Rethinking the Role of AI with Physicians in Oncology

Revealing Perspectives from Clinical and Research Workflows

Verma, H.; Mlynar, Jakub; Schaer, Roger; Reichenbach, Julien; Jreige, Mario; Prior, John; Evéquoz, Florian; Depeursinge, Adrien

DOI
10.1145/3544548.3581506

Publication date
2023

Published in
CHI 2023 - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems

Rethinking the Role of AI with Physicians in Oncology: Revealing Perspectives from Clinical and Research Workflows

Himanshu Verma
Delft University of Technology
Delft, Netherlands
H.Verma@tudelft.nl

Jakub Mlynar
HES-SO Valais-Wallis
Sierre, Switzerland
jakub.mlynar@hevs.ch

Roger Schaer
HES-SO Valais-Wallis
Sierre, Switzerland
roger.schaer@hevs.ch

Julien Reichenbach
HES-SO Valais-Wallis
Sierre, Switzerland
julien.reichenbach@hevs.ch

Mario Jreige
Lausanne University Hospital
Lausanne, Switzerland
Mario.Jreige@chuv.ch

John Prior*
Lausanne University Hospital
Lausanne, Switzerland
John.Prior@chuv.ch

Florian Evéquoz
HES-SO Valais-Wallis
Sierre, Switzerland
florian.evequoz@hes-so.ch

Adrien Depeursinge†
HES-SO Valais-Wallis
Sierre, Switzerland
adrien.depeursinge@hevs.ch

ABSTRACT

Significant and rapid advancements in cancer research have been attributed to Artificial Intelligence (AI). However, AI's role and impact on the clinical side has been limited. This discrepancy manifests due to the overlooked, yet profound, differences in the clinical and research practices in oncology. Our contribution seeks to scrutinize physicians' engagement with AI by interviewing 7 medical-imaging experts and disentangle its future alignment across the clinical and research workflows, diverging from the existing "one-size-fits-all" paradigm within Human-Centered AI discourses. Our analysis revealed that physicians' trust in AI is less dependent on their general acceptance of AI, but more on their contestable experiences with AI. Contestability, in clinical workflows, underpins the need for personal supervision of AI outcomes and processes, i.e., clinician-in-the-loop. Finally, we discuss tensions in the desired attributes of AI, such as explainability and control, contextualizing them within the divergent intentionality and scope of clinical and research workflows.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing methodologies** → *Philosophical/theoretical foundations of artificial intelligence.*

KEYWORDS

AI in Oncology, Clinical Adoption of AI, Imaginaries, Explainability, Contestability, Human-Centered AI, Human-In-The-Loop AI

ACM Reference Format:
Himanshu Verma, Jakub Mlynar, Roger Schaer, Julien Reichenbach, Mario Jreige, John Prior, Florian Evéquoz, and Adrien Depeursinge. 2023. Rethinking the Role of AI with Physicians in Oncology: Revealing Perspectives from Clinical and Research Workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3581506

1 INTRODUCTION

In cancer care, accurately identifying patient-specific cancer subtype, stage and progression as well as response to therapy is crucial for precision (or personalized) medicine, i.e., providing and adapting the most appropriate treatment to the patient at the right time. To achieve this, precision medicine relies on multiple sources of patient data including clinical parameters, histopathology and gene expression of tumoral tissue as well as imaging. Multidisciplinary Tumor Board (MTB) meetings are organized to interpret the various sources of information and achieve consensus on the optimal treatment decision. Recently, the use of quantitative methods for the aforementioned data sources led to the development of *multi-omics* – analysis of multi-modal medical data that can better stratify between cancer sub-types and their patient-specific phenotypes. However, the complexity and the high dimensionality of this multi-omics data cannot be adequately leveraged by a single physician. For instance, *radiomics* is a technique for extracting quantitative biomarkers from a radiology or nuclear medicine image and can result in more than 1000 dimensions from a single tumor region [1]. Therefore, Artificial Intelligence and Machine Learning (AI-ML) methods –which can extract large-scale quantitative features from medical imaging [37]– are required to fully leverage the multi-omics wealth and reveal the high-dimensional pattern signatures of cancer genotypes and their phenotypes.

*Secondary Affiliation: University of Lausanne, Lausanne, Switzerland.
†Secondary Affiliation: Lausanne University Hospital, Lausanne, Switzerland.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581506

Although an increasing number of studies showed the potential of AI approaches for optimizing cancer care [78, 89], sometimes outperforming human observers [26, 77], very few models have made their way into clinical practice [18, 25, 32]. In terms of reliability and generalizability, these AI models were regarded as problematic for a safe clinical application [72]. This discrepancy in adoption of AI in clinical settings has been attributed to varied factors, including, poor contextual fit and mismatch with clinical workflows [9, 25]. To address these fundamental problems, calls have been made for HCI and AI communities to establish deep and meaningful collaborations with healthcare domain, and more importantly, healthcare professionals [28, 57, 70, 83]. In our article, exemplifying one such collaboration, we focus specifically on the unconsidered role of physicians as either *clinicians* or *researchers*, and the overlooked aspects of *clinical* and *research practices* in the development of AI. We hypothesize that the inherent differences in the clinical and research practices are in tension with the current role and focus of AI in oncology. So, a meaningful and more adapted development of AI would entail pursuing a human-centric approach, where the incorporation of AI systems in existing –clinical and research– workflows is not only taken into consideration, but also, is in synergistic agreement with physicians' *imaginaries* [12, 66] of what they want AI to be in oncology. In addition, we argue that HCI's collaborative entanglements with AI research, and its intermediary, yet instrumental, positioning between users and (digital) technology designers, render it a suitable domain to study the aforementioned problem.

To this end, we present a qualitative inquiry where we interviewed imaging experts in oncology, aiming to capture their imaginaries of AI, and how do these differ across their practice as clinicians and researchers. This entailed eliciting fine-grained aspects of physicians' *a)* therapeutic role in the treatment of cancers and the temporality of their engagement with the patients, *b)* divergent aspects of their clinical and research practices, routines, and policies, and *c)* entanglement with AI along with the tensions that emerge due to misaligned needs, experiences, and expectations. In this way, we also aspire to provide a discourse, grounded in existing cancer care ecosystem, on the viable pathways for rethinking AI's focus and vision for augmenting oncological workflows. Our analysis revealed that inherent differences in existing cancer care workflows –with varying intentionality, scope and temporality– engender tensions between AI's imagined role and its impact in oncology. We discuss means of lowering the barriers of AI adoption in clinical settings by eliciting and coalescing desired qualities of AI from the imaginaries of physicians. We observed that physicians' trust in AI is independent of their general acceptance of AI, instead, it is grounded in their contestable experiences with AI. Furthermore, notions of ethics, responsibility, control, decision-making, and explainability with respect to AI are subtly different across a wide range of beneficiaries and stakeholders.

We contribute to the disciplines of HCI and AI, and more broadly to the emerging notions of human-centered and human-in-the-loop AI, by providing empirical insights about physicians'[1] engagements with AI[2], their perceptions and projections regarding the future role and impact of AI in oncology, and reflections about meaningful realignment of AI's role and expectations.

In this paper, after consolidating the research gap following a literature survey in Section 2, we illustrate our method to address the gap in Section 3. Our findings and analyses are described in Section 4 followed by a presentation of broader perspectives in Section 5 discussing our contribution, drawing conclusions and outlining the future discourses for human-centric AI in cancer care.

2 RELATED WORK

In this section, we first review the separate, yet crucial, contributions of AI and HCI in oncology and cancer care. Since our presented research engages with imaging experts in oncology, consequently, our use of AI specifically refers to the *image-based* AI systems – i.e., based on 3-dimensional anatomical and metabolic images at the macroscopic scale which are utilized in the successive phases of cancer therapy. Later, we present a survey of collaborative entanglements of both AI and HCI in supporting cancer care.

2.1 Image-Based AI in Oncology

In the past years, we have seen an exponential interest in using AI on clinically acquired radiology and nuclear medicine images for the purpose of diagnosis and prognosis [18, 90]. More specifically, the systematic review and meta-analysis by Sollini et al. [89] provides an interesting snapshot of the field of research in AI and radiomics in early 2019. The authors filtered articles with the highest quality score (based on the QUADAS-2 criteria). Of the 171 articles selected, 86% (147) focused on oncological applications. Within this subset, the majority, i.e., 56% (83), were concerned with brain and lung cancer, primarily for predictive outcome modeling and biological characterization. Although the authors observed an increase in the overall quality of clinical studies over the years, all approaches were still far from clinical adoption, where the black-box effect, limited sample size, and ethics and liability issues were pointed out as the major hindering factors.

Although the majority of studies have focused on retrospective population/subject level evaluation of the developed models in the clinical research setting, only a few studies have addressed the problem of using the model in a clinical setting with patient-level predictions. Within this small subset, the focus has primarily been on alleviating the black-box effect to improve interpretability, specifically in the context of deep learning models [79, 101]. This increased interpretability has been argued as a significant pathway towards improved –overall– acceptance of AI models among physicians [79].

In addition to enhancing interpretability of AI models, other means of improving their adoption in clinical contexts have been discussed. Pianykh et al. [72] have argued that as the performance of static AI algorithms degrades over time –owing to the naturally occurring changes in local data and environment– it is crucial to adopt an early application of continuous learning for AI in clinical

[1]It is worth noting that our use of the term "*physician*" refers specifically to medical imaging experts within the domain of Oncology (e.g., Radiologists, Radiation Therapists, Nuclear Physicians, and Radiation Oncologists).

[2]Our usage of the terms related to AI (e.g., AI-powered tools, models, classifiers), are specifically meant to be interpreted within the realms of quantitative imaging analysis and image-based (deep) machine learning models. Henceforth, we will use AI as a general term to encapsulate the diverse algorithmic systems, including Machine Learning (ML) and Deep Learning.

radiology routine. In addition, the authors suggest that following the replication of static models, which are imported from clinical research, they should be continually fed with clinical data and their performance and impact should also be continually monitored in the clinical context. In the specific context of predictive models in oncology, Gatta et al. [32] highlighted the importance of "holomics" (also called "medomics"), i.e., integrating radiomics with other *-omics* (e.g., genomics, proteomics) to increase the relevance of models and facilitate their migration to the clinical settings. Considering optimized radiation therapy planning, Thompson et al. [96] pointed out the necessity to adequately make room for AI within the clinical context through programs promoting education of end-users, data availability, and potential changes in clinical workflows for a harmonious integration of AI.

Clinical Decision Support Systems (CDSS), which encapsulate the overarching role and vision of AI within cancer care (and broadly applicable to every aspect of healthcare) have been a subject of significant attention and scrutiny in recent years. Even with CDSS, as with the specific applications of AI discussed earlier, the adoption within clinical settings when compared to clinical research, has been significantly low, if not a failure [57, 109]. This problem of lower clinical adoption has been ascribed to several factors, including, *a)* poor contextual fit within the healthcare ecosystem [45, 104, 109], *b)* mismatch with existing clinical workflows and routines [17, 25], *c)* lack of confidence about decision support interventions among clinicians [25], *d)* misalignment with the needs, expectations, and concerns of clinicians [57], and *e)* challenges with ecological validation of CDSS (particularly image-based decision support) within the clinical context [8].

In our work, we scrutinize this discrepancy between the research and clinical applications of AI-enabled systems (including CDSS) in oncology by capturing physicians' imaginaries, with a particular focus on tensions between *'what AI is'* and *'what physicians want AI to be'*.

2.2 Human-Factors Research in Cancer Care

Seeking to improve the overall quality and experience of cancer care –not just for the patients who are being treated but also for their family and friends– numerous sociological studies have been conducted in the recent past. These contributions have examined the varied facets of cancer care facilities, such as *a)* patients' overall experience with cancer care [93, 94], *b)* perceived sense of stigma, guilt and depression amongst patients and how it impacts their psychological well-being and social interactions [4, 13, 24, 60], *c)* disparities in quality of life between cancer patients from different ethnic backgrounds [4, 16, 91], *d)* the role of doctor-patient interactions and communication strategies in patients' perceived well-being [6, 108], and *e)* the impact of established decision-making practices and the opportunities and challenges for technological interventions [55, 56, 71].

Contributions within the HCI community, similarly, have predominantly focused on patients and patient-centric aspects. Moreover, HCI's (along with its sub-domains) engagement and mission within the realm of oncology has been that of supporting patient's everyday experiences and journeys by alleviating the psychological, social, and logistical seams, which are the inevitable by-products of living with a cancer. Much of the past research within HCI has examined the role of health information systems as enablers for patients to monitor their health, engage with healthcare providers, and seamlessly seek support and assistance through personalized data-centric tools (e.g. [43, 46, 47]), as well as facilitate self-regulation and behavior change through persuasive technologies [50]. In addition, the research on health information systems has also been extended to consolidate training and education of caregivers, for instance, by promoting continuous learning *on-the-job* based on cancer type, treatment modality, and treatment phase [85].

