Reimagining Multidisciplinary Teams: Challenges and Opportunities for LLMs in Cancer MDTs

SORAIA F PAULO, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal
ISABEL NETO, LASIGE, Faculdade de Ciências, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal
NUNO LEITÃO FIGUEIREDO, General Surgery, Hospital Lusíadas, Portugal
DANIEL SIMÕES LOPES, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal
HUGO NICOLAU, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal

Multidisciplinary teams are crucial in tailoring cancer care through collaborative decision-making involving several clinical specialties. The inherent complexity of clinical cases, the increasing abundance of unstructured textual data, and the time restrictions of professionals pose significant challenges to team coordination and patient care. This creates an opportunity for generative AI technologies, such as LLMs, to enhance collaborative work. Despite the growing interest in HCI research to explore LLMs in healthcare, we have yet to understand clinicians' perspectives on this emerging technology in multidisciplinary teams. Our work investigates the challenges, expectations and opportunities for LLMs in this context through a speculative approach. We leveraged the Futures Cone framework and conducted a qualitative study with 11 physicians from different cancer multidisciplinary teams. We contribute with an analysis of themes that emerged from individual interviews and a focus group, highlighting LLMs' potential to enhance and reshape multidisciplinary teams' practices. In addition, we uncover concerns and coping strategies related to LLMs' adoption and provide a set of design opportunities to inform the development of technologies for LLM-enhanced multidisciplinary teams.

CCS Concepts: • **Human-centered computing** → **User studies**; *Collaborative and social computing*.

Additional Key Words and Phrases: Large Language Models, Multidisciplinary Teams, Healthcare, Teamwork

ACM Reference Format:
Soraia F Paulo, Isabel Neto, Nuno Leitão Figueiredo, Daniel Simões Lopes, and Hugo Nicolau. 2025. Reimagining Multidisciplinary Teams: Challenges and Opportunities for LLMs in Cancer MDTs. *Proc. ACM Hum.-Comput. Interact.* 9, 2, Article CSCW157 (April 2025), 22 pages. https://doi.org/10.1145/3711055

1 Introduction

Multidisciplinary Teams (MDTs) have become an essential component of highly specialized decision-making in tailor-made cancer care [7, 27] by engaging multiple clinical specialties to collectively review complex cases or newly diagnosed patients [14]. Oncologic diseases are patient-specific, and adapting the best plan to each case is a lengthy process requiring clinical monitoring and treatment over several years [4]. Cancer MDT meetings (MDTMs) are weekly forums where MDTs

Authors' Contact Information: Soraia F Paulo, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal, soraiafpaulo@tecnico.ulisboa.pt; Isabel Neto, LASIGE, Faculdade de Ciências, INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal, isabel.neto@tecnico.ulisboa.pt; Nuno Leitão Figueiredo, General Surgery, Hospital Lusíadas, Portugal, nuno.leitao.figueiredo@lusiadas.pt; Daniel Simões Lopes, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal, daniel.s.lopes@tecnico.ulisboa.pt; Hugo Nicolau, ITI/LARSyS, Instituto Superior Técnico, Universidade de Lisboa, Portugal, hugo.nicolau@tecnico.ulisboa.pt.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 2573-0142/2025/4-ARTCSCW157
https://doi.org/10.1145/3711055

engage in evidence-based discussions of several patient cases. This deliberation process leverages data presentation, analysis and interpretation to negotiate interactively until the MDT reaches consensus over the best course of treatment [26]. These intend to serve as quality checkpoints, ensuring a thorough evaluation of each case [12, 61] while reducing the wide range of individual decisions made by medical professionals [57]. As a result, MDT-based cancer care has shown a potential to significantly increase survival rates, offering better treatment, follow-up and patient outcomes [2, 57].

While clinical and technological advances have all led to an increase in multimodal approaches, these have also escalated the number of potential points of failure during the continuum of care [71]. Navigating through this spectrum of care involves several providers and requires knowledge about appropriate actions and timing [71], which creates a challenge for team coordination and efficiency. One example is the need to thoroughly prepare MDT meetings by analyzing and producing medical data to allow the team to discuss as many cases as possible in a limited time frame. For instance, radiologists review patient imaging studies before the MDTM to illustrate detail in the image through pointing and annotation while explaining their interpretation to the MDT. Otherwise, patients may be required to undergo a new radiological assessment, often delaying the treatment decision [24, 30, 55, 56]. Another example is the need for timely high-quality medical documentation to ensure effective communication and coordination between clinicians who collaboratively manage increasingly more complex patient cases [28].

Artificial Intelligence (AI) shows new opportunities to contribute to clinical practices by collecting different types of data and developing models to provide the most accurate outputs based on the inputs [15]. Most recently, Generative AI emerged as a new field within AI, enabling the creation of digital data following user instructions, through advanced AI models [8, 22]. This is achieved by extracting and understanding intent information from the instructions provided by humans, and generating the content accordingly, based on its knowledge built upon large databases [8, 47]. Particularly, Large Language Models (LLMs) are models based on Generative AI technologies, to recognise, summarise, translate, predict, and generate text or other forms of content [47]. These powerful pre-trained models can produce novel content in several modalities (*e.g.* OpenAI's ChatGPT and Microsoft's Copilot have been fine-tuned for conversational text-based tasks; Midjourney and OpenAI's DALL-E generate images from text descriptions; Google's Gemini can combine text, audio, image and video). As a result, LLMs have the potential, for example, to improve the production, retrieval, analysis, and communication of medical documentation [16], as well as to assist in disease management by interpreting clinical data and generating accurate diagnosis and treatment suggestions [22, 41]. Despite the growing interest in using LLMs in healthcare, we have a limited understanding of how they can be applied to teamwork contexts, particularly in highly heterogeneous and multidisciplinary teams, like cancer MDTs.

In MDTMs, the output is the result from the individual contributions from each professional and the interactions among MDT members [69], leading to a collective decision on each patient's treatment recommendation grounded on clinical data. However, this process can be challenging, due to the vast amounts of unstructured textual data that clinicians need to manage, which is distributed across multiple systems. They need to retrieve, interpret, summarise, and store this data to inform deliberation [13]. In this context, LLMs could enhance the work of cancer MDTs work by contributing to this process.

In this work, we explore the potential of LLMs to support cancer MDTs' work. Throughout the document we comprehensively address the concept of MDTs with a focus on cancer care. Herein, cancer MDTs will be referred to by MDTs. Key research questions include: What are the challenges and concerns of clinicians in using LLMs in MDTs' work? What are the opportunities for using LLMs in MDT's work? What are the expectations for using LLMs in MDTs? To address these questions,

we conducted semi-structured interviews with 11 clinicians from different MDTs. We grounded our research in the current MDT scenarios and used the Futures Cone model[1] to investigate possible, plausible, probable and preferable futures emerging from the current context. We purposely used this speculative design approach to leverage critical and reflective thinking to explore the future as a means to comment on and provide alternatives to the present [33]. Findings underscore the potential of LLMs to enhance the quality and efficiency of medical documentation, navigate vast medical data and reshape the power dynamics within MDTs. It also highlights the transformative power of LLMs in redefining the future of MDTs, particularly in addressing nuanced and complex cases, and prompting institutional protocol revisions while acknowledging and proposing coping strategies for concerns associated with LLMs adoption in healthcare.

We provide two main contributions. First, we contribute with an analysis of themes that emerged from interviews with members from different MDTs exploring present and future perspectives. Second, we contribute with a set of design opportunities to inform the design of future AI technologies for MDTs relevant to HCI, AI and Biomedical researchers.

2 Related Work

This section provides an overview of existing research and developments in areas crucial to our study. To start, we explore the landscape of MDTs leveraging technological solutions in their collaborative decision-making processes. Subsequently, we delve into the role of LLM technologies in the healthcare domain, examining how LLM tools have shown potential to contribute to various aspects of medical practices. Additionally, we examine research on Human-AI teams, offering insights into the nature of Human-AI collaborations and teamwork.

2.1 MDT Work Practices

The span of cancer care consists of a continuum of different types of care and the transitions between them, including risk assessment, primary prevention, screening, detection, diagnosis, treatment, recurrence surveillance, and end-of-life care. Despite this linear progression, a patient may enter this continuum at different stages, ultimately affecting potential outcomes [71].

MDTs gather a group of professionals from several clinical disciplines, such as surgeons, radiologists, medical oncologists, radiation oncologists, pathologists, psychologists, and nurses, who work together to determine the recommended treatment for each patient throughout the continuum of care [24, 71]. Typically, MDTs meet weekly to discuss a number of individual patients, sequentially. Each case discussion considers examination findings, along with radiology and pathology test results, to reach consensus on diagnosis and stage of disease, followed by the management plan recommendation [28]. Regardless of the specialty, the discussion follows a pattern, starting with a physician presenting the first patient's brief clinical history and findings, according to a list prepared in advance and circulated among the MDT before the MDTM. Then, physicians responsible for this patient's management may intervene and share images, reports or other contents they feel necessary for the evidence-based discussion that follows. The MDT proceeds to ask more questions of the individual presenters, reviewing findings and asking for clarification until consensus on the patient management plan is reached. This is followed by a new discussion for a new patient, following the same pattern, until all patients on the initial list have been discussed [26].

