participant expressed frustration when AI missed details, requiring thorough reviews and negating potential time-saving benefits. Penzkofer et al., aimed to identify prerequisites for the successful implementation of clinically relevant AI in prostate MRI diagnosis.[68] They noted limited adoption of AI in clinical practice despite numerous studies exploring its diagnostic potential. Regarding usability, they suggested tailoring user interfaces to radiologists' workflows for comprehensive MRI evaluation alongside AI results.

Studies on the use of AI for prostate cancer diagnostics primarily focus on the technical perspective, describing the development of AI algorithms. The HCI perspective, which involves understanding user needs and desired interactions with the AI, is often overlooked. Our work addresses this gap by emphasizing strong collaboration with radiologists.

3.2 Human-centered design in healthcare

The integration of AI algorithms has encountered challenges in real-world applications when not considering insights and recommendations arising from the field of HCI.[49] Users who lack an understanding of the system's functionality and capabilities become overwhelmed by the system's output, have limited situational awareness, and may experience a loss of control over the system.[69] Procter et al., also highlighted that AI systems in healthcare often lack an understanding of the organizational context, limiting practitioner trust in their recommendations.[70] However, integrating an HCAI assistant into the radiology workflow enhances result precision, visually simplifies analysis, and significantly reduces assessment time, thus boosting workflow efficiency for radiologists.[71]

Following an HCD process, the current context of use can be fully analyzed to define requirements for the new AI tool, ensuring effective, efficient, and satisfactory user utilization. Particularly in the field of healthcare, where sensitive data is processed, considering human factors and putting humans at the center of an AI design process is the key to good usability and users' acceptance of the system,[72] especially when it comes to designing an eXplainable AI (XAI),[47,73] emphasizing the need for interfaces to offer algorithmic decision descriptions, multi-layer rationalization, and data origin information.[74,75] This supports increased transparency and trust between users and the AI system, which are found to be decisive in the adoption of such systems.[76] The successful deployment of AI technologies hinges on their seamless integration into existing clinical workflows and infrastructure, preserving tasks and processes unaltered.[77–79] Understanding the current context of use is crucial for ensuring positive adoption.[77,80]

Socio-technical, participatory methods have been recommended by an increasing amount of research to effectively involve domain experts throughout the AI development lifecycle.[81–83] Ooge et al.[84] emphasize the vital role of domain experts and end-users in developing visual analytics tools, aligning with recommendations from HCI scholars. Health professionals' feedback underscores the efficacy of participatory and user-centered interaction design methods, particularly in UI drawings, facilitating requirement description and common understanding with system developers.[85] Increased user involvement in design is crucial, with user input identified as pivotal in the success or failure of complex technology.[86] Furthermore, researchers from the community have argued that engaging users in the design process and especially in design decisions, according to the participatory design premises,[87,88] can contribute toward more ethical, adaptable, and useful AI systems.[89]

Usability is crucial for integrating medical software into workflow and adoption.[90] HCD methodologies facilitate achieving usability in medical technology.[91] DT, known for its empathy, creativity, and collaboration, is effective across healthcare domains, offering user-centered solutions.[31] Its success in healthcare stems from considering contextual factors, including user needs and clinical evidence.[92] DT outcomes in healthcare outperform traditional interventions in terms of usability and effectiveness.[93] Given the importance of usefulness and ease of use in medical technology adoption,[94] DT presents a promising approach. Additionally, Chen et al., emphasize a human-centered approach in DT for explainable medical imaging systems.[30]

4 Methodological approach of the design process

Following the design case study framework,[35] we employed an iterative HCD process and adopted a practice-centered computing approach[36] to develop an AI prototype aimed at supporting prostate cancer diagnosis. Our research project underwent ethical review by the institutional review board committee of our university and received clearance.

For this design study, we paid strong attention to the participation of different stakeholders, including radiologists, AI developers, HCI researchers, and UX design professionals, throughout the entire process. Across two DT workshops with a total of 31 participants, we collaboratively explored and refined design ideas that resulted in the development of nine prototypes of varying fidelity (low-, mid-, and high-fidelity). We had 10 evaluation sessions with radiologists to assess the usability of the Medium-Fidelity (Mid-Fi)- and High-Fidelity (Hi-Fi) prototypes. The detailed methodological approach is explained below.

Table 1 presents an overview of the participants in this design study. The irregular numbering reflects that certain pseudonyms (P01, P03, P10, P11) had been assigned in the

Table 1: List of participants in our design study.

Participant	Role	Involvement
P07	Radiologist	DT Workshop
P08	Radiologist	DT Workshop
P12	Radiologist	DT Workshop
P13	AI Developer	DT Workshop
P14	AI Developer	DT Workshop
P15	AI Developer	DT Workshop
P16	UX Professional	DT Workshop
P17	UX Professional	DT Workshop
P18	HCI Researcher	DT Workshop
P19	HCI Researcher	DT Workshop
P20	HCI Researcher	DT Workshop
P21	HCI Master Student	DT Workshop
(…)[a]	HCI Master Student	DT Workshop
P40	HCI Master Student	DT Workshop
P02	Radiologist	Evaluation
P04	Radiologist	Evaluation
P05	Radiologist	Evaluation
P06	Radiologist	Evaluation
P07	Radiologist	Evaluation
P08	Radiologist	Evaluation
P09	Radiologist	Evaluation
P41	Radiologist	Evaluation

[a] P22 to P39 not shown individually for conciseness.

context study, but those participants did not take part in the design study due to their time constraints. Notably, seven of the radiologists (P02, P04, P05, P06, P07, P08, P09) participated in both our context and design studies, while two of them (P06, P07) evaluated both the Mid-Fi and Hi-Fi prototypes.

4.1 Design thinking workshops and low-fidelity prototypes

We organized two dynamic and collaborative DT workshops to leverage interdisciplinary collaboration, fostering creative problem-solving and stimulating out-of-the-box thinking, particularly beneficial due to the time constraints faced by the radiologists and their widespread distribution. Since there are multiple models available illustrating the process of DT,[95,96] we have adapted the DT process according to our available resources, such as time, place, and participants (see Subsection 7.1). Our goal was to generate innovative ideas for an AI-based diagnostic tool tailored for prostate cancer diagnosis and explore visual concepts for different features.

The first workshop was conducted online to integrate participants from different locations. The session lasted for 150 min, bringing together a diverse group of 11 participants considering innovations in healthcare come from diverse user types.[31] The participants included three radiologists experienced in prostate cancer diagnosis, three developers with AI experience, two UX professionals, and three HCI researchers (including one HCI research assistant). For recruiting, we used social media platforms such as LinkedIn[2] and Meetup[3] to advertise the workshop as an event and to explicitly motivate AI developers, radiologists, UX designers, and HCI enthusiasts to participate. While social media effectively reached most roles, radiologists were specifically invited through direct contact. Some radiologists expressed interest in our research project after learning about it through the German Radiological Society, allowing us to extend explicit invitations to the workshop. Most of the participants did not know each other before the workshop. We used Webex[4] to meet online and Miro[5] to gather our insights. We divided the participants according to their roles into two multidisciplinary break-out rooms to work on the tasks together and facilitate diverse discussions.

We structured our workshop into five phases, namely Empathize, Define, Ideation, Prototyping, and Evaluation,[95] to collaboratively engage with our participants. First, we explained the context of use and described the key findings from our context study, which set the focus for the workshop. Then we requested insights from the radiologists in each group to discuss challenges and limitations in their current practices regarding our focus and share their expectations about an AI-driven tool with the group members, so that the participants can gain an empathetic understanding of the user needs. Our previously identified user needs and requirements were confirmed by the radiologists through these steps in the workshop.

We delved into understanding the users' perspectives using an empathy map. Building on that, we crafted actionable problem statements as Point-of-View (POV) statements to help us define specific needs and core challenges of our users. Afterward, we defined How-Might-We (HMW) statements as open-ended questions to generate multiple ideas focusing on desired outcomes. Then we created sketches for rapid ideation by identifying innovative solutions to the problem statements. After finalizing the best parts of the sketches using Dot Voting, we focused on prototyping

2 www.linkedin.com.
3 www.meetup.com.
4 www.webex.com.
5 www.miro.com.

by producing low-fidelity versions of the tool. Participants were instructed to visually map out the interface or key functionalities of the prototype based on the MRIs we provided using the Wireframe tool on Miro. We evaluated the prototypes made by each group with a feedback session with the whole group. Though we discussed together after each of the phases, the final discussion ensured a comprehensive exploration of ideas and the refinement of solutions.

The second workshop was done in person and spanned 90 min, bringing together 20 master's students of HCI from our university. This approach aimed to leverage HCI and design perspectives during the creative phase, potentially fostering even more innovative ideas. The students were all part of an HCI course, so they knew each other. The insights gained from the context study and the feedback received during the first workshop served as the foundation for developing further design ideas. We didn't show them concrete design ideas from the first workshop to avoid bias. We divided the participants into four groups consisting of five students, where at least one of them in each group has professional experience in UX design. Owing to scheduling restrictions, we scaled down the scope of our in-person workshop and concentrated mostly on the ideation and prototyping stages with the participants. An information document detailing the scope of our study was sent to them in advance, and to build empathy with stakeholders, we reaffirmed the context of use and the main findings at the beginning of the workshop. We provided four POVs, which were generated in the previous workshop, to streamline the ideation process, ensuring a focused exploration of specific problem areas within the limited time frame. Each group was given one unique POV, which was used by the participants to create multiple HMWs.

Following that, each participant brainstormed solutions for the HMWs they had chosen, utilizing the Crazy 8 method. Based on the dot voting on the best design parts, participants advanced to the prototyping step by using the provided materials to build four paper prototypes, one for each group.

4.2 Hi-Fi prototypes and evaluation

The six prototypes created by the participants at the end of the two DT workshops, two from the first and four from the second, were not intended as finished solutions; instead, they served as a foundation that our research team, involving HCI researchers and AI developers, further merged and refined through brainstorming sessions while considering participant feedback, translating the design ideas into a clickable Low-Fidelity (Lo-Fi) prototype on Figma.[6]

Derived from that, we developed a Hi-Fi prototype, which was evaluated with five radiologists. Afterwards, we returned to a Mid-Fi Figma prototype to focus on iterative refinement of core features of the solution based on the initial feedback without being constrained by implementation details. This allowed us to rapidly integrate insights from the first evaluation and test design alternatives more flexibly. The Mid-Fi prototype was also evaluated with another five radiologists.

We have conducted initial evaluations through usability tests as a pre-stage to the appropriation study. The evaluation aimed to assess the usability of the AI-based solution by examining user needs and identifying potential flaws. Eight of the usability tests were conducted individually via Zoom,[7] allowing them to participate remotely from their work environment. Three participants (P05, P06, P07) remotely completed approximately 90-min sessions each, accessing the Hi-Fi prototype hosted on a university server through a standard internet browser, along with anonymized MRI images for importing into their Digital Imaging and Communications in Medicine (DICOM) viewer. One 30-min session with two participants (P02, P41) was conducted on-site at a radiology center, where the radiologists accessed the same prototype directly on the researcher's laptop. Additionally, five sessions of approximately 45 min were conducted remotely, where participants (P04, P06, P07, P08, P09) accessed the Mid-Fi prototype through a shared Figma project.

While the remote participants shared their screens and live video during their exploration, participants on-site were directly observed. All participants interacted directly with the prototype and had to simulate a prostate MRI diagnosis using the prototype and provided MRI images, following a specific scenario. They were encouraged to vocalize their thoughts throughout the process using the think-aloud method.[97] To capture these thoughts, all sessions were recorded, transcribed, and analyzed using a thematic analysis approach[98] to identify key topics, issues, and design implications raised by participants. The participants' feedback was helpful to refine the prototypes iteratively. Some design ideas that emerged through the evaluation were already addressed within iterative design cycles; other potential features not implemented in the prototype were discussed. Participants shared insights on the benefits and drawbacks of the features. Familiarity with the prototype

6 https://www.figma.com.

7 https://zoom.us.

likely facilitated participants' ability to envision the planned features as potential additions to the software solution.

5 The HCAI prototypes for the diagnostic tool

In this section, we present the creative ideation and design phase, where solution concepts are generated and visualized in an iterative approach to address user needs and requirements.

5.1 Low-fidelity prototypes

The Lo-Fi prototypes were generated from our workshops and through internal sketching efforts within the research team.[99] The interdisciplinary collaboration in our study fostered a rich exchange of ideas, ranging from user-centric interface designs to interactive features for the prototypes aimed at tackling the identified problems, while our focus remained on exploring how a potential solution could best meet user needs and on uncovering crucial factors that must be considered[100] when designing for actual work practices.

5.1.1 Online design thinking workshop

During the workshop, we expanded upon the user needs and requirements that we gathered and created POV statements to aid in specifying the particular requirements, and employed HMW statements to discover several concepts centered around the intended results. For example, while working on the requirement *"segmentation of distinct prostate regions"* (reference to requirement #2), we crafted the POV statement *"The radiologist needs the segmentation of the individual zones of the prostate for the biopsies because it is time-consuming to annotate them by hand."* which led to the HMW statement of *"How might we provide the segmentation of the individual zones of the prostate (central gland, peripheral zone, transitional zone, seminal vesicle) for the biopsies to save time?"* which was addressed through the prototype that was designed during the workshop.

