Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Corti, L., Oltmans, R., Jung, J., Balayn, A., Wijsenbeek, M., & Yang, J. (2024). “It Is a Moving Process”: Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine. In F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. Toups Dugas, & I. Shklovski (Eds.), *CHI '24: Proceedings of the CHI Conference on Human Factors in Computing Systems* Article 441 Association for Computing Machinery (ACM). https://doi.org/10.1145/3613904.3642551

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

“It Is a Moving Process”: Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine

Lorenzo Corti
l.corti@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Rembrandt Oltmans
rembrandt.oltmans@hotmail.com
Delft University of Technology
Delft, The Netherlands

Jiwon Jung
j.jung-1@tudelft.nl
Delft University of Technology
Delft, The Netherlands
Erasmus MC, University Medical Center Rotterdam
Rotterdam, The Netherlands

Agathe Balayn
a.m.a.balayn@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Marlies Wijsenbeek
m.wijsenbeek-lourens@erasmusmc.nl
Erasmus MC, University Medical Center Rotterdam
Rotterdam, The Netherlands

Jie Yang
j.yang-3@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Figure 1: Salient moments during care for Idiopathic Pulmonary Fibrosis. Clinicians seek support from AI-based Clinical Decision Support Systems (CDSSs) provided an explanation (in boxes) is present. Due to the dynamic uses and purposes of clinicians, different explanations (both in content and visualisation modality) are expected throughout patient care.

ABSTRACT

Clinicians increasingly pay attention to Artificial Intelligence (AI) to improve the quality and timeliness of their services. There are converging opinions on the need for Explainable AI (XAI) in healthcare. However, prior work considers explanations as stationary entities with no account for the temporal dynamics of patient care. In this work, we involve 16 Idiopathic Pulmonary Fibrosis (IPF) clinicians from a European university medical centre and investigate their evolving uses and purposes for explainability throughout patient care. By applying a patient journey map for IPF, we elucidate clinicians’ informational needs, how human agency and patient-specific conditions can influence the interaction with XAI systems, and the content, delivery, and relevance of explanations over time. We discuss implications for integrating XAI in clinical contexts and more broadly how explainability is defined and evaluated. Furthermore, we reflect on the role of medical education in addressing epistemic challenges related to AI literacy.

This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI ’24, May 11–16, 2024, Honolulu, HI, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642551

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Empirical studies in HCI**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Explainable AI, Healthcare, User Needs

ACM Reference Format:
Lorenzo Corti, Rembrandt Oltmans, Jiwon Jung, Agathe Balayn, Marlies Wijsenbeek, and Jie Yang. 2024. “It Is a Moving Process”: Understanding the Evolution of Explainability Needs of Clinicians in Pulmonary Medicine. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI ’24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642551

1 INTRODUCTION

Healthcare providers increasingly pay attention to Artificial Intelligence (AI) to potentially expedite clinical decision-making and administer tailored care to patients. Oftentimes AI applications in healthcare come in the form of Clinical Decision Support Systems (CDSSs) [66]. Despite the promising potential of AI-powered CDSSs, the adoption of these systems is crippled by issues of instability [4, 157] and a lack of transparency [136, 161]. Unlike medical doctors, such systems do not have the same authority in the eyes of patients [40, 161].

As a means to counteract such limitations, Explainable AI (XAI) – by joining efforts from machine learning, human-computer interaction (HCI), and neighbouring fields – has been designing ways to enable AI-based CDSSs to provide explanations for their outputs [15, 69, 71] and promote interpretability [33], reliance [74], and contestability [5, 176]. Despite the ever-increasing body of work within this sphere, research often adopts a positivist, algorithm-centric mindset. It seeks the ideal *explainer* applicable to a breadth of scenarios without pursuing an understanding around *who* the recipients of AI explanations and their needs are. Instead, the HCI community has recently advocated for expanding the notion of explainability by adopting a human-centred viewpoint on it [45, 46, 107]. For instance, prior work investigated medical doctors' information and explainability needs in the early adoption stages [30], around certain tasks [159], and about health-specific datasets [140], and subjective preferences [53] and visualisation modalities [39].

Despite the value of such contributions, existing works still sympathise with the positivist mindset typical of algorithmic research. The pursuit of generalisable insights [39, 53], and experimental studies happening at, and focusing on, a single point in time [30] are the norm. Hence, we have a limited understanding of how explanations for AI-based CDSSs are (1) *used* in practice, (2) for what *purpose* (i.e., the why), and (3) how those might *change over time* in environments as dynamic as healthcare. In particular, in the case of AI-generated treatment suggestions, explanations should help clinicians determine how a patient's trajectory – which might (d)evolve unexpectedly – and past treatment decisions might influence such a suggestion [144]. Accounting for the *temporal dynamics* of user needs is crucial in high-stakes and high-pressure domains like healthcare.

Inspired by categorisations of user needs, and their evolution, in Information Science literature [83, 116], in this work, we seek an understanding of the temporal dynamics of explainability needs of clinicians throughout patient care. We ground our work in the use of CDSSs in pulmonary medicine and, more specifically, in how care for Idiopathic Pulmonary Fibrosis (IPF) is provided at Erasmus MC[1], a large European university medical centre. Given the unknown causes and life-threatening nature of IPF, we echo prior work (subsection 3.2) and embrace a broader definition of *explainability* – encompassing documentation about models, datasets, and processes – to investigate the evolution of clinicians' uses and purposes for explainability. We ask the following questions: (**RQ1**) *What information might clinicians seek in explanations about AI systems?*; (**RQ2**) *When would clinicians engage with explanations about AI systems?*; and (**RQ3**) *To what extent are the properties of explanations from XAI literature aligned with clinicians' purposes?*

To answer such questions, we conducted our study in two phases and used the patient trajectory for IPF as scaffolding for situating clinicians' explainability needs over time. Because patient trajectories are often tight-knit with a country's healthcare system, we combine the national care pathway with the "Patient Community Journey Mapping" design method [88] (hereafter referred to as *journey map*) to outline such a trajectory. First, through an exploratory study (section 4), we (1) validated such a trajectory and (2) gained a preliminary understanding of the study context by engaging a multi-disciplinary pool of participants with backgrounds in computer science, design, and pulmonary medicine. Using the journey map for IPF together with explanations exemplars (Figure 3) , we conducted semi-structured interviews with 12 clinicians, with diverse levels of expertise and specialisations, employed in treating and researching IPF at the Erasmus MC. We enquired about their medical workflows[2], pain points, and uses and purposes for explainable AI-based CDSSs. By treating explanations as a *means*, rather than an *end* (e.g., as in [53]), we investigate the interactions between clinicians and explanations for AI-based CDSSs in pulmonary medicine, the depth of the information being sought, and how these needs evolve throughout a patient's trajectory.

Our results show several tensions around clinicians' explainability needs (Table 3). Given any particular activity, e.g., creating a treatment plan, clinicians seek diverse and mutable explanations – both in content and modality – to cope with the dynamics of patient conditions and the unpredictability of IPF. General explanations (e.g., the patient cohort used to build the system) were preferred to estimate whether a system could be a good fit for them. However, as a consequence of their patient-centric commitment, our participants acknowledged the relevance of patient-specific explanations. Furthermore, clinicians view explanations as tools that support responsible clinical decision-making processes happening on both an individual and équipe level. Finally, epistemic and autonomy challenges were raised in relation to clinicians' capacities to understand and interact with AI systems. As research in XAI progresses and propagates to clinical contexts, we echo prior work [46, 179] around the need for tighter collaboration between its algorithmic and human-centred spheres. To that aim, this work contributes:

- An understanding of how clinicians' explainability needs evolve in relation to the dynamics and uncertainties of IPF. We situate such needs across a patient journey map for IPF.
- Design implications for XAI research in clinical contexts. Taking a longitudinal perspective and tempering common pitfalls of current XAI research [57, 179] are crucial to framing the role of explanations in practice.
- Insights into clinicians' usage of explanations. Explanations for AI systems may constitute tools that clinicians leverage – individually or jointly – throughout their workflows whereas current evaluation criteria and processes are not equipped for that.
- Suggestions for medical education to address epistemic barriers related to AI literacy [112]. Promoting critical thinking about AI-based CDSSs requires close collaborations between healthcare and computer science professionals.

2 STUDY CONTEXT

2.1 Erasmus MC and Patient-centric Care

Erasmus MC began operations in 2013 and grew to be one of the largest and most authoritative scientific university medical centres in Europe, encompassing patient care, higher-level education, and medical research. Erasmus MC includes several research departments and centres covering Pediatric Endocrinology, Allergology,

[1]Erasmus Medical Center, Rotterdam, Netherlands: https://www.erasmusmc.nl/en/

[2]We use *medical workflows* to include both clinical and research workflows [167].

and Hematology. Intertwining patient care and medical research enables Erasmus MC to offer specialised treatments to complex patient cases. At Erasmus MC, healthcare is approached and delivered in a *patient-centric* way, focusing on and respecting individual "patients' personal preferences, desires, and values" to provide high-quality care [37]. To that aim, mutual information exchange between medical doctors and patients is crucial [25]. Ideally, both parties in such interaction should be willing to equally commit to a safe and communicative space to disclose information. In practice, adherence to patient-centric approaches heavily rests on the shoulders and experience of doctors in creating such a space, formulating the right questions, and reducing uncertainty [31, 149, 152].[3] Ultimately, a patient-centric commitment enables doctors – individually or jointly (e.g., in MDOs[4]) – to better inform possible deviations from the established care path depending on patients' needs.

2.2 Idiopathic Pulmonary Fibrosis

Idiopathic Pulmonary Fibrosis (IPF) is a chronic and progressive lung disease causing permanent scarring and breathing difficulties [54]. It is estimated that about 5 million people are affected by IPF globally [120], and it is most common in people in their 70s [54]. After the diagnosis, the average lifespan of patients is between 3 and 5 years [54]. Due to its unknown causes – hence idiopathic – IPF is complex to diagnose and progresses unpredictably [105]. Symptoms include aching muscles, clubbing[5], severe fatigue, and weight loss in addition to shortness of breath. Diagnostic procedures for IPF combine high-resolution computerised tomography (HRCT) scans, chest X-rays, and blood and lung function tests.[6] Several treatments are available to help patients cope with the disease.[7] First, patients are encouraged to adopt healthy lifestyles (e.g., stop smoking or exercise regularly). Prescriptions could include antifibrotic medications – *nintedanib* or *pirfenidone* – and oxygen therapy. Finally, for more severe cases, the existing options cover lung transplants and palliative care (in the late stages of IPF). Cooperation between clinicians, patients, and ongoing research efforts are equally fundamental to providing care to patients affected by IPF. Overall, the intersection of clinical research and patient-centric care makes this a unique context to deeply study the opportunities, challenges, and temporal dynamics around medical AI and XAI.

3 RELATED WORK

We present prior work in Explainable AI from both algorithmic and human-centred viewpoints on the field. Given our focus on idiopathic pulmonary fibrosis, we discuss prior work and applications of (X)AI in (pulmonary) medicine. Finally, we highlight the lack of attention to the temporal evolution of explanations for healthcare.

3.1 Algorithmic XAI

The notion of explainability dates back to research on expert systems [26] and has been reinvigorated by the recent advances of sub-symbolic AI approaches like deep learning; which favour performance (e.g., accuracy) over model transparency. Given the proliferation of these systems in disparate domains (e.g., healthcare [161], and finance [101]), explanations could offer to a variety of stakeholders the means to interpret, evaluate, or contest [5] the output of AI systems.

Nowadays, explainability remains largely algorithm-centred and focuses on describing the outputs of AI systems (i.e., interpretability). Under this interpretation, a plethora of *explainers* have been proposed [15, 69, 71]. Prior work covers *local* (sample-level) or *global* (class-level) explainers, either in post-hoc (i.e., without altering underlying AI models) or self-explaining (i.e., embedded within underlying AI models) [35, 178] fashions. Concretely, common XAI solutions can reflect the importance of individual input features (i.e., feature attribution) [137, 142], select influential [72, 97] or prototypical [34, 126] instances from the training dataset, describe how much a data instance has to change for the model output to change (i.e., counterfactuals) [64, 170], generate human-like concepts [14, 61], or provide rule-based explanations [70, 138]. To cope with the heterogeneity of XAI methods, and to keep track of algorithmic advancements in XAI, prior work has distilled several evaluation *properties* for explanations [7, 33, 100, 122, 147]. Such properties cover both model- or system-specific aspects (e.g., fidelity, stability, or uncertainty) as well as human factors (e.g., comprehensibility, actionability, or coherence).