2.2 Technology for MDTs

Over the last two decades, there has been a growing interest in research on medical teamwork, namely medical team practices and decision-making. In particular, prior work has explored key aspects of the MDTMs workflow, such as achieving diagnosis as a collaborative effort [26], and the impact of information presentation modalities, such as record keeping affecting MDTM's

efficiency, accountability and individual privacy preferences [28]. Others have analyzed MDTs' work methods, protocols and the design of novel technology [29, 31], proposing a framework to evaluate technology appropriateness for MDTMs that consider multiple perspectives: hospital, technology and people [31]. In addition, others explored building technological tools for MDTMs. Sallnas et al. [59] conducted a field study investigating how laser pointing affected collaboration and communication in MDTMs. Also, Frykholm et al. [17] developed a system that gathered information from multiple medical systems to support the discussion and decision process. Finally, support tools for cancer MDTs continue to be an underexplored topic both in the HCI and CSCW fields.

2.3 LLMs in Healthcare

The increasing volume of clinical text data and the intricate nature of clinical decision-making have paved the way for the integration of Natural Language Processing (NLP) tools to aid physicians in making timely, well-informed decisions, thereby enhancing the overall quality and efficiency of healthcare [20, 44]. Notably, the utilization of pretrained language models in the field of medical AI, especially in the analysis of clinical narratives, has been considered crucial [70]. However, it is important to consider that such models may occasionally produce incorrect outputs [19, 58].

Considering the limited availability of clinical language models and their relatively small number of parameters compared to the general domain models, Yang et al. [70] examined the effects of scaling up a dedicated clinical LLM across five clinical NLP tasks, encompassing clinical concept extraction, medical relation extraction, semantic textual similarity, natural language inference, and medical question answering. Results showed improvements across all clinical tasks, with natural language inference and medical question answering demonstrating significantly better outcomes, underscoring the potential of LLMs in complex NLP tasks. Furthermore, Goodman et al. [20] assessed the responses of GPT-3.5 and GPT-4 to medical questions posed by various physicians from different medical specialties. While the responses were largely accurate and comprehensive, they occasionally contained inaccuracies, preventing the model from being considered entirely reliable. Nevertheless, these shortcomings appeared to decrease over a short period, suggesting that these tools, inline with the recent custom GPT trend promoted by OpenAI [1] , could evolve into valuable resources for swift medical information retrieval.

In the context of fine-tuned conversational chatbots [6, 9, 20], researchers have recognised their potential for medical education and clinical decision support [18, 34]. Levkovich and Elyoseph [41] compared evaluations of depressive episodes and recommended treatment protocols generated by ChatGPT-3 and 4 with those of primary care physicians. Results indicated that the models' therapeutic recommendations aligned with accepted guidelines for managing mild and severe depression without displaying gender or socioeconomic biases, which were observed among primary care physicians. Additionally, Liu et al. [44] explored ChatGPT's suggestions for enhancing clinical decision support and compared them to human-generated suggestions. Findings showed that AI-generated suggestions could complement and optimise clinical decision support by identifying improvements in alert logic and supporting their implementation in complex clinical scenarios. Despite many suggestions requiring modification, they were generally deemed highly understandable and relevant. Lastly, Decker et al. [11] compared informed consent documents for six common surgical procedures, one set generated by surgeons and the other by ChatGPT-3.5. Their analysis revealed that the chatbot-generated consents were more readable, complete, and accurate.

Overall, LLMs have positively impacted the performance of text-based clinical tasks. These works target a plethora of medical specialties, acknowledging the potential for enhancing decision

[1]Introducing GPTs (https://openai.com/blog/introducing-gpts). Last visited: 16 January 2024

support, education and democratizing information to patients and medical professionals. However, most research focuses on single-user contexts, leaving medical team settings largely overlooked.

2.4 Human-AI Teams

Teamwork stands as a cornerstone of healthcare approaches, traditionally characterized by medical teams focusing on a specific medical specialty and involving an internal role hierarchy [3, 23, 35–37, 54, 72]. In particular, MDTs represent a distinct paradigm, with individual members often operating independently or as part of other teams within their routine work [25]. In this sense, it is important to contemplate the desirable roles for AI technologies within MDTs and reflect on the potential impact it may have.

Human-AI teams are increasingly prevalent in various domains [48], showing promise to assist human decision-making in high stakes scenarios, such as health contexts [39]. However, Human-AI teams have yet to achieve complementarity, mostly due to humans overreliance on AI [39, 65]. To address this, recent work has offered new insights into how explanations can in fact reduce overreliance based on a cost-benefit perspective [65], and enhance the estimates of an AI model's accuracy in human-AI collaborative clinical decision-making [39]. Human-AI Interaction has leveraged user- and human-centred strategies to enable the development of fair, accountable and useful approaches that increase automation while enhancing and strengthening individuals [60, 62]. Although coordination between humans and AI teammates mainly relies on Human-AI communication, current AI technologies have been found to offer limited participation in natural language communication with humans, especially in team settings [73]. Considering the recent nature of research regarding communication in Human-AI teams [21, 42], dyadic Human-AI teams have been commonly used as a starting point in this field [42, 73]. One example is the work by Zhang et al., who investigated communication tactics that allow AI to effectively coordinate with humans [73]. Findings highlighted the need for proactive communication from AI teammates to foster a collaborative environment, as well as communication styles promoting balanced efficiency and sociability, facilitating coordination. Ultimately, such dyadic team settings may not reflect on more complex team compositions (i.e. more than 2 members), such as MDTs, where there would be multiple heterogeneous human teammates. In addition, findings may also differ between Human-AI teams with and without the same shared responsibility. Moving towards teamwork with several members changes dynamics between team members with different decision-making styles, thereby affecting team efficacy [48].

3 Methodology

Although there is a growing interest in the research community to explore LLMs in healthcare, particularly in MDTs, we have yet to understand clinicians' perspectives on this emerging technology's benefits, limitations, challenges, opportunities and expectations. Our goal is to contribute with design opportunities that are grounded on current working experiences that will guide the creation of novel LLM-enhanced tools. To achieve this, we leveraged the Futures Cone framework [1] and conducted a qualitative study with 11 physicians from different MDTs. Participants were asked to reflect on multiple futures as a means to comment on the present while analyzing opportunities, barriers, and concerns. This study was approved by our institution's ethics committee.

3.1 Participants

We recruited 11 physicians (8 male; 3 female) from six core specialties for MDTs (2 surgeons, 2 radiation oncologists, 2 radiologists, 2 oncologists, 2 gastroenterologists, 1 nuclear medicine physician) ranging from 4 to 32 years of experience (M=14.36; SD=8.43). Eight participants were from the same institution, and the remaining three were from a second institution. Most participants

Table 1. Participants' demographics: Participant ID (ID), Occupational Experience (O_{Exp}), Experience in MDTs (MDT_{Exp}), Current MDT Areas ($MDTs$): Breast, Dermatology (DERM), Gastrointestinal (GI), General, Gynecology (GYN), Hematology (HEMATO), Lung, Neuroendocrine tumours (NETs), Thorax (THORAX) and Urology (URO)

ID	Medical Specialty	$O_{Exp}(years)$	$MDT_{Exp}(years)$	$MDTs$
P01	General Surgery	12	20	GI
P02	Radiology	11	11	GI
P03	Radiation Oncology	15	15	GI, URO
P05	Medical Oncology	7	11	Breast, GI, Lung, NETs, URO
P06	Gastroenterology	14	12	GI
P07	Nuclear Medicine	4	5	Breast, GI, THORAX, URO
P08	Gastroenterology	12	5	GI
P09	General Surgery	9	9	GI
P10	Medical Oncology	32	20	Breast, GI, GYN, URO
P11	Radiation Oncology	28	32	Breast, GI, General, GYN, Lung, URO
P12	Radiology	14	18	DERM, HEMATO

were part of two or more MDTs (54.55%), while others were part of a single MDT (45.45%). They reported professional experiences between 5 and 32 years (M=14.36; SD=7.88) as part of an MDT (Table 1). In addition, participants reported different levels of experience with LLMs (Table 2). While all participants pointed out the existence of ChatGPT, only six had tried it before. Experiences included translation (N=2), text revision (N=3), information retrieval (N=3) and video content generation (N=1). Participants' recruitment followed a snowball sampling approach until reaching thematic saturation.

3.2 Procedure

This study was conducted in two stages. In the first stage, two researchers conducted semi-structured interviews, using a pre-prepared set of questions and allowing time to follow up with spontaneous, in-depth questions on topics that emerged during the interview. In the second stage, due to clinicians' time and availability constraints, we conducted a remote focus group with a subset of six participants (P01, P07, P08, P09, P11, P12) who participated in the interviews. The remote focus group aimed to further explore some of the key findings from the previous stage and allow participants to build on each other's contributions. All participants were asked to sign a consent form at the beginning of the study and received no compensation for their participation.

3.2.1 Semi-structured Interviews. Interviews were conducted over video calls using Zoom. Researchers took interview notes in real-time and also recorded the sessions for posterior analysis.

The interview had four primary categories of questions: 1) demographics, 2) experience with LLMs, 3) general opinions about LLMs, and 4) critical thinking about the future of LLMs in MDTs. The interview guide with all questions is available as supplementary materials.