Through collaborative efforts within two interdisciplinary groups, we successfully crafted two Lo-Fi prototypes. Subsequently, we fine-tuned certain elements of these prototypes, guided by the feedback received during the workshop. To enhance the comprehensiveness of our designs, we incorporated sticky notes containing some HMW statements, insights, and ideas, addressing aspects that couldn't be accommodated within the prototypes due to time constraints. This process ensures that the final prototypes not only reflect the diverse perspectives within the groups but also integrate valuable insights for a more refined outcome.

Figure 2 presents the prototype created by group one during the workshop. Their primary focus was on consolidating crucial information into a single frame, minimizing the need for extensive interaction with various screens and elements to reduce distractions and time. According to the requirements list (see Subsection 2.2 to check the requirements list), the participants demonstrated the visualization of AI-detected zones, lesions, and segmentation through contours (reference to requirements #1, #2, #3). Emphasizing the importance of displaying key information about lesions and the prostate, they showcased overlays. A table was designed to provide an overview of all AI-identified lesions, including the PI-RADS score (requirement #4), with options for result verification and explanations of the score. Also, a detailed view featured extensive information about lesions and patients, offering an option for editing. Participants also streamlined the reporting process, incorporating relevant images directly into the report through checkboxes activated by the "Generate Report" button (requirement #6). However, time constraints hindered the completion of a graphical representation of lesion progress.

In the second prototype presented in Figure 3 made by group two, the radiologist in the group identified comparing AI-based pre-diagnoses with original MRI images as an important user need, which was then focused on. The radiologists' design idea emphasizes a parallel order of the AI results ("annotated images") and the original MRI images ("original images"). These images should be linked to each other and change synchronously while scrolling through the mouse without additional interaction elements. Also, the detection of the prostate borders as well as the segmentation of different zones and lesions were visualized through contours or a heatmap (requirements #1, #2, #3). Participants incorporated an option for corrections of the markings. The right side of the interface shows space for reporting (requirement #6). In a table format, the AI should provide automatic calculations of the prostate and lesions volume (requirement #5). It is intended to ensure the classification of identified lesions following the PI-RADS scheme and to automatically calculate the PI-RADS score (requirement #4). Other HMW statements, which couldn't be addressed in the prototype, refer to a Prostate Sector Map according to PI-RADS, which should be provided and generated automatically, and show the individual lesions, their weighting zone, and the PI-RADS score.

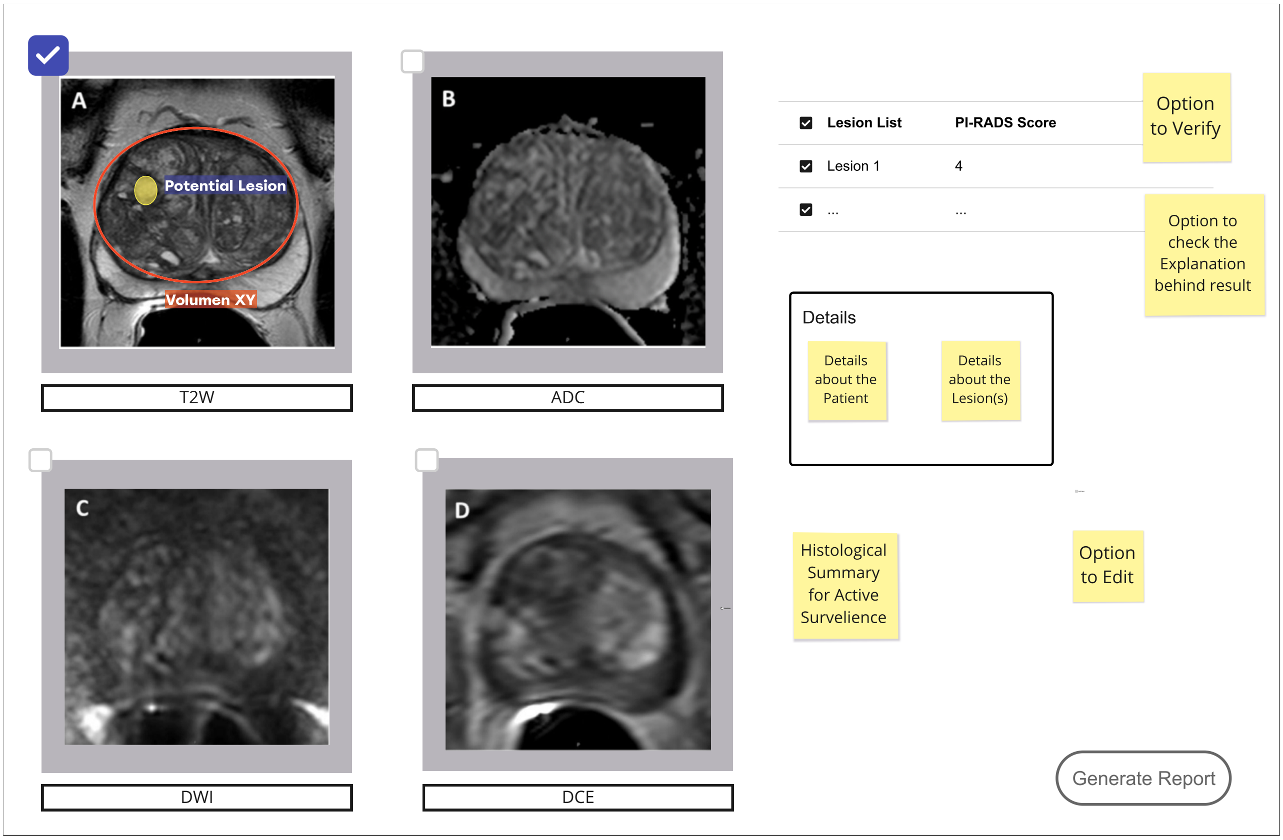


Figure 2: Low-fidelity prototype by Group 1 from online design thinking workshop.

5.1.2 On-site design thinking workshop

A second DT workshop was conducted on-site with 20 participants divided into four groups. Consequently, four paper prototypes were crafted as seen in Figure 4, reflecting the collaborative efforts of the groups and demonstrating the convergence of diverse perspectives into cohesive design solutions for a specific problem.

Group 1 concentrated on showcasing the automatic identification and segmentation of distinct prostate regions using contours, offering a 3D view, and providing an option for radiologist verification (requirements #1 and #2). Group 2 focused on visualizing the automatic detection and localization of potential lesions with classification, emphasizing its potential role in identifying cancer severity among patients and prioritizing cases beforehand (requirements #3 and #4). Group 3 envisioned the automatic calculation of prostate volume and PSA density, emphasizing the importance of allowing radiologists to adjust AI-detected borders for accuracy (requirement #5). Group 4 demonstrated a standardized reporting system that not only provided feedback on AI results but also included a Likert scale option for radiologists to indicate the severity of a misdiagnosis by AI (requirements #6 and #7).

5.1.3 Low-fidelity prototype using figma

Following the prior prototypes and feedback, our research team used quick visual sketching in Figma to support the creative ideation process,[99] which allowed for rapid idea externalization, including the use of images like MRI scans (Figure 5), and proved pivotal in creating a well-informed, user-centered design.

Since our main target users are German radiologists, the UI text for this Lo-Fi and subsequent prototypes was in German to ensure a realistic and familiar user experience. All of the seven Lo-Fi prototypes were discussed and refined within the research team and used for the next iteration of prototyping to create a Hi-Fi prototype.

5.2 High-fidelity prototype

The primary goal in translating the Lo-Fi prototypes into a Hi-Fi prototype was to create a user experience that closely resembled interacting with a finished product. To enhance

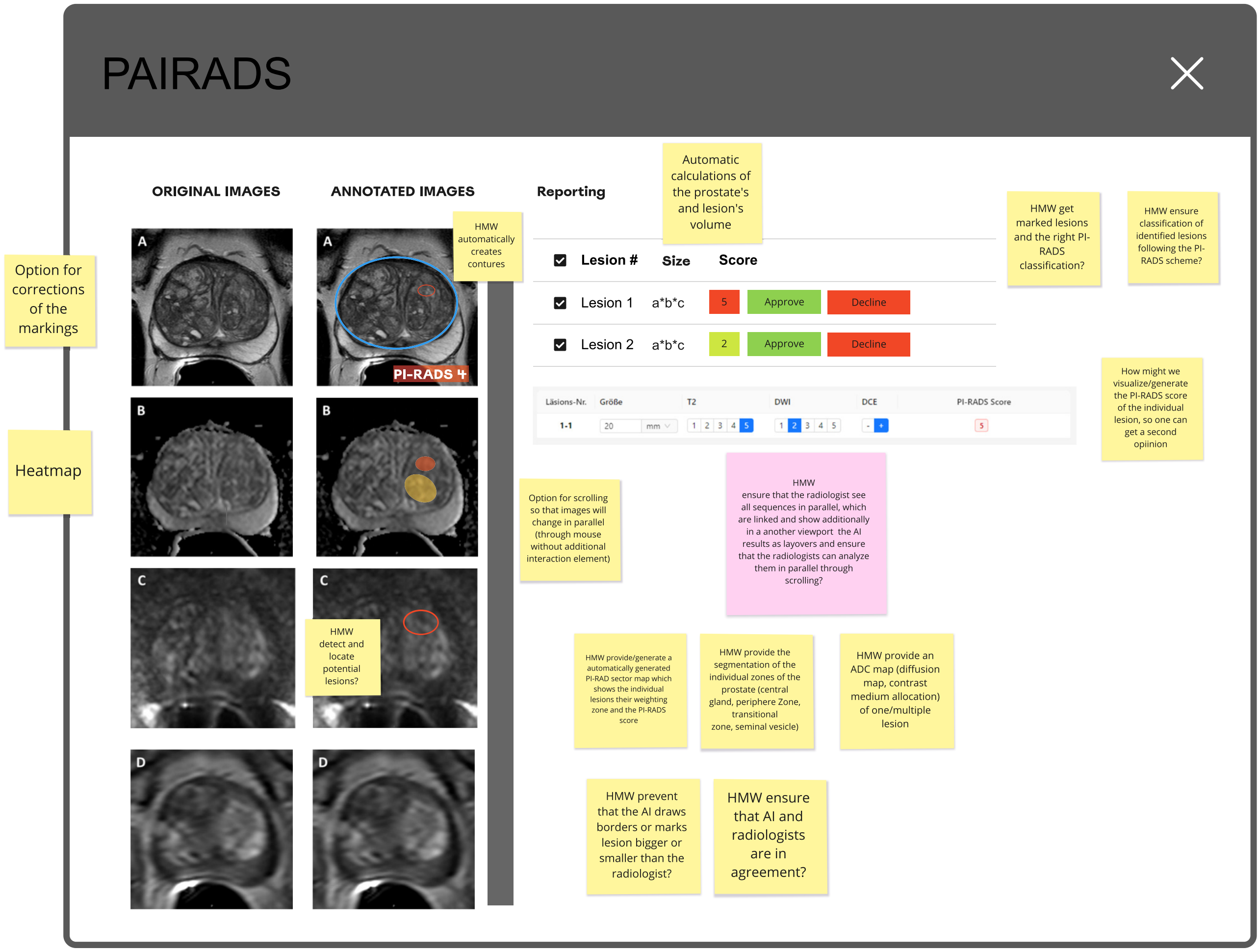


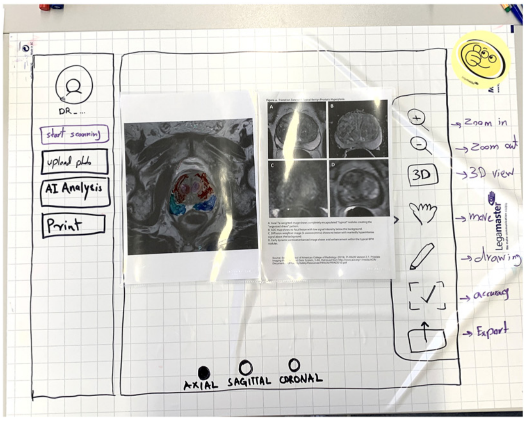
Figure 3: Low-fidelity prototype by Group 2 from online design thinking workshop.

the realism of the prototype interaction, real MRI examination data and actual AI-generated output were utilized and integrated into the Hi-Fi prototype, which was implemented as a front-end web application using AngularJS. The AI developers of our team were concurrently implementing and training the AI algorithm. They used the Semi-Supervised Learning (SSL) technique to train our AI model, since it produces more accurate outcomes by learning from both labeled and unlabeled data,[102,103] and has the potential for medical image segmentation.[104] The technical details of the implementation and training process of the AI algorithm are beyond the scope of this paper. Although the algorithm was not directly linked to the back-end of the Hi-Fi prototype, the AI results were visibly presented within the UI using Adobe Photoshop.[8]

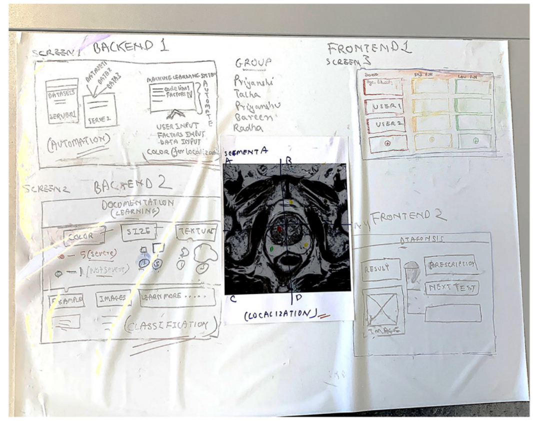
8 https://www.adobe.com/products/photoshop.html.