However, despite the ample body of research, several challenges still remain open. Explainability methods suffer from issues of robustness [145], intra-method disagreement [98], and human understandability [179] – of experts and laypeople alike [6, 13].

3.2 Explainability in HCI

The HCI community argues for and investigates a broader definition of explainability, one that focuses on the recipients of explanations [65, 103, 125] and views AI systems as socially-situated agents [45–47]. To do so, HCI researchers often tie together works and theories from cognitive psychology [110, 111], social sciences [122], design [175], philosophy [27], and – seldom – algorithmic AI [2]. A growing research strand within HCI is that of *Human-centred Explainable AI* (HCXAI).[8] Research within this sphere aims to gain an understanding of *who* the recipients of explanations are [46]. It rests on prior works around framing "XAI stakeholders" [15, 103, 125, 135, 158] and incorporates reflexive practices from design [46] and prior discussions around users and contextfulness of explanations [33, 122, 147]. Furthermore, in contrast with algorithmic XAI research, HCXAI posits a pluralist definition of explanations [51] as different social groups might interpret technological artefacts differently (i.e., interpretive flexibility [19]).

By adopting this lens, a number of prior works connect to the fabric of HCXAI in investigating the technical affordances and end-users of XAI systems [47]. Works targeting developers and practitioners include documentation tools like Model Cards [124] and Datasheets for Datasets [60] as well as an XAI question bank

[3]Uncertainty reduction theory [17] posits that, while interacting, people gather information about the other party to predict behavioural patterns and develop a relationship.
[4]MDO: *multidisciplinair overleg*. Multidisciplinary consultations in which patients' treatments are discussed. These can include professionals from several hospitals.
[5]Clubbing: widening and rounding of the tips of the fingers or toes.
[6]Diagnosing IPF: https://www.nhs.uk/conditions/idiopathic-pulmonary-fibrosis/diagnosis/.
[7]Treating IPF: https://www.nhs.uk/conditions/idiopathic-pulmonary-fibrosis/treatment/.
[8]The CHI community engaged in the discussion through three workshops [49–51].

covering prototypical questions around explainability [107]. Instead, works targeting lay users include data-centric explanations [9] and empirical studies around the relative importance of evaluation properties of explanations [108]. Finally, Langer et al. [103] and Subramonyam et al. [153] propose frameworks to aid interdisciplinary research and communication around AI systems.

In conjunction with such works, others focused on understanding the XAI needs of end-users. Kim et al. [95] enquired about the end-users of a real bird identification app and surfaced needs related to improving human-AI collaboration. Similarly, Cai et al. [29] investigated the needs of pathologists around AI-based diagnostic tools and, later, Cai et al. [30] compiled pathologists' information needs (e.g., capabilities and limitations) during the onboarding phases of prospective AI systems. Instead, Rostamzadeh et al. [140] adapted Datasheets for Datasets [60] for the documentation of healthcare datasets. Finally, Tonekaboni et al. [159] unveiled clinicians' interest in explanations that justify clinical decision-making.

3.3 CDSSs and XAI in Pulmonary Medicine

The application of XAI within healthcare is largely tied to Clinical Decision Support Systems (CDSSs) as the need for explanations is exacerbated by the criticality of medical doctors' decisions, issues of accountability [146], and the proliferation of sub-symbolic AI approaches [43, 118, 141, 166, 171]. CDSSs could "*provide clinicians, staff, patients, or other individuals with knowledge and person-specific information, intelligently filtered, or presented at appropriate times, to enhance health and health care*" [129]. Despite the practical benefits, existing AI-based CDSSs (e.g., Merative[9]) have displayed high false positive rates in real-world settings [161]. Specific to pulmonary medicine, prior work focused around the adoption of AI [89, 93], dedicated support systems [43, 166], diagnostic models [181], and studies comparing CDSSs' performance against pulmonologists' [85, 151, 160]. However, to the best of our knowledge, their wide adoption in pulmonary medicine has not happened yet.

Similarly, while guidelines for implementing XAI in healthcare have been discussed (e.g., [114, 117]), existing surveys [132, 143] show that the application of XAI in pulmonary medicine is sporadically explored. Das et al. [39] highlighted the potential benefits for pulmonologists of using an XAI system to assist in the diagnostic interpretation of pulmonary function tests. Instead, Diprose et al. [41] probed physicians with a hypothetical ML-based risk calculator for pulmonary embolism paired with several explainers [12, 63, 113, 137]. Finally, Evans et al. [53] investigated possible challenges for pathologists in adopting existing explainers [94, 109, 142].

3.4 Longitudinal Perspectives in HCI

The HCI community engaged repeatedly on the topic of longitudinal studies through Special Interest Groups [164], panels [165], and tutorials [82]. Researchers adopted a longitudinal perspective around conversational agents [3, 134], users' behaviours on the Web [58, 156], learning [131], or building specific tools [155]. Despite prior work soliciting longitudinal perspectives and studies (e.g., [48, 115]), one-time data collection is still largely favoured.

[9]Merative (previously IBM Watson Health): https://www.merative.com/.

Related to healthcare, prior works in HCI have approached the problem similarly (subsection 3.2) with longitudinal perspectives being few and far between. For instance, Jardine et al. [84] enquired about end-users' perceptions of internet-delivered therapy over 8 weeks uncovering diverse preferences, uses, and long-term support strategies. Instead, Jo et al. [86] and Blair et al. [21] focused on supporting clinicians when planning and delivering longitudinal health interventions respectively. Such works exemplify the need for longitudinal perspectives when studying clinical settings. It is indeed common for patient conditions and treatments to require clinical progression before actions can be taken – both by clinicians and researchers striving to support clinical workflows.

3.5 Research Gap

Only a dearth of research engaged in understanding end-users' explainability needs [30, 159], preferences [39], and perceptions [11, 41, 53] in clinical settings despite their importance [10]. Furthermore, because such factors are often captured in a single moment in time (e.g., diagnosis [30]), we still lack an understanding of how end-users' explainability needs might evolve over time. Indeed, the temporal dynamics of user needs have been investigated in other fields, e.g., Information Science [83, 116]. We argue this to be a crucial facet of research around AI-based CDSSs: high-pressure situations, uncertain patient trajectories, and doctors' experience can impact the adoption and integration of AI-based CDSSs [40] and the design of explanations they might provide. Specifically for pulmonary medicine and IPF, prior surveys show that research around AI-based CDSSs [89, 150], and explainability [10, 132, 143] is relatively absent.

In this paper, we investigate the temporal dynamics of clinicians' explainability needs within pulmonary medicine. We, particularly, ground our work in the use of CDSSs for providing care for IPF at Erasmus MC (section 2). Unlike prior works that focus on individual points in time [30, 39, 41, 53, 159], we situate such needs throughout patient care for IPF. When doing so, we do not seek to find clinicians' definite preferences for certain explanations (e.g., as in [39, 53]) but rather gain a nuanced understanding of how, and why, their uses and purposes for explanations evolve over time. Finally, we relate our results with literature on Explainable AI and revisit the relevance of the evaluation properties of explanations (subsection 3.1) within pulmonary medicine.

4 EXPLORATORY STUDY

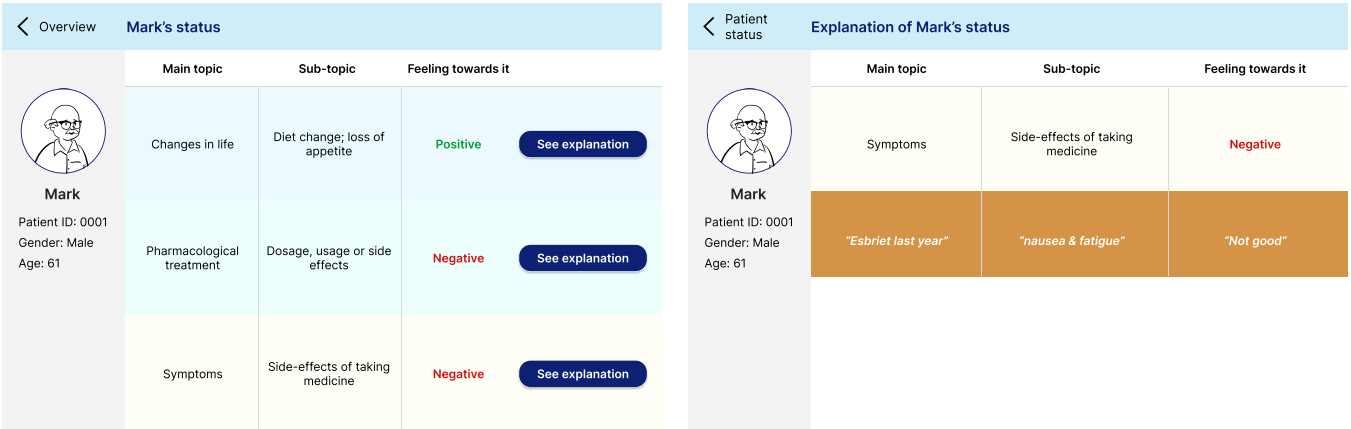
To answer our research questions, and inspired by [173], we conduct our enquiry into IPF clinicians' explainability needs in two steps, namely, an exploratory study and contextual interviews. Here, we describe the exploratory study – a multi-disciplinary co-creation session – to inform the structure and instruments to be used in the contextual interviews with IPF clinicians (section 5).

The exploratory study aimed at informing the design for the main study (section 5), and particularly:

- the interview protocol, by refining our questions to clinicians (e.g., vocabulary)
- the prospective interview prompts, by contrasting a patient journey map and an XAI prototype

Participant	Role	Years of Experience	Familiarity with AI	Knowledge of IPF	Gender
CC-P1	Pulmonologist	22	Basic	Yes	Female
CC-P2	Specialised Nurse	12	Basic	Yes	Female
CC-P3	Resident in training & Postdoc	3	Yes	Yes	Female
CC-P4	Medical PhD Candidate	3	Basic	Basic	Female
CC-P5	Computer Science PhD Candidate	2	Yes	No	Male
CC-P6	Design PhD Candidate	4	Yes	Basic	Female
CC-P7	Computer Science MSc Student	2	Yes	Basic	Male
CC-P8	Design MSc Student	2	Basic	Basic	Female

Table 1: Co-creation participants, their details, and background. *Years of Experience* refers to the years a participant has spent in that role or has had that title.



(a) The prototype shows the main topics, subtopics, and sentiment of that patient's experience.

(b) Extract from a patient's experiences shown as a textual explanation for the main and sub-topic.

Figure 2: Screenshots of the XAI system prototype used in the preliminary study. Here we show the information about Mark, a fictional male, 67-year-old patient.

To structure the exploratory study, we relied on prior literature around the effects that clinicians' past experiences have on assistive tools [30, 40] and how directly testing with existing XAI methods leads to understanding users' preferences and not needs [39, 53].[10] As a confidence check, we opted to query our participants directly to confirm (or set aside) such premises in our specific study context.

4.1 Instruments

Here we describe the instruments used in our exploratory study and tested as prospective prompts for the main study (section 5). These are shared as supplementary material.

Patient Journey. Patient journey mapping [32, 119] is a design method for incorporating patient experiences in healthcare design while providing a bird's eye view of such experiences. In our work, we adopt *patient community journey mapping* [88], a data-driven extension aimed at alleviating the labour-intensive nature of traditional patient journey mapping. We first collected from a US-based platform[11] a large set (140k ca.) of experiences that IPF patients voluntarily shared. Then, we applied topic modelling[12] and manually checked for the validity and reasonableness of the topics. Finally, we aligned our topics with healthcare practice, by combining them with the care path for IPF used at the Erasmus MC.

XAI Prototype. To help participants reflect and envision AI-based CDSSs, we prepared an *XAI system prototype* that allows pulmonologists to inspect patients' experiences at a finer granularity (Figure 2). The design of the prototype was based on the authors' prior knowledge of IPF and existing literature ([29, 67, 173]) but tailored to textual data, i.e., patient experiences. We picked a diverse subset of patients' experiences and associated topics (Figure 2a), and explanations around those topics (Figure 2b) to be displayed in the prototype. The topics aligned with the journey map to reduce friction and cognitive load on participants when moving away from the journey map. Mindful of the time constraints clinicians face in

[10] Prior work [173] has operated similarly when investigating XAI in healthcare.