Another strand of research within HCI has examined patients' collaborative engagement with cancer care and their experiences navigating it. Jacobs et al. [44] studied patients' journey, particularly, their collaborations with caregivers and other stakeholders while managing the overwhelming amount of medical, financial, and emotional challenges. The authors further examined the design space and provided guidelines for improving the patient experience and mitigating logistical challenges by improving access to relevant resources. Gonzales and Riek [34] assessed the role of patient-centered health communication tools to mitigate communication breakdowns (e.g., misunderstandings, alternative beliefs) between oncologists and patients in the clinical settings. Moreover, living with cancer often manifests as a poor quality of psychological well-being for the patients. In this regard, Suh et al. [92] suggest employing an integrative approach that combines cancer and psychological care so that they are perceived indistinguishable by the patients. While a significant amount of HCI research has focused on adult patients, a recent work by Warren [106] investigated the means of leveraging social technologies to improve the emotional and social well-being of child cancer patients. She argues that children with cancer often feel a sense of isolation and loneliness, and develop beliefs of an abnormal childhood, which can be mitigated through "socially-focused" and "playful" technologies.

We observed a significant emphasis on patients and patient-centric inquiries (and interventions) within HCI's collaboration with cancer care. However, HCI's (and CSCW's) engagement with physicians and clinicians, and their respective workflows, routines, and norms is a much recent endeavor. In recent past, there have been growing calls within the HCI and CSCW communities to establish deep and meaningful relationships with the healthcare domain, and particularly, with healthcare professionals (e.g., [49, 69, 95]). Still, only a handful of studies exemplify HCI's collaboration with healthcare professionals (e.g., diabetes [109], mental health [45], stroke rehabilitation [57], prostate cancer [9]). Moreover, in their survey of research at the intersection of CSCW and healthcare, Fitzpatrick and Ellingsen [28] have distilled *four* guiding principles for the HCI/CSCW community recommending *(1)* alignment of HCI/CSCW's agenda with the emerging ICT initiatives in healthcare to enable unique opportunities for cross-fertilization, *(2)* assimilation of HCI/CSCW approaches within diverse clinical practices, contexts, and temporalities, *(3)* consideration of the patient as the key beneficiary, and *(4)* identification of intersections where non-healthcare interventions can be applied to the healthcare domain. In this regard, our work exemplifies one such collaboration between HCI and healthcare professionals (i.e., points *(1)* and *(2)* from aforementioned principles), where we contribute by bringing

contextualized insights about physicians' –clinical and research– practices and their entanglements with AI.

2.3 HCI and AI Collaborations in Cancer Care

In a 2019 workshop at CSCW, Park et al. [70] discussed the opportunities and challenges underlying the collaboration of HCI/CSCW and AI communities in healthcare, its nature and affordances, and the broader –sociological, normative and ethical– implications of integrating AI in healthcare. In cancer care, however, the collaboration between the HCI and AI communities has primarily manifested itself in tools that augment and support diagnostic workflows with AI-powered assistants (e.g., [9, 10, 36, 111]). Although computer-assisted diagnosis has been identified as just one of several application areas where AI can assist in cancer care [79], there is still an inherent misalignment in the focus and implementation of AI in cancer care, as a significant majority of systems/interventions are being developed for cancer diagnosis, overlooking other clinical aspects where the role of AI may be highly desirable (e.g., therapy planning, monitoring, triaging, etc.).

Another strand of work within diagnostic workflows has examined the co-performative nature of physician-AI teams and contrasted it with physician-only diagnosis [10, 36]. For example, Calisto et al. [10] observed a significant reduction in false-positive rates in the AI-assisted physicians compared to the diagnosis prepared without the assistance of AI. Building upon such optimistic evidence and the effectiveness of human-AI teams, AI is being framed as a "supertool" that will enhance human capabilities rather than replace them altogether [87, 107]. Zhou et al. [111], however, argue that the realization of such desired effectiveness of human-AI teams in cancer care is conditional upon a well-aligned and proper integration of AI into clinical practices and routines. In addition to the empirical evidence favoring the effectiveness of physician-AI collaboration, a need for physician's oversight –*physician-in-the-loop*– in the use of AI-powered self-diagnosis apps for end-users was also demonstrated in the survey conducted by Baldauf et al. [5]. The authors found that several aspects underpin the need for physician's oversight in the use of these apps, particularly, the *a)* explanation of domain-specific jargon, *b)* interpretation of (diagnostic) results, and *c)* more importantly, identification of (alternative) treatment possibilities.

The overemphasis in efforts supporting only diagnostic workflows and lack of augmentation for other "*unremarkable*", yet crucial, aspects of clinical work, risks poor integration within clinical settings and rejection of AI among physicians [109]. Furthermore, this imbalanced and biased positioning of AI within cancer care manifests due to the lack of active participation of diverse actors (i.e., physicians, technicians) in AI development, and a better understanding of their practices and needs. This discrepancy is emphasized in the article "Machine Behaviour" by Rahwan et al. [76], who argue that "the scientists who study the behaviours of these [AI agents] are predominantly the same scientists who have created the agents themselves", implying a lack of cross-disciplinary collaboration in AI development. Mlynar et al. [66] further provide grounds for an inclusive, participatory, and rather sociological development of AI, which is different from the "current market-led visions" and is not solely "founded within the [computer and data science] community".

In our contribution, we seek to address the aforementioned discrepancies in the existing focus of AI, and create meaningful routes for a balanced and aligned development as well as a more fitting and desired positioning of AI in cancer care. To this end, we interview physicians who are actively engaged in both research and clinical activities, in order to capture tensions in their engagements with AI across their –research and clinical– practices with divergent intentionality and scope. In addition, such a development should consolidate the interplay between the diverse actors, existing practices, organizational norms, and industry standards [86], in order to lower, if not eliminate, barriers to the adoption of AI in cancer care. Consequently, we also capture the broader collaborative entanglements of physicians in our interviews.

2.4 Research Gaps

As mentioned earlier, there exists a disconnect in the current focus of AI research and the existing practices and protocols within medical institutions with established cancer treatment facilities. In particular, through our literature survey, we have unveiled the subtle underpinnings of this disconnection, i.e., *1)* underwhelming adoption of AI within clinical settings as compared to cancer research [8, 25, 31, 32], *2)* shortage in HCI's collaborations with cancer care professionals which could potentially unravel the contextual intricacies (i.e., clinical practices, protocols, and norms) [28, 69, 70], and *3)* unbalanced and biased positioning of existing clinical AI, which largely focuses on augmenting diagnostic workflows while ignoring other facets of cancer care processes [9, 79, 109, 111]. Furthermore, this disconnect manifests itself in overlooked aspects of medical professionals' practices and experiences that can be better served through the *human-centric* design and development of AI-powered systems. These overlooked aspects are primarily related to the profound differences between *clinical* and *research* practices, and consequently to the divergent needs of physicians, which are currently not supported by existing AI systems. We believe that a nuanced understanding of these differences and tensions that arise in physicians' entanglements with AI on multiple levels (utilitarian, ethical, normative) could pave the way for balanced and fitting AI development.

To address this gap, we extend the approach developed by Mlynar et al. [66], which explores expert groups' imaginaries of AI and accounts for their recurrent patterns, as well as conspicuously absent links between the elicited topics. Borrowed from philosophy and the social sciences, *l'imaginaire* (in French) or the *imaginary* refers to a "shared network of concepts, images, stories, and myths that make possible common practices and provide a sense of legitimacy" [12, p. 2]. Focusing on experts' imaginaries can enhance our understanding of AI as a social phenomenon, rather than a narrowly defined technical "tool". Similarly to the notion of *socio-technical imaginaries* introduced by Jasanoff and Kim [48], we take as our "starting point the resurgence of theoretical interest in the nature of collective self-understandings", but we do not aspire to describe developments at the societal macro-level. We rather aim to explore how AI is made part of the everyday work of oncology

experts, what AI means to them in general, and what aspects of working with such systems remain troublesome.

3 METHODS

We aim to scrutinize the *unbalanced* positioning and impact of AI across *clinical* and *research* practices within cancer care, and to elicit tensions in physicians' entanglements with AI across these divergent facets of their workflows by capturing their imaginaries about *'what AI currently is'* and *'what they want AI to be'*. In this article, our focus is primarily on physicians who use image-based AI in oncology. Their work may include both "handcrafted" radiomics and deep learning models addressing the following five application categories defined by Reyes et al. [79]: *a)* computer-assisted diagnosis and/or staging, *b)* prognosis, *c)* radiation therapy planning, *d)* computer-assisted monitoring of disease progression, and *e)* triaging. Furthermore, these image-based AI systems are increasingly based on the computerized analysis of imaging modalities, in particular Computed Tomography (CT), Positron Emission Tomography (PET), and Magnetic Resonance (MR) images, which is often referred to as "*quantitative imaging analysis*" in the domain of oncology.

To address the research gaps previously illustrated in Section 2.4 about the mismatched alignment of AI within cancer care, and to disentangle its future alignment across the clinical and research workflows, we aim to *(a) develop* a comprehensive understanding of established protocols in cancer treatment facilities, including both "clinical" and "research" aspects; *(b) scrutinize* the role and impact of AI in cancer care, especially its perceived utility, scope, and acceptability by physicians; *(c) bring* insights about physicians' interactions with patients and AI, and the collective engagement of diverse medical specialties in the effective treatment of cancers; and finally *(d) depict* the space of design possibilities where intelligent and interactive tools may support the physicians in meaningful ways. Consequently, we conducted a qualitative –semi-structured interview– study with 7 cancer experts (physicians) belonging to different sub-specialties. Our study draws methodological inspiration from the work of Wang et al. [105], who examined the perspectives and experiences of data science experts while collaborating with automated AI pipelines.

We, initially, intended to conduct face-to-face interviews with physicians along with a field study to gain insight into the established workflows and the various resources (medical images, reports, and instruments) that are conventionally used during the course of cancer treatment. However, the constraints imposed following the onset of the COVID-19 pandemic steered us to conduct the study over video conferencing, and also curtailed our efforts to conduct a field study. Despite these constraints, the design of our interviews sought to capture the aforementioned aspects of physicians' work practices, experiences, and interactions in a fine-grained manner.

3.1 Interview Study

3.1.1 Participants. Initially, a formal email of invitation was sent to 10 physicians working at the Lausanne University Hospital (CHUV) who are involved in the diagnosis, treatment, and rehabilitation of cancer patients, and are simultaneously engaged in research activities. Seven physicians agreed to participate in our study. Lausanne University Hospital (CHUV) is one of the 5 academic hospitals in Switzerland with over 12,000 employees, and plays a key role in healthcare, research, and training. It has 1,000 beds and serves a population of approximately 800,000 inhabitants. It has several departments dedicated to cancer care, amongst which *(a)* the **Nuclear Medicine and Molecular Imaging** department has over 30 personnel (physicians, fellows, technicians), 2 PET/CT scanners, and performs around 7,000 PET/CT scans per year on approximately 4,500 patients; *(b)* the department of **Diagnostic and Interventional Radiology** has 24 senior physicians and 35 fellows with a clinical activity of about 33,000 CT and 26,000 MRI scans per year (on 4 CT and 4 MRI scanners); *(c)* the department of **Radio-Oncology** employs 42 personnel and provides over 19,000 treatments to approximately 1,550 patients; and *(d)* the **Precision Oncology Center**, which provides personalized approaches to cancer treatment based on real-world clinical data, has over 20 personnel and access to data of over 90,000 patients. We obtained an ethics approval from our institution's Human Research and Ethics Committee (HREC) before the commencement of the study.

We reached out to medical professionals whose expertise lie within the specialties of medical oncology, nuclear medicine, radiology, and radiation oncology. These specialists predominantly use different imaging modalities to track the development of tumors and metastases. They also incorporate algorithmic means to analyze large amounts of image data during the treatment process. Moreover, these experts are also accustomed to leveraging AI in their clinical and research activities. This was the primary rationale behind involving them in our study. The invitation letter contained *1)* the general objective of our research without revealing the specifics of the questions we intended on asking to prevent biasing their responses, and *2)* the description of our research approach including the approximate time required on the participants' part. In addition, participants were informed that we would acknowledge them properly in our research publications and that there would be no financial compensation.

In addition to the aforementioned specialties, surgery, immunotherapy, chemotherapy, pathology, etc. are other sub-specialties which are involved in the treatment of cancers. However, they were not included in our study because they are less likely to interact with AI systems in the course of treatments. Even though these domains were not part of our study, still it is worth noting that the established medical protocols require a collaborative effort (e.g., weekly Multidisciplinary Tumor Board meetings) amongst diverse range of specialties to decide on the most effective and personalized treatment for each patient. Capturing these collaborative aspects was one of the key focus of our inquiry.

Upon receiving an affirmative response, we requested that our participants agree to an informed consent document that clearly stated how we would use and analyze the collected data, and also sought their permission to record video of the discussions during the interviews. Our interview study involved cancer experts who are highly busy professionals, with additional responsibilities besides cancer care during the (COVID-19) pandemic. Six interviews were conducted over Zoom, and one expert responded to the interview questions in writing, over Google Docs, due to availability

Participant	Sub-specialization	Role	Experience (years)
P01	Nuclear Medicine & Molecular Imaging	Full Professor & Director	>20
P02	Nuclear Medicine & Radiology	Resident & Clinic Head	5
P03	Radiology	Full Professor	24
P04	Medical Oncology	Full Professor	15
P05	Radiation Oncology	Fellow & Instructor	5-10
P06	Radiation Oncology	Associate Professor	>20
P07	Nuclear Medicine & Radiology	Fellow & Senior Lecturer	11

Table 1: The table illustrates the sub-specialties of our interviewees along with their role and experience (i.e., years of practice). When asked about their experience, some of our interviewees gave us with an exact duration, while others gave us a range. It is worth noting that all of our interviewees are involved in both the clinical and research activities.

constraints. Table 1 lists the participants, their sub-specialties, role within their respective departments and their experience.