Once the AI algorithms have processed the MRI images, the results are made available via a web-based front-end application, allowing users to examine each image in greater detail. The AI primarily focuses on detecting clinically significant lesions and accurately delineating both the whole prostate and the lesions, which are presented to the user in a simplified DICOM viewer. Additionally, the solution generates a structured and informative graphic to be included in the radiological report.

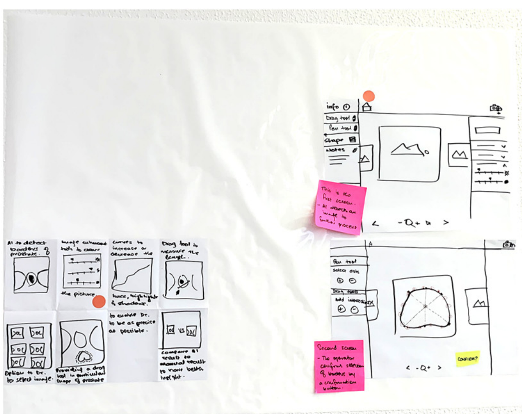
Here, we will provide a detailed description of our UI and highlight key features. At first, radiologists start to select an MRI exam from an overview page listing all exams (scans) processed with patient details, exam date (newest first), and the AI's processing status. The details page, as shown in Figure 6, is split into a left side for textual data and a right-side viewer for graphical data, such as the MRI images and lesion graphics. This simultaneous display allows users to easily make connections and comparisons,



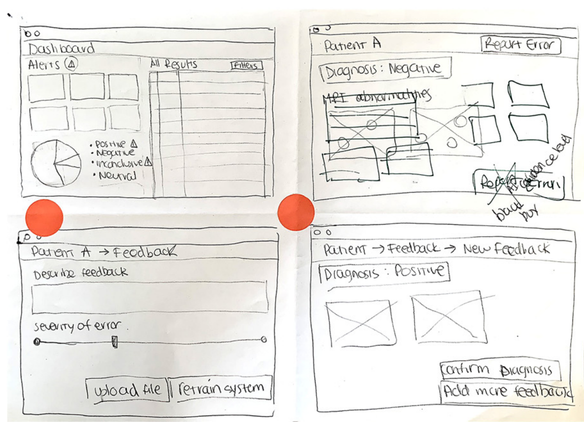
(a) Paper Prototype of Group 1



(b) Paper Prototype of Group 2



(c) Paper Prototype of Group 3



(d) Paper Prototype of Group 4

Figure 4: Low-fidelity paper prototypes from Group 1 to Group 4 from on-site design thinking workshop. (a) Paper prototype of Group 1. (b) Paper prototype of Group 2. (c) Paper prototype of Group 3. (d) Paper prototype of Group 4.

which in turn reduces cognitive load by eliminating manual view switching.

The viewer includes the MRI series relevant to PI-RADS scoring. Users can navigate between series via top buttons, the scroll bar, or direct mouse-wheel scrolling, a common interaction element in the DICOM viewer and explicitly identified as a user need during the workshop. The viewer also displays AI-generated delineations. Through mask options, the prostate segmentation, PZ, and any lesion detected are presented and classified as significant by the AI algorithm (Figure 7). Activated masks remain visible throughout the image stack during scrolling. For orientation, the scroll bar thumbnails were color-coded according to active masks, with distinct colors differentiating organ/zone from lesion delineations.

Data transfer features are included (copy to clipboard, download, or export to Picture Archiving and Communication System (PACS)), enabling users to view edited series in their DICOM viewer or integrate them directly into reports. The viewer can display the AI-generated structured graphic (Figure 8), reducing input errors, increasing efficiency, and ensuring consistent outputs transferable to final reports.

In Figure 9, the inputs are structured into sections aligned with the diagnostic workflow observed in our context study, reducing cognitive effort by maintaining familiar procedures. Checkboxes in each section support process documentation and help radiologists resume work after interruptions. The first section addresses diagnostic image quality, which is pre-filled by AI according to the concept. The prostate measurements section shows PSA, prostate volume, and PSA density, automatically calculated from AI segmentation to improve efficiency and reduce errors. The option to validate the values is available by assessing the right-side images. The lesion-finding section lists AI-detected lesions, which remain editable by the user. While the PI-RADS score is automatically derived from registered values, users can input a manual score for

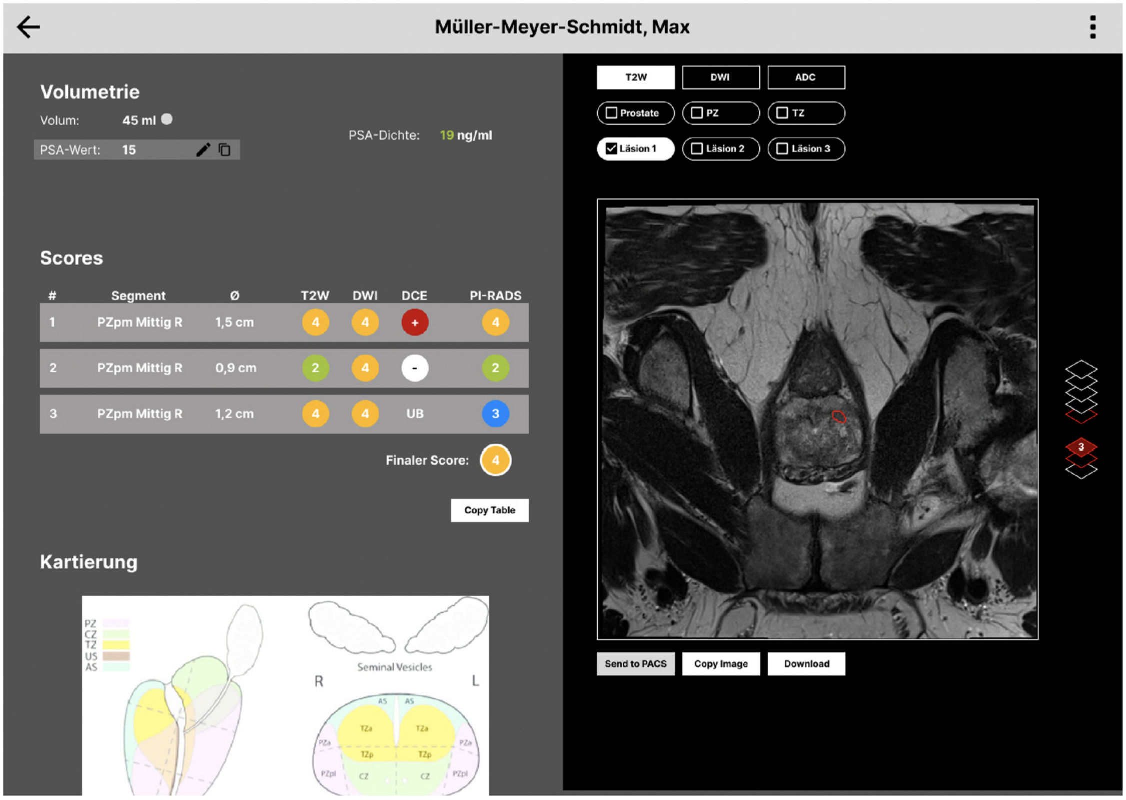
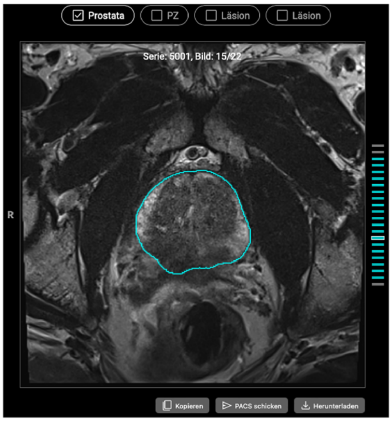


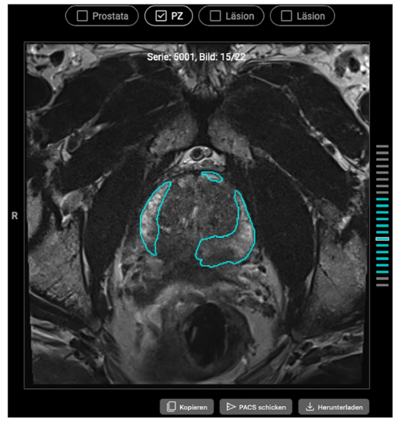
Figure 5: Low-fidelity prototype: figma sketch of diagnostic tool UI with the PI-RADS sector map[101] and a T2W image.

Figure 6: High-fidelity prototype: detailed UI with examination-details page of a patient.

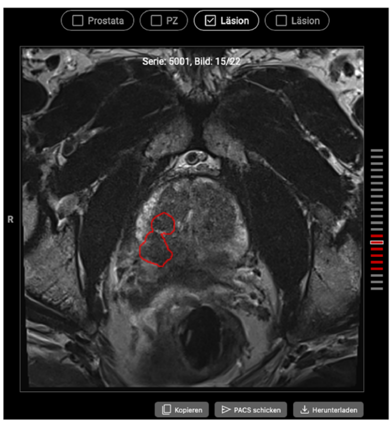
ambiguous cases. The lesion-finding section allows checking corresponding delineations, adding new lesions, or hiding registered ones to manage graphics and dismiss AI suggestions. The documentation section covers the local and external cancer spreading and additional findings, using structured fields (oriented on the PI-RADS v2.1 benign



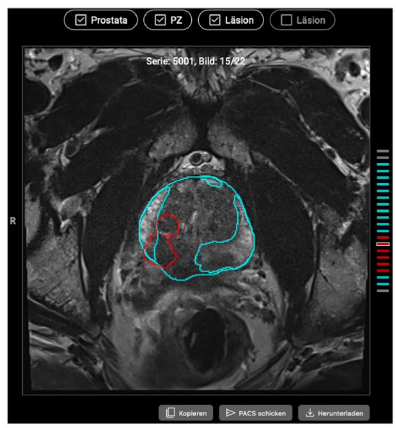
(a) Prostate



(b) Peripheral Zone (PZ)



(c) Lesion



(d) Prostate, PZ, and Lesion

Figure 7: High-fidelity prototype: viewer options with four different kinds of segmentation masks. (a) Prostate. (b) Peripheral zone (PZ). (c) Lesion. (d) Prostate, PZ, and Lesion.

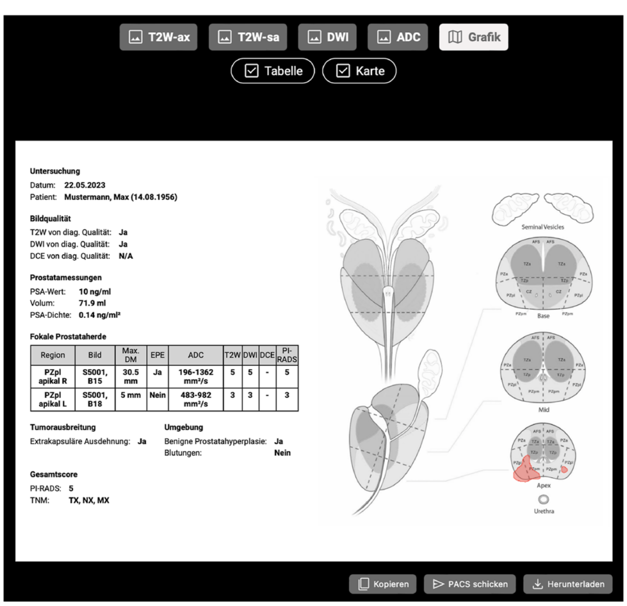


Figure 8: High-fidelity prototype: AI-generated sector map with structured table.

findings list[43]). The final part registers the overall assessment score. Each section is structured as a table where the current exam is a row.

5.3 Middle-fidelity prototype

Within our iterative process, multiple design and evaluation cycles were conducted. After evaluating the Hi-Fi version, a Mid-Fi prototype was created in Figma (Figure 10).

The Mid-Fi prototype aims to reduce perceptual and cognitive load by enabling radiologists to make quick, pre-defined decisions on AI-suggested abnormalities. This reflects the radiologists' request to mark easily categorizable lesions using their expertise (P09). The AI suggestions can validate decisions or be overridden in case of disagreement, serving purely as supportive augmentation. Confidence bars promote diagnostic caution, and an on/off switch enables AI use either as pre-diagnostic guidance or as a confirmatory tool.