First, participants were asked to fill in a demographics questionnaire. Second, we explored participants' experience with LLMs by asking them to explain in their own words what LLMs were and describe any previous experiences with publicly available models (*e.g.* ChatGPT). Researchers proceeded to provide a concise explanation according to the script prepared in advance. Third, to

Table 2. Participants' experience with LLMs: Participant ID (ID), Experiences reported with different features from ChatGPT (Experience), Explanation of the concept of LLMs (Definition)

ID	Experience	Definition of LLMs
P01	Translation, Revision, Retrieval	-
P02	None	*"It predicts the next word based on previous words?"*
P03	Translation, Video content generation	*"Algorithms that connect words within a giant database to find what what we are looking for"*
P05	None	*"I think it is like something that exists in the police, with all (...) the occurrences. I have a case that I want to find out about, I put a clue in there, (...) and similar things or links between that person and other cases come up, as the basis (...) for the investigation."*
P06	Retrieval	*"Specific AI models that (...) calculate the probability of certain phrases being associated, based on undirected training in a very large amounts of online text. They (...) infere about the most appropriate response (...)."*
P07	Revision, Retrieval, Coding	-
P08	Retrieval	-
P09	None	-
P10	Revision	*"(...) Something like a language identification standard and with that you can search a text and be able to look for relevant themes or aggregate data."*
P11	None	-
P12	None	*"Models that allow us to recognise language, our language, I guess. Then (...) more complex models emerge from that (...) and that are called AI."*

elicit participants' views, we used a 2-minute video [2] to illustrate how an LLM tool can be used for efficient document production during a medical appointment. Our objective was to establish a uniform baseline among all participants, ensuring they could offer well-informed opinions about LLMs. After showing the video, participants were asked about the challenges and opportunities associated with integrating LLMs into the work of MDTs.

To delve into participants' views on possible, plausible, probable and desirable futures, we strategically employed the Futures Cone framework [1]. This speculative technique served as both a guiding tool and elicitation method to gain insight into participants' perspectives on opportunities, challenges, concerns, and expectations in future scenarios. In particular, physicians were asked about the role of LLMs in the far and near future of MDTs (*"In 10 to 20 years' time, what role do you think LLMs could play in this context?"*; *"And in the near future, what role will LLMs play in the work of MDTs?"*). Both questions looked into exploring the larger set of possibilities physicians could think of, and naturally start to associate them with different levels of likelihood. In addition, physicians were explicitly asked about the most likely and desirable features offered by LLMs in this context (*"Taking into account the possibilities offered by LLMs, which are the most likely to happen in the future? Why/why not?"*; *"And which are the most desirable? Why/why not?"*). The last question had a dual purpose. First, it aimed to uncover any options that participants considered impossible but still desirable. Second, it sought to clarify which of the options were more or less desirable.

3.2.2 Remote Focus Group. To better accommodate physicians' availability and conflicting schedules, one researcher conducted a remote focus group following the Asynchronous Remote Community (ARC) method [45, 46]. Based on ARC's categorization of activities, we asked participants to

[2]EmpowerMD: Medical conversation to medical intelligence by Microsoft Research (https://www.youtube.com/watch?v=jnGlOCBK3kM). Last visited: 15 January 2024

engage in a semi-synchronous discussion, generating new ideas *(generative)* about different topics. We found out from previous observations that physicians are highly active on WhatsApp groups, so we used a private WhatsApp group *(tool)* to facilitate study activities. Participants exchanged all of their messages in the same group*(shared activity)* during a period of two weeks *(duration)*, and were reminded they could use either text or voice input *(media)* for commodity and ease of use.

First, the researcher invited participants to join the WhatsApp group and waited for all accepting physicians to join before introducing the first topic, considering participants can only keep track of the group's contents produced after they joined. Over the course of the remote focus group, and after any period of 24 hours without any responses, the researcher asked a new question either to clarify and deepen the ongoing discussion or to introduce a new topic. In addition, the researcher's questions avoided targeting specific participants to maintain a shared sense of discussion. The remote focus group targeted two main topics: 1) How could LLMs change the role of the MDT?, 2) How could LLMs affect the MDT's dynamics?

3.3 Data Analysis

Two researchers performed a deductive analysis of the interview transcripts following an iterative coding process. The codes were deducted from our research questions and the Futures Cone framework. We explored opportunities, challenges and expectations, and related it to different stages of the MDT's work (pre-MDTM, during MDTM, and post-MTDM). In addition, we looked at the Futures Cone concepts of possible, plausible, probable and desirable futures, as well as long and short-term perspectives. First, researchers independently coded two interviews, which were then used to discuss and refine the codes to create the initial codebook. Then, both researchers coded a random subset of 31 excerpts, followed by two iterations of refinement and comparison. Cohen's kappa was computed to assess inter-rater reliability. After reaching an agreement (cohen's kappa = 0.85), the two researchers independently coded a subset of the remaining interviews. Later, one of the researchers analyzed the focus group transcripts following the same set of codes.

4 Findings

Findings generated three key themes: LLMs' Role (subsection 4.1), Reimagining the Future of MDTs (subsection 4.2) and Concerns and Coping Strategies Towards LLMs (subsection 4.3). We start by exploring clinicians' perspectives about the role LLMs can play in contributing to MDTs' work. Then, we outline the potential roles MDTs and clinicians could take on following the adoption of LLMs. We finish by presenting clinicians' concerns and coping strategies to deal with LLMs uptake in such healthcare contexts.

4.1 LLMs Role

Clinicians considered LLMs potential to process, generate and interact with vast amounts of medical data. In addition, their perspectives explored how LLMs can evolve into more autonomous team members, and break down power structures by levelling clinical knowledge among the MDT.

4.1.1 Enhancing the Quality and Efficiency of Medical Documentation. In the short-term, participants expressed an immediate interest in these technologies' abilities to automatically generate summaries to prepare each clinical case prior to the MDTM, as well as summaries of the discussions that occur during the MDTM. Moreover, participants actively discussed the plausibility of using MDTM voice recordings as input for generating comprehensive summaries. This additional feature would enhance the MDTMs by capturing and summarising verbal discussions, providing an audio-based perspective to complement the written documentation. These summaries could be incorporated in the MDTMs' reports, possibly combined with seamless integration of different

types of medical data and data sources. This included a strong desire for structuring, organizing and producing high-quality medical documentation efficiently, which was described as key to improve coordination between medical specialties. In particular, participants set a preference for documents following a standardised format over free text, leveraging LLMs as desirable tools to transform clinical notes and free-text reports into documents that offer them all relevant clinical data quickly and clearly.

4.1.2 Interacting with Vast Amounts of Medical Data. Participants emphasised LLMs' probable potential to streamline workflows by leveraging the models' information-handling abilities. Considering the large amounts of medical data generated throughout each patient's continuum of care, clinicians described how MDTs could benefit from LLMs for efficient information retrieval using queries, text-based prompts and visual representations to explore clinical data more efficiently. This included having quick access to a patient's clinical pathway, which was highly desirable, as well as exploring the entire MDTs' dataset to inform the discussion of similar cases. P10 claimed *"(...) Everything that can be done so that we can have information that is accessible more quickly...this allows us to evaluate what is transcribed more quickly and with this can come alerts or entries in flowcharts, pathways, whatever... that enables us, in a simple way, to make queries in the future to things that are, in the meantime, stored or structured, without needing much intervention. (...) How many situations [were there], such as this problem we are now seeing? Being able to have a visualisation and learn from it, I think that's an asset, and I think that's immediate.*". Ultimately, participants described how MDTs could probably feed large volumes of patient data to LLMs to identify trends and correlations to assist the MDTs decision-making process using data-driven insights. P7 explored the concept of LLMs as tools to *"integrate more and more data (...) to stratify risks and analyse the best treatment options"*, emphasizing that *"We [MDT] always take action based on probability, not certainty."*.