Although, from a perspective of quantitative research, the number of interviews in our study could be considered a limitation, we ground our work in the standards of qualitative research as formulated, for instance, in Burawoy's assertion that "[o]ne cannot dismiss [qualitative researcher's method] because she alters the world she studies, because her data are idiosyncratic, because she extends out from the local to the extra-local, or because she only has a single case! The method simply dances to another tune" [7]. In other words, the absolute number of individual interviewees is not necessarily relevant in qualitative research – what *counts* is the variety, regularity and integrity of themes elicited from the interviews conducted. We focused on physicians who frequently use AI in their clinical and research routines and are actively contributing to the frontier research in medical imaging and oncology. With this professional context as our research subject, and in line with established principles of interview research [30], we have achieved data saturation in terms of gathering concordant experiences with the use of AI, emerging needs and expectations, and recommendations for a ethical and balanced development of AI.

3.1.2 Procedure. The lead author conducted the interviews in a semi-structured manner. Moreover, the interview questions were organized into the following *three* themes:

(1) *Role and Background:* The initial set of questions were intended to "*break the ice*" and asked participants about their medical speciality, their role in the diagnosis and treatment of cancers, and the approaches they employ in the treatment process.
(2) *Interactions with Patients and Relevant Protocols:* The next set of questions regarded the participants' interactions with the patients, and the overall journey of a patient from being diagnosed with a cancer to treatment and rehabilitation. Particularly, at what stage of the treatment do they (participants) get involved, and the nature of their involvement. How do they take decisions about the most effective course of treatment? Besides them, what other sub-specialties are involved in the decision making process? What is the nature of this collaboration, and how often does it happen? What kind of artefacts (reports, visualizations, medical images, etc.) are shared and discussed during these collaborations? Moreover, we also gathered participants' opinions and experiences regarding how technology –in particular AI– has influenced the evolution of their domain and the ingrained constructs and protocols. Where necessary, we asked our participants to further elucidate how AI's role and impact has manifested across their clinical and research workflows.
(3) *Impact and Scope of AI:* The last part of the interview was focused on the participants' interactions and experiences with AI (manifesting as either algorithmic procedures or tangible instruments), and how they are achieved in their everyday work – either clinical or research oriented. We asked questions which could provide us with fine-grained insights about participants' practices which are currently supported by AI, and the diverse ways in which AI impedes the seamless attainment of their objectives. Finally, the participants were asked to project and reflect on the ways in which AI may shape the future of cancer care.

It should be noted that despite our interest in the aforementioned topics, we did not directly ask interviewees about existing tensions or conflicts in their interactions with AI between their clinical and research workflows. This was done to prevent our interviewees from projecting their preferential judgements into their responses. Moreover, our methodological preference was to capture their perspectives, experiences and opinions impartially, and avoid setting an expectation that the interview focused on the comparative analysis of the aforementioned tensions between workflows. Still, the interviewer asked follow-up questions to elicit the discrepancies in the use and impact of AI across the two workflows.

The interviews lasted for approximately one hour and were video recorded, except for one interview where the participant wrote the responses in a shared document. All the interviews were transcribed by one researcher. After the transcription process, three researchers performed the open coding to determine recurrent topics that were identified through the repeated reading of the transcripts. The same three researchers, later, collectively interpreted and consolidated codes during axial coding of 4 (out of 7) interviews into relevant categories, following which, two researchers completed the axial coding of remaining interviews.

3.2 Interview Coding and Identified Themes

The approach to coding and analysis employed in this paper is aligned with the inherent characteristics of qualitative research,

Code	Description
`Clinical vs. Research Practices`	Passages of the interviews in which the experts describe their clinical and/or research procedures as part of their everyday work.
`Decision Instruments`	Established decision-making processes and institutions, such as multi-disciplinary tumor boards (MTBs, see [61]).
`Improving Clinical Decision Making`	Interviewee's commentary on decision-making processes regarding potential improvement of practices.
`Promoting AI in Cancer Care`	Ideas on what can be done to advocate and promote AI-powered tools in oncology.
`Attitudes towards AI`	Opinions and approaches to AI in general, including positive/negative assessments and evaluations.
`AI Projections`	Expressed opinions about AI that were formulated as positive, future-oriented assessments.
`Ruptures and Issues with AI`	Zones of potential failure of AI in cancer care, e.g., issues surrounding the use of black boxes, explainability, poor clinical adoption.
`Legal Aspects`	Problems related to accountability for diagnostic decisions based on AI-based tools.
`Ethics of AI`	Topics related to the ethical aspects of AI, such as safety and privacy issues, social justice, or the reproduction of inequalities.
`Humanitarian Aspects`	Topics related to the overall organization of cancer care, health care in general, and its societal functions.
`Education and Digital Literacy`	AI as a teachable element in education; learning about algorithms, data, main principles of AI and their important limitations.
`Data and Data Governance`	Discussions pertaining to the collection and maintenance of datasets containing quantified information about patients and their personal/health situations.

Table 2: The code-book that was created from the analysis of interviews though discussions and consensus in two sessions organized amongst the co-authors. In addition to the codes, the table also provides descriptions of them.

i.e., the fact that its "design, fieldwork, and data collection are most often provisional, emergent, and evolutionary processes" [81]. Our analysis was informed by the principles of content analysis [23, 100]. Epistemologically, our research also took guidance from other related approaches in qualitative research [82], regarding the way interviews are treated and analysed in order to progressively arrive at conclusions about the studied social phenomena. This included taking into account the reflexive relationship between "the words" and "the world" [42], or between the "inside" and the "outside" of the interview [64]. In our research approach, we gradually moved from reading the verbatim transcripts of the *interview* recordings, to working with inductively derived *codes*, to assembling broader recurring *themes* that structured our analysis [99].

In coding the transcribed interviews, we followed the essential methodological premises [14], aiming to develop codes that describe the content of the speakers' talk, aspects that are taken for granted, and the intertwining of structure and context in monitoring, maintaining, preventing, or transforming actions and utterances produced within the interview (according to [33]). After the open coding phase, the researchers organized two sessions to consolidate the codes into relevant categories and themes [22]. This was done to reflect on how meaning is produced on the basis of interview transcripts through the coding process, considering recent critiques that caution against equating qualitative analysis with "coding data" in a simplistic and formalizing way [73]. During this phase, relevant segments of the interviews were aggregated into categories and further compared to identify common approaches to discussed topics among the participants [11]. Table 2 illustrates our code book which describes the topical structure and content of the interviews.

4 FINDINGS

In this section, we describe our findings based on the analysis of the semi-structured interviews. In sections 4.1 and 4.2, we illustrate the inherent differences in the clinical and research workflows, particularly decision making, which currently influence the underwhelming adoption of AI in clinical contexts (see Research Gaps in Section 2.4). In sections 4.3 and 4.4, we present analyses related to the research gap concerning the unbalanced and biased positioning of AI and the ensuing tensions in clinical and research practices.

4.1 Clinical *vs.* Research Routines

We asked the interviewees to describe their role and responsibilities within the university hospital (see Section 3.1.2). They provided a detailed account of the dual nature of their work – i.e., *1)* to ensure an optimal and customized care path for cancer patients which is "*highly reliant and derived from the national and international guidelines*" (P05), and *2)* the research component, which aims to contribute to the development of new knowledge and the advancement of the domain of oncology through novel means. Although

entwined, the clinical and research practices essentially differ in their **intentionality** and **scope**.

4.1.1 Differences in "Intentionality" and "Scope". The consolidated verbal accounts of our interviewees reveal that clinical practices are recognized for their "curative intent" and an "individualistic scope" that puts the spotlight on a single patient whose treatment is very specific and based on numerous attributes such as the type of cancer, anatomical and physiological peculiarities, and medical history. Depicting the humane nature of the clinical activities, P05 stated that "*we are not treating a disease [but] we are treating a human being*", where "*we try to be very proactive and very preemptive in the management of side effects*". Moreover, the temporality of interaction with the patient is significantly higher and may range from a few weeks to several months, depending on the type of treatment (surgery, radiation therapy, or systemic treatment involving chemotherapy or immunotherapy) and the severity of the disease. This high level of interactivity with the patients further necessitates the establishment of an inclusive environment which does not just impose the most suitable treatment onto the patient, but engages them in the decision making process – "*the patient is always part of the solution*" (P01).

On the other hand, research practices are more dynamic in terms of outcomes and impacts, and their scope is broad and aimed at the respective research communities of our interviewees. They are manifested in comprehensive analysis of patient cohorts, involving the development of statistical and predictive models that amalgamate multiple streams of data coming from different diagnostic and treatment modalities. While a significant proportion of research activities are retrospective, these activities require minimal to no patient contact. In addition, although the research outcomes serve the long-term objective of advancing the domain of cancer therapy in the form of published articles, they can also serve a more immediate purpose of "*turning these [predictive models and data-centric insights] into clinical action*" (P04). However, this latter case corresponds to specific patients "*who are failing standard of care therapies*" (P04), and may require novel treatments which are highly personalized for the patients and based on the analysis of their molecular and microscopic profiles (such as genomics, proteomics, histopathology).

4.1.2 Differences in AI Needs and Analytical Practices. According to our interviewees, the fundamental differences between clinical and research practice also translate into different needs in terms of the nature and type of digital tools used by physicians, particularly their interactions and relationship with AI. The research activities primarily entail the analysis of "*hundreds or let's say thousands of features which are obtained from one exam or one organ*", and consolidate the features corresponding to "*imaging, metabolic, and morphological data*" (P02). To conduct the analysis of this immense volume of data, P04 elaborated that his team "*writes a lot of [their] own code in R and other languages, and [they] use a little bit of Google tool box*", and employs a "*mixed bag approach*", where "*there is not one major library (publicly available APIs) that [they] haven't used*".

On the contrary, the clinical practices employ "*very simple tools that enable a kind of home-made analysis*", which "*is strange because when we go to [medical conferences] we always see 3D images, in color, with multi-parametric quantitative analysis, and it's really not how it is in the clinical routine*" (P03). She further illustrates that although analytical practices on the clinical side integrate data from multiple sources (e.g., CT, PET, MRI), still "*it's a visual analysis and not a quantitative one*" (P03).

> "*We have some very basic parameters that probably are still the most important. Size of the tumor is very basic but it works. The tumor border is very important in oncology and quite difficult to assess because it's very subjective, and currently we don't have the possibility to do an automatic segmentation of the whole tumor. Another parameter is vascularization for some specific treatments, and currently we define vascularization subjectively, only through our eyes, and say if it is hyper-vascular or not. Similarly, by looking at images we say if the tumor is aggressive or not.*" (P03)

Underpinning the previous argument, P07 added that "*[diagnostic] decisions are mainly taken based on morphological analysis of images, while quantitative parameters serve as complementary information*", and since the "*interpretation of quantitative parameters may be challenging, [he] uses cutoff values that have been validated in the literature*". Additionally, P07 acknowledged the limitations with the use of thresholds and accounts for them in his diagnosis because "*cutoff values are often determined in very specific situations and cannot be extrapolated to all the cases we see in daily practice*". Finally, despite advances in image analysis approaches, the clinical analysis of medical imaging data is still "*archaic*" and done "*in a cross-sectional manner*" because "*[physicians] know the anatomy, physiology, and their normality*" (P01). Our interviewees' illustration of substantial differences in analytical practices between the research and clinical workflows is consistent with previous literature, which suggests misalignment with the local context [104] and lack of ecological validity in clinical settings [8] as the key contributors to this discrepancy.

Owing to the large amount of multi-modal data that needs to be analyzed for each patient on the clinical side (e.g., each PET/CT scan generates approximately 2000 images), combined with the need to serve a high volume of patients (40-50 patients per day), adds a significant amount of workload for the physicians and the support staff. Still, the preference to rely on one's own prior knowledge and experience, as opposed to autonomous AI-powered tools, when preparing a diagnosis and prognosis can be attributed in part to the notion of responsibility.

> "*The directors of hospitals and national health agencies require that a patient has the same care in every hospital, and we cannot base a diagnosis or treatment on two technologies that are not available everywhere ... That's why we cannot use some very innovative technology that is only available in one part of the world. So, it's a political health question and not a question of performance or utility.*" (P03)

In this section, our intention to provide a detailed account of the differences between clinical and research routines is to establish that these diverse work practices open up different spaces of design possibilities that HCI and AI researchers should consider when designing intelligent tools for this context. In the following sections, we will utilize this difference as a contextual lens to ground the

collaborative and decision making constructs, as well as the notion of ethics and responsibility in relation to AI.

4.2 Clinical Decision Support: Reality *vs.* Expectation

In clinical practice, deciding on the appropriate course of treatment for each patient entails the assessment of numerous aspects and the synthesis of multiple sources of information, as previously illustrated in Section 4.1. In addition, choosing amongst the various treatment modalities (surgery, radiation therapy, etc.) and, more importantly, deciding on their order or combination, which is highly customized for each patient, is a difficult task. "*We know now that cancer is a chronic disease, that means that the patient will live a long time with it. So, we need to decide which kind of treatment should be the first, second, or third.*" (P03). As a result, Multidisciplinary Tumor Boards (MTBs) have become a norm, and an essential practice for taking decisions in cancer treatment facilities worldwide.