[11] Inspire, associated with the American Lung Association, offers forums for patients to discuss their experiences. https://www.inspire.com/groups/living-with-pulmonary-fibrosis/

[12] BERTopic: https://maartengr.github.io/BERTopic/index.html

practice, the explanations we generated (using [154]) consisted of salient excerpts from patients' experiences (Figure 2b).

4.2 Method of the Exploratory Study

We opted for a participatory approach to include diverse perspectives and foster a fruitful discussion around the use of XAI within pulmonary medicine. We organised a 1.5-hour-long co-creation session that engaged a multidisciplinary team with expertise in IPF, design, and XAI. Participants were recruited through the professional networks of the authors. Table 1 summarises their details.

Structure. The co-creation session started with a small introduction to the research project and its goals. Then, participants engaged in discussing the journey map in a *think aloud* fashion. Questions covered data collection, rationales around topics, and how those related to clinical practice. Afterwards, following initial familiarisation, participants engaged in using the XAI system prototype (Figure 2) and the explanations it included. Participants were tasked to create short profiles of patients based on the information displayed through the system prototype. The session closed with an open discussion about the perceived usefulness and understandability of the two instruments.

Analysis. The session was recorded with participants' consent and analysed by the first and second authors. Participants' comments were mainly clustered in relation to the journey map and the XAI prototype to decide which instrument to use in the main study. Additional comments were coded inductively and served to inform the interview protocol.

4.3 Outcomes

Together with a rough outline for the interview protocol, the main, and concrete, outcome of the exploratory study (and prompt for the main study) is a validated patient journey map (simplified in Figure 4). The journey map begins with patients experiencing the first symptoms consulting additional (e.g., online) resources. Around the same time, consultations with general practitioners take place. After being referred to a hospital, patients go through physical examinations and tests (e.g., lung function) with lead practitioners and specialised nurses. The results are then discussed by a cohort of medical doctors in MDOs to reach a diagnosis. Then, patients and doctors discuss the definition of a treatment plan. Patients receive continuous support in recurring consultations and treatment revisions. Lastly, patients might opt for a better quality of life and decide on hospice or palliative care.

Overall, clinicians engaged in the co-creation session found the journey map to be comprehensive and aligned with their experience. They, however, pointed out that the journey map represents the "*the ideal situation*" as the timeline can be blurrier. Conversely, participants engaged with the XAI prototype on a surface level, barely interacting with it "*We still need to level it and test it. But it gives a crude impression of what AI is and what it could do.*" (CC-P1). Motivated by the general agreement on the phases and actions portrayed in the journey map by clinicians (CC-P1 - CC-P4), we settled on it as the contextual prompt for our main study.

5 METHOD FOR THE MAIN STUDY

5.1 Recruitment

Interviews were held by leveraging the authors' professional networks to seek out participants with diverse medical roles, years of experience, and familiarity with AI. Given the tight and busy schedules of our participants, we used a combination of purposive and convenience sampling for our study. Concretely, recruitment was carried out through email and in person during selected unit-wide meetings at the Erasmus MC in which the authors were given authorisation to partake. We conditioned further reaching out based on the potential to provide rich insights around XAI in pulmonary and IPF care. Data collection stopped when additional interviews failed to contribute relevant, new information. Overall, we spoke with 12 clinicians whose details are summarised in Table 2.

5.2 Conducting Interviews

Interviews[13] were scheduled from May to July 2023 and lasted on average 35 minutes, depending on clinicians' availabilities, and were recorded using videoconferencing software. Respondents were sent an informed consent form beforehand.[14] Interviews started with an off-the-record introduction about the goals and outline of the interview. After that, with the participants' consent, we started the recording. We prepared an interview guide to provide a flexible structure for the conversations. Initially, respondents were asked "grand tour questions" [106] about their medical role, and familiarity with AI and IPF. Depending on the latter, participants were then shown the journey map (section 4) as an initial prompt to establish meaningful communication [44] and to discuss their practice and knowledge. Thereafter, participants had access to the journey map as a reference for sharing their experiences. The interviews then proceeded to discuss the challenges they currently face in practice, e.g., creating treatment plans. Once a common vocabulary was established, we started shifting the attention to AI in pulmonary medicine. We enquired about their perceptions of AI, what role they see it taking, and how it could affect the scenarios disclosed thus far. Finally, we delved deep into clinicians' needs and uses of explanations in medical workflows, in the context of AI systems, and how these two domains compare or (mis)align. Given the breadth of the potential insights around explanations, we are guided by the framework from Xu et al. [174] and probed participants on *what* they would like to be explained (e.g., data features), *when* (e.g., disease diagnosis), and *how* (e.g., numerically) that should be explained. In this last segment, we relied on explanation exemplars (Figure 3; subsubsection 5.2.1) to surface insights specific to pulmonary medicine.

5.2.1 Exemplar Explanations. To help participants reflect on the kind of explanations they might look for, we hand-crafted a selection of *exemplar explanations*[15] drawn from algorithmic Explainable AI literature. We used the exemplars in a "what if?" fashion only after querying clinicians about their envisioned use for explainability. Furthermore, we refrained from preparing a large pool of exemplars

[13] Interview guide available as supplementary material.
[14] The research and informed consent materials received approval from the Human Research Ethics Committee of our institution.
[15] A similar idea is that of *conceptual artefacts* [59] from speculative design.

Participant	Medical Role	Background	Years of Experience	Familiarity with AI	Knowledge of IPF	Gender
P1	Pulmonologist	VLD	20	Yes	Yes	Male
P2	Pulmonologist	Surgery	12	Yes	Yes	Male
P3	Pulmonologist	Thoracic Oncology	1	No	Basic	Female
P4	Resident in training	Epidemiology	6	No	Basic	Male
P5	Resident in training & Postdoc	Oncology	4	No	Basic	Male
P6	Resident in training & Postdoc	ILD	3	Yes	Yes	Female
P7	Physician-researcher	Thoracic Oncology	3	No	No	Female
P8	Physician-researcher	Radiology	1	Yes	No	Female
P9	PhD Candidate	ILD	1	No	No	Female
P10	PhD Candidate	Oncology	2	No	Basic	Male
P11	PhD Candidate	OPD	3	Yes	Yes	Female
P12	PhD Candidate	Pharmacology	2	Yes	Yes	Female

Table 2: Interview participants' details and background. *Years of Experience* refers to the years they have spent in that role or have had that title. VLD: vascular lung diseases; ILD: interstitial lung diseases; OPD: obstructive pulmonary diseases.

but rather selected a variety of visualisation modalities to elicit rich and contextualised responses rather than inquiring about their specific preferences around existing XAI methods. Inspired by Kim et al. [95] and Vilone and Longo [168], we prepared 6 exemplar explanations (Figure 3) based on existing XAI methods and ascribing to real tools and visualisations customary to pulmonologists [39], e.g., pulmonary function tests. Out of the 6 exemplars, 4 are single-modality (Figures 3a – 3d): numerical (e.g., [137]), rule-based (e.g., [138]), textual (e.g., [14]), and visual (e.g., [142]). The remaining two combine visual and textual elements (Figure 3e), and rule-based, visual, and textual elements (Figure 3f) respectively.

5.2.2 Data Processing. Interviews were conducted in English (by the first author) and in Dutch (by the second author) according to participants' preferences. Dutch-spoken interviews were manually transcribed and later translated into English.[16] Instead, English-spoken interviews were automatically transcribed.[17]

5.3 Analysis

We analyse participants' (anonymised) responses through *codebook* thematic analysis (TA) [96, 121, 139]. This declension of TA, situated between *reflexive* [23, 36] and *coding reliability* [22, 68, 87] approaches to TA, provides scaffolding to answer our research questions in an integrative manner [24, 36]. From an epistemological perspective, we adopt a contextualist account [77, 81] and consider responses to be valid knowledge within pulmonary medicine (**RQ1**, **RQ2**). Instead, from an ontological perspective, we embrace a critical realist account [56, 65] in the attempt to expose latent information about the evaluation of XAI within pulmonary medicine (**RQ3**). We adopt as a reference point the criteria for evaluating explanations – both model- and human-centred ones – surveyed by Liao et al. [108]. Practically, we engaged in a combination of deductive and inductive coding to identify initial central concepts – based on literature and interview structure – and then build meaning around those and emergent concepts. The first author took the lead in the data analysis, first familiarising themself with the data (by reading transcripts and creating preliminary descriptive memos [62]), and then coding the data. The second and third authors contributed with partial coding, review of the codes, and definition of themes – ultimately mitigating individual positionalities. Coding was conducted using Atlas.ti[18] while groups and themes were delineated and refined through in-person meetings between the authors. We identified 270 codes, organised into 25 clusters, further refined into 6 groups, and finally distilled into 3 themes.

5.4 Authors' Positionality and Perspective

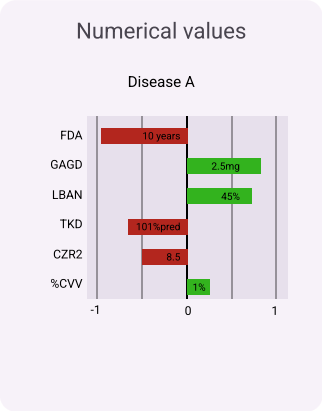
To provide more clarity to readers, we disclose how the authors' perspectives and assumptions shaped the analysis. The authors (all based in the Netherlands) work in diverse fields. Authors 1 (Italian male), 2 (Dutch male), 4 (French female), and 6 (Chinese male) research in computer science. Author 3 (South Korean female) researches in design and healthcare. Author 5 (Dutch female) researches in pulmonary medicine. Despite some unfamiliarity with the study context, our interest in exploring the intersection between healthcare, AI, and XAI led to the willingness to deeply investigate a single, relatively less explored context as a means to gain focused insights. The construction of this paper was mostly shaped by author 1's views on XAI and reflections with the co-authors.

We acknowledge that, due to our background and occupation, we approach the domain from a position of privilege. However, despite that and the introduction of external theories in our study ([88, 108]), we commit to giving up the belief that our prior knowledge is superior to that of the involved clinicians (Krogh and Koskinen [99]) and commit to a careful and contextual interpretation of clinicians' responses around their experiences, perception of XAI, and patient-specific examples that were brought up throughout the interviews.

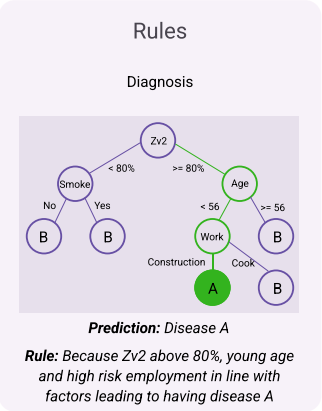
6 RESPONSES FROM CLINICIANS

We now discuss the themes resulting from our interviews with clinicians, organised in relation to our research questions: information needs (**RQ1**), moments and conditions in which doctors

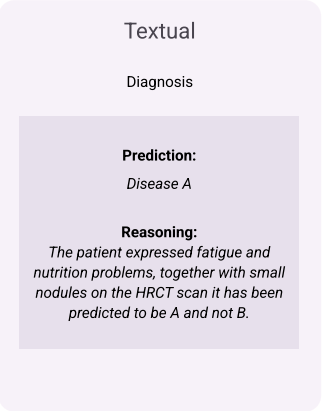
[16] DeepL Translate: https://www.deepl.com/translator
[17] Microsoft Teams: https://www.microsoft.com/microsoft-teams/group-chat-software
[18] Atlas.ti: https://atlasti.com/



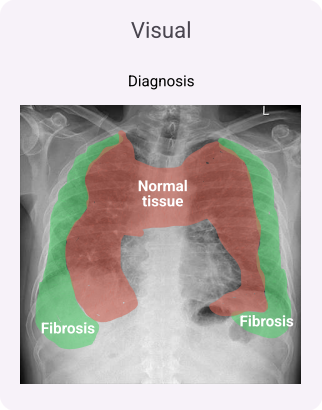
(a) Exemplar for numerical explanations referring to lab test results.