4.1.3 From Tools to Autonomous Team Agents. Clinicians portrayed the evolving role of LLMs in the far and near future. In the near future, participants believed LLMs are probably going to be auxiliary to the MDT, describing them as *"tools"*, *"subordinates"* and *"consultants"* which will respond to MDTs' requests. In the long-term, clinicians described different roles in which LLMs could possibly evolve from this subordinate role, subject to being consulted by MDTs, and become more autonomous team agents, potentially serving as clinical managers, supervisors, moderators, collaborators and research assistants. Clinical managers' role targeted LLMs dedicated to monitoring patient care to *"standardise and speed up"* the process. P6 described the plausibility of such systems focusing on informing the MDT about missing data or procedural actions that are required: *"When we are going to discuss a case, it will say to us, "Look out for this patient, brought in because of X, who's missing [data]... and needs to book an appointment in a fortnight' time"*. Participants added LLMs would possibly be able to access the patients' exam results in real-time and help the MDT optimise patients' management, namely by being integrated in multiple systems that allow them to automatically schedule follow-up appointments or exams. In this line, LLMs were also perceived as potential supervisors, or *"alert systems"*. In P6's words *"And then, in a dynamic way, (...) the system itself will know if this information isn't there and ask the question, "So, shouldn't this have been done?", for example"*. Here, LLMs would take on a more active role in supporting MDTs' decision-making, learning from previous cases and making suggestions to physicians about information that is missing and that is not being considered in current case discussions. Similarly, P10 described the possibility of LLMs collaborating in the decision-making process by capturing and processing clinical data presented during the MDT meeting, and presenting its own suggestion: *"While we're discussing, the algorithm is doing this [processing] too. And it [MDT's discussion] can come to an end, and we will have complementary decisions."*. Besides fostering standardised decision-making, LLMs were recognised as collaborators due to their skills in promoting innovative ways to address a

case and enhancing the discovery process. P10 identified a possible problem with standardization, expressing that MDTs might not be evolving or actively looking to improve their decision-making, saying that *"There's also the risk of there being some crystallisation of decisions, let's call it that, according to what is a bit of the [MDT's] practice. Sometimes, you have to think a bit outside the box. (...) I think that they can perfectly make suggestions that open up new horizons and that better solutions can be found, and so I think that this is the horizon towards which we are heading."*. Furthermore, clinicians emphasised that LLMs could probably enhance the quality of medical decisions by constantly being updated with data from the latest medical publications and guidelines. P11 explained *"I think there are a lot of decisions that are the result of poor professional updating. (...) It's impossible for you to read every article. (...) Basically, there's a lot of outdatedness. If doctors are increasingly in demand, the level of information and updating is impossible. (...) So, frankly, the decision could be of better quality. Better quality because it's more up-to-date."*. Finally, clinicians described LLMs working as meeting moderators and possibly ensuring the consistency of the discussion process. P3 explained that *"There are keywords or key aspects of a clinical case that lead to a therapeutic proposal. (...) I think a system like this can systematise the discussion and homogenise it so that decision-making is... more homogeneous (...)"*.

4.1.4 Breaking Power Structures in MDTs. Clinicians suggested that it is possible that LLMs will lead to the democratization of information. This included creating a more levelled discussion between medical specialties, as well as avoiding bias coming from core medical specialties or stronger personalities during MDTMs. P11 explained that *"The truth is that, as far as I'm concerned, I realise that other colleagues have a poor grasp of what my area can offer. And it's possible that I suffer from the same problem in other areas. It's also true that the guidelines are sometimes very clear and unambiguous and sometimes allow for alternative choices. When this is the case, the "alpha" personality present tends to be the most successful. There can and sometimes is a bias. From this point of view, a helper, accepted as a credible, unbiased source without second interests, can and possibly will be a valuable discussion and decision-making aid!"*. In this context, participants highlighted two main advantages. On the one hand, breaking power structures in MDTs would make it possible for discussions to be more objective. On the other hand, clinicians with different levels of expertise would benefit from LLMs that improve inter-specialty communication by providing a holistic view of a clinical case's problem and solution space. P3 summarised this idea by saying *"Bad MDTs are when you have a doctor who often occupies some position of leadership and who has the absolute truth of all the cases and the others are... attendees. This is a very classic MDTM (...), there's one guy who speaks, the others keep quiet and when others disagree (...) something doesn't end up working very well, but this has to do with the pyramid of power within an MDT. And when the specialties aren't all placed on the same level, if there is one who's in charge, who thinks he knows everything, and the others who simply, through some form of passivity, or acceptance or resignation, we could say, simply clam up... This is a bad MDT. A good MDT is when everyone's opinion counts equally within the group, where everyone listens to and respects each other's opinions and a consensus is reached, even a democratic consensus of what is best for the patient."* However, participants argued that despite gaining access to quality information, this would not allow them to make a critical analysis of an area that is not their own, which would probably maintain the need for MDT's discussions.

4.2 Reimagining the Future of MDTs

Considering potential LLMs' contributions, clinicians reflected on how this could reflect on the future of MDTs. P10 explained *"I'm convinced that it [LLM tool] won't replace the role of the MDT, but it will change it"* highlighting that exploring LLMs' role will lead MDTs to transform their own.

This included MDTs focusing on nuanced and complex cases, and revising institutional practices and protocols.

4.2.1 Focusing on Nuanced and Complex Cases. Participants believed that while LLMs show potential for enhancing MDT's workflows, these will probably be seen as complementary rather than substitutes of the MDTs role. On the one hand, clinicians described the effects of democratizing knowledge and contributing to individualised tasks as a plausible way to improve and streamline MDTs' approaches. Participants discussed MDTs' ability to focus on discussing unique medical cases and subjective aspects they do not find desirable or fitting for LLMs' considerations. On the contrary, MDTs are prepared to deal with this information, which was found to be one of the main factors making it possible for LLM-aided MDTs to prevail over complete LLM replacement. P6 provided one example of subjective information they believe LLM tools are not prepared to deal with appropriately, *"It has a lot to do with patients' wishes about what should be treated for their illness. Patients' wishes often don't coincide with what appears in computers, books and algorithms. There are patients who wish not to be treated; there are patients who want to be operated on. (...) In today's medical practice, these wishes are sovereign, while patients are autonomous in making decisions. Another thing is that there are many grey areas within medicine, where it's not necessarily clear that certain options are superior to others. And so sometimes there are flowcharts that are ambivalent"*. Thus, participants highlighted that one of the MDTs' strengths is to take advantage of this collaborative approach to discuss the best options in these grey areas. P3 added *("...) If we make decisions [only] in a structured way without getting into a subjective and individualised debate about each patient, then... MDTs can quietly disappear because any algorithm can make the decision for the group. That's not true, because there are never two identical patients, but there are patients with identical stages, with identical images. But then there's the human being, isn't there? And the only person who knows them [besides themselves] ends up being the doctor who has seen them and who has seen the frailties, the individualities of each person who comes into our consultations, right?"*. Even though consistent decision-making is desirable, clinicians admitted that personalised treatment is key, which involves considering subjective data to make the best clinical decisions for each particular patient. P6 clarified that determining the optimal treatment options involves considering several components varying in importance and between patients: *"The patient should be operated on, but he's looking after his mother. But he's his mother's caregiver. He's the sole carer for an elderly relative. What I mean is that there are many human components related to individualities that are...very predominant. And that every patient has something in particular. We just don't know what it is."*. Participants debated whether these *"human components"* should be registered or not, hampering the LLMs ability to consider it and contribute to such cases. P9 suggested that *"It may eventually speed up less complex decisions, but for more complex ones, it will still need specialist input. Who knows if MDTMs will only be held for a certain subset of patients that the LLM finds more difficult or less accurate?"*. This reinforced previous ideas that LLMs could probably contribute to improving medical decision-making by streamlining less complex cases and opening the way for MDTs to focus on more complex and nuanced decisions.

4.2.2 Revising Institutional Practices and Protocols. Clinicians indicated that while they believe in the probability of LLMs being integrated into MDT's work, they are not entirely sure if or how they will contribute to this process. P11 mentioned that *"It [LLMs] gives us quality decision-making, democratises access, brings more knowledge... and then yes, then we'll see how it ends up"*. Still, participants explained that to take advantage of this technology, they will need to revise institutional processes to avoid inefficiency or unnecessary redundancy. P9 clarified that *"This technology is so disruptive that wanting to see it in processes based on the current paradigm will necessarily be wasting its potential. In other words, it is first necessary to determine very clearly what*

LLMs can contribute to MDTs and to review the role of each player in order to restructure the process. (...) The necessary corollary will be to define the human role in the process as well as we do for LLMs. And, in my view, the goal will have to be to identify what we can offer the process to increase its quality.". Overall, clinicians suggested they would be willing to revise current practices to accommodate and leverage the advantages offered by LLM tools, other than just trying to shape such tools to their needs. P1 stressed the importance of revising and improving institutional protocols by saying: *"It is essential to have a pre-defined institutional protocol (dynamic, changing with the evidence, of course) in the MDT meeting. It must be realised and altered by the team members. This central idea changes everything"*.

4.3 Concerns and Coping Strategies Towards LLMs

While clinicians showed interest in leveraging LLMs potential and enhancing their practices, this was allied with concerns and coping strategies to deal with LLMs' adoption.

4.3.1 Sharing sensitive data. Another common concern was finding a balance between the need for data to enhance and fine-tune LLMs, while ensuring that data remains private. Participants explained MDTs operate within an inherently sensitive context, where cancer patients are vulnerable because of their disease, emphasizing their personal or clinical data must be kept confidential. Thus, participants discussed which aspects must be included in institutional protocols considering LLMs integration in MDTs. They found it plausible for patients' religion, physical and mental state, and obligations to others included, while financial circumstances must not be registered, even though they are accounted for during MDT's discussions. Nonetheless, others believed there were no reasons for concern, as P7 stated that *"If we already let our bank details be used for so many purposes, surely we can make it work for our health data"*.