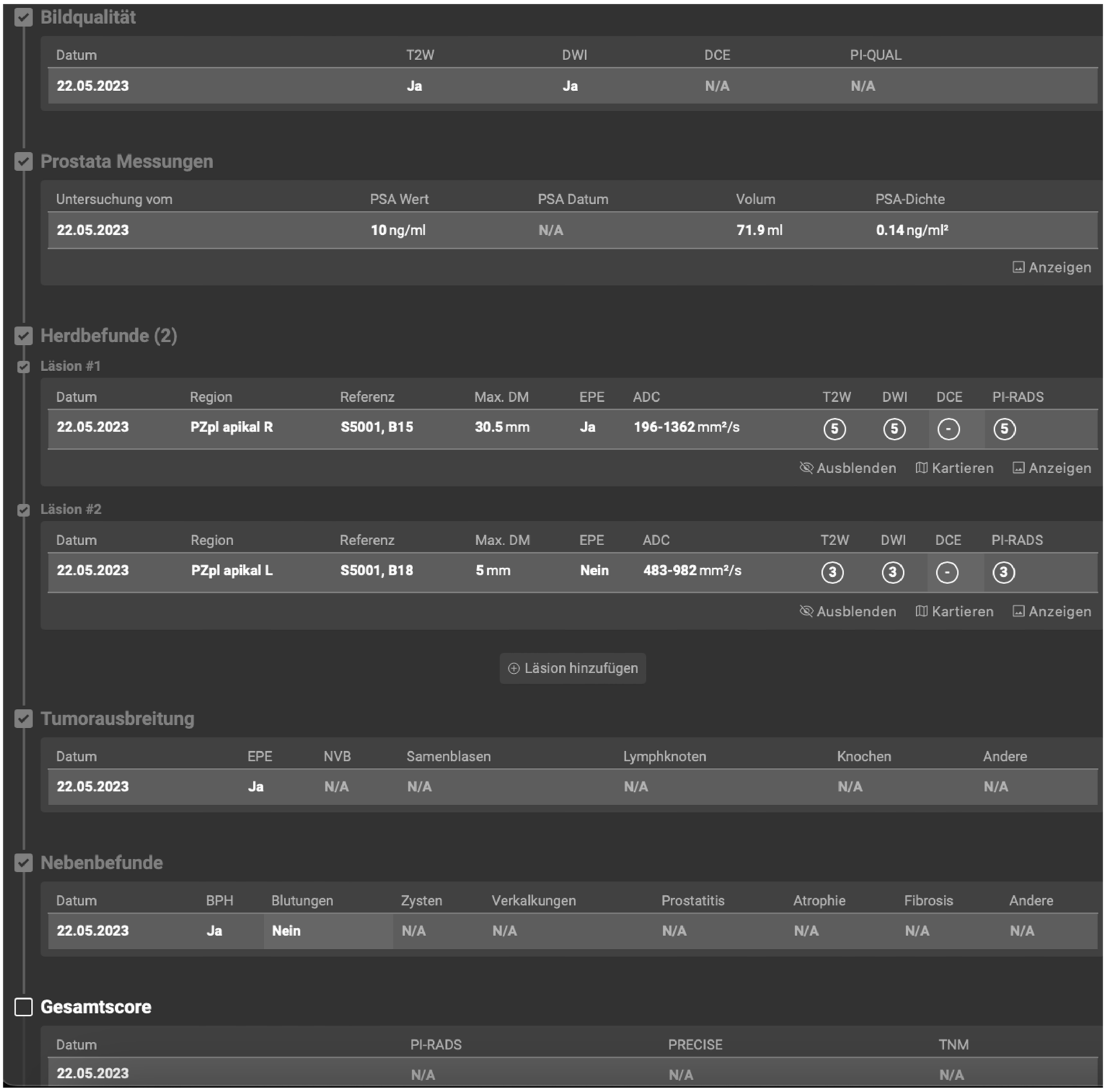


Figure 9: High-fidelity prototype: detailed UI with complete input form.

6 Initial evaluation of the HCAI prototypes

The evaluation of the Hi-Fi and Mid-Fi prototypes provided valuable insights into integrating AI-based solutions into prostate MRI diagnosis. Alongside supporting the design concepts' efficacy in meeting user needs, as reflected in overall positive feedback, the evaluation also highlighted key considerations essential for developing AI solutions in this domain.

For both prototypes, greater transparency about how AI works appeared to influence trust in the results. P09 emphasized that, if the system works reliably, it could substantially reduce workload and save time, as current tasks such as reviewing hundreds of images, manual reporting, and volume calculation are highly time-consuming.

Moreover, the evaluation highlighted the importance of supporting a smooth and efficient workflow as radiologists reported visual strain toward the end of the workday. Therefore, features such as quickly jumping to the respective lesion delineation from the input form were appreciated. However, it was also emphasized that the solution's effectiveness hinges on the performance of the AI algorithm. P09 explains that the AI system is particularly effective at detecting small lesions that he might have overlooked, empowering him to complete a more accurate diagnosis.

Regarding the workflow order, most radiologists prefer to briefly review MRI images themselves before activating the AI. But also parallel viewing – with and without AI – for

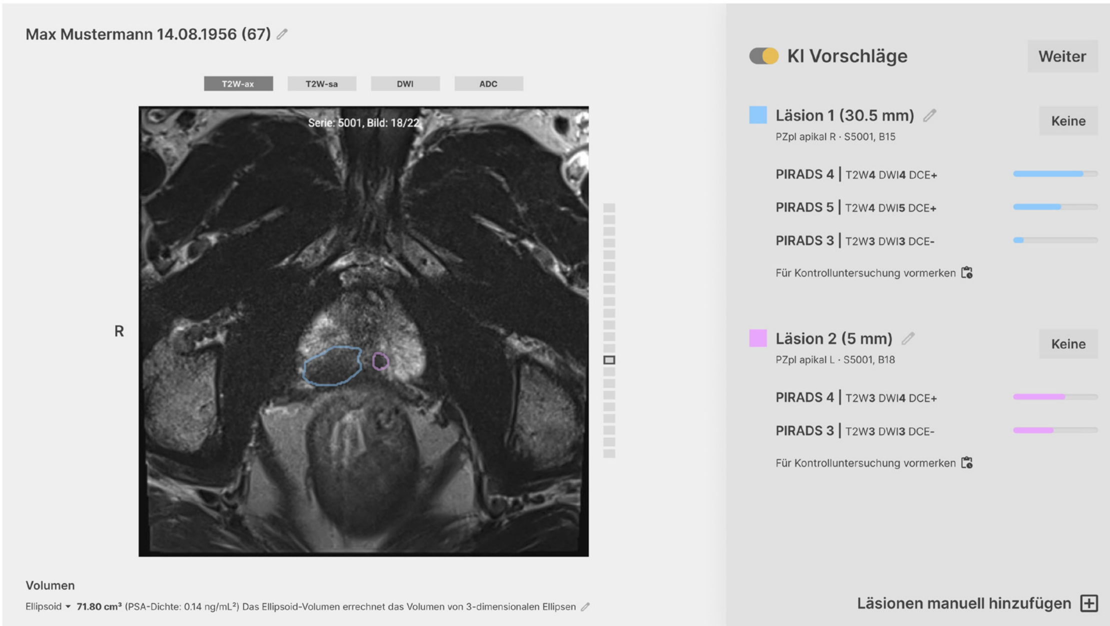


Figure 10: Middle-fidelity prototype after iteration: diagnostic tool UI with a more streamlined design.

direct comparison was mentioned. As P09 noted, a *"tool that significantly supports me or even identifies these lesions itself, and I simply verify"* would be helpful. So, individual preferences must be addressed. Hence, the Mid-Fi version includes a feature to enable or disable AI support for individual preferences, which was appreciated by radiologists, as it enabled them to switch seamlessly between AI mode and original mode to support a dynamic workflow.

6.1 Usability evaluation: first iteration

The Hi-Fi prototype was evaluated first, with radiologists generally endorsing its overall concept during the study. P06 highlighted his recognition by saying *"I would take it how it is right now, with the points we discussed, the changes, improvements, etc. I would take it, let it be certified, and bring it to the market."* The UI was perceived as easy to use, enabling users to orient themselves quickly; however, this applies mainly to individuals *"who are familiar with the subject matter"* (P02).

The presentation of prostate delineations and identified lesions on the T2W images was considered beneficial by all participants, enabling users to examine the AI algorithm's outcomes easily. Highlighting suspicious regions facilitated the assessment of lesion number, size, and location, while navigation through the full image stack supported a comprehensive understanding of AI conclusions. Active mask visualization *"[.] facilitates and fastens a quick overview. If I want to see where the algorithm finds something, I don't need to scroll through the whole image set but can approach it directly"* (P05).

Participants acknowledged the importance of the prostate measurement section and the relevance of each displayed value for the graphic. Especially, the automated calculation of prostate volume and PSA density was valued. As P06 noted, *"This is in principle what facilitates one's work, that you don't have to type into your calculator or ask Siri."*

Incorporating the PI-RADS score for each sequence, as well as the overall final score, was considered essential for the assisted workflow. The automated calculation of the final lesion score and the option for manual input were regarded as practical. The potential inclusion of visual alerts in cases where the manually assigned score contradicts the PI-RADS algorithm was regarded as a beneficial safeguard.

During task completion, participants used separate screens for the prototype and their DICOM viewer, as intended. The concept of assessing images within the DICOM viewer while simultaneously utilizing the AI solution on a distinct screen was deemed logical. According to P07, he does the diagnosis in the DICOM viewer anyway, and then he cross-checks on the other side. Conversely, P05 highlighted the potential advantages of direct integration into the existing PACS, which could reduce cognitive load.

The input form on the left side received a unanimous positive response from all participants, appreciating the structure's alignment with their current workflow. This arrangement allowed the participants to methodically progress through individual steps and seamlessly proceed once they were completed. P07 said, *"This already represents my approach very nicely, how I would work through the diagnosis. That's what I need. That I can go through this and mentally tick my things. And then I have my finished graphic and report."* All radiologists (also in the second evaluation cycle) appreciated automatic reporting, valuing any functionality that saves time and reduces workload. The ability to manually add lesions and correct values is deemed beneficial. However, P02 mentions a desire to have more flexibility when selecting secondary findings, e.g., by hiding irrelevant options. As this is a minor issue, we were able to change that directly in the Hi-Fi prototype.

The prostate sector map and the marked lesions also received positive feedback, along with suggestions for improvement. Especially, marking lesions based on their delineation results in a more precise representation was acknowledged. All participants recognized the advantages of automatically generating a lesion graphic, such as its time-saving and enhanced efficiency. Furthermore, it eradicates the potential for transfer errors and guarantees a consistent graphic outcome.

Although participants agreed on most features, variations exist between the values, preferences, and workflows of individual radiologists. Divergences emerged regarding aspects such as the importance of the lesion volume, information included in the report graphic, and the use of color for highlighting. This represents the challenge that design decisions valued by some can be inconvenient for others. As suggested by participants, customizable settings could mitigate this issue.

6.2 Usability evaluation: second iteration

In our case, the final version is the Mid-Fi prototype, which was created after evaluating the Hi-Fi prototype. In general, the AI is viewed as effective not only when making accurate assessments but also for not significantly disrupting the workflow when predictions are incorrect. While P08 expects high accuracy, he recognizes false positives as inevitable and suggests a threshold to limit displayed lesions, though he wants to see enough abnormalities to avoid missing cancerous lesions. P04 and P06 readily override AI assessments, appreciating that the prototype supports human oversight to compensate for AI fallibility. P07 highlighted the importance of quickly and intuitively dismissing false positives, valuing the ability to easily delete AI suggestions, particularly in the TZ where they are frequent. P07 found it more efficient to correct an inaccurate AI suggestion with minimal adjustments than to manually input findings from scratch. Though he acknowledged the risk of diminishing intuition, which could be particularly problematic given the relatively frequent occurrence of edge cases.

P09 appreciates the AI by saying, *"Of course, I think it is good that the system shows you where an abnormal finding is and allows you to see boundaries, so I can get an idea of how large the finding is"*. He also appreciated the AI-suggested lesions in the list, praising the clear presentation of their details, such as numbering, zone assignment, and size.

The confidence bars were positively received as they offer a more natural experience (P08) and allow participants to choose among multiple AI suggestions rather than relying on a single prediction. P04 compared this to differential diagnosis, where physicians weigh several possibilities before deciding, and P06 highlighted its novelty for prostate cancer, *"I find it helpful. [It] is already used in other fields [.], where diagnostic suggestions are provided along with a probability. I think it is an interesting feature. I have never seen anything like this in prostate cancer diagnosis before"*.

As P09 evaluated a pre-version of the Mid-Fi prototype, he found it challenging to link lesions in the MRI sequence to the corresponding AI-generated entries in the list, particularly when multiple lesions were present, *"On the left, I have two lesions labeled as 'Lesion 1' and 'Lesion 2.' When I see the red-highlighted lesion in the MRI image on the right, I cannot tell which of the two it is"* (P09). He further suggested sorting AI-identified lesions by size, from largest to smallest, reflecting his typical workflow in which larger lesions are examined first before identifying smaller ones. That was already implemented in the final version, which can be seen in Figure 10.

6.3 Design recommendations for redesign

During the evaluation, several design recommendations for improvement were suggested. While we will address key points here, we will not delve into suggestions for minor changes, as they are not the primary focus.

Similar to other AI-generated data, the ability to modify map markings should always be adjustable to integrate smoothly into clinical workflows. This includes the ability to remove or add lesions and, as P06 and P07 proposed, to manually delineate new ones directly on the MRI. This would allow the system to auto-populate values based on the image data within the delineated region. We believe that incorporating this feature may empower radiologists to have more control over the mapping process, while also addressing concerns regarding the practicality and efficiency of manual lesion delineation expressed by P05.