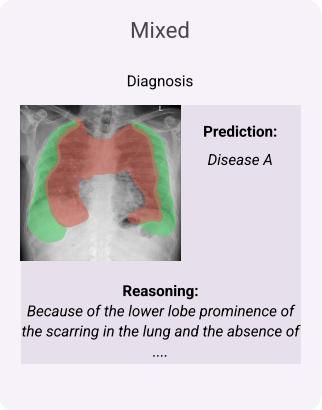
(b) Exemplar of rule-based explanations with individual patient parameters.



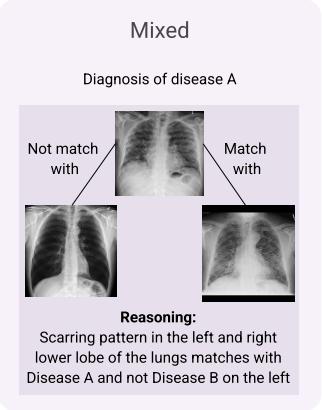
(c) Exemplar of textual explanations reporting the rationale for a diagnosis.



(d) Exemplar of visual explanations based on a chest X-ray.



(e) Multi-modal exemplar combining visual and textual elements.



(f) Multi-modal exemplar combining rules, visual, and textual elements.

Figure 3: Explanation exemplars we prepared for the interviews with clinicians working on IPF. Inspired by [95, 168]. The contents of explanations are plausible but fictional.

seek explanations (**RQ2**), and the alignment between properties of *goodness* of explanations and clinicians' purposes (**RQ3**). We relate our results to the journey map for IPF in Figure 4 and summarise key insights in Table 3.

6.1 Theme 1: "*With paracetamol, you don't know exactly how it works*" – About Explanatory Depth

The "unnatural" feeling (P11) of communicating decisions to colleagues (P9, P12) and patients (P3, P10) without explaining motivated them to look for explanations with AI-based CDSSs too. Particularly, participants expressed the need for both *general system explanations* about the affordances of such systems (particularly around validation) and *local, patient-specific explanations* that would highlight patient-specific factors. Regardless, multi-modal *visualisation modalities* seemed to align better with clinicians' needs.

> "*I think we need some explanation, it won't be sufficient to only say, well, "it's pneumonia". [...] We are used to explaining how we got to a certain answer [...] it would feel very unnatural to only get one diagnosis without any explanation.*" (P11).

6.1.1 General System Explanations. First and foremost, participants highlighted the need for general explanations that would surface *1)* the capabilities and limitations of AI-based CDSSs as well as *2)* details about the patient cohort data used during development. Our participants saw AI-based CDSSs primarily as enhancing tools for which they do not need to know the underlying technicalities. P4 exemplified this need by equating AI systems, and their explanations, with paracetamol, i.e., a medication about which not everything is known but it is used because of its benefits.

> "*With paracetamol, you don't know exactly how it works.*" (P4).

In this sense, participants expressed uneasiness around technical jargon which felt irrelevant in practice (P1, P3, P7). The only outlier was P6 who, given prior experience with AI algorithms, wanted explanations to cover concrete implementation details such as models

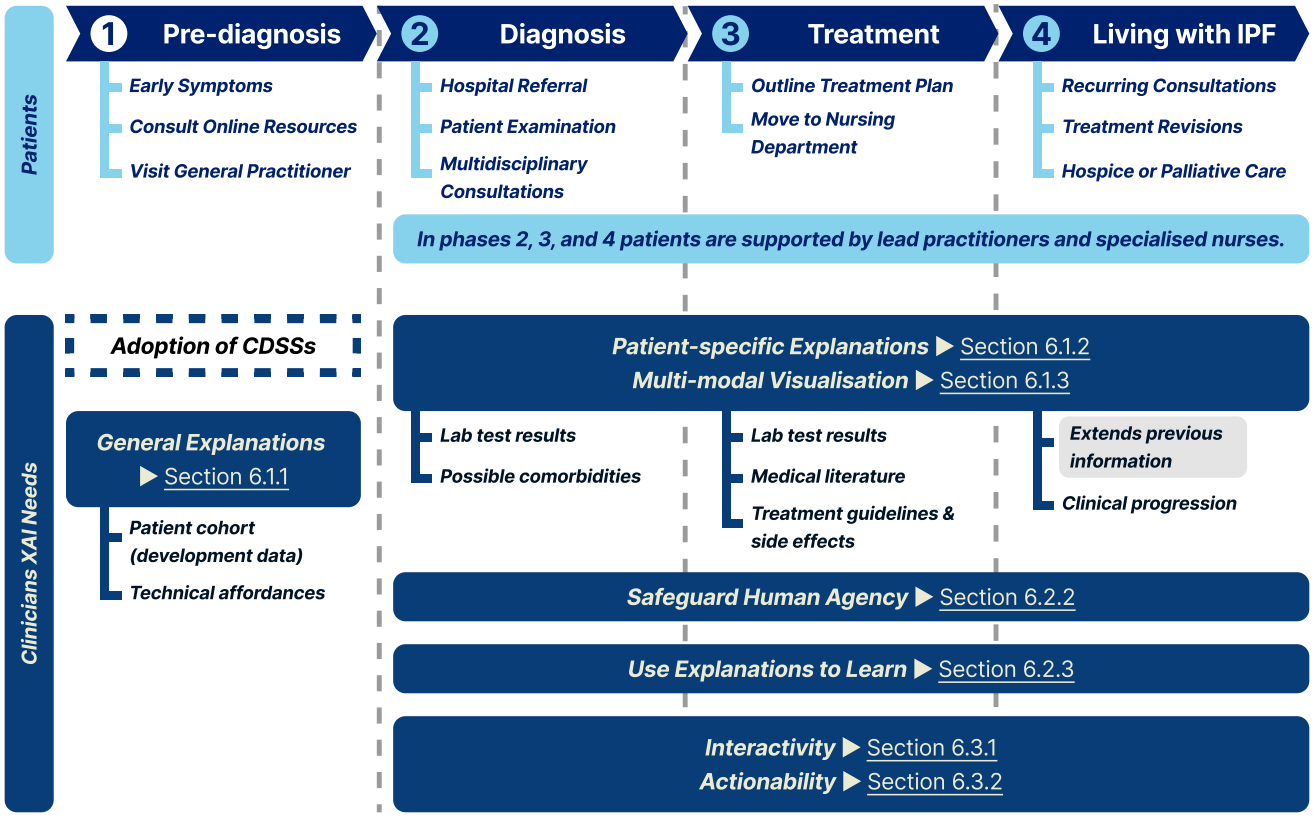


Figure 4: Salient phases of the IPF journey map (the complete journey map is shared as supplementary material). For each phase, we highlight our results and indicate the corresponding section. Additionally, we specify what content our participants found relevant at each phase of the journey map. Initially, clinicians seek general explanations that cover, e.g., the demographics of the patients whose data was used to build the CDSS. Then, clinicians desire explanations that cover patient-specific factors. For instance, if an AI-based CDSS is meant to support diagnosing IPF, its explanations should cover lab test results and possible comorbidities. Overall, clinicians see explanations as means to safeguard agency over clinical decision-making and learn to use AI-based CDSSs. Specifically, clinicians found interactivity (e.g., being able to ask further questions based on the explanation) and actionability (i.e., explanations that contribute to their clinical practice) to be key properties of explanations.

and scoring functions. Most of all, participants focused on the data used to build such systems, its clinical relevance, and the demographics of the patient cohort represented in such data (P3, P5, P6, P7, P8, P9, P10). Participants specified that such explanations could give them an early impression (P6) of whether a prospective AI support system would be applicable and useful for them in practice (e.g., as a physician (P8)), given the type of patients under care.

> "*It has to be validated [in] real-life with a patient population similar to the one you have in your own hospital. So, if it's not, then you can't even test it for that. I think it's important to monitor that*" but "*[...] it doesn't necessarily have to be in our department, as long as that population matches. You just have to look: is this the normal IPF patient that we have? It also has to be done in the same kind of hospital.*" (P6)

6.1.2 Local, Patient-specific Explanations. In relation to their patient-centric commitment (P1), participants acknowledged the value of local explanations to surface patient-specific factors that, in turn, would help them provide high-quality care.

> "*I think we have to remember that [the] goal of us, walking in this building here, is that we still want to provide the best care.*" (P1).

As such, local explanations were deemed useful if related to factors such as test results, comorbidities, or unforeseen disease declines (P6, P7) instead of AI-specific features or mechanisms.

In light of this, almost all participants (excl. P10 and P12) denoted how having access to lower-granularity explanations would provide more credibility to clinical decision-making which otherwise would feel "unnatural" (P11). Particularly when communicating with patients, local explanations could reflect information around risk factors for IPF (P9), test results (P3), side effects of treatments (P7), and historic trends in similar patients (P2).

> "*I wouldn't just tell people that I do it this way because the AI says so [...] We sometimes deal with very rare things, and then I come across something again, and*

> *then I think 'why did we do this?' [...] It's also man's nature [to ask] why, so I think that's also what is needed.*" (P3)

6.1.3 Visualising Explanations. Overall, participants (P1, P2, P3, P4, P5, P7, P10, P11) gravitated towards multi-modal explanations (Figures 3e and 3f). These were perceived as instruments allowing for clinicians' discretion, and enabling them to quickly glance over explanations and dive deeper if needed (e.g., in the presence of a rare mutation (P3)) in their day-to-day practice.

Instead, single-modality explanations (Figures 3a – 3d) yielded mixed reactions and were perceived as highly situational instruments, dependent on the nature of the question being asked.

> "*That very much depends on what question you ask. Is it what percentage of fibrosis does a patient have? Then of course you want a picture like this [points at Figure 3d]. If you ask what diagnosis your patient should give, then the factors and rules are more helpful [points at Figure 3b]. If you want to decide on treatment, then you also want to see a decision tree like that [points at Figure 3f].*" (P9)

Numerical and rule-based explanations were considered customary for clinicians – "*[...] that's kind of the way how I think, [how] many doctors would also apply their train of thought.*" (P5) – as they resemble lab tests and reasoning processes respectively. Often, they related this information to their training in universities and hospitals (P3, P4, P5, P7, P12). On the other hand, textual explanations were considered passable (P1, P10) but "*disappointing*" (P5) if not related to medical literature. Participants also found this delivery modality to be useful in communicating the regulations or guidelines an AI system might follow or refer to (P5, P7, P8, P11). Instead, for disciplines like radiology and oncology where imaging techniques are more common, visual explanations would allow clinicians to visually identify elements they recognise (P9) and formulate a preliminary understanding (P1, P2, P4, P10, P11).

While these comments emphasise the nature of the information our participants seek from AI explanations, we explore their relation to medical workflows and tasks in subsection 6.2.

6.2 Theme 2: "*So that you can see the clinical progression*" – About Explanation Dynamics

Participants' needs over *what* explanations should include are tightly connected to their medical workflows. Here, we organise such comments and rationales around the patient journey map (Figure 4) to highlight how *explanations translate to medical practice*. In parallel to this, two major considerations emerged. First, the desire to *safeguard human agency* and second, the role of explanations in *learning* how and when an AI-based CDSS should be used.

6.2.1 Translating Explanations to Medical Practice. While discussing explainability, participants focused on selected phases of the journey map (Figure 4). Namely, *Diagnosis* (Phase 2), *Treatment* (Phase 3), and *Living with IPF* (Phase 4). Regarding instead **Phase 1** (Pre-diagnosis), participants did not discuss it in depth as it mainly occurs outside Erasmus MC and access to information about patients is difficult or absent.

Phase 2: *Diagnosis* – Patients that reach this phase of the journey, usually arrive at the hospital after being referred by a general practitioner. After some preliminary checks, e.g., pulmonary function, the first consultations are scheduled. In this setting, participants saw explanations and AI as support tools for diagnosing IPF, acting as a second set of eyes to first compare and contrast their judgement with that of AI, and dive into the explanation in case of disagreement (P7, P9, P11).

> "*Well, you can have something that was already decided without AI and whether there is a match between those two [the AI's and clinician's decision]. [...] But if it doesn't match, then you can look at the explanation as to why that is. Then I would look at the explanation [to see] if something else comes out that I overlooked.*" (P7)

Critically, P6 commented on explanations possibly playing a bigger part within MDOs – their gold standard for diagnosing IPF:

> "*If it's [the explanation] integrated also with the MDO then it's just part of the prediction. Then you're not going to say I say this, the AI says this. Then you also use that explanation with it [the prediction] in the MDO.*" (P6)

Phase 3: *Treatment* – After IPF has been diagnosed, doctors and patients engage in the definition of a treatment plan. Participants commented that for AI-generated suggestions to be meaningful, patient-specific characteristics (e.g., mutational stages of a tumour (P7), or kidney function (P3)) and treatment side-effects should be clarified in the explanation (P5, P7, P10, P11).