All our participants unanimously affirmed the centrality of MTBs in clinical decision making concerning treatments and follow-ups. "*Since the last 10 years, all decisions are made in MTB meetings*" (P06). P03 further added that "*we cannot decide alone on a treatment plan, so the MTB meeting is mandatory for each new diagnosis of cancer, for each recurrent disease, and for each modification of the treatment*". In addition, "*in big institutions, [MTB] meetings are something that's very traditional and well established*" (P05). Furthermore, depending on the patient's status, the nature of the discussions may vary but their multi-disciplinarity is maintained, as illustrated by P07:

> "*Immediate informal discussions happen during the emergency phase to speed up the initial management and treat potential life threatening situations, however, when the patient is stabilized, and after the initial workup has been completed, the discussion is more formal in one or several MTBs.*" (P07)

Our interviewees stated that owing to the multi-disciplinary and collaborative nature of MTBs, different specialties are represented within them, such as surgery, oncology, radiology, radiation oncology, internal and nuclear medicine, and pathology. In addition, depending on the particular nature of the disease, other specialists such as neurologists, pulmonologists, cardiologists and pediatricians may be present at these meetings. Furthermore, "*major disease groups have their meetings weekly*" (P06). Moreover, "*[MTB] meetings are not patient based, and we generally discuss 10-20 patients in each meeting*" (P05). In response to a question about the established practices within these MTB meetings, our participants provided us with the following account. Prior to each meeting, participating members prepare a diagnostic workup of the patients in question, grounded in their medical specialty and based on the specific examinations they have performed. During the MTB meeting, the members present these analyses, either in the form of a report or by assembling a set of images from the hospital's Picture Archiving and Communication System (PACS). An assistant is also present in these meetings to register the minutes of the ongoing discussions and the collective decision.

In response to our questions aimed at comprehending the rationale behind the practice of MTB meetings, our interviewees furnished several insights. Illustrating an example, P01 stressed the importance of 'collective validity' of the most efficacious therapy for the patients: "*to make the MTB most efficient, each speciality has to say that this (referring to an example) therapy will work ... in a way, every speciality is then hand-in-hand doing the best they can for the patient*". This collective endeavor may also lead to "*reduced errors*" (P06) since multi-disciplinary collaborative analysis could complement individual assessments and assist in the development of a robust treatment plan. Extending his argument further, P06 explained that if each specialist pursued his/her therapy – "*a surgeon would like to operate every time, a radiation oncologist always wants to irradiate, and medical oncologist always wants to do chemotherapy*", the likelihood of reaching a better judgement would be low as this individualistic approach does not consolidate all aspects of a patient's well-being and medical history. Also, the MTB meetings provide a framework for enabling standardized and accepted treatments for the patients while mitigating the adverse effects of some investigational treatments: "*I am a radiation oncologist for 25 years, so, I mean, I can feel things, OK! And I can sometimes exaggerate. So, MTB meetings lets you to do, let's say, standard treatments*" (P06). Moreover, P05 maintained that the MTB meetings also provide a "*great educational opportunity*" that is afforded by "*the cross-talk between disciplines and to know how others approach this problem*". Finally, in terms of logistics, the MTB meetings also make things easier for the patients:

> "*Because if you don't have this common language or cross-talk, it's gonna be very easy for the patient to be lost in between sub-specialties. And the patient's appointments will not be coordinated. So, we are trying to spare the patients this hassle, the logistics of coming back-and-forth to the hospital.*" (P05)

The prevalence of MTBs, combined with the collective validity of their outcomes, confers on them the status of legitimacy. Consequently, decisions made in MTBs are recognized and often required by insurance companies, as illustrated by P06: "*In some countries, if there is no report of a MTB decision, the patient care, the insurance will not pay the treatment*".

In terms of clinical decision-making practices, particularly with regard to MTBs, our interviewees acknowledged that the role and impact of AI is currently limited and minimal. Although some form of predictive modeling or pattern recognition could be employed to prepare an analysis for presentation within MTBs, the scope is still relatively small and applies to rare diseases and novel treatment modalities. As a result, hospitals are constantly "*coalescing and curating huge datasets of hundreds, maybe thousands of patients*" (P04, P05) to allow for retrospective analysis and to support future decision making. In order to *a) extend* this data-informed decision support within MTBs, *b) increase* its accessibility to a higher number of patients, and *c) foster* the development of a synergistic educational environment for both physicians and AI experts, P04 who is leading the Precision Oncology program in his institution, is involving AI researchers as an additional speciality within the MTBs.

In summary, our findings exhibit tensions between the reality of clinical decision making within MTBs and the expectations and enthusiasm for the use of AI to support clinical decisions. Moreover, as previously discussed in section 2.1, the lack of adoption of AI

in clinical workflows –particularly in clinical decision making– can also be attributed to the recognition of MTBs as a legitimate decision-making instrument that has yet to assimilate AI as a new specialty.

4.3 Revealing the AI Adoption Gap

We asked the interviewees about their perceptions regarding the current role and impact of AI in cancer care, and how it might evolve in the future. In particular, how AI-powered technologies will shape their work practices and their relationship with patients. They expressed a positive disposition towards the utility of AI and its potential to transform cancer treatment and clinical practices in the future – "*the most interesting [AI] tools are going to come and help us in the future*" (P01) which will "*facilitate faster screenings of patients in non-invasive manner*" (P06). P04 also held a similar attitude regarding AI's potential: "*I think AI has clearly revolutionized oncology already, but the perspective for the future is much bigger. I think we are really in the infancy of what AI can deliver for such a domain*". Despite their affirmative outlook about the future potential of AI, our interviewees also reflected on several factors that may influence the existing disparities in the positioning of AI and its subsequent adoption across their research and clinical workflows, as illustrated below.

4.3.1 Difficulties with AI Integration and Scalability. The role of AI has evidently been more pronounced on the research side as compared to the clinical practice, especially with regards to the realization of precision (or personalized) medicine. However, P05 argued that "*we have seen a lot of papers which promise that AI will be personalizing medicine, but to be honest, so far it has been far from practicality*". Additionally, despite their promised performance, these AI models have failed to scale on the clinical side, even when "*applied to very simple and basic problems such as diagnosing whether a pulmonary nodule is benign or malignant*" (P05). Jacobs et al. [45] have also underlined this problem (within the context of mental health, and not cancer care) of discrepancies between AI predictions and existing standards of care. As a result, these models do not inspire confidence amongst our interviewees owing to their failure to address simple clinical tasks, and subsequently, expose conflicts with physician's diagnosis and judgements:

> "*Although it's tempting to publish papers on some sophisticated cool stuff and important on a conceptual level, in my opinion, I want something that I can use in the clinic and can talk to our patients with confidence that this model can be trusted.*" (P05)

Another problem with seamlessly amalgamating AI into the clinical workflow is the lack of effective means to integrate multimodal data and the absence of meaningful ontologies. P04 illustrated this problem by stating that:

> "*When you have to integrate the treatment plan, response to treatment, imaging, genomics, proteomics, and pathology data, we simply don't have a system ... and to say that a system will tell patients what to do and it's done, is a bit of science fiction. I have worked from the data to the patient with every little step, and it's still impossible to plug an AI system and let it spill out something new because you have to know where the data is, how was it captured, and what are the semantics.*" (P04)

In the context of our interviewees, the collaborative nature of decision making within MTB meetings requires analyses that may incorporate data from multiple tests, medical histories, and comparisons with various benchmarks for a single patient. This inherent heterogeneity in data sources, and the need for dynamism and flexibility in diagnostic analysis, adds a level of complexity that cannot be adequately addressed by an independent AI system.

The aforementioned challenges to seamlessly integrating AI into clinical workflows, as described by our interviewees, were also identified by Cabitza et al. [8] as the barriers to the "last mile of implementation" of AI. The authors argue that the lack of standardized protocols for generating, combining, processing, and validating high-quality clinical data that can be used to train clinical AI systems, a neglected practice in both cancer care and AI research, creates barriers to integrating AI into clinical settings. Although these problems are pronounced in clinical practice and are sensitive to the local context (see [25, 104]), our interviewees did not report similar problems in their research practice. Furthermore, our interviewees reported previously in Section 4.2 that hospitals are constantly coalescing large clinical datasets, which may indicate efforts to bridge research and clinical practice by facilitating retrospective analysis and enabling personalised treatment planning.

4.3.2 Disparity in Mutual Knowledge between Physicians and AI. All of our interviewees acknowledged their familiarity and knowledge of AI (due to their strong background in statistics), and they also apply advanced AI concepts in their research practices. Still, they emphasized the need for embodying the fundamental knowledge of AI in the education of physicians within clinical settings: "*you've probably heard that AI will not replace physicians, but imaging specialists who are using AI will replace the ones who are not*" (P01). P05 further added that "*I think it's an unmet need and a must for radiation oncologists to have a minimum acceptable understanding of advanced statistical and AI methods*". Extending this line of argumentation, P04 stated that medical and AI communities have to learn to understand each other better, in order to establish grounds for aligning mutual expectations and knowledge:

> "*It's very important that doctors understand what is AI, because it's fair to say that the level of familiarity of the medical system with AI is not huge, and people can be fooled in thinking that AI is actually smart, but it's clearly not. It's extremely powerful, but not intelligent and can go wrong in every wrong corner possible. Secondly, the problem is that AI does not know enough about doctors.*" (P04)

This two-way educational initiative is expected to result in "*super doctors*" because the physicians will be more equipped to employ AI "*as a decision help*" in their daily clinical work "*to have a more meaningful impact*" (P04). Simultaneously, such an approach will also pave the way for human-centric design of AI-powered technologies that are more attuned to physicians' practices and needs (P01, P02).

In addition, AI was referred as a "*tool to improve our basic understanding because we are reaching a plateau in our understanding in the cancer field*" (P02), which can enable physicians to "*push [their] biological reasoning quite far*" (P04). Furthermore, physicians are trained to examine images visually for diagnostic purposes, and the massive amount of images they encounter on the clinical side makes a "*part of [their] work very repetitive*" (P01). Therefore, AI tools will be much needed in this space because "*algorithms can look directly at signals and get to some kind of diagnosis*" (P01), as well as "*help [physicians] to see things which [they] are not seeing because [they] cannot put as much information in [their] analyses*" (P02). In this way, "*machines can help [physicians] by showing them where to look for new and diverse features*" (P01), and "*stimulate [them] to think more*" (P02).

These findings reinforce previous research in HCI about a more fitting position for AI in clinical workflows, which involves assisting with repetitive and tedious tasks and improving physician performance by showing them salient and latent aspects of patient attributes [57, 109].

4.3.3 Lack of Contestability and Validation Studies. Our interviewees reported that another way to build trust in AI is through extensive validation studies, which could enable them to develop contestable experiences with AI in clinical settings. Both P03 and P07 suggested that in order to be comfortable with AI, and to develop trust in its capabilities, they need to compare the results coming out of an AI system with their own analysis. Moreover, the use of black-boxes in the clinical context was seen by interviewees as an impediment to the development of trust in AI. P07 stated that: "*I am personally less confident with black-boxes*" because drawing conclusions about how a certain outcome was produced based on raw data is not straightforward. In addition, AI systems when trained on a certain population might fail to work for other geographical regions owing to the differences in population. P03 exemplified this aspect by citing her own research –an imaging technique for diagnosing breast cancer called "contrast mammography"– and stated that

> "*I was very comfortable to use it and to believe in it, but when I went to [Asia] and I tried this approach, it failed, because [Asian] people don't have the same breasts as [Europeans] and we could not extrapolate the results we obtained in [Europe].*" (P03)

These geographical differences can be further reproduced by AI, especially deep learning models, and exacerbate the problem of bias in predicted outcomes, which was also observed by Kaushal et al. [53]. In addition, the authors revealed in their study a troubling aspect that the majority of existing healthcare AI models are trained on patient data from a handful of geographic regions [53], which could further impact their application in local contexts.

Furthermore, P01 and P02 expressed their concern about the mismatch in the speed of advancements in AI domain and the time it takes to conduct validation studies: "*AI algorithms are changing so fast that no one can really take the time and make the validation studies which would be necessary to know the performance and to be able to use it*" (P01) and "*it's going so fast that by the time you use [AI systems], they are already outdated*" (P02). Consequently, to address this problem of mismatch in temporality between AI advancements and their validation, our interviewees suggested that algorithms be subjected to the same certification standards (e.g., FDA) as other medical appliances (P01, P06). Further extending the argument to support the value of validation studies, P02 cited the example of Mars Rover, which employs "*decade-old technologies, which are now outdated, but still used because they have been extensively tested and can be trusted*".

In summary, in our interviewees' imaginaries, there is an urgent need for proper standardized instruments to *a)* develop contestable personal experiences with AI while fostering means to assimilate local data in the continuous training and testing of AI models, and *b)* construct synergistic pathways between the clinical and research contexts which could remove existing boundaries between the global – *'outward'*– focus of AI research and its local – *'inward'*– deployment in clinical practices (i.e., generalizability vs. applicability).