4.3.2 Broken Promises. Clinicians reported previous acquaintances with AI's promises of improvement in healthcare contexts as a way of justifying this should be a longer rather than shorter integration process. Participants mentioned how AI has long been investigated in the context of image diagnostics in the field of Radiology, but radiologists still need to carry out most processes manually. P3 mentioned, *"I see my colleagues in Radiology opening up the images and continuing to dictate what they see, and continuing to write down what they see... I don't know; five or six years ago, I was at meetings where leading groups in AI and research were asking us what was important to analyse in those scans to feed the system with a lot of information so that it could one day be analysed automatically. (...) Now, if you ask me in the short future if something like this is going to change our clinical practice, I don't think so."*. Another participant, P10, recalled promoting IBM's Watson in its initial stages, and described how it did not meet his expectations *"IBM tried to do this a few years ago with Watson and it didn't go well. (...) Now, it [AI] may have a new impetus, but at the time, there was so much publicity, so much around Watson. I myself promoted meetings and all that, and then nothing came of it. (...) It didn't fulfil the objectives."*.

4.3.3 Multiverse of Competing LLMs. Participants imagined a possible future where multiple competing LLMs would contribute to the MDT's work, discussing how this could affect the liability of their decisions. As P11 put it: *"It can give us quality decision-making, democratise access, improve decision-making, bring in more knowledge and then... Then we'll see how it ends up. (...) When this becomes a reality one day, there won't be one program, there will be many programs rivalling each other and suggesting different things. So another problem is that the decision, down the line, is going to be tainted, you know? Without us realising it"*. P11 explained that LLMs could cater to different industrial partners and companies, or present options based on *"merely financial"* aspects, such

as *"these are options up to 10,000 dollars, here up to 20,000 dollars..."*, which would compromise the quality of the MDT's decisions.

4.3.4 Fostering Therapeutic Obstinacy. Despite LLMs potential to contribute to MDT's decision-making, participants speculated about how it could also promote therapeutic obstinacy by continuously suggesting and pushing for treatment options without knowing *"where to stop"*. As a result, they reinforced the importance of ensuring the MDT has *"the final word"* in all clinical decisions.

4.3.5 Affecting Human Competence. Clinicians expressed concern about the increased reliance on LLMs possibly leading to reduced and deteriorated human interactions, affecting both the MDT and doctor-patient relations. P8 explained that *"Typically, cancer patients end up having a therapeutic approach and then it's follow-up and surveillance. This is often standardised and very regular. If it's well defined and there are no surprises in this follow-up, it [LLMs] could also help. But that would be automatic (...)and then there would be less need, perhaps, for consultations (...) And you could lose a bit of medical contact, which I think is important, even to clarify doubts"*. Participants highlighted that by heavily relying on LLMs, they could compromise human interactions, *"lacking humanisation and sensitivity"* (P6) which are key for cancer patients. Besides, they were worried that LLMs would degrade the interpersonal relation among the MDT by introducing suggestions that can bias MDT members who will follow *"God all "righty", the machine"* (P11) more easily than their peers. Overall, participants feared that the enhancements introduced by LLMs would result in possible loss of human competence.

4.3.6 Fear of the Unknown. Clinicians shared diverging perspectives towards the challenges and concerns they foresee for LLM adoption. While there were several worries, there were also participants who had entirely optimistic views and said they had *"no great concern"* (P7). However, such participants added they only fear the issues they cannot currently predict, as P7 clarified that would not pose a challenge to LLM adoption *"I think that's the big fear, but I'm not normally one to let that stop me, you have to go out there and realise what it is, what's happening."*

4.3.7 Gradual Integration in MDTs' workflows. Despite their interests in different features offered by LLMs, participants described a gradual integration process as a strategy to address their concerns, validate the LLM tools' contributions and give the LLMs time to collect data and learn from the MDTs, as P10 claimed *"A system is only as good as its training"*. In particular, some clinicians highlighted that while LLMs' outputs would need to be verified and validated in the near future, this would probably be temporary, admitting that after a given period of time *"the system would be fine-tuned enough"* (P1), and the MDT could be more trusting of the system reaching a satisfactory level of performance.

5 Discussion

This section discusses our research findings, design opportunities for LLM-enhanced MDTs, our methodological approach, and the limitations of the work. The clinicians' perspectives on the challenges, concerns, and expectations can guide researchers in leveraging and studying the effects of LLMs in MDTs practice. Moreover, designers and researchers can benefit from the design opportunities identified and benefit from our asynchronous research experience when engaging in participatory activities with time-restricted user groups.

5.1 Challenges and Concerns

Overall, participants' concerns reflect their cautious approach towards integrating LLMs in the medical decision-making process. They emphasise the importance of maintaining a balance between technological advancements and ensuring that decisions ultimately serve patients' best interests. To

this end, clinicians explored the need to gradually validate LLMs' outputs as a strategy to mitigate concerns.

Overlooking Patient-specific Factors Participants worry about LLMs introducing bias in the decision-making process and their potential to overlook critical patient-specific factors. Considering the phenomena of overreliance that has been largely reported in Human-AI research [65], this could lead to relevant data being disregarded, especially during nuanced and complex case discussions where LLMs contribute with suggestions. Additionally, in a scenario where LLMs would be heavily trained on a particular institution's dataset, clinicians acknowledged that this could also be a source of bias, considering the limited variability of such training data.

Responsibility and Accountability There is a concern about the responsibility and accountability associated with decisions potentially made by LLMs. While the MDT's approach is based on a decision made by a collective considering several perspectives, clinicians worry about systems being given a role where they actually weigh in or even hold power over treatment decisions in an attempt to streamline patient management. Participants emphasised the importance of having a clear framework for attributing responsibility for decisions, especially in cases where the LLM provides recommendations. Besides being deeply related to the need for validation to ensure such suggestions were correct, this also indicated the need for verification within the MDTs' and the clinical institution's protocols.

Implementation Challenges The practical implementation of suggestions made by LLMs also raised concerns. First, clinicians need to ensure that decisions are feasible, considering dynamic personal and institutional factors like available resources, financial considerations, and existing healthcare infrastructure. While such objective factors could be considered by an LLM-based system, clinicians also mentioned that not all are suitable for registering in the protocol or patient files, hampering LLMs ability to learn and improve their sensitivity in such contexts.

Ethical and Legal Considerations Participants recognised the importance of establishing ethical and legal frameworks to guide the use of LLMs in healthcare. They emphasise the need for these systems to align with established norms, standards, and patients' best interests. This included LLMs' potential for fostering therapeutic obstinacy, considering these models would not know where to stop and would exacerbate unnecessary clinical interventions that no longer benefit the patient.

Ultimately, the primary challenges and issues revolved around LLMs assuming a more prominent role within the MDTs' work. This extended beyond a supervisory capacity, as LLMs began to function as decision-makers with substantial influence over clinical decisions, which does not align with clinicians' perspectives for the desirable future, nor does it align with literature arguing for doctor-AI collaboration in which humans keep the decision-making role [32] .

5.2 MDTs' Expectations

There is a high level of expectation for LLMs to play a transformative role in optimising the MDT process, from documentation to decision support and administrative tasks. The technology is anticipated to enhance efficiency, accuracy, and overall patient care in this context.

Automation of Documentation: Participants expect LLMs to automate the process of documenting discussions and decisions during MDT meetings. This includes keeping records of each medical specialties' input in the discussion, particularly in relation to examination data. The goal is to create accurate and detailed records of clinical opinions, which in turn could be fed to the LLM.

Decision Support: Participants anticipate that LLMs can provide decision support during MDT meetings. For example, if a specific diagnosis or condition is discussed, the system might suggest additional tests, treatments, or follow-up procedures based on available data.

Integration of Data Sources: The technology is expected to integrate multimodal medical data seamlessly. This includes information from various diagnostic tests, imaging studies, pathology

reports, and more. The integrated data would inform clinical decisions, providing quick and easy access to one document or interface.

Impact on Administrative Tasks: Participants foresee LLMs significantly reducing the need for manual administrative tasks. This could include creating reports, scheduling follow-up appointments and exams, and even sending notifications or reminders to patients. This would help optimise the patient's management and reduce administrative burdens, leading to increased efficiency and cost-effectiveness in healthcare operations.

Enhanced Data Analysis and Presentation: LLMs are expected to assist in analysing vast amounts of multimodal data and presenting it in a structured and understandable format. This would provide a clear overview of patient information and facilitate more informed discussions.

Improved Efficiency and Time Management: LLMs are anticipated to save time for healthcare professionals by streamlining patient's case discussions that align with standard institutional protocols and guidelines, where the MDT would only validate the system's management recommendation. This would allow the MDT to focus more on complex and nuanced clinical assessments and decision-making.

Real-time Monitoring and Alerts: There is an expectation that LLMs can monitor patient progress in real-time and provide alerts or notifications for necessary follow-up actions. This could include rescheduling appointments or ordering additional tests based on evolving patient conditions.

In general, clinicians' expectations reflect a transformative vision for the future of healthcare, leveraging a collaborative synergy between advanced LLM systems and MDTs.

5.3 Design Opportunities

We explore a set of design opportunities aimed at enhancing MDTs' efficiency and collaboration.