> "*[...] even going as far as proposing treatments – that is more in oncology and especially [for] me in pulmonary oncology [...] pinpointing specifics based on mutational stages of the tumour and some other biological processes*" (P5)

Further, participants juxtaposed the thoughtfulness that is required for some decisions (P1, P2) to the transactional nature of most AI-based CDSSs – ask, and be told. By referring to his occupation as a lung transplant surgeon, P2 highlighted the unfitness and lack of care and understanding similar systems might exhibit in certain scenarios. Carefully crafted AI explanations, could better assist clinicians in assuring credible clinical decision-making processes.

> "*Then you also do justice for this type of care product [lungs], to avoid missing people, or transplanting them too early or unnecessarily. [...] Lung transplantation is a big grey area it's really a small group of patients.*" (P2)

Phase 4: *Living with IPF* – In this last phase, treatment is already underway and recurring consultations ensure that patients react correctly to it, that the (possible) side effects of medications are bearable, or that treatments are correctly revisited. Here, participants (P6, P10, P11) independently focused on the temporal dimension and the importance of adopting a longitudinal view around explanations by, for instance, having AI-generated suggestions explained through trends (P2, P9, P10). P2 exemplified this by referring to the moment following a lung transplant:

> "*Initially, lung function rises, and then it stabilizes. But it can also be a rejection, then there's a drop in lung function. Then we perform a number of steps: a CT scan, a bronchoscopy where we culture for bacteria and virus, and where we take morsels of tissue for diagnostics, and send those to the pathologist and they see, for example, an A1-B0 reaction, an A1 you can also have if there's a viral infection at play. At that point we wait to see: if the cultures are negative, we can still decide to give a rejection treatment.*" (P2)

Although P2 is an outlier in our participant pool, this comment symbolises the breadth of contextual information that explanations for AI-based CDSSs should, in their view, relate to.

Instead, participants (P5, P10, P12) commented on re-assessing treatment plans highlighting both the value of explanations that relate to temporal dynamics and their possible initial absence due to the lack of clinical progression (P12).

> "*You then put the cures in an interval, so that you can look at the side effects. Those are linked to the lab, if you see that there is an increase or decrease in lab values then we know that maybe we should make an adjustment.*" (P10).

6.2.2 Safeguarding Human Agency. While participants recognised the performance of CDSSs to be crucial towards their adoption (P6, P10), they also stressed the need for a human controller throughout patient care given the high diversity of IPF patients they attend to (P3, P5, P7). Nonetheless, our participants envisioned diverse uses for those systems – as additional data points (akin to lab test results (P11)), as an additional pair of eyes (P4, P12), or as artefacts meant to give suggestions (P5) – but always with the idea that "*[the AI] has to add something in practice*" (P8). Such viewpoints are not surprising. Our participants – often facing complex patients' needs – rightfully consider their training and interactions with colleagues foundational to their modus operandi whereas AI could introduce unwarranted roadblocks.

Despite these interpretations, participants expressed the possibility for AI explanations to be integrated with existing workflows (e.g., in MDOs) and become aids towards more "*substantive discussions*" (P2), information sharing (P6, P10), and diagnostics (P1).

> "*I think you should always be able to discuss it [the explanation] with a colleague.*" (P10).

Concretely, explanations could provide the means to better evaluate AI-generated suggestions related to, e.g., adjust treatment plans (P8); or ignore them altogether (P10) based on patient examinations.

On this last point, several participants (P3, P7, P8, P10, P11) stressed that having a human controller present throughout patient case does not necessarily signify attempting to become better collaborators with AI itself (P12). Instead, critical thinking needs to be exercised. To that aim, P1 directly challenged responsibilities clinicians might have in the future as AI systems get more and more prevalent:

> "*Do we see our own role as [some] sort of interpreters of the information and having a good conversation with the patient? Or do we see that we do still have a role to see if the AI is still correct with our own ideas?*"

6.2.3 Explain to Learn. The final facet of explanations that surfaced during our interviews is related to explanations serving as learning tools on how to use an AI system. Provided that an AI system has been appropriately validated and some guarantees are given beforehand (subsection 6.1), our participants discussed more concrete positions on learning to use AI systems in practice and testing whether they hold up to the initial expectations.

Initially, doctors might look at explanations more frequently (P7, P8, P12) as a way to "*to go into a little depth*" (P1) into what an AI-based CDSS might be doing and get accustomed to it. During this probation period for AI, doctors can formulate a mental model of how that system might operate and come to an understanding of what that system could do for them *concretely*.

> "*So [in] a complex system where you have a lot of patients [...] it's really nice that at least in the beginning when you start using it you understand what exactly counts because everybody has in their head an algorithm how you aggregate all those patients characteristic to a product, treatment A or B.*" (P8)

Additionally, this can be combined with a *prospective validation* of the system (P8) in which a system is tested against a large backlog of historical data (e.g., CT scans) and compared with the suggestions and rationales of radiologists.

Only after clinicians learn the capabilities and shortcomings of an AI system, they might start taking into account acceptance (P8, P10). However, participants hinted at difficulties around the acceptance of an AI system by referencing howprolonged collaborations help them get a sense of who the *more knowledgeable others*[19] are when in need of a second opinion (P1, P3, P11). In this vein, the practical viewpoints of our participants highlighted the conceivable decay in the utility of explanations in the case of an AI-based CDSS that displays consistent behaviour alignment with their own judgement (P3, P4, P9, P11, P12).

> "*At some point, you trust someone's knowledge and ability when you consult someone who you know is very knowledgeable about something. That is of course more difficult in such a large automated system.*" (P1).

Concurrently, more experienced participants acknowledged the consequences of AI explanations and AI-generated advice for inexperienced clinicians. These might be both "*enlightening*" and foster learning, or detrimental and provide convincing motivations for what they do to the point they "*do not know any different*" (P2).

6.3 Theme 3: "*Then it doesn't have as much value to me*" – About the Goodness of Explanations

Concerning properties of *goodness* of explanations from XAI literature (subsection 5.3), participants naturally focused on *interactivity* as a means to *personalise explanations* to their needs while retaining agency (subsection 6.2). On top of that, participants underscored the necessity for those explanations to display *actionable insights* that help them chart the next course of action.

[19]The concept of *more knowledgeable other* from Vygotsky's theory of cognitive development [169] refers to someone who has a better understanding or ability about a particular task, process, or concept.

6.3.1 Interactivity for Personalised Explanations. Participants viewed interactivity of explanations as a key property for them to flexibly query explanations and retain agency (subsection 6.2). They repeatedly underscored their interest in explanations that could convey clear and concise information. Explanations that are too extensive could lead to high cognitive load [6, 179], or be completely disregarded: "*Then it doesn't have as much value to me.*" (P8). Some participants framed the *compactness* [100, 147] (i.e., the amount of detail) of explanations as an upstream design choice dependent on the concrete task or application: "*If it's too detailed then of course people aren't going to look anymore. It's really per-application how detailed it should be.*" (P6). In addition to this, some participants mentioned the long-term effects of explanations, e.g., building end-users' trust, and the potential benefits brought by detailed *justifications*[20] for AI systems: "*That may be a lot of reading but that's what's going to help build trust eventually*" (P3 on Figure 3c).

In attempting to strike such a balance, participants highlighted the epistemic barrier that AI explanations might create. Our participants desired explanations to be slim and free from technical jargon so as to not hinder their *comprehension* [33] of AI's affordances. P1 – echoed by P6, P8, and P11 – stated that "*some degree of knowledge would be necessary, but you don't need to exactly know how the system works in the background.*". Similarly, P5 expanded this by pointing to the need to assess "*the critical steps*" an AI takes towards a decision and communicate those to clinicians. Given such accounts, visual modalities of explanations (subsection 6.1) play a fundamental role in the way information is conveyed to clinicians. For instance, rule-based (Figure 3b) and textual (Figure 3c) explanations should be of "*manageable size*" (P3, P7) for clinicians to be willing to engage with them. Oftentimes, multi-modal explanations seemed more beneficial for our participants: "*Text I don't like. Visual is too little. 'It should speak', mixed is best.*" (P3).

In this sense, some participants saw the *interactivity* [18, 147] of explanations as a plausible solution to their concerns about the comprehensibility and accessibility of explanations, allowing them to further query the AI and its explanations (P5, P9).

> "*I think it's important that it's visual at a glance but that if you want more information you can zoom in for more information [...] so that you can still ask questions.*" (P9)

6.3.2 Obtaining Actionable Insights. Participants also expressed the need for explanations to provide actionable[21] insights that help them chart the next course of action, e.g., coming to a diagnosis, or escalating the discussion to an MDO. Explanations were perceived as companions to their own decisions (subsection 6.2), particularly, as preliminary checks while waiting for more educated judgements and rationales from colleagues or lab tests. Thus, for explanations to be actionable, they should refer to information that relates to doctors' practice (subsection 6.1). While including percentages or probability values within explanations might communicate the *(un)certainty* of an AI system's answer, these were perceived as de-contextualised and unclear – if not useless.

> "*You never know for sure. Suppose within a certain patient category the system doesn't work 90% but 70%, and the output is yes or no. Something comes out [it], but you don't know which group that falls into. It's hard to look at even with [the] probability of whether that advice is right or wrong.*" (P8).

In view of this, respondents instead longed for explanations that would be *coherent* [122, 147] with external sources of knowledge: prior patients (P8), expertise shared within the group, and medical literature (P3, P5). P3 considered this to be a much-needed basis for comparison given the unpredictability of IPF, the state-of-the-practice at Erasmus MC, and the credibility required in clinical decision-making.

> "*References, what it [the advice] is based on, from the scientific literature, to see what the basis of the advice is. We sometimes deal with very rare things, and then I come across something again, and then I think 'why did we do this?'. If there are references there, then I understand why we did that.*" (P3)

Related to this, several participants (P1, P2, P4, P5, P7, P12) seemed aware of the possible aversion toward suggestions and explanations from AI stemming from how they have been operating in the past. Despite their propensity for research and the frequent need to re-assess their decisions, they reflected on how they might judge more harshly disagreeing information (explanation or not), and by whom it is given.

> "*We are terribly opinionated of course. Often I had a colleague ask me 'What would you do?'. I then say 'I would do that', and then they ask 'Why would you do that?'. But then they go back and do it their own way anyway. So, we are stubborn after all.*" (P2)

7 DISCUSSION

Explainability is critical for the integration of AI-based CDSSs in pulmonary medicine. By interviewing clinicians working on IPF, we identified several tensions around clinicians' explainability needs. While general, non-technical explanations were preferred, their patient-centric commitment called for patient-specific explanations that could enable them to maintain agency over clinical decision-making. Furthermore, clinicians might face challenges in engaging and understanding explanations. Results from the interviews are summarised in Table 3. We now discuss the implications of our results for future research.

7.1 Integrating (X)AI in Medical Workflows

Erasmus MC is a university medical centre where research and clinical workflows are intertwined to provide high-quality care. Because of their commitment to patient-centric care, our respondents were more welcoming of technological advances and cutting-edge treatments found in pulmonary medicine literature. Despite the presence of healthcare protocols at the national level and within Erasmus MC, our respondents were not afraid to stray from such predefined "*ideal*" pathways if beneficial to patients. In this sense, our respondents often referred to their 1-on-1, or group (e.g., in

[20] The notion of justifications for AI systems is enquired in prior work at the intersection of law and AI [76, 78].
[21] We adopt the broader definition of actionability by Liao et al. [108] instead of the one from algorithmic recourse [91, 147].