4.4 Notion of Ethics and Responsibility Regarding AI

The consequence of physicians' interactions with AI systems, particularly in relation to the notion of ethics and responsibility, often manifested in our interviews in the form of a phrase "*in case there is an error, who will be responsible*" (P06). This conditional expression underpins *two* nuanced but entwined observations: *1)* legal and normative principles underpin the notion of responsibility in healthcare, and *2)* the impact of failures on the part of AI, particularly black-boxes, risks violating the core values of healthcare (see [3, 65]). In this section, we present insights from our interviews which highlight these tensions with regards to the use of AI in clinical contexts.

4.4.1 On Tensions Between AI and Responsibility. In response to our question about responsibility, and who bears it in case of a mistake, P01 answered that "*a physician is legally always responsible*" and "*each hospital protects itself*" (P06). Moreover, basing his argument in human rights and elaborating the patient's perspective P02 stated that "*a patient has the right to be here and to be treated, and we will do our best to provide the maximum attention for medical care*", and as a consequence you cannot "*put your trust in something that is not trustful*" or "*bypass the doctors*". P05 also expressed similar concerns regarding the autonomous use of AI systems on the clinical side:

> "*... from the legal point-of-view and from ethical standpoint, if I am not 100% sure that this [AI model] will be able to independently choose which treatment modality works best for this particular patient – a human being, I will not be able to tell the patient that I trust this model.*" (P05)

It is noteworthy, that our interviewees' concerns about "bypassing the doctors" signify a rather extreme scenario about responsibility, where the use of AI in the clinical context is free from the oversight and discretion of physicians and manifests in a state of indeterminacy regarding responsibility. Still, our interviewees presented a more realistic vision of AI in clinical contexts where it is properly supervised, and discussed the nuances in interpreting ethical and legal notions within healthcare. P02 argued that:

> "*[AI] can be present at all times, but [it] should always be supervised by the [physicians] ... I cannot imagine, I prefer not to imagine a system where we put the patient on the scanner, AI does the diagnosis, and robot does the surgery.*" (P02)

In this regard, P01 presented an analogy from the airline industry: "*it's a bit like autopilots in modern airplanes which are perfectly capable of taking-off and landing, but a pilot is still there to supervise*". Furthermore, implying that the aforementioned is an open-ended question that spans beyond the mere outcomes –whether successful or not– of AI, and concerns the ethical nature of AI-assisted clinical decision making, P01 provided another example of Autonomous Vehicles (AV): "*If you would have one AV hitting another one. Who is responsible for the crash? The vehicle which had the latest update, or which does not have the latest update?*". Furthermore, P06 exemplified a different scenario involving doctors who crowdsource medical images of patients to developing countries for analysis and reporting. In this case, P06 argued that legally this practice is similar to delegating decision making to AI. While these examples are emblematic of the problems surrounding the ethics of using AI systems to treat humans, they also underscore the complexity of attributing responsibility when accounting for the collective outcome of a human-AI team, which in turn has implications for broader acceptance and trust in these systems.

The research practices, on the other hand, are retrospective in nature and often dissociated from individual patients. In addition, our interviewees remarked that medical research bodies have established regulatory instruments and frameworks (e.g., Human Research and Ethics Committees, Journal Editorial Boards) which ensure ethical and responsible behavior on part of physicians, and prevent them from conducting "*harmful, investigative, and unethical research*" (P05, P06).

4.4.2 On Tensions Between AI and Explainability. Sustained interactions between the physicians and patients embody transparent communication of diagnostic findings, engaging patients in a discussion about the effective treatment plan, and explaining every detail of the treatment and its impact (also to "*debunk some myths*" (P05)). However, with regards to AI systems, in particular the ones employing deep neural networks, the desired notions of explainability and transparency are in conflict with the design and functioning of these systems as illustrated by P03: "*we are not confident using a black box because we don't like it and we don't understand how it works*". In order to bridge this gap in physicians' understanding of the underlying mechanisms and outcomes, explainable AI has been recommended as a significant milestone towards increased adoption of AI and to maintain underlying values in healthcare [3, 74]. Here, as well, the desired notion of explainability is subtly different across the clinical and research workflows.

On the clinical side, our interviewees highlighted difficulties in understanding some of the features used to train the deep learning algorithms and how they relate to biological processes. P05 stated that:

> "*Papers which use generated deep learning radiomics features have a lot of features like wavelet transformations, which as a clinician I struggle to understand, what is the significance and how to correlate this to some tumor features [...] Hand-crafted features, for instance the ones related to the texture can be easily correlated to tumor heterogeneity or the central necrosis, that is something [physicians] understand and can verbalize and explain to the patient.*" (P05)

Owing to these concerns around the use of black boxes in the clinical routines, AI-powered tools are perceived as auxiliary tools meant to re-examine physicians' assessment. "*Currently, AI is an additional tool which can provide an additional parameter to help confirm something that we have already assessed subjectively*" (P03). This argument is aligned with P04's assertion that AI's future on the clinical side is that of a decision help (see Section 4.3.2). However, "*the problem arises when [physicians'] assessment is opposite to that of AI, in such cases it is not easy to believe in the outcome of AI*" (P03). This statement further underlines the need for contestable experiences with AI in clinical settings (see Section 4.3.3), not only to develop understanding of its outcomes, but more importantly, to construct comprehensive mental models about the entire AI pipeline and its relationship to biological reasoning [65, 74].

To address the aforementioned problem of selecting meaningful and reproducible features to train AI algorithms, which are comprehensible for the physicians and correlate to biological functioning of tumors, P05 reported contributing to a standardization initiative known as 'Imaging Biomarkers Standardization Initiative (IBSI)'. He elaborated that "*I think we need to understand the features more, and make sure, for instance, that the features we are extracting are reliable and reproducible*" (P05). Similarly, P04 stressed that causality and not correlation must be accounted for when selecting the features:

> "*There is a lot of biology to be understood here. I mean, you can look at basically an indefinite number of covariances, but the secret is to really look at strong signals and consider those that you can hopefully connect with a reasonable biology ... to not be fooled just by correlation but to seek for causality.*" (P04)

These arguments surrounding the standardization of features and their relationship with biological processes were recognized by our interviewees as essential, and some (P02, P03, P04, and P07) justified that they fall under the banner of 'quality control' – "*not just of the data quality, but also of what the algorithm is predicting*" (P04).

On the research side, however, the explainability of AI is not directed to the patients, but towards the respective research communities of our interviewees, or locally, towards other physicians (e.g., in MTBs). Despite these differences in perceptions of desired explainability across clinical and research workflows, our interviewees' efforts to standardize feature spaces and elicit causal relationships between them and biological processes underline the need to align and homogenize the currently divergent notions of explainability throughout.

In summary, our findings demonstrate the divergent directionality of explainability perceptions and needs across research and clinical practices. These findings add to the existing research within AI and HCI domains on explainable AI by providing contextualized

empirical insights that deviate from the existing unilateral, and rather universal, conception of explainability.

5 DISCUSSION

In this section, we discuss the presented findings and explain their implications for oncological context and the overarching role of human-centric AI research in cancer care. We discuss viable means of bridging the AI adoption gap in clinical contexts and mitigating, if not eliminating, the disconnect between the current role and impact of AI in oncology and existing workflows, protocols, and imaginaries (see section 2.4). In section 5.1, we consolidate our interviewees' imaginaries into relevant categories that define a refocused and balanced vision of AI in cancer care. Section 5.2, particularly, disentangles the notions of explainability and contestability across the clinical and research workflows. Finally, we realign the position of AI within collaborative decision making constructs in Section 5.3, where we discuss AI's instrumental and constitutive conception within clinical contexts.

5.1 Disentangling AI Across Clinical and Research Workflows

In this section, we begin the discussion of our analytical findings by employing the notion of *imaginaries*, and its relation to HCI, as elaborated by Mlynar et al. [66]. Particularly, we use this conceptual lens to ground our findings within the dichotomous nature of clinical and research practices. According to the authors, who drew their conclusions from interviews with urban experts, AI in the cities should be *contextual*, *collaborative*, *controlled*, and *conscious*. We use this framework to investigate some broader implications of our findings and also test its utility when applied to a completely different expert group – physicians in oncology. This has been also suggested by Mlynar et al., who emphasize that "[i]t is still necessary to gain a clearer understanding in further research on how the imaginary of our expert group differs from the imaginaries of other groups" (p. 11). We draw inspiration from the "four Cs" to elaborate on the empirical findings of the present study, focusing on how the notion of imaginaries applies in research and in clinical workflows, allowing us to specify the tension between the two.

First, the *contextuality* of AI. In [66], this conveys that any non-human intelligence requires human beings and their knowledge of the local context, setting, culture, terminology, and workplace practices. Indeed, this aspect is already inherently built into the AI tools, which are produced as highly elaborate platforms that serve the specific needs of physicians. Sendak et al. [83] describe 4 values to consider when designing clinical AI – *1)* "define problem in context", *2)* "build relationship with stakeholders", *3)* "respect professional discretion", and *4)* "create ongoing feedback loops with stakeholders". Also, Fitzpatrick and Ellingsen [28] present similar suggestions – need for HCI/CSCW to be more engaged with healthcare professionals to develop a better understanding of the local context. For a "tool" to be efficiently incorporated into the everyday practices in the hospital, it must be designed and produced with these practices in mind, otherwise it is of limited purpose. Given the variety of working procedures, technical equipment, and established ways of working, the working cultures can be very locally specific. The trust of physicians in AI does not result from a technology becoming generally accepted, but from a personal relationship to the piece of technology that can be embraced and relied upon as a sensible working instrument (see Section 4.3.3): "*... to be comfortable with this kind of tool, we need to make our own experience ... and to to make sure that the software works pretty well, and each time it gives a good result, and you test it yourself in your patient, in your true life. ... I do not believe any in any kind of tools that I don't test myself on my patient*" (P03). Moreover, in the clinical context, an important aspect is the explainability of AI – quite literally so, as the diagnostic and clinical judgements need to be explainable to their patients (P05). In the imaginary of our experts, AI is present mostly as a tool that is to be used to arrive at accountable, reasonable, and reputable decisions, which can be defended not only in front of the patient, but also in the community of other experts in the same domain (see also Section 5.2).

Second, the *collaborativeness* of AI. It is formulated by [66] as an assemblage of instruments and information, of tangible, digital and social objects, including human individuals, who participate collectively in a "democratic process" that produces the AI outcomes. Such development is often seen as a requirement for research on human-AI collaboration in healthcare [10, 36]. In our case, a collaborative, democratic process is present in decision-making instruments such as the MTB meetings (see Section 4.2). Public and private spheres intertwine in these environments, as physicians' imaginary of AI is literally delimited by the responsibility to the patient as an individual. Furthermore, our interviewees projected optimism in the considered role of AI as a collaborator, which will enable them to become "*super doctors*" (P04) because they will be able to assimilate more multi-dimensional information in their analysis and visualize latent and subtle aspects of underlying biology in their clinical practices. This anticipated role of AI is in line with the findings of Oh et al. [68], where the role of AI as a collaborator was appreciated for its capability to reveal more details and take initiatives to drive collaborative efforts; this aspect is given further attention in Section 5.3 with regard to collaborative decision-making. As noted above, the notion of AI as a tool for specific tasks permeates the imaginary of our experts in oncology. A complex autonomous assemblage of information, algorithms, and tangible technologies is something that seems outside the scope of the imaginable and the possible as part of their everyday work – if not technologically, then ethically. Current AI is viewed as an instrument for the achievement of particular tasks, rather than as a transformative development in medicine.

Third, physicians are expecting AI to be *controlled*. In addition to the dynamics of collaborative social structures, this point, according to [66], concerns the possibility of human overruling the non-human intelligence and making visible any tacit ideologies inherently built into AI-based technologies. In the clinical context, the ideological aspect is of lesser relevance. The "ideology" of medicine is quite straightforward and well regulated by law. The problem is not purely ethical, but also legal, i.e., related to the formulation of "guidelines" and their efficiency [38]. The issue of control over AI was often stressed as important by the interviewees, especially with regard to trust and responsibility (see Section 4.4.1). Here, one may consider that while AI in our study is employed in the form of intangible algorithms, previous research has shown that "anthropomorphism and [general] intelligence" (tangible and/or human-like

devices) enhance trust in AI [97]. Nevertheless, rather than control as a mechanism protecting the society against ulterior motives and hidden ideologies of the AI creators, the control over AI in oncology seems to be a matter of service to patients, and physicians thus have "*human rights to respect*" (P02). This is both true for research and the clinical contexts.

Fourth, and finally, the *consciousness* of AI. In [66], this aspect has to do primarily with problems related to data as a representation of reality, and with the production of intelligent technological solutions based on creativity and progressiveness rather than on the reproduction of earlier states of the world. Physicians take into consideration their patients' privacy and the usability of the data for diagnostic and research procedures. The concern is that the algorithm grounds its outcomes in data that they can trust – both in terms of quality and quantity. The distinction that seems to be operative is a distinction between "*one's own data*" and "*institutional data*" (P05), with differing layers of responsibilities and engagement. Given the issues with patient privacy, even local access to institutional data for research purposes can be problematic. This shows that the boundary between statistical categories of populations (in research) and their application to concrete human beings for diagnostic purposes (in clinical practice) produces not only functional synergies within the institution, but also issues in permeability.