5.3.1 Automating Medical Documentation. In recent years, the documentation burden involved in clinicians' tasks has created an opportunity for NLP approaches to increase efficiency by alleviating documenting tasks [10, 16, 43, 49, 64]. MDTs rely on documentation from several sources (*i.e.* scattered among different systems) and formats, and medical specialties, which is produced throughout the whole clinical pathway and possibly during long periods of time. Thus, clinicians' expectations and anticipations align with the idea of leveraging LLMs to automate documentation tasks within MDTs. This includes collecting verbal notes and reflections, summarisation and reporting of discussions and broader administrative and data-related functions, producing high-quality medical documentation in a standardised format. Implementing automation in documentation can lead to increased efficiency, accuracy, and consistency in maintaining records of MDT discussions and decisions. This could contribute to all stages of MDTs' work. Before MDTMs, preparing of each patient's case to be discussed at the MDTM, and preventing missing data that can cause delays in discussions. During MDTMs, enhancing evidence-based decision-making. And optimising follow-up actions by maintaining the entire MDT informed of the process.

5.3.2 Levelling Knowledge to Enhance Team Dynamics. There is a notable opportunity for LLMs to level the knowledge across different medical specialties within MDTs. While all clinical information is already available, it is not easily interpreted by other medical specialties, which put expert-presenter roles in a privileged position to influence the discussion [26]. In addition, the MDT approach considers some specialties as playing a core role in the process [38], which ultimately creates a hierarchy and power-structure where some perspectives weigh more than others in decision-making. Considering LLMs abilities to query and present information in different ways, this creates the chance to democratise information, not only by making data readily available but understandable to professionals with different levels of expertise. Also, LLMs could provide a holistic perspective on the problem and solution space of treatments. Moving towards a levelled

playing field fosters a culture of shared knowledge and collective decision-making, breaking down information silos and leading MDTs to re-shape their team dynamics, reduce bias, and create a more collaborative and informed decision-making process.

5.3.3 Enabling complex discussions. Medical decision-making relies on a combination of clinical data, patients' history and wishes, leading to nuanced and complex case discussions. By leveraging international and institutional guidelines, coupled with the ability to assist in data-driven decision-making based on trends and correlations within patient data, LLMs show potential to aid and streamline MDTs discussions, creating opportunities for more complex discussions to take place. Such discussion includes both challenging and intricate medical cases, highlighting the importance of subjective and individualised debate dealing with the grey areas of Medicine, where unique and sensitive data prevalence varies from case to case. Beyond promoting data-driven discussions, LLMs also allow MDTs to explore novel perspectives and approaches to complex cases, leveraging insights and suggestions from LLMs. In addition, the promise of offering a holistic view of the problem also promotes a more engaging discussion that tackles complex cases comprehensively.

5.4 Reflection on Asynchronous Methodology

Over two months, researchers made several attempts to schedule synchronous focus groups, which proved to be extremely challenging, considering clinicians' highly demanding positions and conflicting schedules. To adapt to clinicians' time restrictions, we used the ARC method and facilitated discussion via WhatsApp, which was a familiar platform to all. While this facilitated the discussion, it also created other challenges which are important to reflect on.

First, the group implicitly defined a more experienced clinician as their leader, even though they belonged to different teams and institutions. This practice is embedded in MDTs' structures [50], and even in this asynchronous remote discussion the effect prevailed. This leadership assignment became evident through time, as participants would only comment on something after this leader's response. Moreover the nominated leader explicitly acknowledged his role to the researcher, showing he realized the effect of his answers to guide the remaining participants answers.

Second, the asynchronous discussions were less active with less interactions between participants than in the traditional face-to-face format. The way each question was raised had also an effect on the discussion. A question that started by addressing individual participants and emphasizing the potential contribution of each medical specialty promoted more discussion. Additionally, the conversation flow changed through time. In the beginning, the moderator researcher started the discussions by posting a question, which was followed by individual answers that did not necessarily build on each others' opinions. Over time, participants appeared more willing to reply to each other and take on each other's ideas. At the end, discussions proved more dependent on moderation to keep the exchanges going by adding new questions to clarify some of the mentioned aspects or introducing new subjects that maintained participants' engaged.

Lastly, all participants opted for texting despite being prompted to use the form of input they found more suitable for themselves, including voice input. This might have contributed to less interaction between participants, as they tended to present long answers, which took time and effort to prepare, instead of quicker voice input messages.

5.5 Limitations and Future Work

In this section, we recognise our work's limitations and suggest directions for future research. *Cultural Influence:* We worked with MDTs from two clinical institutions in one country. While our findings aim to be generalised, it is essential to consider the impact of cultural and institutional

backgrounds on the MDTs perspectives. Even so, the international careers of these medical professionals, as well as access to international scientific publications may have mitigated this effect to a certain extent.

MDT's Context: Our study involved clinicians from cancer MDTs. While there may be analogous practices and workflows in other MDTs, their perspectives are significantly shaped by the specific context in which they operate. Future work should consider MDTs working in different areas to validate and build-on our findings. Moreover, our study targeted clinical MDTs and described some of the unique challenges faced by these MDTs. Nonetheless, we argue that some of our findings could be extended and explored in non-clinical settings in the future. This includes the use of LLM-based tools to democratise access to information, leveraging LLMs information-handling abilities, as well as exploring different types of roles LLMs can play as part of different teams.

MDT's Stakeholders: We focused on LLMs potential to contribute to physicians' tasks within MDT contexts, contributing to the collaboration and coordination between medical specialties. While MDTs in cancer care may include other roles, such as nurses, pharmacists and clinical secretaries, these did not play a role in the MDTs we worked with. However, similarly to other works focusing on physicians perspectives in contexts with a variety of stakeholders [68], we acknowledge that considering a wider range of stakeholders would be a necessary complement to our work, to create a more holistic approach.

Despite the MDT model's aim to encourage holistic treatment planning by incorporating a range of expertise, studies have shown this remains a medical-dominated context [5, 67]. For example, nurses tend to participate significantly less in MDTMs than medical professionals, not only due to time and availability constraints, but also because their contributions can be overlooked or ignored. Nurses typically spend more time with patients than other members of the MDT and are therefore more likely to hold patient-centred insights. They are expected to lead on patient and carer communication issues, ensuring that deliberation considers the needs, values and priorities of individual patients [5, 67]. Prior research has shown that nurses' contributions in MDTMs influenced key processes in deliberation, namely prompting discussions, influencing the treatment plan and facilitating teamwork, especially through sharing tacit, patient-centred information [67]. Due to the implicit nature of this knowledge, it may be challenging to articulate or document, creating both obstacles and opportunities for LLMs contribution. Given our findings on the importance of considering patients' preferences, we believe that including nurses' perspectives would further enhance our results. First, leveraging LLMs to create a more levelled and balanced discussion, giving nurses and other healthcare professionals a greater voice in the deliberation process. Second, as nurses take on the role of patient-advocates, this would highlight MDTs' need to focus on nuanced and complex cases, exploring LLMs as complementary tools to streamline processes, but also ensuring consistency and standardisation in the process. Third, the need to revise institutional practices, not only considering how LLMs could play a role among physicians, but also their potential to foster inter-specialty collaboration. Additionally, this could also introduce a new finding, looking into the effects of LLMs enhancing the nurses' role within the MDT, potentially expanding their involvement in communicating MDTM outcomes to patients.

Video-based Elicitation: We used a video to establish a baseline among participants, considering their varying levels of experience with LLMs. We introduced the video as one of many examples showing LLMs' potential in healthcare scenarios. Researchers also highlighted that the video was merely meant as an inspiration, that allowed participants to speculate about other scenarios, which should not limit or constrain their ideas or views about the subject. Nonetheless, this elicitation technique could have created a bias towards the type of examples the video provides within the healthcare context. Future work should explore different elicitation techniques and examine participants' responses based on their experience with LLMs to validate and expand our findings.

Implementation and Ethical Considerations: In this study, our goal was to focus on clinicians' views and uncover a range of ideas, rather than the techniques and processes required to implement them. Therefore, we only present the implementation and ethical challenges mentioned during our interviews, which are associated with the challenges clinicians foresee in the near and distant future. Future work exploring design opportunities in this context should consider analysing and discussing implementation and ethical implications to inform their design process. Still, the integration of AI within healthcare contexts has been challenged by many factors, including the need for human oversight, explainability and privacy [63]. For LLM-based tools to be effectively implemented in hospitals, they must adhere to strict data protection and privacy standards, and data would need to be thoroughly checked for the presence of bias, to ensure reliability and usefulness. Additionally, explainability poses a significant hurdle for the adoption of generative AI technologies, particularly in healthcare [66]. Healthcare professionals depend on transparent communication and the ability to explain clinical decisions, which contrasts with the often black-box nature of the underlying mechanisms and outcomes of LLM systems. To bridge this gap, explainable AI is seen as crucial for fostering adoption while upholding core healthcare values [66].

Moreover, one of digital health's main challenges has been the lack of medical data interoperability [40]. Data is distributed across isolated databases, incompatible systems and proprietary software. To streamline the MDTs work processes, LLMs would need to be integrated within the hospital's digital ecosystem, communicating with a plethora of systems to gather all necessary data.

Other challenges include dealing with data sensitivity and complexity, namely by sharing certain data with the appropriate roles only. If during a consultation a patient shares that they have attempted suicide in the past, this may (or may not) impact the best clinical treatment. This raises the question of whether this information should be shared and with whom. Future systems will need to be designed to consider if LLMs can access such data and use it to generate documentation beyond the individual workspace of the healthcare professional with whom it was shared.