Theme	Key Insights
Theme 1: "*With paracetamol, you don't know exactly how it works*" – About Explanatory Depth (subsection 6.1)	*1)* General explanations about system affordances are preferred, if free from technical jargon. Confirms [30]. *2)* Uncertainties around patients and disease motivate the need for local explanations that surface patient-specific factors. Confirms [173]. *3)* Clinicians gravitate towards multi-modal explanations that cover multiple information sources.
Theme 2: "*So that you can see the clinical progression*" – About Explanation Dynamics (subsection 6.2)	*1)* High variability in IPF pushes clinicians to seek diverse explanations throughout patient care, even when re-examining patients. *2)* Explanations should support group dynamics within the medical équipe and promote clinicians' agency of AI-based CDSSs. *3)* Explanations' relevance can diminish in time as doctors learn when and how they can use a CDSS.
Theme 3: "*Then it doesn't have as much value to me*" – About the Goodness of Explanations (subsection 6.3)	*1)* Compactness of the explanations affects willingness to engage with them and the comprehension process around the capabilities of a CDSS. *2)* Interactivity can modulate the details included in explanations, allow for follow-up queries, and contribute to clinicians' sense of agency. *3)* Actionability of the explanations relates to charting the next course of action, not to altering the output of the system.

Table 3: Summary of the themes and insights obtained by interviewing IPF clinicians.

MDOs), exchanges with colleagues as the benchmark for sharing and contrasting their perspectives.

Their patient-centric commitment appears, however, to be in tension with issues related to the working environment: slow-moving (albeit trustworthy) administrative processes, shortage of staff, and hurdles in securing funds. Integrating explanations could spur further roadblocks in such an environment. Indeed, regardless of the presence of explanations, participants often underscored the idea of letting the AI generate its output and then enabling them to take ownership of disregarding that suggestion or testing it first-hand with patients (P5). Prior work has also highlighted similar behaviour in non-expert end-users of AI systems who, while valuing interpretability, prioritized accuracy [127].

7.1.1 Taking a Longitudinal Perspective. Prior research in HCI has proposed and investigated a plethora of tools aimed at supporting healthcare professionals in their activities [11, 29, 30]. Despite their valuable insights, those works often report about snapshots in time and do not account for the temporal dimension of supportive tools for clinical decision-making. While our work only provides qualitative pointers toward how clinicians' needs around XAI might evolve over time, we believe future research around developing and testing explainable CDSSs should adopt a longitudinal angle (e.g., [133]). Designing and conducting longitudinal studies is resource-intensive. However, they could give a broader perspective and grounding around users' explainability needs around CDSSs in addition to specific preferences – enquired in [39, 41, 53] – on existing explainability methods. For instance, for diseases as uncertain as IPF and according to our respondents, it would be very easy for a hypothetical explainable system to be incorrect. Upstream system validation, either from a technical ("training", "validation", and "test" approach) or clinical (controlling for patient cohort) standpoint, would only provide partial reassurance. Directly testing a CDSS in practice with a range of real patients (e.g., as in [21, 84, 86]), would instead supply clinicians with enough information to determine the usefulness of such a CDSS. We stress that we do not argue for fully automating some of clinicians' activities, but rather resonate with our respondents in concealing CDSSs as recommenders over which clinicians maintain full agency (e.g., through reject options [75]) – both during and after testing a CDSS. Finally, on a general note, journey maps can provide an informative scaffolding for enquiring about the temporal dynamics of user needs. For instance, they have been proven useful in retail to understand customers' behaviour, feelings, and attitudes [180]. However, journey maps are context-specific, potentially challenging to create (or dependent on data availability; subsection 4.1), and that warrant attentive validation.

7.1.2 Avoiding Shiny Objects. It is clear that nowadays advances in AI happen at breakneck speed. The same can not be said for healthcare, and for good reasons. Even at Erasmus MC new discoveries and tools from medical research do not immediately alter, or disrupt, existing practices. Reproducibility and clarity of evidence are foundational for clinical adoption. As Topol [161] said (later echoed by Antoniadi et al. [10]), AI-based tools in medicine are "*high on promise and relatively low on data and proof*". While the promise of better-performing AI systems sounds enticing, we argue it is important not to fall pray of the *Fear of a Better Option*[22] when evaluating prospective CDSSs. We concur with our participants in viewing human agency and alignment with clinical practice as more important than accuracy-based metrics on datasets which, despite being purposed for similar tasks (e.g., classifying nodules malignancy (P5)), might include a sample of patients with (very) different demographics. This also holds for explanations as the majority of XAI research regularly focuses on a small subset of criteria when evaluating explanations, or devises ad-hoc benchmarks that obfuscate potential pitfalls of the explainers being proposed [57].

We suggest that future researchers investigating explainable systems for healthcare first gain a clear understanding of the needs

[22]Fear of a Better Option: https://patrickmcginnis.com/blog/meet-fobo-the-evil-brother-of-fomo-that-can-ruin-your-life/

and requirements of end-users – something that disciplines like software engineering [80] have been advocating for decades – and then seek a balance between such requirements and the technical prowess (i.e., raw performance) of prospective systems. Prior HCI works [29, 144, 163] shed light on some of these aspects. Indeed, we acknowledge this to be a perilous but worthwhile path to follow given the unstable nature of the current generation of AI systems and their diverse and contextualised interpretations [19].

7.2 Explanations are Part of Conversations

Our study showed the importance for clinicians to access explanations that have a translation to clinical practice, which they regularly referred to when discussing their expectations for useful CDSSs. Our participants, particularly, saw explanations as a means to provide credibility to clinical decision-making. Despite the partial overlap in results with prior work [41, 53, 159], our participants also viewed explanations as support tools within medical workflows. That is, something clinicians could bring up in individual discussions with colleagues and larger multidisciplinary meetings as additional data, evidence, or doubt on which to deliberate.

7.2.1 Informational and Transactional Needs. The participants, unsurprisingly, saw CDSSs as additional tools at their disposal capable of surfacing relevant information either corroborating their viewpoint or novel and insightful. The information-seeking process of our participants resembles the one outlined by Sivaraman et al. [144]. If a CDSS' recommendation, and associated explanation, are aligned with clinicians' judgement, they would treat it as evidence and move on with it – *similarly* to how they interpret lab test results. Conversely, if misalignment is present, they would ignore the machine recommendation if under pressure. Instead, if partial alignment is present, clinicians might postpone the final decision and seek the opinion of a *more knowledgeable* colleague (e.g., with more years of experience, or a different specialisation). While the final goal might remain the same (e.g., making a diagnosis), in the latter case the nature of the information-seeking process shifts from *informational* [83] – clinicians intend to satisfy their information needs – to *transactional* [83] – clinicians intend to locate a different source of information to satisfy their information needs [52].

Future research could further investigate this phenomenon and connect with ongoing efforts around explanations that provide both evidence and criticisms for a machine recommendation [28, 123] or that can be selected based on users' input and goals [102]. Our participants indirectly underscored this aspect when discussing data used to build AI systems. If a particular patient is under-represented in a cohort, an AI-based CDSS might exhibit, e.g., popularity bias [1, 148], skewing its recommendations and generating incongruency with the patient-centric commitment of our participants. Furthermore, while our participants displayed reluctance to blindly trust AI-based CDSSs, prior work uncovered issues of anchoring bias [162] in medical settings [11] related to when both AI suggestions and explanations are served (e.g., before clinicians' decision-making process). While we did not use a prototype system, it is conceivable that depending on the nature of the AI-based CDSS (e.g., proactive monitoring or reactive diagnosis support) different delivery strategies, and their timing, should be further investigated.

7.2.2 Opportunities for the Design of Explanations. Comparatively to the extensive work on AI-supported clinical decision-making (e.g., [29, 73, 92]), the design of explanations for clinical scenarios has received little attention. Oftentimes, prior works make use of explainers that are readily available [39] which, however, are not necessarily aligned with clinicians' needs. In this context, we echo prior studies around HCXAI (subsection 3.2) around relaxing the predominant techno-centric view on XAI and extending the definition of "explanation" beyond AI system output to include users' needs and purposes for explanations.

Besides the broad need for explanations connected to clinical practice, our results show that the temporal dimension of users' needs and purposes largely affects how explanations are used, if at all. Throughout patient care, clinicians drift from general explanations about AI-based CDSSs (sought during the early adoption phase, confirming [30]) toward local, patient-specific explanations that benefit their practice more directly (subsection 6.1). For the former, several artefacts already exist in the form of documentation for the underlying models [124] and the training datasets [60]. However, those works target a different audience (i.e., developers) and not clinicians. Future research in this area could focus on tailoring such artefact for clinicians similar to how Rostamzadeh et al. [140] adapted 'Datasheets for Datasets' [60] for healthcare or how Anik and Bunt [9] explain training data to end-users.

Instead, for local explanations, while our participants gravitated towards multi-modal explanations, several factors condition their use (subsection 6.3). Naively, multi-modal explanations could be achieved as a combination of existing XAI methods that is later rigorously tested [42] with clinicians to ensure its ecological validity. On a deeper level, we advise future research to be directed towards interactive explanations [125, 174] and, particularly, *selective* and *mutable* [18] explanations. Selective explanations would enable clinicians to decide when to interact with them, change visualisation modality, and tweak the granularity of the information (subsubsection 6.3.1). This would allow clinicians to quickly glance over an explanation and, if necessary, expand to view additional details, e.g., medical literature and pose more questions. Selectivity directly relates to clinicians' perspectives on the visualisation modalities (subsection 6.1) and their desire to maintain agency (subsection 6.2). For instance, our participants desire explanations grounded in medical literature (e.g., when creating a treatment plan). Mutable explanations expand these ideas to encompass testing hypotheses and comparison of different circumstances. Clinicians could use these explanations to inquire about how a patient would react to treatments by tweaking and inspecting the explanation. Furthermore, in case multiple AI models are implemented within the same CDSS, mutable explanations would allow clinicians to reconfigure the system and get a variety of suggestions and explanations. In turn, mutable explanations support clinicians in navigating diverse sources of information (subsection 6.1), e.g., patient data, and medical literature. Finally, we stress the importance of adopting a longitudinal view to both complement existing work in the area and better inform how selectivity and mutability should be designed, implemented, and evaluated (subsection 6.3) given clinicians' specific needs.

7.3 AI Literacy: an Absentee in Medical Curricula?

Our results surfaced a diverse spectrum of explanations, both in terms of depth of the information (subsection 6.1) and interaction moments throughout patient care (subsection 6.2). Despite the desire for explanations, participants often mentioned that such information should not be too technical as there is no need, at all times, to know how a CDSS might work in the background. However, regardless of the inclusion of technical information or jargon, this raises the question of whether clinicians (even beyond pulmonary medicine) possess the background knowledge to properly interpret explanations of AI systems. While specific institutions and domains (e.g., radiology) may provide some level of AI training during residencies, prior work in other disciplines [104, 112, 172], to the best of our knowledge, has yet to address AI literacy in healthcare staff and students beyond self-reported measurement scales [90]. Oftentimes, there is no direct connection between medical knowledge and the decisions, or decision-making process, of an AI-based CDSS. For instance, causality is a crucial factor in medicine for an effective, efficient, and satisfactory clinical decision-making process [79]. However, despite ongoing efforts from researchers in causal ML and XAI [20, 69], it is still an elusive concept within the current generation of AI systems.

We concur with some of our participants on the impending need for a broader educational support around AI literacy [112]; seemingly missing from several medical curricula. We note that the need to include AI literacy was explicitly voiced (P1), and asked to us (P2), by participants with more medical experience. Their concerns were related to the over-reliance that younger clinicians could manifest when using CDSSs during training. They worried about inexperienced clinicians leaning too much on those tools rather than learning from more experienced colleagues and potentially reaching a point where they do not know any different. We do not argue for a radical shift towards a technical imprint in medical curricula but emphasise the need for introducing basic notions of AI early – additionally to what is provided via residency training – so that clinicians are better equipped to critically evaluate the outputs and explanations of AI systems. Close interdisciplinary collaborations between healthcare and computer science professionals (in spirit, similar to [153]) could assist such an endeavour.

AI systems, in general, do not hold any communicative intent [16] and, as such systems (technically) advance, so does the risk of plausible-looking decisions, recommendations, and explanations. As Bender and Koller [16] argue, systems (e.g., large language models) purely built on *form*, do not have a way to produce *meaning*. In this sense, prior HCI work has investigated the potential effects of biases and misunderstandings of AI's capabilities [11, 38, 55]. As AI-based support systems manifest within healthcare, developing education around AI for healthcare can be beneficial for clinicians and patients alike. However, special consideration should be taken as that could come at the cost of longer medical studies or compromise with existing courses and training activities.