To summarize, we can say that the framework of imaginaries of AI offered by Mlynar et al. [66] provides a useful starting point for an inquiry into our interviewees' stance towards AI. Nevertheless, it can be applied to the reality of oncology as lived and discussed by our interviewees only due to considerable generality of the "four Cs". Context, collaboration, control, and consciousness work as descriptive glosses for the analysis of our experts' imaginaries, but they have to be worked out on the basis of empirical material with regard to the specificity of work in oncology. This issue will receive more attention in the following section.

5.2 Explainability and Contestability

Our interviewees' trust in AI, particularly on the clinical side, was not observed to be resulting from their general acceptance of AI. They unanimously acknowledged their use of AI-ML algorithms to support their diagnostic and (mostly) research activities, and responded affirmatively to AI's potential in cancer care. Still, in their discourses, their trust, or rather the lack of it, was grounded within the notions of *explainability* (in relation to 'black-boxes') and *contestability* (with regards to developing personal experiences with AI, and running validation studies). Both these notions have been a subject of active deliberation in diverse disciplines in recent years. Owing to HCI's collaborative endeavors with the AI community, these notions have been intensely scrutinized, as is evident from the organization of workshops in venues such as ACM CHI [20, 21] (on explainability) and CSCW [98] (on contestability).

5.2.1 Explainability. In the imaginary of our interviewees, the lack of *explainability*, in relation to the use of black-box systems referred to two specific attributes of *1) mechanics of operation*, i.e., knowledge of how the system works and produces specific outcomes, and *2) intelligibility of features*, i.e., how do the input (radiomics) features which are used to train the model, relate to meaningful biological processes. Moreover, the consequences of using AI have implications for diverse actors with different scopes and temporalities within the clinical ecosystem, which are different from those on the research side. In research, the explanations regarding the use of specific AI-ML models and complex multi-layered feature spaces are directed *outwards*, aimed at the respective academic community. However, on the clinical side, the explanations are directed both *externally* and *internally* – *1) externally* towards the patient, their family, and in some cases, the patient care (insurance), *2) internally* aimed at the physician preparing the diagnosis, and *3)* subsequent discussions within the Multidisciplinary Tumor Boards (MTBs).

Consequently, to consolidate these diverse needs and directions for explanations, "opening the black-box" would translate into the supervised use of AI, where its role and impact are monitored consistently and throughout. The peculiarities of the disease across different patients, and the resulting differences in treatment types and plans, necessitate that AI is constantly monitored and its outcomes are scrutinized throughout the duration of the therapy to preemptively evade any harm. Such an extensive supervision, then, may eventually result in the development of trust amongst physicians as discussed by Ferrario and Loi [27]. However, it can be argued that despite trust in AI, the supervised use of these systems will not cease over time due to the nature of clinical practices and the established protocols that encapsulate the institutionalized constructs of collective legitimacy of decisions (in MTBs), liability, and human-rights.

Furthermore, amongst the aforementioned attributes, 'mechanics of operation' has been the subject of extensive investigation in recent years within HCI (e.g., [15, 19, 52, 59, 88]), and arguments have been made to operationalize holistic explainability of AI [74] and foster interactive attainment of explanations [65], which are aligned with the context of use, involved stakeholders, organizational values, and industrial standards. The second attribute of 'intelligibility of features' is grounded in the clinician's need to offer valid and transparent justifications to the patient about the disease, its diagnosis and therapy, and may at times also require them to go beyond the conventional explanations by checking the influence of misinformation through "*debunking myths*" (P05). Hamon et al. [39] have examined the challenges in providing "human legible explanations" in practice, with respect to the use of complex and high-level features, which also make it difficult to draw "clear causal links between data and the final decisions". These challenges were emphasized by our interviewees as well, and perhaps underpin their reliance on hand-crafted features that can reasonably relate to tumor morphology and underlying biological processes.

5.2.2 Contestability. To overcome the challenges posed by the functional realization of explainability, recommendations have been made to calibrate (clinical) practices surrounding the use of AI, so that their decisions and impact can be contested [2, 40, 62, 74, 103]. In addition, recent research has provided positive evidence in favor of clinical application of AI, where physicians can actively "co-create with AI" [31] and "spar [with AI] like their colleagues" [9]. For our participants, contestability, like explainability, manifests at different levels – *1)* by empowering patients to weigh in on the treatment planning, and *2)* by personally experiencing AI, with their own data, and extensive validation of its performance. Our interviewees explained that the mismatch between the speed of

advances in AI and the time required to conduct thorough validations hinders contestability. Consequently, physicians still use standardized, tested, and relatable indicators in their diagnosis and decision making. Furthermore, this mismatch is manifested in the different rates at which the research and clinical sides are adopting AI systems, and the lack of seamless transfer of high-end AI models from the research side to clinical practice.

Yurrita et al. [110] propose a value-based AI assessment framework, where they have explicated the tensions in the operationalization of these high-level AI qualities. The authors argue, that both the qualities of contestability ('individual empowerment') and explainability ('openness') are in opposition to performance of AI, which may further explain its limited –or sporadic experimental use in case of failing standards of care– use in clinical treatments. These findings, though, emphasize the value of incorporating these qualities into AI development for clinical use. At the same time, they paradoxically highlight the difficulties of embodying them in medical AI systems.

5.3 Imaginaries of Collaborative Decision-Making

Decision-making procedures are central components of clinical practice. In a sense, the link between research and clinical practice is in the former domain's ability to support decision-making in the latter, and to enable the shift from diagnostic categories to individual patients that is at the background of all healthcare. In this section, we discuss our empirical results with regard to AI's role in collaborative decision-making procedures, building on the collaborative aspect of AI imaginaries discussed above in Section 5.1. Moreover, this section responds to research done in CDSS (Clinical Decision Support Systems, e.g. [25, 57]), that has been a dominant theme at the intersection of HCI and AI, while our results indicate that the imagined conception of decision making is different from the reality. We discuss how, according to the imaginary of AI identified in our interview analysis, AI can be made part of decision-making, outlining a *scale* of social imaginaries whose two poles are delimited by what we call an *instrumental* and *constitutive* conception of AI. This conceptual pair can be relevant not only to design technological solutions, but –perhaps more importantly– to evaluate them *ex ante* and examine the possible social implications of the implementation of AI in clinical practice and in research, keeping in mind their alternative orientations to individual patients in clinical work, and to statistically constructed categories in research.

Our interviewees have recurrently underscored the importance of Multidisciplinary Tumor Board (MTB) meetings [61] as a shared decision-making procedure (see Section 4.2). During our research, we have realized the importance of understanding the role of MTBs in the larger healthcare ecosystem, as well as comprehending their own internal structure and organization. MTBs are instances of collaborative multidisciplinary teamwork which promote "an optimal environment for collaborative decision-making in which patients are key stakeholders and all relevant cancer care professionals are actively involved" [67]. Our identification of the centrality of MTBs is in line with other recent studies, confirming that "we must have system understanding of how the system works with respect to concepts and the relationships between them before we can model or use formal decision science approaches" [54], and with attempts to increase the use of shared decision-making in healthcare [58] that also take into account the specific uncertainty of such activities as social processes [51].

The findings reported in this paper show that AI can be –and to some extent already is– implemented in *collaborative decision-making*, defined as "an in-depth personalized iterative assessment of patient's medical, psychological and social status", where the patient has a "proactive role as a key stakeholder" [67]. However, if AI is to be adequately and fully incorporated into MTBs as a trustworthy and unremarkable device or participant, it must respond and align itself with the local procedures that take place in the meetings, and it must be crafted to their own temporality and orderliness as social environments of collaborative decision-making [29]. Verma et al. [102] distinguish two alternate conceptions of (digital) technology. This distinction can also be usefully applied to the position of AI in clinical practice. According to Verma et al., "*instruments* and *environments* are two opposite conceptions of technology, exemplified ... in the tension between 'means to an end' on the one hand, and, on the other hand, conditions structuring more fundamentally the social activities that constitute [the] work [of the studied disciplinary domain]. The central difference is that tools as instruments are tied to solving clearly outlined problems ... while tools as environments transform the horizon of possibilities in a more essential sense" [102]. Consequently, we propose that there are two ways of implementing AI in collaborative decision-making: *instrumental* and *constitutive*. The remainder of this section provides further considerations of these two approaches and their broader implications.

As an *instrument*, on the one hand, AI could provide support to the MTB member, offering her a possible decision informed by the algorithmic model; it is one of the tools that the experts may use and take into account in their own assessment of the cases discussed at MTB meetings: "*we really need to have tools that help to decide what could be the best treatments for the patient and in the different line of treatment*" (P03). The trouble in implementing AI is less about consensus and more about legitimacy, which ultimately relates to the issue of trust. Of course, problems appear when the recommendation of the algorithm is in disagreement with the recommendations of the MTB experts. More research is therefore needed into the social mechanisms of dealing with disagreement, and arriving at consensual decision, within MTBs. Only by explaining the inner interactional workings of MTBs as social settings can one develop AI systems that can be used efficiently as part of these workings – for every successful technology is ultimately successful because it can be "made at home with the rest of our world" [80] and incorporated into existing procedures once put into practice. Such knowledge would be valuable for designing specific AI systems that could be incorporated in the communicative processes that are part and parcel of decision making, while informing the design through the distinctive local interactional organization of such processes (cf. [75]).

On the other hand, as a *constituent*, AI itself becomes a competent member of the MTB meetings, being taken seriously by the other (i.e., human) members of the board, and offering its "opinion" to be considered as one of the expert opinions. This would involve many requirements on the competences of AI (cf. [41, 84]). Nevertheless,

its position would eventually be different from a human participant, since the AI would only have an advisory role, and would not actively take part in the treatment procedure, thus escaping some professional biases as described by P01: "*Every specialty is then hand in hand doing the best they can do for the patient. But each is considering the patient from his angle.*" Concurrently, this position would largely relieve the AI "advisor" from the accountability to the patient, thus leaving the responsible decision on the medical actors who implement the treatment. Without doubt, the constitutive conception of AI's role in MTB meetings is far from current technological possibilities. Nevertheless, it is the *imaginaires* and possible futures that are at the center of our discussion in this section. A clear formulation of the social *imaginaires* that frame –all too often only tacitly and implicitly– engineering's *goals and aims* is necessary to carefully examine the ethical and societal implications of their technical realization. This must be done well in advance, and assessment of the possible impact is needed *before* a technology is produced and marketed (see [35, 63, 76]). Our proposition of considering the two ways of involving AI in MTB meetings is to be taken as outlining a scale of the imaginary, which can assist in taking conscious and collaborative decisions about the future technological reality of AI's role in medical practice, both clinical and research-oriented.

6 CONCLUSION

The forces driving rapid advancements in AI are inherently disconnected from the study of their impact on individuals, societies, and organizations [76]. This discrepancy manifests in a biased focus of AI development and deployment. The role and impact of AI in cancer care is similarly unbalanced – significant and rapid advancements have been made on the research side, however, AI's role in the clinical side has been minimal and limited. These differences can be attributed to the overlooked, yet profound, differences in the clinical and research practices in oncology. We contribute by scrutinizing physicians' current engagements with AI by interviewing 7 physicians –who are involved in both cancer treatment and research– and disentangling its future alignment across the clinical and research workflows. In this way, we have essentially diverged from the ostensible "*one-size-fits-all*" paradigm of AI development, and aimed to adjust AI's position and impact while coalescing the dichotomy in cancer treatment and research. Our analysis reveals that physicians' trust in AI, on the clinical side, is independent of their general acceptance of AI. Instead, it is grounded in their contestable experiences with AI, and their preferential disposition towards a supervised employment of AI – *clinician-in-the-loop*. Furthermore, we elicit the desired qualities of AI which are grounded in our experts' imaginaries [66], and examine how divergent intentionality and scope of clinical and research workflows engender tensions between practice and principle. Particularly, we provide justifications anchored in practices and norms about the possible causes for these tensions, and pragmatic and contextualized means of diffusing them, especially in relation to globally accepted notions about the ethical- and responsible-development of AI (such as control, collaboration, explainability, contestability, etc.). At a more general level, we propose that AI can be included in collective decision-making processes in oncology either *instrumentally* (as a "tool") or *constitutively* (as a "member"), each of these alternatives generating different sets of ethical, societal and technological issues to be solved in future research.

ACKNOWLEDGMENTS

This work was partially supported by the Hasler Foundation (via the EPICS project number 20004), the Swiss National Science Foundation (SNSF, grant 205320_179069), the Swiss Personalized Health Network (SPHN, via the IMAGINE and QA4IQI projects), and the HES-SO internal fund RCSO "Radiomics in higher education: *QuantImage* as a teaching tool" (RADHED). In addition, this work was also supported by the EU Horizon 2020 grant (number 101016233) via the PERISCOPE (Pan-European Response to the Impacts of COVID-19 and Future Pandemics and Epidemics) project.

We thank all our interviewees, including those who agreed to be identified, and others who preferred to remain anonymous. We, particularly, thank Prof. Clarisse Dromain, Prof. Esat-Mahmut Ozsahin, Dr. Hesham ElHalawani, Prof. Olivier Michielin, and Dr. Vincent Dunet for sharing their perspectives, expertise, and ideas generously. Finally, we thank the anonymous reviewers for their constructive comments, as well as Mireia Yurrita Semperena for copy editing and for vital suggestions on improving the manuscript.