6 Conclusion

In this paper, we explored the perspectives of clinicians regarding the integration of LLMs into MDTs. Findings revealed a nuanced understanding among clinicians, highlighting both opportunities and challenges associated with adopting LLMs in MDT workflows.

Clinicians expressed high expectations for LLMs, envisioning their role as transformative in optimising various aspects of the MDT process. Anticipated benefits included the automation of documentation, summarisation of discussions, decision support, integration of diverse medical data sources, and enhanced efficiency in administrative tasks. The overarching vision was that LLMs could significantly improve patient management, reduce administrative burdens, and facilitate more informed and efficient decision-making. However, clinicians also voiced concerns and challenges. Key apprehensions included the potential for bias and subjectivity in decision-making, issues related to responsibility and accountability, challenges in the practical implementation of LLMs' suggestions, and the need for robust ethical and legal frameworks. Notably, the study highlighted the potential for LLMs to level the playing field among different medical specialties within MDTs. By providing a holistic view of the problem and solution space in cancer treatments, LLMs were seen as a means to democratise knowledge, fostering improved communication and collaboration among specialists.

Based on these findings, we contributed with a set of design opportunities to explore LLMs to improve MDT's efficiency and collaborative practices. In particular, we envision opportunities in the design of automated medical documentation, levelling clinical knowledge among medical specialties, and facilitating complex discussions. Future research and iterative design processes

should address clinicians' concerns, ensuring a seamless and beneficial integration of LLMs into the complex landscape of treatment-related decision-making within MDTs.

Acknowledgments

The authors would like to thank the physicians from Hospital dos Lusíadas and Fundação Champalimaud for taking part in our study. This work was supported by the European project DCitizens: Fostering Digital Civics Research and Innovation in Lisbon (GA 101079116), the Portuguese Recovery and Resilience Program (PRR), IAPMEI/ANI/FCT under Agenda C645022399-00000057 (eGamesLab), and the Foundation for Science and Technology (FCT) through funds SFRH/BD/136212/2018, UIDP/50009/2020 (ITI/LARSyS) [52], UID/000408/2025 (LASIGE) [53] and UIDB/50021/2020 (INESC-ID) [51].

References

[1] 1994. Possible futures, preferable futures.. In *The Healthcare Forum Journal*, Vol. 37. 23–29.

[2] 2000. ACS Cancer Facts and Figures. *Cancer Practice* 8, 3 (2000), 105–108. https://doi.org/10.1046/j.1523-5394.2000.83001.x

[3] Katherine Ann Zellner, Matt Coates, Alex Lee, Swathi Jagannath, Aleksandra Sarcevic, Emily C. Alberto, Allison Harvey, Randall S. Burd, and Ivan Marsic. 2021. Characterizing Speech in Life Saving Interventions to Inform Computerized Clinical Decision Support for Complex Medical Teamwork. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 199–202.

[4] VP Bastiaenen, I Hovdenak Jakobsen, R Labianca, A Martling, DG Morton, JN Primrose, PJ Tanis, S Laurberg, Research Committee, and the Guidelines Committee of the European Society of Coloproctology (ESCP). 2019. Consensus and controversies regarding follow-up after treatment with curative intent of nonmetastatic colorectal cancer: a synopsis of guidelines used in countries represented in the European Society of Coloproctology. *Colorectal Disease* 21, 4 (2019), 392–416.

[5] George TF Brown, Hilary L Bekker, and Alastair L Young. 2022. Quality and efficacy of multidisciplinary team (MDT) quality assessment tools and discussion checklists: a systematic review. *BMC cancer* 22, 1 (2022), 286.

[6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[7] Markus Brunner, Sinclair M Gore, Rebecca L Read, Ashlin Alexander, Ankur Mehta, Michael Elliot, Chris Milross, Michael Boyer, and Jonathan R Clark. 2015. Head and neck multidisciplinary team meetings: effect on patient management. *Head & Neck* 37, 7 (2015), 1046–1050.

[8] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. 2023. A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. arXiv:2303.04226 [cs.AI] https://arxiv.org/abs/2303.04226

[9] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems* 30 (2017).

[10] Enrico Coiera, Baki Kocaballi, John Halamka, and Liliana Laranjo. 2018. The digital scribe. *NPJ digital medicine* 1, 1 (2018), 58.

[11] Hannah Decker, Karen Trang, Joel Ramirez, Alexis Colley, Logan Pierce, Melissa Coleman, Tasce Bongiovanni, Genevieve B Melton, and Elizabeth Wick. 2023. Large Language Model- Based Chatbot vs Surgeon-Generated Informed Consent Documentation for Common Procedures. *JAMA Network Open* 6, 10 (2023), e2336997–e2336997.

[12] Harish Dharmarajan, Jennifer L Anderson, Seungwon Kim, Shaum Sridharan, Umamaheswar Duvvuri, Robert L Ferris, Mario G Solari, David A Clump, Heath D Skinner, James P Ohr, et al. 2020. Transition to a virtual multidisciplinary tumor board during the COVID-19 pandemic: University of Pittsburgh experience. *Head & neck* 42, 6 (2020), 1310–1316. https://doi.org/10.1002/hed.26195

[13] Soraia F Paulo, Nuno Figueiredo, Daniel S Lopes, and Hugo Nicolau. 2024. Enhancing Cancer Care through Design: Understanding Multidisciplinary Teams. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 135–139. https://doi.org/10.1145/3656156.3663718

[14] Mary L Fennell, Irene Prabhu Das, Steven Clauser, Nicholas Petrelli, and Andrew Salner. 2010. The organization of multidisciplinary care teams: modeling internal and external influences on cancer care quality. *Journal of the National Cancer Institute Monographs* 2010, 40 (2010), 72–80.

[15] Eleonore Fournier-Tombs. 2023. *Machine Learning and Collective Unintelligence*. Springer Nature Switzerland, Cham, 147–158. https://doi.org/10.1007/978-3-031-41390-2_9

[16] David Fraile Navarro, A Baki Kocaballi, Mark Dras, and Shlomo Berkovsky. 2023. Collaboration, not Confrontation: Understanding General Practitioners' Attitudes Towards Natural Language and Text Automation in Clinical Practice. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–34.

[17] Oscar Frykholm, Marcus Nilsson, Kristina Groth, and Alexander Yngling. 2012. Interaction design in a complex context: medical multi-disciplinary team meetings. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. 341–350.

[18] Aidan Gilson, Conrad Safranek, Thomas Huang, Vimig Socrates, Ling Chi, Richard Andrew Taylor, and David Chartash. 2022. How well does ChatGPT do when taking the medical licensing exams? The implications of large language models for medical education and knowledge assessment. *medRxiv* (2022), 2022–12.

[19] Rachel S Goodman, J Randall Patrinely, Travis Osterman, Lee Wheless, and Douglas B Johnson. 2023. On the cusp: Considering the impact of artificial intelligence language models in healthcare. *Med* 4, 3 (2023), 139–140.

[20] Rachel S Goodman, J Randall Patrinely, Cosby A Stone, Eli Zimmerman, Rebecca R Donald, Sam S Chang, Sean T Berkowitz, Avni P Finn, Eiman Jahangir, Elizabeth A Scoville, et al. 2023. Accuracy and Reliability of Chatbot Responses to Physician Questions. *JAMA Network Open* 6, 10 (2023), e2336483–e2336483.

[21] Andrea L Guzman and Seth C Lewis. 2020. Artificial intelligence and communication: A human–machine communication research agenda. *New media & society* 22, 1 (2020), 70–86.

[22] Jonathan Huang, Luke Neill, Matthew Wittbrodt, David Melnick, Matthew Klug, Michael Thompson, John Bailitz, Timothy Loftus, Sanjeev Malik, Amit Phull, et al. 2023. Generative artificial intelligence for chest radiograph interpretation in the emergency department. *JAMA network open* 6, 10 (2023), e2336100–e2336100.

[23] Swathi Jagannath, Neha Kamireddi, Katherine Ann Zellner, Randall S Burd, Ivan Marsic, and Aleksandra Sarcevic. 2022. A Speech-Based Model for Tracking the Progression of Activities in Extreme Action Teamwork. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–26.

[24] Bridget Kane and Kristina Groth. 2014. Multidisciplinary Work Practices: A Comparison of Three Major European Hospitals. In *Proceedings of the 2014 IEEE 27th International Symposium on Computer-Based Medical Systems (CBMS '14)*. IEEE Computer Society, USA, 369–375. https://doi.org/10.1109/CBMS.2014.14

[25] Bridget Kane and Saturnino Luz. 2006. Multidisciplinary medical team meetings: An analysis of collaborative working with special attention to timing and teleconferencing. *Computer Supported Cooperative Work (CSCW)* 15 (2006), 501–535.

[26] Bridget Kane and Saturnino Luz. 2009. Achieving Diagnosis by Consensus. *Comput. Supported Coop. Work* 18, 4 (aug 2009), 357–392. https://doi.org/10.1007/s10606-009-9094-y

[27] Bridget Kane and Saturnino Luz. 2011. Information sharing at multidisciplinary medical team meetings. *Group Decision and Negotiation* 20, 4 (2011), 437–464.

[28] Bridget Kane and Saturnino Luz. 2013. "Do no harm": Fortifying MDT collaboration in changing technological times. *International Journal of Medical Informatics* 82, 7 (2013), 613–625.