7.4 Limitations

Our study has several limitations but future work could bring triangulation to our results [130]. First, we limited our enquiry to Idiopathic Pulmonary Fibrosis. While this helped in contextualising participants' responses, the themes we constructed might not apply to other diseases within pulmonary medicine. For instance, clinicians specialised in lung cancer are familiar with IPF because the two diseases can co-occur. However, they can follow different care and treatment paths. Second, we conducted the study in a European university medical centre. Our results may not be transferable to other countries in the Global North with potentially different healthcare systems. Additional inconsistencies could arise when attempting to replicate a similar study in the Global South as prior work discussed differences in healthcare systems [8] and inequalities exacerbated by the use of AI [128, 177]. Finally, while we found grounding for the XAI exemplars in the literature (subsubsection 5.2.1) and showed them only after participants disclosed their explainability needs, it is conceivable that those prompts generated anchoring bias in our participants. Similarly, while we found the journey map to align well with our participants, the participatory nature of our preliminary study could have exacerbated power dynamics between the participants and how they perceived, and agreed, on the journey map.

8 CONCLUSION

In this work, we involved 16 clinicians, from a European university medical centre, working on Idiopathic Pulmonary Fibrosis to enquire about their uses and purposes for Explainable AI and how these evolve over time throughout patient care. First, with the help of 4 clinicians, we outlined a patient journey map for IPF to provide scaffolding for our research. Then, we conducted 12 semi-structured interviews to outline the evolution throughout patient care of clinicians' uses and purposes for XAI. We showed that several tensions arise in relation to their needs around explainability for AI-based CDSSs. Clinicians seek diverse explanations – both in content and modality – to cope with patients' dynamics and uncertainties of IPF. While general explanations of the affordances of AI-based CDSSs are valued in early adoption phases, local explanations – especially multi-modal ones – are anticipated throughout patient care to surface patient-specific features. By adopting properties of goodness of explanations as an interpretative lens, we corroborate ongoing efforts by the HCI community around extending the scope of XAI beyond AI system outputs and how it is evaluated. Particularly, we found compactness, interactivity, and actionability of explanations to be key drivers for clinicians. However, our results also highlight the diminishing relevance of explanations as clinicians learn when and how to use such systems. We concluded by reflecting on the lack of longitudinal perspectives in researching XAI for CDSSs, implications for the design of explanations in clinical settings, and the role of medical education in further promoting AI literacy.

ACKNOWLEDGMENTS

We thank our participants from Erasmus MC for making space in their busy schedules to talk with us. We also thank Ruixuan Zhang for visualising the patient community journey map, and Mireia Yurrita Semperena and Sole Pera for the insightful discussions. This research has been partially supported by the TU Delft Design@Scale AI Lab.

REFERENCES

[1] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, Bamshad Mobasher, and Edward Malthouse. 2021. User-Centered Evaluation of Popularity Bias in Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (Utrecht, Netherlands) *(UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 119–129. https://doi.org/10.1145/3450613.3456821

[2] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/3173574.3174156

[3] Saul Albert, Magnus Hamann, and Elizabeth Stokoe. 2023. Conversational User Interfaces in Smart Homecare Interactions: A Conversation Analytic Case Study. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) *(CUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/3571884.3597140

[4] Michael A. Alcorn, Qi Li, Zhitao Gong, Chengfei Wang, Long Mai, Wei-Shinn Ku, and Anh Nguyen. 2019. Strike (With) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[5] Marco Almada. 2019. Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (Montreal, QC, Canada) *(ICAIL '19)*. Association for Computing Machinery, New York, NY, USA, 2–11. https://doi.org/10.1145/3322640.3326699

[6] Ahmed Alqaraawi, Martin Schuessler, Philipp Weiß, Enrico Costanza, and Nadia Berthouze. 2020. Evaluating Saliency Map Explanations for Convolutional Neural Networks: A User Study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 275–285. https://doi.org/10.1145/3377325.3377519

[7] David Alvarez Melis and Tommi Jaakkola. 2018. Towards Robust Interpretability with Self-Explaining Neural Networks. In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Vol. 31. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/3e9f0fc9b2f89e043bc6233994dfcf76-Paper.pdf

[8] Lameck Mbangula Amugongo, Nicola J. Bidwell, and Caitlin C. Corrigan. 2023. Invigorating Ubuntu Ethics in AI for Healthcare: Enabling Equitable Care. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 583–592. https://doi.org/10.1145/3593013.3594024

[9] Ariful Islam Anik and Andrea Bunt. 2021. Data-Centric Explanations: Explaining Training Data of Machine Learning Systems to Promote Transparency. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 75, 13 pages. https://doi.org/10.1145/3411764.3445736

[10] Anna Markella Antoniadi, Yuhan Du, Yasmine Guendouz, Lan Wei, Claudia Mazo, Brett A. Becker, and Catherine Mooney. 2021. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences* 11, 11 (2021). https://doi.org/10.3390/app11115088

[11] Anne Kathrine Petersen Bach, Trine Munch Nørgaard, Jens Christian Brok, and Niels van Berkel. 2023. "If I Had All the Time in the World": Ophthalmologists' Perceptions of Anchoring Bias Mitigation in Clinical AI Support. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 16, 14 pages. https://doi.org/10.1145/3544548.3581513

[12] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert Müller. 2010. How to Explain Individual Classification Decisions. *Journal of Machine Learning Research* 11, 61 (2010), 1803–1831. http://jmlr.org/papers/v11/baehrens10a.html

[13] Agathe Balayn, Natasa Rikalo, Christoph Lofi, Jie Yang, and Alessandro Bozzon. 2022. How Can Explainability Methods Be Used to Support Bug Identification in Computer Vision Models?. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 184, 16 pages. https://doi.org/10.1145/3491102.3517474

[14] Agathe Balayn, Panagiotis Soilis, Christoph Lofi, Jie Yang, and Alessandro Bozzon. 2021. What Do You Mean? Interpreting Image Classification with Crowdsourced Concept Extraction and Analysis. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 1937–1948. https://doi.org/10.1145/3442381.3450069

[15] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[16] Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 5185–5198. https://doi.org/10.18653/v1/2020.acl-main.463

[17] Charles R. Berger and Richard J. Calabrese. 1975. Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication. *Human Communication Research* 1, 2 (03 1975), 99–112. https://doi.org/10.1111/j.1468-2958.1975.tb00258.x arXiv:https://academic.oup.com/hcr/article-pdf/1/2/99/22344108/jhumcom0099.pdf

[18] Astrid Bertrand, Tiphaine Viard, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2023. On Selective, Mutable and Dialogic XAI: A Review of What Users Say about Different Types of Interactive Explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 411, 21 pages. https://doi.org/10.1145/3544548.3581314

[19] Wiebe E. Bijker, Thomas P. Hughes, and Trevor Pinch. 1987. *The Social Constructions of Technological Systems*. MIT Press.

[20] Shreyan Biswas, Lorenzo Corti, Stefan Buijsman, and Jie Yang. 2022. CHIME: Causal Human-in-the-Loop Model Explanations. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 10, 1 (Oct. 2022), 27–39. https://doi.org/10.1609/hcomp.v10i1.21985

[21] Johnna Blair, Dahlia Mukherjee, Erika F. H. Saunders, and Saeed Abdullah. 2023. Knowing How Long a Storm Might Last Makes It Easier to Weather: Exploring Needs and Attitudes Toward a Data-Driven and Preemptive Intervention System for Bipolar Disorder. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 496, 12 pages. https://doi.org/10.1145/3544548.3581563

[22] Richard E Boyatzis. 1998. *Transforming qualitative information: Thematic analysis and code development*. sage.

[23] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[24] Virginia Braun and Victoria Clarke. 2021. *Thematic Analysis : A Practical Guide*. SAGE Publications Ltd, London. http://digital.casalini.it/5282292 Developed and adapted by the authors of this book, thematic analysis (TA) is one of the most popular qualitative data analytic techniques in psychology and the social and health sciences. Building on the success of Braun & Clarke's 2006 paper first outlining their approach - which has over 100,000 citations on Google Scholar - this book is the definitive guide to TA, covering: - Contextualisation of TA - Developing themes - Writing TA reports - Reflexive TA It addresses the common questions surrounding TA as well as developments in the field, offering a highly accessible and practical discussion of doing TA situated within a clear understanding of the wider terrain of qualitative research. Virginia Braun is a Professor in the School of Psychology at The University of Auckland, Aotearoa New Zealand. Victoria Clarke is an Associate Professor in Qualitative and Critical Psychology in the Department of Social Sciences at the University of the West of England (UWE), Bristol. [Publisher's text]..

[25] Gerald-Mark Breen, Thomas T. H. Wan, Ning Jackie Zhang, Shriram S. Marathe, Binyam K. Seblega, and Seung Chun Paek. 2009. Improving Doctor–Patient Communication: Examining Innovative Modalities Vis-à-vis Effective Patient-Centric Care Management Technology. *Journal of Medical Systems* 33, 2 (01 Apr 2009), 155–162. https://doi.org/10.1007/s10916-008-9175-3

[26] Bruce G Buchanan and Edward H Shortliffe. 1984. *Rule based expert systems: the mycin experiments of the stanford heuristic programming project (the Addison-Wesley series in artificial intelligence)*. Addison-Wesley Longman Publishing Co., Inc.

[27] Stefan Buijsman. 2022. Defining Explanation and Explanatory Depth in XAI. *Minds and Machines* 32, 3 (01 Sep 2022), 563–584. https://doi.org/10.1007/s11023-022-09607-9

[28] Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The Effects of Example-Based Explanations in a Machine Learning Interface. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 258–262. https://doi.org/10.1145/3301275.3302289

[29] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300234

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

[31] Joseph N Cappella, ML Knapp, and GR Miller. 1994. The management of conversational interaction in adults and infants. *Handbook of interpersonal communication* (1994), 380–418.

[32] Pascale Carayon, Abigail Wooldridge, Peter Hoonakker, Ann Schoofs Hundt, and Michelle M. Kelly. 2020. SEIPS 3.0: Human-centered design of the patient journey for patient safety. *Applied Ergonomics* 84 (2020), 103033. https://doi.org/10.1016/j.apergo.2019.103033

[33] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. 2019. Machine learning interpretability: A survey on methods and metrics. *Electronics* 8, 8 (2019), 832.

[34] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. 2019. This Looks Like That: Deep Learning for Interpretable Image Recognition. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/adf7ee2dcf142b0e11888e72b43fcb75-Paper.pdf

[35] Jianbo Chen, Le Song, Martin Wainwright, and Michael Jordan. 2018. Learning to Explain: An Information-Theoretic Perspective on Model Interpretation. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 883–892. https://proceedings.mlr.press/v80/chen18j.html

[36] Victoria Clarke and Virginia Braun. 2013. Successful qualitative research: A practical guide for beginners. *Successful qualitative research* (2013), 1–400.

[37] Committee on Quality of Health Care in America. 2002. Crossing the quality chasm: A new health system for the 21st century. *J. Healthc. Qual.* 24, 5 (Sept. 2002), 52.

[38] Pat Croskerry. 2013. From mindless to mindful practice—cognitive bias and clinical decision making. *N Engl J Med* 368, 26 (2013), 2445–2448.

[39] Nilakash Das, Sofie Happaerts, Iwein Gyselinck, Michael Staes, Eric Derom, Guy Brusselle, Felip Burgos, Marco Contoli, Anh Tuan Dinh-Xuan, Frits ME Franssen, et al. 2023. Collaboration between explainable artificial intelligence and pulmonologists improves the accuracy of pulmonary function test interpretation. *European Respiratory Journal* 61, 5 (2023).

[40] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.

[41] William K Diprose, Nicholas Buist, Ning Hua, Quentin Thurier, George Shand, and Reece Robinson. 2020. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *Journal of the American Medical Informatics Association* 27, 4 (02 2020), 592–600. https://doi.org/10.1093/jamia/ocz229 arXiv:https://academic.oup.com/jamia/article-pdf/27/4/592/34153285/ocz229.pdf

[42] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretable Machine Learning. arXiv:1702.08608 [stat.ML]

[43] Pierre Durieux, Rémy Nizard, Philippe Ravaud, Nicolas Mounier, and Eric Lepage. 2000. A Clinical Decision Support System for Prevention of Venous ThromboembolismEffect on Physician Behavior. *JAMA* 283, 21 (06 2000), 2816–2821. https://doi.org/10.1001/jama.283.21.2816 arXiv:https://jamanetwork.com/journals/jama/articlepdf/192759/joc00041.pdf

[44] Pelle Ehn. 1990. *Work-Oriented Design of Computer Artifacts.* L. Erlbaum Associates Inc.