The evaluation revealed the importance of preserving overwritten AI-generated results while ensuring clear visual differentiation. Retaining original outputs allows radiologists to compare their input with AI suggestions, track discrepancies over time, and monitor AI performance. To support this, a dedicated dashboard might facilitate this monitoring. Visual contrast between rectified and AI-generated values, such as overlaying both delineations, as suggested by P07, was considered essential to avoid confusion and make discrepancies transparent. This differentiation aids radiologists in indicating AI-generated results and understanding the process leading to outcomes. Moreover, it facilitates crafting reports, as discrepancies from AI output can be communicated to the referrer. Moreover, maintaining access to original values supports retroactive traceability, particularly for cases requiring re-evaluation.

For the Hi-Fi prototype, concerns were raised regarding the accuracy of markings on the prostate sector map. *"The graphic does not match the anatomy"* (P02). In this case, the lesion would have to be corrected manually. Participants further expected large lesions to be consistently marked across all relevant planes (base, middle, apex) and criticized the absence of markings in sagittal and coronal views or alternative sequences (e.g., DWI), though opinions on their relevance varied. P02 suggested color-coding lesions (*"[.] to distinguish between them, with the target lesion in red and secondary lesions in a different color"*). While this was then implemented in the Mid-Fi prototype, P07 and P08 misinterpreted the colors as indicators of cancer severity, which might be an unintended but potentially useful feature, as such coding is familiar from other software.

7 Discussion

We have learned that new technologies, such as AI, can be particularly useful in the medical sector to achieve more efficient workflows. However, such technologies can also be imperfect,[105,106] which is critical, especially in healthcare, as sensitive data are processed and misdiagnosis can cause patients harm.[107] As demonstrated by both the literature and our study, seamless integration into radiologists' workflows is essential for ensuring the efficiency and usability of AI systems.

Conducting a DCS and following the structured, three-phase process allowed us to continuously involve radiologists and gain an in-depth understanding of their real-world workflows, challenges, and expectations toward AI support. Focusing on a specific use case enabled targeted empirical inquiry and the derivation of contextually grounded design requirements.[28] The DCS framework further guided the translation of these insights into concrete design solutions for the seamless integration of the AI system into existing diagnostic workflows. This approach proved particularly valuable for aligning technological innovation with clinical practice through iterative, user-centered refinement. While our study was situated in the German healthcare context and focused on prostate cancer diagnostics, many identified insights and design principles extend beyond this scope. Both the literature and our empirical data[28] included references to other cancer types and international approaches, underscoring the broader relevance of our findings. The proposed prototype and interaction concepts can be adapted to other cancers and healthcare systems, provided they align with local workflows and standards. More broadly, the underlying principles, like interdisciplinary collaboration, human oversight, and transparency through visualization, are transferable across medical domains and national contexts, supporting human-centered AI in healthcare.

While AI has the potential to significantly enhance radiological practices, challenges such as limited interpretability and transparency remain critical in high-stakes environments.[50,108,109]

Addressing these issues through explainability requires interdisciplinary collaboration among AI experts, radiologists, designers, and researchers to develop systems that are both accurate and explainable, fostering trust and enabling seamless integration into radiological workflows.[25,70]

While AI may provide accurate results,[110,111] understanding the AI's conclusion remains unclear[112–114] which can lead to mistrust and hinder the broader adoption of AI technologies.[49,113] Though in our study, radiologists generally expressed trust in AI, with confidence increasing through repeated use and demonstrated accuracy,[28] consistent with prior work,[115] they would also accept a certain margin of error, noting that human interpretation is likewise prone to mistakes, particularly under visual strain.

However, addressing such issues through explainability requires an interdisciplinary approach[116] with participatory design approaches[117] to minimize the likelihood of creating undesirable technology.[118]

One of the greatest strengths of our contribution is our interdisciplinary collaboration with radiologists, AI developers, UX designers, and HCI researchers during the context study, as well as in the design study to consider different perspectives and to thus achieve a HCAI solution, that is both accurate and explainable,[25] fostering trust and enabling seamless integration into radiological workflows.[70] Our study highlights that integrating AI into medical practice requires more than technical expertise; it depends on continuous interdisciplinary collaboration and genuine co-design with practitioners.[83,116] Creating prototypes at varying levels of fidelity to iteratively refine the product is fundamental to the practice of HCI.[119] In this way, our work demonstrates how different stakeholders without prior design experience can effectively contribute their own ideas for advanced AI solutions through DT, ensuring that radiologists' perspectives remain central and are translated into concrete design ideas. Concerning feasibility in terms of implementation, the perspectives of the AI developers were also leveraged. The group dynamics proved to be an ideal mix to achieve the optimal outcome in our context.

Our study confirms the potential of AI in prostate cancer diagnostics but emphasizes that decision-making responsibility must remain with radiologists. AI should be framed as a collaborative partner that augments, rather than replaces, human expertise by offering a second opinion, for instance, by highlighting suspicious regions, distinguishing prostate zones, and suggesting PI-RADS scores. Such an approach reflects principles of Human-AI Collaboration, where the AI directs attention and provides complementary input while the radiologist retains final responsibility. This aligns with Griesshaber et al., who emphasize the importance of AI systems capable of guiding users toward relevant findings.[120] The role of the second opinion also highlights its effectiveness in bringing attention to aspects that busy radiologists might overlook due to time constraints and burnout.[121] Moreover, it is important to have a sanity check by the radiologist to ensure the accuracy and reliability of the AI findings. Our collaboration with radiologists indicates that the AI can serve as an assistant, supporting their daily practice by offering suggestions that they can choose to accept, reject or correct, which can increase AI acceptance rates.[122,123] Striking the right balance between human judgment and AI recommendations is crucial to ensure optimal outcomes. Our design, using approximate confidence bars, resembles human estimation, preserves professional agency, and encourages critical engagement with AI suggestions according to Human Augmentation, arguing for high human control over the technology.[74] Hence, we suggest that HCAI systems should support, rather than replace, radiologists' judgment in uncertain diagnostic contexts.

Our study confirms findings from the literature suggesting that the decision of the AI system will be made more understandable and transparent through visualizations, which will lead to a fair and responsible perception of the human-system decisions.[27] This can lead us forward in a discussion of whether visualization can be the first step of explainability. All radiologists in our evaluation study also agreed that it was helpful that the prostate delineations and detected lesions were presented on the MRI images, making it simple and understandable for them to detect and review the results of the AI system.

Although AI models can demonstrate a high and consistent level of accuracy based on their training data, they are still subject to inherent limitations, uncertainties, and biases.[26,124] These models rely on patterns within the data they were trained on and may not always account for unforeseen scenarios or unique cases. Moreover, if the training data is not representative or lacks diversity, the AI model may perform well in some instances but poorly in others.[50,124] We have learned in our study that obtaining prostate cancer data is challenging, and additionally, most patients are male, older, and predominantly white due to the German medical data we work with, leading to an inherent bias in the training process that we must be aware of. Consistent with previous literature, our study indicates that users require a comprehensive understanding of the AI's strengths and limitations.

An extensively explored aspect of the state-of-the-art revolves around determining the optimal point at which the diagnostic AI output should be introduced to the radiologist's process.[125–127] Some studies show that people blindly trust collaborative assistants, regardless of their own ability.[128,129] On the other hand, Berkel et al., found that the participants were focused on relying solely on their skills and judgment, and they disregarded any visual indicators that could potentially disrupt their capability to identify abnormalities.[130] However, in our study, most radiologists prefer to first analyze the MRI images independently of the AI and then use the AI results as support so as not to be too influenced. When evaluating the AI prototype, a significant requirement that emerged was the need for verification. Despite the absence of specific XAI techniques in our prototype to enhance the understanding of AI outputs, participants still utilized available information to interpret the automatically generated results. Researchers found out that the AI-driven system could improve clinicians' trust in AI suggestions by offering evidence of their reliability and relevance to individual cases and by relying on a consistent validation process, similar to how clinicians validate each other's suggestions in practice,[131] which we are also planning to incorporate in the future steps.

7.1 Limitations

While our contribution offers valuable insights into our research goal, it is important to acknowledge its limitations, particularly concerning our sample. Although we have invested great effort in recruiting radiologists, we faced considerable challenges in the process due to the scarcity of available professionals. Nevertheless, we were

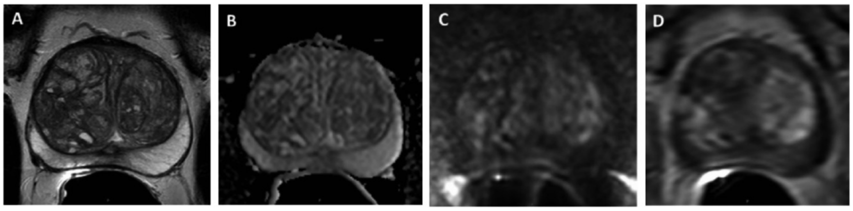


Figure 11: Images of MRI sequences for A) T2W, B) ADC, C) DWI, D) DCE of prostate.[43]

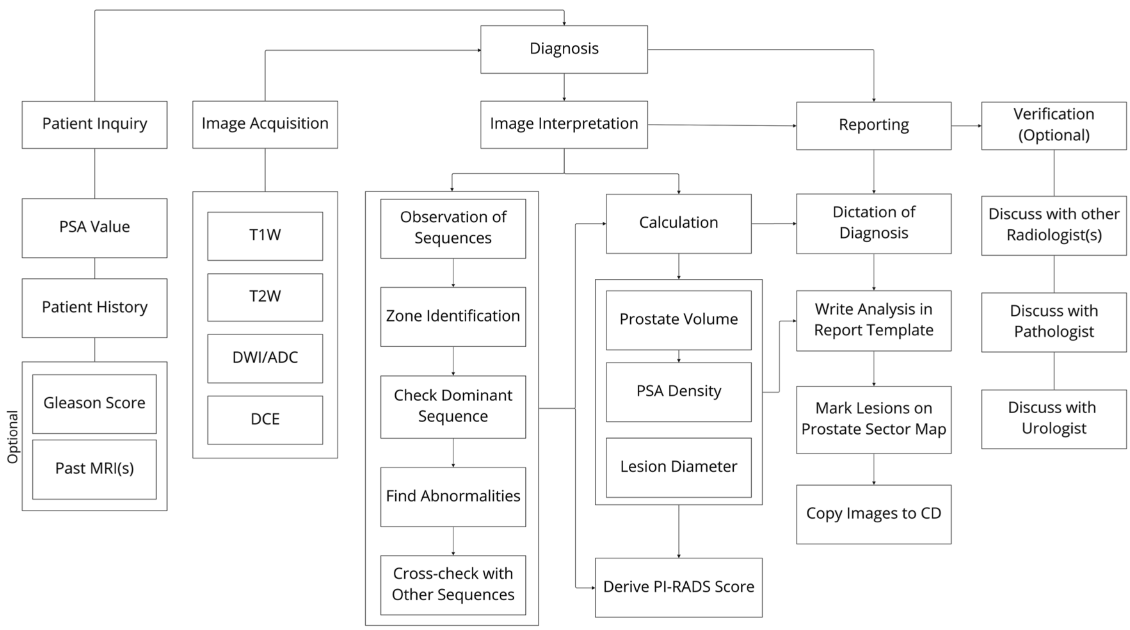


Figure 12: Workflow of a radiologist in prostate cancer diagnosis.

able to engage experienced individuals in our workshops and evaluation sessions, providing valuable insights.

Appropriation phase in a Design Case Study (DCS) primarily focuses on the long-term adoption of the technical artifact. However, due to our current stage, we have not yet implemented the new tool within radiological organizations over an extended period. Moving forward, we plan to integrate the final AI solution into real-world settings and incorporate radiologists' feedback through a human-in-the-loop approach, necessitating further evaluation and appropriateness studies with larger participant pools and long-term evaluations.

An additional constraint concerned our DT workshops, which had to be conducted within a few hours due to participants' limited availability. Therefore, we used an adapted design sprint with DT as the methodological framework and, instead of requiring participants to conduct their own research, provided them with insights from our prior empirical study. Radiologists explained their daily practices and challenges to non-medical participants in the first workshop, and we shared these insights with students in the second workshop to help them empathize with real users. However, some HCI methods proved difficult for radiologists to grasp, leading to critical discussions about the purpose of DT and its structured approach. With little time to bridge these knowledge gaps, it was also challenging for us as researchers to convey the value of HCD, highlighting the need for more time on foundational understanding in future workshops.

8 Conclusions

In this paper, we presented a HCAI solution to support radiologists in diagnosing prostate cancer, designed through a combination of practice-centered and HCD approaches. We worked closely and collaboratively with radiologists, AI developers, HCI researchers, and UX professionals, who were directly involved in our design study through DT workshops and evaluation sessions. We present how we maintained a focus on real users through their strong involvement while paying close attention to their practices. This, we argue, has enabled us to address their needs and to design an AI tool that can support current radiological practice.

Our iterative path for designing the prototype, from Lo-Fi to Hi-Fi and back to Mid-Fi, proved to be an effective methodological choice that refined interactions and strengthened the final design. The workshops and evaluation sessions enabled interaction with different prototypes, playing a significant role in facilitating practical discussions among different stakeholders about specific features and their alignment with user needs.

Our interdisciplinary approach to designing a HCAI solution, incorporating the expertise and practical insights of end users and other stakeholders, provides a valuable framework for future AI developments in healthcare. Our study highlights that collaborative and iterative HCD processes are essential to developing meaningful HCAI solutions applicable across the HCI community and beyond.

Our future work will build on these findings by further refining the prototype based on user feedback, testing its effectiveness in real-world clinical settings, and investigating the long-term impacts of HCAI solutions on workflow efficiency and patient outcomes.