REFERENCES

[1] Hugo J W L Aerts, Emmanuel Rios Velazquez, Ralph T H Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, Frank Hoebers, Michelle M Rietbergen, C René Leemans, Andre Dekker, John Quackenbush, Robert J Gillies, and Philippe Lambin. 2014. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications* 5 (June 2014), 4006. Issue 1.

[2] Kars Alfrink, Ianus Keller, Gerd Kortuem, and Neelke Doorn. 2022. Contestable AI by Design: Towards a Framework. *Minds and Machines* (2022), 27. https://doi.org/10.1007/s11023-022-09611-z

[3] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, and Vince I Madai. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making* 20, 1 (2020), 1–9.

[4] Kimlin Tam Ashing-Giwa, Geraldine Padilla, Judith Tejero, Janet Kraemer, Karen Wright, Anne Coscarelli, Sheila Clayton, Imani Williams, and Dawn Hills. 2004. Understanding the breast cancer experience of women: a qualitative study of African American, Asian American, Latina and Caucasian cancer survivors. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* 13, 6 (2004), 408–428.

[5] Matthias Baldauf, Peter Fröehlich, and Rainer Endl. 2020. Trust Me, I'm a Doctor – User Perceptions of AI-Driven Apps for Mobile Health Diagnosis. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM '20)*. Association for Computing Machinery, New York, NY, USA, 167–178. https://doi.org/10.1145/3428361.3428362

[6] Christine A Barry, Fiona A Stevenson, Nicky Britten, Nick Barber, and Colin P Bradley. 2001. Giving voice to the lifeworld. More humane, more effective medical care? A qualitative study of doctor–patient communication in general practice. *Social science & medicine* 53, 4 (2001), 487–505.

[7] Michael Burawoy. 1998. The Extended Case Method. *Sociological Theory* 16, 1 (1998), 4–33.

[8] Federico Cabitza, Andrea Campagner, and Clara Balsano. 2020. Bridging the "last mile" gap between AI implementation and operation:"data awareness" that matters. *Annals of translational medicine* 8, 7 (2020), 9.

[9] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 104 (nov 2019), 24 pages. https://doi.org/10.1145/3359206

[10] Francisco Maria Calisto, Carlos Santiago, Nuno Nunes, and Jacinto C Nascimento. 2022. BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artificial Intelligence in Medicine* 127 (2022), 102285.

[11] Ashley Castleberry and Amanda Nolen. 2018. Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning* 10, 6 (2018), 807–815. https://doi.org/10.1016/j.cptl.2018.03.019

[12] Cornelius Castoriadis. 1997. *The imaginary institution of society.* MIT Press, Cambridge.

[13] Alison Chapple, Sue Ziebland, and Ann McPherson. 2004. Stigma, shame, and blame experienced by patients with lung cancer: qualitative study. *Bmj* 328, 7454 (2004), 1470.

[14] Kathy Charmaz. 2003. Grounded theory. In *Qualitative psychology: A practical guide to research methods*, J. A. Smith (Ed.). Sage, London, 81–110.

[15] Michael Chromik, Malin Eiband, Felicitas Buchner, Adrian Krüger, and Andreas Butz. 2021. I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 307–317. https://doi.org/10.1145/3397481.3450644

[16] Jenifer L Culver, Patricia L Arena, Michael H Antoni, and Charles S Carver. 2002. Coping and distress among women under treatment for early stage breast cancer: comparing African Americans, Hispanics and non-Hispanic Whites. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* 11, 6 (2002), 495–504.

[17] Srikant Devaraj, Sushil K Sharma, Dyan J Fausto, Sara Viernes, Hadi Kharrazi, et al. 2014. Barriers and facilitators to clinical decision support systems adoption: A systematic review. *Journal of Business Administration Research* 3, 2 (2014), 36.

[18] Haoran Ding, Chenzhou Wu, Nailin Liao, Qi Zhan, Weize Sun, Yingzhao Huang, Zhou Jiang, and Yi Li. 2021. Radiomics in Oncology: A 10-Year Bibliometric Analysis. *Frontiers in Oncology* 11 (2021), 3677.

[19] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. https://doi.org/10.1145/3411764.3445188

[20] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O. Riedl. 2021. Operationalizing Human-Centered Perspectives in Explainable AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 94, 6 pages. https://doi.org/10.1145/3411763.3441342

[21] Upol Ehsan, Philipp Wintersberger, Q. Vera Liao, Elizabeth Anne Watkins, Carina Manger, Hal Daumé III, Andreas Riener, and Mark O Riedl. 2022. Human-Centered Explainable AI (HCXAI): Beyond Opening the Black-Box of AI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 109, 7 pages. https://doi.org/10.1145/3491101.3503727

[22] Victoria Elliott. 2018. Thinking about the Coding Process in Qualitative Data Analysis. *The Qualitative Report* 23, 11 (11 2018), 2850–2861. https://www.proquest.com/scholarly-journals/thinking-about-coding-process-qualitative-data/docview/2155621346/se-2 Name - Oxford University; Copyright - Copyright The Qualitative Report Nov 2018; Last updated - 2019-11-11; SubjectsTermNotLitGenreText - New York; United Kingdom–UK; England; Los Angeles California.

[23] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of Advanced Nursing* 62, 1 (2008), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2648.2007.04569.x

[24] Nicole M Else-Quest, Noelle K LoConte, Joan H Schiller, and Janet Shibley Hyde. 2009. Perceived stigma, self-blame, and adjustment among lung, breast and prostate cancer patients. *Psychology and Health* 24, 8 (2009), 949–964.

[25] Glyn Elwyn, Isabelle Scholl, Caroline Tietbohl, Mala Mann, Adrian GK Edwards, Catharine Clay, France Légaré, Trudy van der Weijden, Carmen L Lewis, Richard M Wexler, et al. 2013. "Many miles to go…": a systematic review of the implementation of patient decision support interventions into routine clinical practice. *BMC medical informatics and decision making* 13, 2 (2013), 1–10.

[26] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542 (January 2017), 115.

[27] Andrea Ferrario and Michele Loi. 2022. How Explainability Contributes to Trust in AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1457–1466. https://doi.org/10.1145/3531146.3533202

[28] Geraldine Fitzpatrick and Gunnar Ellingsen. 2013. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)* 22, 4 (2013), 609–665.

[29] Cecilia E Ford and Joshua Raclaw. 2015. Meetings as interactional achievements: A conversation analytic perspective. In *The Cambridge handbook of meeting science*. Cambridge University Press, Cambridge, UK, 247–276.

[30] Patricia I. Fuchs and Lawrence R. Ness. 2015. Are We There Yet? Data Saturation in Qualitative Research. *The Qualitative Report* 20, 9 (2015), 1408–1416. https://doi.org/10.46743/2160-3715/2015.2281

[31] Astrid Galsgaard, Tom Doorschodt, Ann-Louise Holten, Felix Christoph Müller, Mikael Ploug Boesen, and Mario Maas. 2022. Artificial intelligence and multidisciplinary team meetings; a communication challenge for radiologists' sense of agency and position as spider in a web? *European Journal of Radiology* 155 (2022), 110231. https://doi.org/10.1016/j.ejrad.2022.110231

[32] Roberto Gatta, Adrien Depeursinge, Osman Ratib, Olivier Michielin, and Antoine Leimgruber. 2020. Integrating radiomics into holomics for personalised oncology: from algorithms to bedside. *European Radiology Experimental* 4, 1 (December 2020), 11.

[33] Graham R. Gibbs. 2007. *Analyzing Qualitative Data*. Sage, London.

[34] Michael J. Gonzales and Laurel D. Riek. 2013. Co-Designing Patient-Centered Health Communication Tools for Cancer Care. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare* (Venice, Italy) *(PervasiveHealth '13)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 208–215. https://doi.org/10.4108/icst.pervasivehealth.2013.252109

[35] Thomas Grote and Philipp Berens. 2020. On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics* 46, 3 (2020), 205–211. https://doi.org/10.1136/medethics-2019-105586 arXiv:https://jme.bmj.com/content/46/3/205.full.pdf

[36] Hongyan Gu, Jingbin Huang, Lauren Hung, and Xiang 'Anthony' Chen. 2021. Lessons Learned from Designing an AI-Enabled Diagnosis Tool for Pathologists. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 10 (apr 2021), 25 pages. https://doi.org/10.1145/3449084

[37] Julien Guiot, Akshayaa Vaidyanathan, Louis Deprez, Fadila Zerka, Denis Danthine, Anne-Noelle Frix, Philippe Lambin, Fabio Bottari, Nathan Tsoutzidis, Benjamin Miraglio, Sean Walsh, Wim Vos, Roland Hustinx, Marta Ferreira, Pierre Lovinfosse, and Ralph T.H. Leijenaar. 2022. A review in radiomics: Making personalized medicine a reality via routine imaging. *Medicinal Research Reviews* 42, 1 (2022), 426–440. https://doi.org/10.1002/med.21846 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/med.21846

[38] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30, 1 (2020), 99–120.

[39] Ronan Hamon, Henrik Junklewitz, Gianclaudio Malgieri, Paul De Hert, Laurent Beslay, and Ignacio Sanchez. 2021. Impossible Explanations? Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 549–559. https://doi.org/10.1145/3442188.3445917

[40] Clément Henin and Daniel Le Métayer. 2022. Beyond explainability: justifiability and contestability of algorithmic decision systems. *AI & SOCIETY* 37, 4 (2022), 1397–1410. https://doi.org/10.1007/s00146-021-01251-8

[41] Marjan Huisman. 2001. Decision-making in meetings as talk-in-interaction. *International Studies of Management & Organization* 31, 3 (2001), 69–90.

[42] Bogdana Huma and Jack B. Joyce. 2022. 'One size doesn't fit all': Lessons from interaction analysis on tailoring Open Science practices to qualitative research. *British Journal of Social Psychology* 00 (2022), 1–15. https://doi.org/10.1111/bjso.12568

[43] Maia Jacobs, James Clawson, and Elizabeth Mynatt. 2013. Cancer Compass: Examining Personal Health Record Usage among Breast Cancer Survivors. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare* (Venice, Italy) *(PervasiveHealth '13)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 264–265. https://doi.org/10.4108/icst.pervasivehealth.2013.252036

[44] Maia Jacobs, James Clawson, and Elizabeth D. Mynatt. 2014. Cancer Navigation: Opportunities and Challenges for Facilitating the Breast Cancer Journey. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1467–1478. https://doi.org/10.1145/2531602.2531645

[45] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 659, 14 pages. https://doi.org/10.1145/3411764.3445385

[46] Maia L. Jacobs, James Clawson, and Elizabeth D. Mynatt. 2014. My Journey Compass: A Preliminary Investigation of a Mobile Tool for Cancer Patients. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 663–672. https://doi.org/10.1145/2556288.2557194

[47] Maia L. Jacobs, James Clawson, and Elizabeth D. Mynatt. 2015. Comparing Health Information Sharing Preferences of Cancer Patients, Doctors, and Navigators. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 808–818. https://doi.org/10.1145/2675133.2675252

[48] Sheila Jasanoff and Sang-Hyun Kim. 2015. *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. University of Chicago Press,

Chicago, USA.

[49] Peter Jones, David Cronin, Dean Karavite, Ross Koppel, Prudence Dalrymple, Kai Zheng, Michelle Rogers, and Bob Schumacher. 2011. Designing for Whole Systems and Services in Healthcare. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 359–362. https://doi.org/10.1145/1979742.1979530

[50] Kazi Sinthia Kabir, Erin L. Van Blarigan, June M. Chan, Stacey A. Kenfield, and Jason Wiese. 2019. "I'm Done with Cancer. What Am I Trying to Improve?": Understanding the Perspective of Prostate Cancer Patients to Support Multiple Health Behavior Change. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare* (Trento, Italy) *(PervasiveHealth'19)*. Association for Computing Machinery, New York, NY, USA, 81–90. https://doi.org/10.1145/3329189.3329207

[51] Heather L. Kane, Michael T. Halpern, Linda B. Squiers, Katherine A. Treiman, and Lauren A. McCormack. 2014. Implementing and evaluating shared decision making in oncology practice. *CA: A Cancer Journal for Clinicians* 64, 6 (2014), 377–388. https://doi.org/10.3322/caac.21245

[52] Harmanpreet Kaur, Eytan Adar, Eric Gilbert, and Cliff Lampe. 2022. Sensible AI: Re-Imagining Interpretability and Explainability Using Sensemaking Theory. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 702–714. https://doi.org/10.1145/3531146.3533135

[53] Amit Kaushal, Russ Altman, and Curt Langlotz. 2020. Geographic distribution of US cohorts used to train deep learning algorithms. *Jama* 324, 12 (2020), 1212–1213.

[54] Craig Kuziemsky. 2016. Decision-making in healthcare as a complex adaptive system. *Healthcare Management Forum* 29, 1 (2016), 4–7. https://doi.org/10.1177/0840470415614842 arXiv:https://doi.org/10.1177/0840470415614842 PMID: 26656389.

[55] Benjamin W Lamb, Katrina F Brown, Kamal Nagpal, Charles Vincent, James SA Green, and Nick Sevdalis. 2011. Quality of care management decisions by multidisciplinary cancer teams: a systematic review. *Annals of surgical oncology* 18, 8 (2011), 2116–2125.