[29] Bridget Kane and Saturnino Luz. 2017. Trust, Ethics and Access: Challenges in Studying the Work of Multi-disciplinary Medical Teams. In *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 527–528.

[30] Bridget Kane, Saturnino Luz, D Sean O'Briain, and Ronan McDermott. 2007. Multidisciplinary team meetings and their impact on workflow in radiology and pathology departments. *BMC medicine* 5, 1 (2007), 15.

[31] Bridget Kane, Saturnino Luz, and Pieter Toussaint. 2013. Developing a framework for evaluation of technology use at multidisciplinary meetings in healthcare. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. IEEE, 355–360.

[32] A Baki Kocaballi, Kiran Ijaz, Liliana Laranjo, Juan C Quiroz, Dana Rezazadegan, Huong Ly Tong, Simon Willcock, Shlomo Berkovsky, and Enrico Coiera. 2020. Envisioning an artificial intelligence documentation assistant for future primary care consultations: A co-design study with general practitioners. *Journal of the American Medical Informatics Association* 27, 11 (2020), 1695–1704.

[33] Sandjar Kozubaev, Chris Elsden, Noura Howell, Marie Louise Juul Søndergaard, Nick Merrill, Britta Schulte, and Richmond Y Wong. 2020. Expanding modes of reflection in design futuring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[34] Tiffany H Kung, Morgan Cheatham, Arielle Medenilla, Czarina Sillos, Lorie De Leon, Camille Elepaño, Maria Madriaga, Rimel Aggabao, Giezel Diaz-Candido, James Maningo, et al. 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS digital health* 2, 2 (2023), e0000198.

[35] Diana S Kusunoki and Aleksandra Sarcevic. 2015. Designing for temporal awareness: The role of temporality in time-critical medical teamwork. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 1465–1476.

[36] Diana S Kusunoki, Aleksandra Sarcevic, Nadir Weibel, Ivan Marsic, Zhan Zhang, Genevieve Tuveson, and Randall S Burd. 2014. Balancing design tensions: iterative display design to support ad hoc and multidisciplinary medical teamwork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3777–3786.
[37] Diana S Kusunoki, Aleksandra Sarcevic, Zhan Zhang, and Randall S Burd. 2013. Understanding visual attention of teams in dynamic medical settings through vital signs monitor use. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 527–540.
[38] Benjamin Lamb, Heather Payne, Charles Vincent, Nick Sevdalis, and James SA Green. 2011. The role of oncologists in multidisciplinary cancer teams in the UK: an untapped resource for team leadership? *Journal of evaluation in clinical practice* 17, 6 (2011), 1200–1206.
[39] Min Hun Lee and Chong Jun Chew. 2023. Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–22.
[40] Moritz Lehne, Julian Sass, Andrea Essenwanger, Josef Schepers, and Sylvia Thun. 2019. Why digital medicine depends on interoperability. *NPJ digital medicine* 2, 1 (2019), 79.
[41] Inbar Levkovich and Zohar Elyoseph. 2023. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. *Family Medicine and Community Health* 11, 4 (2023), e002391.
[42] Claire Liang, Julia Proft, Erik Andersen, and Ross A Knepper. 2019. Implicit communication of actionable information in human-ai teams. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[43] Steven Y Lin, Tait D Shanafelt, and Steven M Asch. 2018. Reimagining clinical documentation with artificial intelligence. In *Mayo Clinic Proceedings*, Vol. 93. Elsevier, 563–565.
[44] Siru Liu, Aileen P Wright, Barron L Patterson, Jonathan P Wanderer, Robert W Turer, Scott D Nelson, Allison B McCoy, Dean F Sittig, and Adam Wright. 2023. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *Journal of the American Medical Informatics Association* 30, 7 (2023), 1237–1245.
[45] Haley MacLeod, Ben Jelen, Annu Prabhakar, Lora Oehlberg, Katie Siek, and Kay Connelly. 2017. A guide to using asynchronous remote communities (ARC) for researching distributed populations. *EAI Endorsed Transactions on Pervasive Health and Technology* 3, 11 (2017).
[46] Juan F. Maestre, K. Cassie Kresnye, Julia C. Dunbar, Ciabhan L. Connelly, Katie A. Siek, and Patrick C. Shih. 2020. Conducting HCI Research with People Living with HIV Remotely: Lessons Learned and Best Practices. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3375202
[47] Meredith Ringel Morris. 2023. *Scientists' Perspectives on the Potential for Generative AI in their Fields*. Technical Report. https://arxiv.org/abs/2304.01420
[48] Imani Munyaka, Zahra Ashktorab, Casey Dugan, J Johnson, and Qian Pan. 2023. Decision Making Strategies and Team Efficacy in Human-AI Teams. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–24.
[49] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Available at SSRN 4375283* (2023).
[50] Eivor Oborn and Sandra Dawson. 2010. Knowledge and practice in multidisciplinary teams: Struggle, accommodation and privilege. *Human Relations* 63, 12 (2010), 1835–1857.
[51] Fundação para a Ciência e a Tecnologia. 2020. Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa. https://doi.org/10.54499/UIDB/50021/2020
[52] Fundação para a Ciência e a Tecnologia. 2020. Laboratório de Robótica e Sistemas de Engenharia. https://doi.org/10.54499/UIDP/50009/2020
[53] Fundação para a Ciência e a Tecnologia. 2025. Laboratório de Sistemas de Grande Escala. https://doi.org/10.54499/UID/000408/2025
[54] Hannah RM Pelikan, Amy Cheatle, Malte F Jung, and Steven J Jackson. 2018. Operating at a distance-how a teleoperated surgical robot reconfigures teamwork in the operating room. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–28.
[55] Helen A. Powell and David R. Baldwin. 2014. Multidisciplinary team management in thoracic oncology: more than just a concept? *European Respiratory Journal* 43, 6 (2014), 1776–1786. https://doi.org/10.1183/09031936.00150813
[56] Stella Rowlands and J Callen. 2013. A qualitative analysis of communication between members of a hospital-based multidisciplinary lung cancer team. *European Journal of Cancer Care* 22, 1 (2013), 20–31.
[57] Thomas Ruhstaller, Helen Roe, Beat Thürlimann, and Jonathan J. Nicoll. 2006. The multidisciplinary meeting: An indispensable aid to communication between different specialities. *European Journal of Cancer* 42, 15 (2006), 2459–2462. https://doi.org/10.1016/j.ejca.2006.03.034
[58] Malik Sallam. 2023. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In *Healthcare*, Vol. 11. MDPI, 887.

[59] Eva-Lotta Sallnäs, Jonas Moll, Oscar Frykholm, Kristina Groth, and Jonas Forsslund. 2011. Pointing in multi-disciplinary medical meetings. In *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 1–6.

[60] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. 59–68.

[61] Thomas D Shellenberger and Randal S Weber. 2018. Multidisciplinary team planning for patients with head and neck cancer. *Oral and Maxillofacial Surgery Clinics* 30, 4 (2018), 435–444.

[62] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.

[63] Karl Stöger, David Schneeberger, and Andreas Holzinger. 2021. Medical artificial intelligence: the European legal perspective. *Commun. ACM* 64, 11 (2021), 34–36.

[64] Marieke M van Buchem, Hileen Boosman, Martijn P Bauer, Ilse MJ Kant, Simone A Cammel, and Ewout W Steyerberg. 2021. The digital scribe in clinical practice: a scoping review and research agenda. *NPJ digital medicine* 4, 1 (2021), 57.

[65] Helena Vasconcelos, Matthew Jörke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S Bernstein, and Ranjay Krishna. 2023. Explanations can reduce overreliance on ai systems during decision-making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–38.

[66] 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*. 1–19.

[67] Isla Wallace, Helen Barratt, Sarah Harvey, and Rosalind Raine. 2019. The impact of Clinical Nurse Specialists on the decision making process in cancer multidisciplinary team meetings: A qualitative study. *European Journal of Oncology Nursing* 43 (2019), 101674.

[68] Lauren Wilcox, Robin Brewer, and Fernando Diaz. 2023. AI Consent Futures: A Case Study on Voice Data Collection with Clinicians. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30.

[69] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.

[70] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, et al. 2022. A large language model for electronic health records. *NPJ Digital Medicine* 5, 1 (2022), 194.

[71] Jane G Zapka, Stephen H Taplin, Leif I Solberg, and M Michele Manos. 2003. A framework for improving the quality of cancer care: the case of breast and cervical cancer screening. *Cancer Epidemiology Biomarkers & Prevention* 12, 1 (2003), 4–13.

[72] Katherine A Zellner, Aleksandra Sarcevic, Megan A Krentsa, Travis M Sullivan, and Randall S Burd. 2023. Understanding Delay Awareness and Mitigation Mechanisms through an Iterative Design and Evaluation of a Prototype Alert System for Complex Teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–30.

[73] Rui Zhang, Wen Duan, Christopher Flathmann, Nathan McNeese, Guo Freeman, and Alyssa Williams. 2023. Investigating AI Teammate Communication Strategies and Their Impact in Human-AI Teams for Effective Teamwork. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–31.

Received January 2024; revised July 2024; accepted October 2024