Acknowledgments: We would like to thank our project partners Nils Blaumer, David Matuschek, and Demster Bijl from Gemedico GmbH for a great collaboration in our research project. We would particularly like to thank all our participants for their engagement in the workshops, the design activities and evaluation sessions as well as the constant exchange with our field partners from the radiology centers. Additionally, we thank our student assistant, Jessica Fraas for her support in the workshop organization as well as our students Paul Schreiber und Kim Philipp Stahl for contributing in the evaluation part.

Research ethics: Our study design underwent ethical review by our institution's IRB and received clearance.

Informed consent: All Participants received study information, privacy and data security details, and signed a consent form before participating.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. The first two authors, Ms. Plinz-Saßmannshausen and Ms. Ontika, were primarily responsible for advancing this study, developing the research design, defining the methodology, leading and conducting research activities, analyzing the data, and shaping and writing the paper. Mr. Tran and Mr. Ukleja collected additional data, advanced the prototyping process, and conducted evaluations. Mr. Pinatti framed the contribution of the paper, reviewed all parts, and helped to shape it. Mr. Rohde supported the research in an advisory capacity, drawing on his expertise in the field of socioinformatics.

Use of Large Language Models, AI and Machine Learning Tools: Use of Grammarly and ChatGPT for grammar check and DeepL for translations.

Conflict of interest: The authors state no conflict of interest.

Research funding: This paper presents the results of the research project PAIRADS, funded by the Bundesministerium für Bildung und Forschung – BMBF (funding code: 16SV8651).

Data availability: Not applicable.

A: Appendix

Glossary

AFS	The Anterior Fibromuscular Stroma (AFS) is a region within the prostate gland consisting of fibrous and muscular tissues located anteriorly, playing a crucial role in prostate anatomy
AIMDSS	AI-based Medical Diagnosis Support Systems
CAD	Computer-Aided Detection
CAID	Computer-Aided Intelligent Diagnosis
CDS	Clinical Decision Support
CDSS	Clinical Decision Support Systems
CNN	Convolutional Neural Network
CZ	The Central Zone (CZ) located between the transition zone and peripheral zone, is involved in the ejaculatory ducts and constitutes a small portion of the prostate gland
DCE	Dynamic Contrast-Enhanced (DCE) imaging involves continuously acquiring images before, during, and after contrast agent injection. It helps evaluate tissue perfusion and vascularity, providing insights into tissue characteristics
DCS	Design Case Study
DICOM	Digital Imaging and Communications in Medicine
DT	Design Thinking
DWI	Diffusion-Weighted Imaging (DWI) measures the random motion of water molecules in tissues. It aids in identifying tissue characteristics and abnormalities based on the diffusion of water molecules
HCAI	Human-Centered Artificial Intelligence
HCD	Human-Centered Design
HCI	Human Computer Interaction
Hi-Fi	High-Fidelity
HMW	How-Might-We
Lo-Fi	Low-Fidelity
Mid-Fi	Medium-Fidelity
mpMRI	Multiparametric Magnetic Resonance Imaging
MRI	Magnetic Resonance Imaging
PACS	Picture Archiving and Communication System
PI-RADS	Prostate Imaging Reporting and Data System (PI-RADS) is a structured reporting scheme for mpMRI in the evaluation of suspected prostate cancer

POV Point-of-View
PSA Prostate Specific Antigen (PSA) is a protein made by cells in the prostate gland (both normal cells and cancer cells). Elevated levels in blood tests can indicate various prostate conditions, including prostate cancer, but it is not solely diagnostic and needs further evaluation
PZ The Peripheral Zone (PZ) refers to the outer region of the prostate gland and is the most common location for prostate cancer to develop
SSL Semi-Supervised Learning
T2W A T2-Weighted (T2W) image is a type of MRI sequence that emphasizes differences in the spin-spin relaxation time (T2) of tissues. It enhances contrast based on tissue water content and is valuable in imaging anatomical structures, especially in visualizing fluid-filled spaces and abnormalities within tissues
TZ The Transitional Zone (TZ) refers to a region within the prostate gland that surrounds the urethra
UI User Interface
UX User eXperience
XAI eXplainable Artificial Intelligence

References

1. Zentrum für Krebsregisterdaten. Prostate cancer, 2022. https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Prostatakrebs/prostatakrebs_node.html (accessed 2025-07-24).
2. Erdmann, F.; Spix, C.; Katalinic, A.; Christ, M.; Folkerts, J.; Hansmann, J.; Kranzhöfer, K.; Kunz, B.; Manegold, K.; Penzkofer, A.; Treml, K.; Vollmer, G.; Weg-Remers, S.; Barnes, B.; Buttmann-Schweiger, N.; Dahm, S.; Fiebig, J.; Franke, M.; Gurung-Schönfeld, I.; Haberland, J.; Imhoff, M.; Kraywinkel, K.; Starker, A.; von Berenberg-Gossler, P.; Wienecke, A. Krebs in Deutschland für 2017/2018, 2021. https://doi.org/10.25646/8353.
3. Dunn, M. W.; Kazer, M. W. Prostate Cancer Overview. *Semin. Oncol. Nurs.* **2011**, *27* (4), 241−250.
4. Loeb, S.; Bjurlin, M. A.; Nicholson, J.; Tammela, T. L.; Penson, D. F.; Carter, H. B.; Carroll, P.; Etzioni, R. Overdiagnosis and Overtreatment of Prostate Cancer. *Eur. Urol.* **2014**, *65* (6), 1046−1055.
5. Bundesärztekammer. Ärztestatistik Zum 31. Dezember 2021, 2021. Available from: https://www.bundesaerztekammer.de/fileadmin/user_upload/BAEK/Ueber_uns/Statistik/2021/2021_Statistik.pdf.
6. Sorantin, E.; Grasser, M. G.; Hemmelmayr, A.; Tschauner, S.; Hrzic, F.; Weiss, V.; Lacekova, J.; Holzinger, A. The Augmented Radiologist: Artificial Intelligence in the Practice of Radiology. *Pediatr. Radiol.* **2022**, *52* (11), 2074−2086.
7. Quinsten, A. S.; Apel, M.; Oliveira, S. Remote MR scanning-A Solution for Shortage of Skilled Radiographers. *J. Med. Imag. Radiat. Sci.* **2023**, *54*, 410−414.
8. Malamateniou, C.; Knapp, K. M.; Pergola, M.; Woznitza, N.; Hardy, M. Artificial Intelligence in Radiography: Where are we now and what does the Future Hold? *Radiography (Lond.)* **2021**, *27* (Suppl 1), S58−S62.
9. Nazar, M.; Alam, M. M.; Yafi, E.; Su'ud, M. M. A Systematic Review of human−computer Interaction and Explainable Artificial Intelligence in Healthcare with Artificial Intelligence Techniques. *IEEE Access* **2021**, *9*, 153316−153348.
10. Huisman, M.; Ranschaert, E.; Parker, W.; Mastrodicasa, D.; Koci, M.; de Santos, D. P.; Coppola, F.; Morozov, S.; Zins, M.; Bohyn, C.; Koç, U.; Wu, J.; Veean, S.; Fleischmann, D.; Leiner, T.; Willemink, M. J. An International Survey on AI in Radiology in 1041 Radiologists and Radiology Residents Part 2: Expectations, Hurdles to Implementation, and Education. *Eur. Radiol.* **2021**, *31* (11), 8797−8806.
11. Waymel, Q.; Badr, S.; Demondion, X.; Cotten, A.; Jacques, T. Impact of the Rise of Artificial Intelligence in Radiology: What do Radiologists Think? *Diagn. Interv. Imaging* **2019**, *100* (6), 327−336.
12. Jungmann, F.; Jorg, T.; Hahn, F.; Dos Santos, D. P.; Jungmann, S. M.; Düber, C.; Mildenberger, P.; Kloeckner, R. Attitudes Toward Artificial Intelligence Among Radiologists, IT Specialists, and Industry. *Acad. Radiol.* **2021**, *28* (6), 834−840.
13. von Wedel, P.; Hagist, C. Physicians' Preferences and Willingness to Pay for Artificial Intelligence-Based Assistance Tools: A Discrete Choice Experiment Among German Radiologists. *BMC Health Serv. Res.* **2022**, *22* (1), 398.
14. Yang, L.; Ene, I. C.; Belaghi, R. A.; Koff, D.; Stein, N.; Santaguida, P. L. Stakeholders' Perspectives on the Future of Artificial Intelligence in Radiology: A Scoping Review. *Eur. Radiol.* **2022**, *32* (3), 1477−1495.
15. Egevad, L.; Swanberg, D.; Delahunt, B.; Ström, P.; Kartasalo, K.; Olsson, H.; Berney, D. M.; Bostwick, D. G.; Evans, A. J.; Humphrey, P. A.; Iczkowski, K. A.; Kench, J. G.; Kristiansen, G.; Leite, K. R. M.; McKenney, J. K.; Oxley, J.; Pan, C.-C.; Samaratunga, H.; Srigley, J. R.; Takahashi, H.; Tsuzuki, T.; van der Kwast, T.; Varma, M.; Zhou, M.; Clements, M.; Eklund, M. Identification of Areas of Grading Difficulties in Prostate Cancer and Comparison with Artificial Intelligence Assisted Grading. *Virchows Arch.* **2020**, *477* (6), 777−786.
16. Goldenberg, S. L.; Nir, G.; Salcudean, S. E. A New Era: Artificial Intelligence and Machine Learning in Prostate Cancer. *Nat. Rev. Urol.* **2019**, *16* (7), 391−403.
17. Cabitza, F.; Campagner, A.; Balsano, C. Bridging the "Last Mile" Gap Between AI Implementation and Operation: "Data Awareness" that Matters. *Ann. Transl. Med.* **2020**, *8* (7), 501.
18. Abowd, G. D.; Dey, A. K.; Brown, P. J.; Davies, N.; Smith, M.; Steggles, P. Towards a Better Understanding of Context and Context-Awareness. In *Handheld and Ubiquitous Computing*; Gellersen, H.-W., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1999; pp. 304−307.
19. Garibay, O. O.; Winslow, B.; Andolina, S.; Antona, M.; Bodenschatz, A.; Coursaris, C.; Falco, G.; Fiore, S. M.; Garibay, I.; Grieman, K.; Havens, J. C.; Jirotka, M.; Kacorri, H.; Karwowski, W.; Kider, J.; Konstan, J.; Koon, S.; Lopez-Gonzalez, M.; Maifeld-Carucci, I.; McGregor, S.; Salvendy, G.; Shneiderman, B.; Stephanidis, C.; Strobel, C.; Holter, C. T.; Xu, W. Six Human-Centered Artificial Intelligence Grand Challenges. *Int. J. Hum. Comput. Interact.* **2023**, 1−47.
20. Schmidt, A. Implicit Human Computer Interaction Through Context. *Pers. Technol.* **2000**, *4* (2−3), 191−199.
21. Baxter, G.; Sommerville, I. Socio-Technical Systems: From Design Methods to Systems Engineering. *Interact. Comput.* **2010**, *23* (1), 4−17.
22. Petersson, L.; Larsson, I.; Nygren, J. M.; Nilsen, P.; Neher, M.; Reed, J. E.; Tyskbo, D.; Svedberg, P. Challenges to Implementing Artificial Intelligence in Healthcare: A Qualitative Interview Study