[56] Benjamin W Lamb, Nick Sevdalis, Sonal Arora, Anna Pinto, Charles Vincent, and James SA Green. 2011. Teamwork and team decision-making at multidisciplinary cancer conferences: barriers, facilitators, and opportunities for improvement. *World journal of surgery* 35, 9 (2011), 1970–1976.

[57] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2020. Co-Design and Evaluation of an Intelligent Decision Support System for Stroke Rehabilitation Assessment. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 156 (oct 2020), 27 pages. https://doi.org/10.1145/3415227

[58] France Légaré, Rhéda Adekpedjou, Dawn Stacey, Stéphane Turcotte, Jennifer Kryworuchko, Ian D Graham, Anne Lyddiatt, Mary C Politi, Richard Thomson, Glyn Elwyn, and Norbert Donner-Banzhoff. 2018. Interventions for increasing the use of shared decision making by healthcare professionals. *Cochrane Database of Systematic Reviews* 7, CD006732 (2018), 380.

[59] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. 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, 1–15. https://doi.org/10.1145/3313831.3376590

[60] Noelle K LoConte, Nicole M Else-Quest, Jens Eickhoff, Janet Hyde, and Joan H Schiller. 2008. Assessment of guilt and shame in patients with non–small-cell lung cancer compared with patients with breast and prostate cancer. *Clinical lung cancer* 9, 3 (2008), 171–178.

[61] Claudio Luchini, Rita T Lawlor, Michele Milella, and Aldo Scarpa. 2020. Molecular Tumor Boards in Clinical Practice. *Trends in Cancer* 6, 9 (2020), 738–744. https://doi.org/10.1016/j.trecan.2020.05.008

[62] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 106 (apr 2021), 25 pages. https://doi.org/10.1145/3449180

[63] Tamra Lysaght, Hannah Yeefen Lim, Vicki Xafis, and Kee Yuan Ngiam. 2019. AI-Assisted Decision-making in Healthcare: The Application of an Ethics Framework for Big Data in Health and Research. *Asian Bioethics Review* 11, 3 (2019), 299–314.

[64] Jody Miller and Barry Glassner. 2016. The 'inside' and the 'outside': finding realities in interviews. In *Qualitative Data Analysis*, David Silverman (Ed.). Sage, London.

[65] Brent Mittelstadt, Chris Russell, and Sandra Wachter. 2019. Explaining Explanations in AI. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 279–288. https://doi.org/10.1145/3287560.3287574

[66] Jakub Mlynar, Farzaneh Bahrami, André Ourednik, Nico Mutzner, Himanshu Verma, and Hamed Alavi. 2022. AI beyond Deus Ex Machina – Reimagining Intelligence in Future Cities with Urban Experts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 370, 13 pages. https://doi.org/10.1145/3491102.3517502

[67] Catherine Ménard, Isabelle Merckaert, Darius Razavi, and Yves Libert. 2012. Decision-making in oncology a selected literature review and some recommendations for the future. *Current Opinion in Oncology* 24, 4 (2012), 381–390.

[68] Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174223

[69] Tom Ongwere, Andrew B.L. Berry, Clara Caldeira, Rosa I. Arriaga, Amid Ayobi, Eleanor R. Burgess, Kay Connelly, Patricia Franklin, Andrew D Miller, Aehong Min, and Nervo Verdezoto. 2022. Challenges, Tensions, and Opportunities in Designing Ecosystems to Support the Management of Complex Health Needs. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 71, 7 pages. https://doi.org/10.1145/3491101.3503714

[70] Sun Young Park, Pei-Yi Kuo, Andrea Barbarin, Elizabeth Kaziunas, Astrid Chow, Karandeep Singh, Lauren Wilcox, and Walter S. Lasecki. 2019. Identifying Challenges and Opportunities in Human-AI Collaboration in Healthcare. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 506–510. https://doi.org/10.1145/3311957.3359433

[71] Vivek Patkar, Dionisio Acosta, Tim Davidson, Alison Jones, John Fox, and Mohammad Keshtgar. 2011. Cancer Multidisciplinary Team Meetings: Evidence, Challenges, and the Role of Clinical Decision Support Technology. *International Journal of Breast Cancer* 2011 (2011), 1–7. https://doi.org/10.4061/2011/831605

[72] Oleg S. Pianykh, Georg Langs, Marc Dewey, Dieter R. Enzmann, Christian J. Herold, Stefan O. Schoenberg, and James A. Brink. 2020. Continuous Learning AI in Radiology: Implementation Principles and Early Applications. *Radiology* 297, 1 (October 2020), 6–14.

[73] Elizabeth A. St. Pierre and Alecia Y. Jackson. 2014. Qualitative Data Analysis After Coding. *Qualitative Inquiry* 20, 6 (2014), 715–719. https://doi.org/10.1177/1077800414532435

[74] Thomas Ploug and Søren Holm. 2020. The four dimensions of contestable AI diagnostics-A patient-centric approach to explainable AI. *Artificial Intelligence in Medicine* 107 (2020), 101901.

[75] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[76] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. 2019. Machine behaviour. *Nature* 568, 7753 (2019), 477–486.

[77] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, and Andrew Y. Ng. 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. https://doi.org/10.48550/ARXIV.1711.05225

[78] Erik R. Ranschaert, Sergey Morozov, and Paul R. Algra (Eds.). 2019. *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks* (1 ed.). Springer, Cham, Switzerland. 1–373 pages. https://doi.org/10.1007/978-3-319-94878-2

[79] Mauricio Reyes, Raphael Meier, Sérgio Pereira, Carlos A. Silva, Fried-Michael Dahlweid, Hendrik von Tengg-Kobligk, Ronald M. Summers, and Roland Wiest. 2020. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiology: Artificial Intelligence* 2, 3 (May 2020), e190043.

[80] Harvey Sacks, Gail Jefferson, and Emanuel A. Schegloff. 1992. *Lectures on Conversation*. Vol. 1 & 2. Blackwell, Oxford, UK.

[81] Johnny Saldaña. 2014. Coding and Analysis Strategies. In *The Oxford Handbook of Qualitative Research*, Patricia Leavy (Ed.). Oxford University Press, Oxford, 581–598.

[82] Margrit Schreier. 2012. *Qualitative content analysis in practice*. Sage Publications, Thousand Oaks, CA, USA.

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

[84] Lucas M Seuren, Wyke Stommel, Dieneke van Asselt, Özcan Sir, Martijn Stommel, and Yvonne Schoon. 2019. Multidisciplinary meetings at the emergency

department: A conversation-analytic study of decision-making. *Social Science & Medicine* 242 (2019), 112589.

[85] Parth V. Shah and Kathryn E. Ringland. 2019. Health Information Technologies for Cancer Care: Characterizing Developments for Precision and Personalized Medicine. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work & Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 362–366. https://doi.org/10.1145/3311957.3359504

[86] Ben Shneiderman. 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems. *ACM Trans. Interact. Intell. Syst.* 10, 4, Article 26 (oct 2020), 31 pages. https://doi.org/10.1145/3419764

[87] Ben Shneiderman. 2022. *Human-Centered AI.* Oxford University Press, Oxford, UK.

[88] Andrew Silva, Mariah Schrum, Erin Hedlund-Botti, Nakul Gopalan, and Matthew Gombolay. 2022. Explainable Artificial Intelligence: Evaluating the Objective and Subjective Impacts of XAI on Human-Agent Interaction. *International Journal of Human–Computer Interaction* 0, 0 (2022), 1–15. https://doi.org/10.1080/10447318.2022.2101698

[89] Martina Sollini, Lidija Antunovic, Arturo Chiti, and Margarita Kirienko. 2019. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *European Journal of Nuclear Medicine and Molecular Imaging* 46, 13 (2019), 2656–2672.

[90] Jiangdian Song, Yanjie Yin, Hairui Wang, Zhihui Chang, Zhaoyu Liu, and Lei Cui. 2020. A review of original articles published in the emerging field of radiomics. *European Journal of Radiology* 127 (jun 2020), 108991.

[91] Stacie M Spencer, Jessica M Lehman, Christina Wynings, Patricia Arena, Charles S Carver, Michael H Antoni, Robert P Derhagopian, Gail Ironson, and Neil Love. 1999. Concerns about breast cancer and relations to psychosocial well-being in a multiethnic sample of early-stage patients. *Health Psychology* 18, 2 (1999), 159.

[92] Jina Suh, Spencer Williams, Jesse R. Fann, James Fogarty, Amy M. Bauer, and Gary Hsieh. 2020. Parallel Journeys of Patients with Cancer and Depression: Challenges and Opportunities for Technology-Enabled Collaborative Care. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1, Article 038 (May 2020), 36 pages. https://doi.org/10.1145/3392843

[93] Kadri Suija, Kadi Ilves, Pille Ööpik, Heidi-Ingrid Maaroos, and Ruth Kalda. 2013. Patients' experience with cancer care: A qualitative study in family practice. *The European journal of general practice* 19, 2 (2013), 111–116.

[94] Kadri Suija, Tanel Kordemets, Kadi Annuk, and Ruth Kalda. 2016. The role of general practitioners in cancer care: a mixed method design. *Journal of Cancer Education* 31, 1 (2016), 136–141.

[95] Svetlena Taneva, Waxberg Sara, Goss Julian, Rossos Peter, Nicholas Emily, and Cafazzo Joseph. 2014. The Meaning of Design in Healthcare: Industry, Academia, Visual Design, Clinician, Patient and Hf Consultant Perspectives. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 1099–1104. https://doi.org/10.1145/2559206.2579407

[96] Reid F. Thompson, Gilmer Valdes, Clifton D. Fuller, Colin M. Carpenter, Olivier Morin, Sanjay Aneja, William D. Lindsay, Hugo J.W.L. Aerts, Barbara Agrimson, Curtiland Deville, Seth A. Rosenthal, James B. Yu, and Charles R. Thomas. 2018. Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? *Radiotherapy and Oncology* 129, 3 (December 2018), 421–426.

[97] Indrit Troshani, Sally Rao Hill, Claire Sherman, and Damien Arthur. 2021. Do We Trust in AI? Role of Anthropomorphism and Intelligence. *Journal of Computer Information Systems* 61, 5 (2021), 481–491. https://doi.org/10.1080/08874417.2020.1788473 arXiv:https://doi.org/10.1080/08874417.2020.1788473

[98] Kristen Vaccaro, Karrie Karahalios, Deirdre K. Mulligan, Daniel Kluttz, and Tad Hirsch. 2019. Contestability in Algorithmic Systems. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. Association for Computing Machinery, New York, NY, USA, 523–527. https://doi.org/10.1145/3311957.3359435

[99] Mojtaba Vaismoradi, Jacqueline Jones, Hannele Turunen, and Sherrill Snelgrove. 2016. Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice* 6, 5 (2016), 100–110. https://doi.org/10.5430/jnep.v6n5p100

[100] Mojtaba Vaismoradi, Hannele Turunen, and Terese Bondas. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences* 15, 3 (2013), 398–405. https://doi.org/10.1111/nhs.12048 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/nhs.12048

[101] Bas H.M. van der Velden, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, and Max A. Viergever. 2022. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis* 79 (jul 2022), 102470.

[102] Himanshu Verma, Jakub Mlynář, Camille Pellaton, Matteo Theler, Antoine Widmer, and Florian Evéquoz. 2021. "WhatsApp in Politics?!": Collaborative Tools Shifting Boundaries. In *Human-Computer Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 655–677.

[103] Joel Walmsley. 2021. Artificial intelligence and the value of transparency. *AI & SOCIETY* 36, 2 (2021), 585–595.

[104] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 697, 18 pages. https://doi.org/10.1145/3411764.3445432

[105] Dakuo Wang, Justin D. Weisz, Michael Muller, Parikshit Ram, Werner Geyer, Casey Dugan, Yla Tausczik, Horst Samulowitz, and Alexander Gray. 2019. Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 211 (nov 2019), 24 pages. https://doi.org/10.1145/3359313

[106] Jillian L. Warren. 2019. Exploring Socially-Focused Technologies That Can Help Children with Cancer Feel More Like Children Despite Their Disease, Treatment and Environment. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3290607.3299078

[107] Fons Wijnhoven. 2022. Organizational Learning for Intelligence Amplification Adoption: Lessons from a Clinical Decision Support System Adoption Project. *Information Systems Frontiers* 24, 3 (2022), 731–744. https://doi.org/10.1007/s10796-021-10206-9

[108] Emma Burkitt Wright, Christopher Holcombe, and Peter Salmon. 2004. Doctors' communication of trust, care, and respect in breast cancer: qualitative study. *Bmj* 328, 7444 (2004), 864.

[109] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300468

[110] Mireia Yurrita, Dave Murray-Rust, Agathe Balayn, and Alessandro Bozzon. 2022. Towards a Multi-Stakeholder Value-Based Assessment Framework for Algorithmic Systems. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 535–563. https://doi.org/10.1145/3531146.3533118

[111] Sharon Zhou, Henrik Marklund, Ondrej Blaha, Manisha Desai, Brock Martin, David Bingham, Gerald J Berry, Ellen Gomulia, Andrew Y Ng, and Jeanne Shen. 2020. Deep learning assistance for the histopathologic diagnosis of Helicobacter pylori. *Intelligence-Based Medicine* 1 (2020), 100004.