with Healthcare Leaders in Sweden. *BMC Health Serv. Res.* **2022**, *22* (1), 850.
23. Zając, H. D.; Li, D.; Xiang, D.; Carlsen, J. F.; Kensing, F.; Andersen, T. O. Clinician-Facing AI in the Wild: Taking Stock of the Sociotechnical Challenges and Opportunities for HCI. *ACM Trans. Comput.-Hum. Interact.* **2023**, *30* (2), 1−39. Article 33, 39 pages. https://doi.org/10.1145/3582430.
24. Greenhalgh, T.; Wherton, J.; Papoutsi, C.; Lynch, J.; Hughes, G.; A'Court, C.; Hinder, S.; Procter, R.; Shaw, S. Analysing the Role of Complexity in Explaining the Fortunes of Technology Programmes: Empirical Application of the NASSS Framework. *BMC Med.* **2018**, *16*, 1.
25. Ontika, N. N.; Saßmannshausen, S. M.; Abid, H. S.; de Carvalho, A. F. P.; Pipek, V. Towards Human-Centered AI: Learning from Current Practices in Radiology. Trust, Professional Vision, and Diagnostic Work. In *Exploring Human-Centered AI in Healthcare: A Workshop Report*; International Institute for Socio-Informatics (IISI), Vol. *19*, 2022; pp. 33−38.
26. Rubin, D. L. Artificial Intelligence in Imaging: The Radiologist's Role. *J. Am. Coll. Radiol.* **2019**, *16* (9 Pt B), 1309−1317.
27. Lee, M. K. Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. *Big Data Soc.* **2018**, *5* (1), 1−16.
28. Saßmannshausen, S. M.; Ontika, N. N.; De Carvalho, A. F. P.; Rouncefield, M.; Pipek, V. 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 (Honolulu, HI, USA) (CHI '24)*; Association for Computing Machinery: New York, NY, USA, 2024. Article 468, 20 pages.
29. Bender-Salazar, R. Design Thinking as an Effective Method for Problem-Setting and Needfinding for Entrepreneurial Teams Addressing Wicked Problems. *J. Innov. Entrep.* **2023**, *12* (1), 24.
30. Chen, H.; Gomez, C.; Huang, C.-M.; Unberath, M. Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence from a Systematic Review. *NPJ Digit. Med.* **2022**, *5* (1), 156.
31. Ku, B.; Lupton, E. *HEALTH DESIGN THINKING: Creating Products and Services for Better Health*; MIT Press: Cambridge, MA, 2022.
32. Bartlett, R.; Boyle, J. A.; Smith, J. S.; Khan, N.; Robinson, T.; Ramaswamy, R. Evaluating Human-Centred Design for Public Health: A Case Study on Developing a Healthcare App with Refugee Communities. *Res. Involv. Engagem.* **2021**, *7* (1), 32.
33. Catalani, C.; Green, E.; Owiti, P.; Keny, A.; Diero, L.; Yeung, A.; Israelski, D.; Biondich, P. A Clinical Decision Support System for Integrating Tuberculosis and HIV Care in Kenya: A Human-Centered Design Approach. *PLoS One* **2014**, *9* (8), 1−9.
34. Steen, M.; Manschot, M.; De Koning, N. Benefits of Co-Design in Service Design Projects. *Int. J. Des.* **2011**, *5* (2), 53−60.
35. Stevens, G.; Rohde, M.; Korn, M.; Wulf, V. 23Grounded Design: A Research Paradigm in Practice-Based Computing. In *Socio-Informatics*; Oxford University Press, 2018. arXiv: https://academic.oup.com/book/0/chapter/193131333/chapter-ag-pdf/44626518/book_25681_section_193131333.ag.pdf.
36. Wulf, V.; Pipek, V.; Randall, D.; Rohde, M.; Schmidt, K.; Stevens, G., Eds. In *Socio-informatics: A Practice-based Perspective on the Design and Use of IT Artifacts*; Oxford University Press: United Kingdom, 2018.
37. Kim, E.; Krishnan, J. A. Redesigning Healthcare to Fit with People. *BMJ* **2016**, *354*, i4536.
38. Maguire, M.; Bevan, N. *User Requirements Analysis*; Springer: US, Boston, MA, 2002; pp. 133−148.
39. Wulf, V.; Müller, C.; Pipek, V.; Randall, D.; Rohde, M.; Stevens, G. *Practice-Based Computing: Empirically Grounded Conceptualizations Derived from Design Case Studies*; Springer: London, London, 2015; pp. 111−150.
40. Beyer, H.; Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1997.
41. Ontika, N. N.; Saßmannshausen, S. M.; Syed, H. A.; de Carvalho, A. F. P. Exploring Human-Centered AI in Healthcare: A Workshop Report. *Int. Rep. Socio-Inf. (IRSI)* **2022**, *19* (2), 1−54.
42. Hussy, W.; Schreier, M.; Echterhoff, G. *Forschungsmethoden in Psychologie und Sozialwissenschaften*; Springer-Verlag, 2013.
43. American College of Radiology, 2019. Prostate imaging − reporting and data system, version 2.1. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS (accessed 2024-1-13).
44. Ontika, N. N.; Plinz-Saßmannshausen, S. M.; de Carvalho, A. F. P.; Hadid, F.; Müller, C. Communication Matters: Technology for Information Exchange Needs in the Context of Prostate Cancer Diagnosis. *Proc. ACM Hum.-Comput. Interact.* **2025**, *9* (7), 1−37. Article CSCW387, 37 pages. https://doi.org/10.1145/3757568.
45. Harper, R. *Being Human: Human-Computer Interaction in the Year 2020*; Microsoft Research Ltd: Cambridge, England, 2008.
46. Inkpen, K.; Chancellor, S.; De Choudhury, M.; Veale, M.; Baumer, E. P. S. Where is the Human? In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow Scotland Uk)*; ACM: New York, NY, USA, 2019.
47. Förster, M.; Klier, M.; Kluge, K.; Sigler, I. Fostering Human Agency: A Process for the Design of User-Centric XAI Systems. In *Proceedings of the 41st International Conference on Information Systems, ICIS 2020, Making Digital Inclusive: Blending the Locak and the Global, Hyderabad, India, December 13-16, 2020*; George, J. F.; Paul, S.; De', R.; Karahanna, E.; Sarker, S.; Oestreicher-Singer, G., Eds.; Association for Information Systems, 2020. Available from: https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/12.
48. Mashinchi, M. I.; Ojo, A. K.; Sullivan, F. J. Towards a Theoretical Model of Dashboard Acceptance and Use in Healthcare Domain. In *Hawaii International Conference on System Sciences*, 2021.
49. Cai, C. J.; Winter, S.; Steiner, D.; Wilcox, L.; Terry, M. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.* **2019**, *3*, 1−24. CSCW, Article 104, 24 pages. https://doi.org/10.1145/3359206.
50. Pouyan, E. Use of AI-Based Tools for Healthcare Purposes: A Survey Study from Consumers' Perspectives. *BMC Med. Inf. Decis. Making* **2020**, *20* (1), 1−19.
51. Strachna, O.; Asan, O. Reengineering Clinical Decision Support Systems for Artificial Intelligence. In *2020 IEEE International Conference on Healthcare Informatics (ICHI)*; IEEE, 2020; pp. 1−3.
52. Xie, Y.; Chen, M.; Kao, D.; Gao, G.; Chen, X. A. CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '20)*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1−13.

53. Grüning, M.; Trenz, M. Me, You and AI—Managing Human AI Collaboration in Computer Aided Intelligent Diagnosis. In *Proceedings of the Twentieth Annual Pre-ICIS Workshop on HCI Research in MIS.* International Conference on Information Systems (ICIS), Austin, TX, 2021.
54. Narmatha, C.; Eljack, S. M.; Tuka, A. A. R. M.; Manimurugan, S.; Mustafa, M. A Hybrid Fuzzy Brain-Storm Optimization Algorithm for the Classification of Brain Tumor MRI Images. *J. Ambient Intell. Hum. Comput.* **2020**, 1—9; https://doi.org/10.1007/s12652-020-02470-5.
55. Choy, G.; Khalilzadeh, O.; Michalski, M.; Do, S.; Samir, A. E.; Pianykh, O. S.; Geis, J. R.; Pandharipande, P. V.; Brink, J. A.; Dreyer, K. J. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* **2018**, *288* (2), 318—328.
56. Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L. H.; Aerts, H. J. W. L. Artificial Intelligence in Radiology. *Nat. Rev. Cancer* **2018**, *18* (8), 500—510.
57. Meyer, A.; Rakr, M.; Schindele, D.; Blaschke, S.; Schostak, M.; Fedorov, A.; Hansen, C. Towards Patient-Individual PI-Rads v2 Sector Map: Cnn for Automatic Segmentation of Prostatic Zones from T2-Weighted MRI. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019; pp. 696—700.
58. Chui, K. T.; Gupta, B. B.; Chi, H. R.; Arya, V.; Alhalabi, W.; Ruiz, M. T.; Shen, C. W. Transfer Learning-Based Multi-Scale Denoising Convolutional Neural Network for Prostate Cancer Detection. *Cancers* **2022**, *14*, 15.
59. Cuocolo, R.; Stanzione, A.; Castaldo, A.; De Lucia, D. R.; Imbriaco, M. Quality Control and Whole-Gland, Zonal and Lesion Annotations for the PROSTATEx Challenge Public Dataset. *Eur. J. Radiol.* **2021**, *138*, 109647.
60. Ghavami, N.; Hu, Y.; Gibson, E.; Bonmati, E.; Emberton, M.; Moore, C. M.; Barratt, D. C. Automatic Segmentation of Prostate MRI Using Convolutional Neural Networks: Investigating the Impact of Network Architecture on the Accuracy of Volume Measurement and MRI-Ultrasound Registration. *Med. Image Anal.* **2019**, *58*, 101558.
61. Huang, X.; Liu, B.; Hao, J.; Nguyen, H.; Knopp, M. V.; Miao, Q.; Jia, G. Automatic Segmentation of Prostate Structures Using a Convolutional Neural Network from Multiparametric MRI. *ACM Int. Conf. Proc. Ser.* **2019**, 92—96. https://doi.org/10.1145/3364836.3364855.
62. Lapa, P.; Rundo, L.; Gonçalves, I.; Castelli, M. Enhancing Classification Performance of Convolutional Neural Networks for Prostate Cancer Detection on Magnetic Resonance Images: A Study with the Semantic Learning Machine. In *GECCO 2019 Companion - Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion*, 2019; pp. 381—382.
63. Ma, L.; Guo, R.; Zhang, G.; Tade, F.; Schuster, D. M.; Nieh, P.; Master, V.; Fei, B. Automatic Segmentation of the Prostate on CT Images Using Deep Learning and Multi-Atlas Fusion. In *Medical Imaging 2017: Image Processing*, Vol. *10133*, 2017; p. 101332O.
64. Yoo, S.; Gujrathi, I.; Haider, M. A.; Khalvati, F. Prostate Cancer Detection Using Deep Convolutional Neural Networks. *Sci. Rep.* **2019**, *9* (1), 1—10.
65. Zhong, P.; Wu, J.; Yuan, Z.; Zhang, Y.; Tang, X. 3D Fully Convolutional Network Incorporating Savitzky-Golay Filtering for Prostate Segmentation. *ACM Int. Conf. Proc. Ser.* **2019**, 88—91. https://doi.org/10.1145/3364836.3364854.
66. Litjens, G.; Debats, O.; Barentsz, J.; Karssemeijer, N.; Huisman, H. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Trans. Med. Imag.* **2014**, *33* (5), 1083—1092.
67. Buck, C.; Hennrich, J.; Kauffmann, A. L. Artificial Intelligence in Radiology: A Qualitative Study on Imaging Specialists' Perspectives. In *Proceedings of the 42th International Conference on Information Systems (ICIS)*: Austin, USA, 2021. Available from: https://aisel.aisnet.org/icis2021/is%5fhealth/is%5fhealth/20/. Paper Number 2515.
68. Penzkofer, T.; Padhani, A. R.; Turkbey, B.; Haider, M. A.; Huisman, H.; Walz, J.; Salomon, G.; Schoots, I. G.; Richenberg, J.; Villeirs, G.; Panebianco, V.; Rouviere, O.; Logager, V. B.; Barentsz, J. ESUR/ESUI Position Paper: Developing Artificial Intelligence for Precision Diagnosis of Prostate Cancer Using Magnetic Resonance Imaging. *Eur. Radiol.* **2021**, *31* (12), 9567—9578.
69. Naiseh, M.; Al-Thani, D.; Jiang, N.; Ali, R. Explainable Recommendation: When Design Meets Trust Calibration. *World Wide Web* **2021**, *24* (5), 1857—1884.
70. Procter, R.; Tolmie, P.; Rouncefield, M. Holding AI to Account: Challenges for the Delivery of Trustworthy AI in Healthcare. *ACM Trans. Comput.-Hum. Interact.* **2023**, *30* (2), 1—34. Article 31, 34 pages. https://doi.org/10.1145/3577009.
71. Calisto, F. M.; Santiago, C.; Nunes, N.; Nascimento, J. C. Introduction of Human-Centric AI Assistant to Aid Radiologists for Multimodal Breast Image Classification. *Int. J. Hum. Comput. Stud.* **2021**, *150*, 102607.
72. Schwalbe, N.; Wahl, B. Artificial Intelligence and the Future of Global Health. *Lancet* **2020**, *395* (10236), 1579—1586.
73. Arrieta, A. B.; Rodríguez, N. D.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; García, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; Herrera, F. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI, 2019. *CoRR* abs/1910.10045. arXiv:1910.10045. Available from: http://arxiv.org/abs/1910.10045.
74. Shneiderman, B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *Int. J. Hum. Comput. Interact.* **2020**, *36* (6), 495—504.
75. Xu, W. Toward Human-Centered AI: A Perspective from Human-Computer Interaction. *Interactions* **2019**, *26* (4), 42—46.
76. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138—52160.
77. Beede, E.; Baylor, E.; Hersch, F.; Iurchenko, A.; Wilcox, L.; Ruamviboonsuk, P.; Vardoulakis, L. M. 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 (Honolulu, HI, USA) (CHI '20)*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1—12.
78. Bellio, M.; Furniss, D.; Oxtoby, N. P.; Garbarino, S.; Firth, N. C.; Ribbens, A.; Alexander, D. C.; Ann, B. Opportunities and Barriers for Adoption of a Decision-Support Tool for Alzheimer's Disease. *ACM Trans. Comput. Healthcare* **2021**, *2* (4), 1—19. Article 32, 19 pages. https://doi.org/10.1145/3462764.
79. McIntosh, C.; Conroy, L.; Tjong, M. C.; Craig, T.; Bayley, A.; Catton, C.; Gospodarowicz, M.; Helou, J.; Isfahanian, N.; Kong, V.; Lam, T.; Raman, S.; Warde, P.; Chung, P.; Berlin, A.; Purdie, T. G. Clinical Integration of Machine Learning for Curative-Intent Radiation

Treatment of Patients with Prostate Cancer. *Nat. Med.* **2021**, *27* (6), 999−1005.
80. Yang, Q.; Steinfeld, A.; Zimmerman, J. *Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes (CHI '19)*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1−11.
81. Cooper, N.; Horne, T.; Hayes, G. R.; Heldreth, C.; Lahav, M.; Holbrook, J.; Wilcox, L. A Systematic Review and Thematic Analysis of Community-Collaborative Approaches to Computing Research. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022. Available from: https://api.semanticscholar.org/CorpusID:248419416.
82. Tahaei, M.; Constantinides, M.; Quercia, D.; Kennedy, S.; Muller, M.; Stumpf, S.; Liao, Q. V.; Baeza-Yates, R.; Aroyo, L.; Holbrook, J.; Luger, E.; Madaio, M.; Blumenfeld, I. G.; De-Arteaga, M.; Vitak, J.; Olteanu, A. Human-Centered Responsible Artificial Intelligence: Current & Future Trends. In *CHI 2023 - Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (Conference on Human Factors in Computing Systems − Proceedings)*; Schmidt, A.; Väänänen, K.; Goyal, T.; Kristensson, P. O.; Peters, A., Eds.; Association for Computing Machinery, 2023; pp. 1−4. Publisher Copyright: © 2023 Owner/Author.; 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023 ; Conference date: 23-04-2023 Through 28-04-2023.
83. Yildirim, N.; Zlotnikov, S.; Sayar, D.; Kahn, J. M.; Bukowski, L. A.; Shah Amin, S.; Riman, K. A.; Davis, B. S.; Minturn, J. S.; King, A. J.; Ricketts, D.; Tang, L.; Sivaraman, V.; Perer, A.; Preum, S. M.; McCann, J.; Zimmerman, J. 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 (CHI '24)*; ACM, 2024.
84. Ooge, J.; Stiglic, G.; Verbert, K. Explaining Artificial Intelligence with Visual Analytics in Healthcare. *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* **2021**, *12*, e1427.
85. Martikainen, S.; Ikävalko, P.; Korpela, M. Participatory Interaction Design in User Requirements Specification in Healthcare. *Stud. Health Technol. Inf.* **2010**, *160* (PART 1), 304−308.
86. Kushniruk, A.; Nøhr, C. Participatory Design, User Involvement and Health IT Evaluation. *Evid.-Base Health Inf.: Promot. Saf. Effic. Sci. Methods Ethical Pol.* **2016**, *222*, 139−151.
87. Bjerknes, G.; Bratteteig, T. User Participation and Democracy: A Discussion of Scandinavian Research on System Development. *Scand. J. Inf. Syst.* **1995**, *7* (1), 73−98.
88. Bratteteig, T.; Wagner, I. Unpacking the Notion of Participation in Participatory Design. *Comput. Support. Coop. Work* **2016**, *25* (6), 425−475.
89. Bond, R. R.; Mulvenna, M.; Finlay, D.; Wong, A.; Koene, A.; Brisk, R.; Boger, J.; Adel, T. Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making. 2−9. In *RoCHI 2019: International Conference on Human-Computer Interaction, RoCHI; Conference date: 17-10-2019 Through 18-10-2019*, 2019.
90. Ratwani, R. M.; Savage, E.; Will, A.; Fong, A.; Karavite, D.; Muthu, N.; Rivera, A. J.; Gibson, C.; Asmonga, D.; Moscovitch, B.; Grundmeier, R.; Rising, J. Identifying Electronic Health Record Usability and Safety Challenges in Pediatric Settings. *Health Aff. (Millwood)* **2018**, *37* (11), 1752−1759.
91. Grots, A.; Pratschke, M. Design Thinking — Kreativität Als Methode. *Mark. Rev. St Gallen* **2009**, *26* (2), 18−23.
92. Oliveira, M.; de Senzi Zancul, E.; Fleury, A. L. Design Thinking as an Approach for Innovation in Healthcare: Systematic Review and Research Avenues. *BMJ Innov.* **2020**, *7*, 491−498.
93. Altman, M.; Huang, T. T. K.; Breland, J. Y. Design Thinking in Health Care. *Prev. Chronic Dis.* **2018**, *15* (180128), E117.
94. Gagnon, M.-P.; Ngangue, P.; Payne-Gagnon, J.; Desmartis, M. m-Health Adoption by Healthcare Professionals: A Systematic Review. *J. Am. Med. Inform. Assoc.* **2016**, *23* (1), 212−220.
95. Sarah Gibbons. Design Thinking 101, 2016. https://www.nngroup.com/articles/design-thinking/ (accessed 2025-9-29).
96. Wolniak, R. The Design Thinking Method and its Stages. *Systemy Wspomagania w Inzynierii Produkcji* **2017**, *6*, 247−255.
97. Jaspers, M. W. M.; Steen, T.; van den Bos, C.; Geenen, M. The Think Aloud Method: A Guide to User Interface Design. *Int. J. Med. Inf.* **2004**, *73* (11), 781−795.
98. Braun, V.; Clarke, V. Thematic Analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*; American Psychological Association: Washington, 2012; pp. 57−71.
99. Verstijnen, I. M.; van Leeuwen, C.; Goldschmidt, G.; Hamel, R.; Hennessey, J. M. Sketching and Creative Discovery. *Des. Stud.* **1998**, *19* (4), 519−546.
100. Houde, S.; Hill, C. Chapter 16 - what do Prototypes Prototype? In *Handbook of Human-Computer Interaction*; Helander, M. G.; Landauer, T. K.; Prabhu, P. V., Eds.: North-Holland, Amsterdam, 1997, 2nd ed.; pp. 367−381.
101. American College of Radiology (ACR), ESUR, AdMeTech. PI-RADS. Prostate Imaging − Reporting and Data System. *Prostate Imaging-Report. Data Syst.* **2019**, 1−69.
102. Li, X.; Yu, L.; Chen, H.; Fu, C.-W.; Xing, L.; Heng, P.-A. Transformation-Consistent Self-Ensembling Model for Semisupervised Medical Image Segmentation. *IEEE Transact. Neural Networks Learn. Syst.* **2021**, *32* (2), 523−534.
103. Xia, Y.; Yang, D.; Yu, Z.; Liu, F.; Cai, J.; Yu, L.; Zhu, Z.; Xu, D.; Yuille, A.; Roth, H. Uncertainty-Aware Multi-View Co-Training for Semi-Supervised Medical Image Segmentation and Domain Adaptation. *Med. Image Anal.* **2020**, *65*, 101766.
104. Luo, X.; Chen, J.; Song, T.; Chen, Y.; Wang, G.; Zhang, S. Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency, 2020. *CoRR* abs/2009.04448. arXiv:2009.04448. Available from: https://arxiv.org/abs/2009.04448.
105. Chitra, B.; Kumar, S. S. Recent Advancement in Cervical Cancer Diagnosis for Automated Screening: A Detailed Review. *J. Ambient Intell. Hum. Comput.* **2022**, *13* (1), 251−269.
106. Gu, H.; Huang, J.; Hung, L.; Chen, X. A. Lessons Learned from Designing an AI-Enabled Diagnosis Tool for Pathologists. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 1−25. CSCW1, Article 10, 25 pages. https://doi.org/10.1145/3449084.
107. Balogh, E. P.; Miller, B. T.; Ball, J. R. *Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, & the National Academies of Sciences*; National Academies Press: US, 2015.
108. Bleher, H.; Braun, M. Diffused Responsibility: Attributions of Responsibility in the Use of AI-Driven Clinical Decision Support Systems. *AI Ethics* **2022**, *2*, 1−15.
109. Chakrobartty, S.; El-Gayar, O. Explainable Artificial Intelligence in the Medical Domain: A Systematic Review. *Americas Conference on Information Systems (AMCIS) 2021 Proceedings*, 2021.

110. Barnett, A. J.; Schwartz, F. R.; Tao, C.; Chen, C.; Ren, Y.; Lo, J. Y.; Rudin, C. A Case-Based Interpretable Deep Learning Model for Classification of Mass Lesions in Digital Mammography. *Nat. Mach. Intell.* **2021**, *3* (12), 1061–1070.
111. Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M. F. Artificial Intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework and Future Research Agenda. *J. Ambient Intell. Hum. Comput.* **2022**, *14*, 1–28.
112. Bahl, M. Artificial Intelligence in Clinical Practice: Implementation Considerations and Barriers. *J. Breast Imaging* **2022**, *4* (6), 632–639.
113. Gamkrelidze, T.; Zouinar, M.; Barcellini, F. Working with Machine Learning/Artificial Intelligence Systems: Workers' Viewpoints and Experiences. In *Proceedings of the 32nd European Conference on Cognitive Ergonomics (Siena, Italy) (ECCE '21)*; Association for Computing Machinery: New York, NY, USA, 2021. Article 14, 7 pages.
114. Ghassemi, M.; Oakden-Rayner, L.; Beam, A. L. The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care. *Lancet Digit. Health* **2021**, *3* (11), e745–e750.
115. Bossen, C.; Pine, K. H. Batman and Robin in Healthcare Knowledge Work: Human-Ai Collaboration by Clinical Documentation Integrity Specialists. *ACM Trans. Comput.-Hum. Interact.* **2023**, *30* (2), 1–29. Article 26, 29 pages. https://doi.org/10.1145/3569892.
116. Ayobi, A.; Hughes, J.; Duckworth, C. J.; Dylag, J. J.; James, S.; Marshall, P.; Guy, M.; Kumaran, A.; Chapman, A.; Boniface, M.; O'Kane, A. A. Computational Notebooks as Co-Design Tools: Engaging Young Adults Living with Diabetes, Family Carers, and Clinicians with Machine Learning Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*; Association for Computing Machinery: New York, NY, USA, 2023. Article 300, 20 pages.
117. Walsh, G.; Wronsky, E. AI + Co-Design: Developing a Novel Computer-Supported Approach to Inclusive Design. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing (Austin, TX, USA) (CSCW '19 Companion)*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 408–412.
118. Yildirim, N.; Pushkarna, M.; Goyal, N.; Wattenberg, M.; Viégas, F. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*; Association for Computing Machinery: New York, NY, USA, 2023. Article 356, 13 pages.
119. Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2007.
120. Grießhaber, N. M.; Mörike, F. Von Bedrohung bis kollaborative Partnerschaft: Eine qualitative Studie zu den Rollen KI-basierter Technologie aus der Perspektive niedergelassener Radiolog*innen in Deutschland Threat or Collaborative Partnership? A Qualitative Study on Roles of AI-Based Technology from the Perspective of Practicing Radiologists in Germany. In *Proceedings of Mensch Und Computer 2021* (Ingolstadt, Germany) (MuC'21). Association For Computing Machinery, New York, NY, USA, 2021; pp. 69–78.
121. Patel, R. S.; Bachu, R.; Adikey, A.; Malik, M.; Shah, M. Factors Related to Physician Burnout and its Consequences: A Review. *Behav. Sci. (Basel)* **2018**, *8* (11), 98.
122. Blezek, D. J.; Olson-Williams, L.; Missert, A.; Korfiatis, P. AI Integration in the Clinical Workflow. *J. Digit. Imaging* **2021**, *34* (6), 1435–1446.
123. Calisto, F. M.; Santiago, C.; Nunes, N.; Nascimento, J. C. BreastScreening-AI: Evaluating Medical Intelligent Agents for human-AI Interactions. *Artif. Intell. Med.* **2022**, *127*, 102285.
124. Friedrich, A. B.; Mason, J.; Malone, J. R. Rethinking Explainability: Toward a Postphenomenology of Black-Box Artificial Intelligence in Medicine. *Ethics Inf. Technol.* **2022**, *24* (1), 1–9.
125. Alberdi, E.; Povykalo, A.; Strigini, L.; Ayton, P. Effects of Incorrect Computer-Aided Detection (CAD) Output on Human Decision-Making in Mammography. *Acad. Radiol.* **2004**, *11* (8), 909–918.
126. Cabitza, F.; Campagner, A.; Sconfienza, L. M. Studying human-AI Collaboration Protocols: The Case of the Kasparov's Law in Radiological Double Reading. *Health Inf. Sci. Syst.* **2021**, *9* (1), 8.
127. Fogliato, R.; Chappidi, S.; Lungren, M.; Fisher, P.; Wilson, D.; Fitzke, M.; Parkinson, M.; Horvitz, E.; Inkpen, K.; Nushi, B. Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (Seoul, Republic of Korea) (FAccT '22)*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 1362–1374.
128. Herse, S.; Vitale, J.; Johnston, B.; Williams, M.-A. Using Trust to Determine User Decision Making & Task Outcome During a Human-Agent Collaborative Task. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (Boulder, CO, USA) (HRI '21)*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 73–82.
129. Hong, S. R.; Hullman, J.; Bertini, E. Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs. *Proc. ACM Hum.-Comput. Interact.* **2020**, *4*, 1–26. CSCW1, Article 68, 26 pages. https://doi.org/10.1145/3392878.
130. van Berkel, N.; Ahmad, O. F.; Stoyanov, D.; Lovat, L.; Blandford, A. Designing Visual Markers for Continuous Artificial Intelligence Support: A Colonoscopy Case Study, *ACM Trans. Comput. Healthcare*, **2021**, *2* (1), 1–24. Article 7, 24 pages. https://doi.org/10.1145/3422156.
131. Yang, Q.; Hao, Y.; Quan, K.; Yang, S.; Zhao, Y.; Kuleshov, V.; Wang, F. Harnessing Biomedical Literature to Calibrate Clinicians' Trust in AI Decision Support Systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*; Association for Computing Machinery: New York, NY, USA, 2023. Article 14, 14 pages.