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

[46] Upol Ehsan and Mark O. Riedl. 2020. Human-Centered Explainable AI: Towards a Reflective Sociotechnical Approach. In *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, Constantine Stephanidis, Masaaki Kurosu, Helmut Degen, and Lauren Reinerman-Jones (Eds.). Springer International Publishing, Cham, 449–466.

[47] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O. Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 34 (apr 2023), 32 pages. https://doi.org/10.1145/3579467

[48] Upol Ehsan, Pradyumna Tambwekar, Larry Chan, Brent Harrison, and Mark O. Riedl. 2019. Automated Rationale Generation: A Technique for Explainable AI and Its Effects on Human Perceptions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 263–274. https://doi.org/10.1145/3301275.3302316

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

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

[51] Upol Ehsan, Philipp Wintersberger, Elizabeth A Watkins, Carina Manger, Gonzalo Ramos, Justin D. Weisz, Hal Daumé Iii, Andreas Riener, and Mark O Riedl. 2023. Human-Centered Explainable AI (HCXAI): Coming of Age. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 353, 7 pages. https://doi.org/10.1145/3544549.3573832

[52] Michael D. Ekstrand, Maria Soledad Pera, and Katherine Landau Wright. 2023. Seeking Information with a More Knowledgeable Other. *Interactions* 30, 1 (jan 2023), 70–73. https://doi.org/10.1145/3573364

[53] Theodore Evans, Carl Orge Retzlaff, Christian Geißler, Michaela Kargl, Markus Plass, Heimo Müller, Tim-Rasmus Kiehl, Norman Zerbe, and Andreas Holzinger. 2022. The explainability paradox: Challenges for xAI in digital pathology. *Future Generation Computer Systems* 133 (2022), 281–296. https://doi.org/10.1016/j.future.2022.03.009

[54] Fred F Ferri and MD Facp. 2023. *Ferri's Clinical Advisor 2024, E-Book*. Elsevier Health Sciences, Chapter Idiopathic Pulmonary Fibrosis, 775–777.

[55] Riccardo Fogliato, Shreya Chappidi, Matthew Lungren, Paul Fisher, Diane Wilson, Michael Fitzke, Mark Parkinson, Eric Horvitz, Kori Inkpen, and Besmira Nushi. 2022. Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1362–1374. https://doi.org/10.1145/3531146.3533193

[56] Christopher Frauenberger. 2016. Critical Realist HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 341–351. https://doi.org/10.1145/2851581.2892569

[57] Timo Freiesleben and Gunnar König. 2023. Dear XAI Community, We Need to Talk! Fundamental Misconceptions in Current XAI Research. arXiv:2306.04292 [cs.AI]

[58] Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro. 2013. Contributor Profiles, Their Dynamics, and Their Importance in Five Q&a Sites. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (San Antonio, Texas, USA) *(CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 1237–1252. https://doi.org/10.1145/2441776.2441916

[59] Bill Gaver and Heather Martin. 2000. Alternatives: Exploring Information Appliances through Conceptual Design Proposals. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. Association for Computing Machinery, New York, NY, USA, 209–216. https://doi.org/10.1145/332040.332433

[60] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. *Commun. ACM* 64, 12 (nov 2021), 86–92. https://doi.org/10.1145/3458723

[61] Amirata Ghorbani, James Wexler, James Y Zou, and Been Kim. 2019. Towards Automatic Concept-based Explanations. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/77d2afcb31f6493e350fca61764efb9a-Paper.pdf

[62] Barney G Glaser, Judith Holton, et al. 2004. Remodeling grounded theory. In *Forum qualitative sozialforschung/forum: qualitative social research*, Vol. 5.

[63] Alex Goldstein, Adam Kapelner, Justin Bleich, and Emil Pitkin. 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics* 24, 1 (2015), 44–65. https://doi.org/10.1080/10618600.2014.907095 arXiv:https://doi.org/10.1080/10618600.2014.907095

[64] Yash Goyal, Ziyan Wu, Jan Ernst, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. Counterfactual Visual Explanations. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, 2376–2384. https://proceedings.mlr.press/v97/goyal19a.html

[65] Ben Green and Salomé Viljoen. 2020. Algorithmic Realism: Expanding the Boundaries of Algorithmic Thought. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 19–31. https://doi.org/10.1145/3351095.3372840

[66] Robert Greenes. 2014. *Clinical decision support: the road to broad adoption.* Academic Press.

[67] Hongyan Gu, Chunxu Yang, Mohammad Haeri, Jing Wang, Shirley Tang, Wenzhong Yan, Shujin He, Christopher Kazu Williams, Shino Magaki, and Xiang 'Anthony' Chen. 2023. Augmenting Pathologists with NaviPath: Design and Evaluation of a Human-AI Collaborative Navigation System. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 349, 19 pages. https://doi.org/10.1145/3544548.3580694

[68] Greg Guest, Kathleen M MacQueen, and Emily E Namey. 2011. *Applied thematic analysis.* sage publications.

[69] Riccardo Guidotti. 2022. Counterfactual explanations and how to find them: literature review and benchmarking. *Data Mining and Knowledge Discovery* (28 Apr 2022). https://doi.org/10.1007/s10618-022-00831-6

[70] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. 2018. Local Rule-Based Explanations of Black Box Decision Systems. arXiv:1805.10820 [cs.AI]

[71] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* 51, 5, Article 93 (aug 2018), 42 pages. https://doi.org/10.1145/3236009

[72] Han Guo, Nazneen Rajani, Peter Hase, Mohit Bansal, and Caiming Xiong. 2021. FastIF: Scalable Influence Functions for Efficient Model Interpretation and Debugging. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 10333–10350. https://doi.org/10.18653/v1/2021.emnlp-main.808

[73] Mark Hartswood, Rob Procter, Mark Rouncefield, Roger Slack, James Soutter, and Alex Voss. 2003. 'Repairing' the Machine: A Case Study of the Evaluation of Computer-Aided Detection Tools in Breast Screening. In *ECSCW 2003*, Kari Kuutti, Eija Helena Karsten, Geraldine Fitzpatrick, Paul Dourish, and Kjeld Schmidt (Eds.). Springer Netherlands, Dordrecht, 375–394.

[74] Gaole He, Lucie Kuiper, and Ujwal Gadiraju. 2023. Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 113, 18 pages. https://doi.org/10.1145/3544548.3581025

[75] Kilian Hendrickx, Lorenzo Perini, Dries Van der Plas, Wannes Meert, and Jesse Davis. 2023. Machine Learning with a Reject Option: A survey. arXiv:2107.11277 [cs.LG]

[76] Clément Henin and Daniel Le Métayer. 2022. Beyond explainability: justifiability and contestability of algorithmic decision systems. *AI & SOCIETY* 37, 4 (01 Dec 2022), 1397–1410. https://doi.org/10.1007/s00146-021-01251-8

[77] Karen Henwood and Nick Pidgeon. 1994. Beyond the qualitative paradigm: A framework for introducing diversity within qualitative psychology. *Journal of Community & Applied Social Psychology* 4, 4 (1994), 225–238.

[78] Mireille Hildebrandt. 2019. Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20, 1 (2019), 83–121. https://doi.org/doi:10.1515/til-2019-0004

[79] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller. 2019. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, 4 (2019), e1312.

[80] Michael Jackson. 1995. *Software Requirements & Specifications: A Lexicon of Practice, Principles and Prejudices.* ACM Press/Addison-Wesley Publishing Co., USA.

[81] Marianne E Jaeger and Ralph L Rosnow. 1988. Contextualism and its implications for psychological inquiry. *British Journal of Psychology* 79, 1 (1988), 63–75.

[82] Anthony Jameson, Silvia Gabrielli, and Antti Oulasvirta. 2009. Users' Preferences Regarding Intelligent User Interfaces: Differences among Users and Changes over Time. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA) *(IUI '09)*. Association for Computing Machinery, New York, NY, USA, 497–498. https://doi.org/10.1145/1502650.1502734

[83] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management* 44, 3 (2008), 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015

[84] Jacinta Jardine, Caroline Earley, Derek Richards, Ladislav Timulak, Jorge E. Palacios, Daniel Duffy, Karen Tierney, and Gavin Doherty. 2020. The Experience of Guided Online Therapy: A Longitudinal, Qualitative Analysis of Client Feedback in a Naturalistic RCT. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376254

[85] Kwang Nam Jin, Eun Young Kim, Young Jae Kim, Gi Pyo Lee, Hyungjin Kim, Sohee Oh, Yong Suk Kim, Ju Hyuck Han, and Young Jun Cho. 2022. Diagnostic effect of artificial intelligence solution for referable thoracic abnormalities on chest radiography: a multicenter respiratory outpatient diagnostic cohort study. *European Radiology* 32, 5 (01 May 2022), 3469–3479. https://doi.org/10.1007/s00330-021-08397-5

[86] Eunkyung Jo, Myeonghan Ryu, Georgia Kenderova, Samuel So, Bryan Shapiro, Alexandra Papoutsaki, and Daniel A. Epstein. 2022. Designing Flexible Longitudinal Regimens: Supporting Clinician Planning for Discontinuation of Psychiatric Drugs. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>New Orleans</city>, <state>LA</state>, <country>USA</country>, </conf-loc>) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 352, 17 pages. https://doi.org/10.1145/3491102.3502206

[87] Helene Joffe. 2011. Thematic analysis. *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners* (2011), 209–223.

[88] Jiwon Jung, Ki-Hun Kim, Tess Peters, Dirk Snelders, and Maaike Kleinsmann. 2023. Advancing Design Approaches through Data-Driven Techniques: Patient Community Journey Mapping Using Online Stories and Machine Learning. *International Journal of Design; Vol 17, No 2 (2023)* (2023 2023). http://www.ijdesign.org/index.php/IJDesign/article/view/4671

[89] Alan Kaplan, Hui Cao, J. Mark FitzGerald, Nick Iannotti, Eric Yang, Janwillem W.H. Kocks, Konstantinos Kostikas, David Price, Helen K. Reddel, Ioanna Tsiligianni, Claus F. Vogelmeier, Pascal Pfister, and Paul Mastoridis. 2021. Artificial Intelligence/Machine Learning in Respiratory Medicine and Potential Role in Asthma and COPD Diagnosis. *The Journal of Allergy and Clinical Immunology: In Practice* 9, 6 (2021), 2255–2261. https://doi.org/10.1016/j.jaip.2021.02.014

[90] Ozan Karaca, S. Ayhan Çalışkan, and Kadir Demir. 2021. Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – development, validity and reliability study. *BMC Medical Education* 21, 1 (18 Feb 2021), 112. https://doi.org/10.1186/s12909-021-02546-6

[91] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic Recourse: From Counterfactual Explanations to Interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 353–362. https://doi.org/10.1145/3442188.3445899

[92] Saif Khairat, David Marc, William Crosby, and Ali Al Sanousi. 2018. Reasons For Physicians Not Adopting Clinical Decision Support Systems: Critical Analysis. *JMIR Med Inform* 6, 2 (18 Apr 2018), e24. https://doi.org/10.2196/medinform.8912

[93] Danai Khemasuwan, Jeffrey S. Sorensen, and Henri G. Colt. 2020. Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19. *European Respiratory Review* 29, 157 (2020). https://doi.org/10.1183/16000617.0181-2020 arXiv:https://err.ersjournals.com/content/29/157/200181.full.pdf

[94] Been Kim, Martin Wattenberg, Justin Gilmer, Carrie Cai, James Wexler, Fernanda Viegas, and Rory sayres. 2018. Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 2668–2677. https://proceedings.mlr.press/v80/kim18d.html

[95] Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 250, 17 pages. https://doi.org/10.1145/3544548.3581001

[96] Nigel King and Joanna M Brooks. 2016. *Template analysis for business and management students.* Sage.

[97] Pang Wei Koh and Percy Liang. 2017. Understanding Black-box Predictions via Influence Functions. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, 1885–1894. https://proceedings.mlr.press/v70/koh17a.html

[98] Satyapriya Krishna, Tessa Han, Alex Gu, Javin Pombra, Shahin Jabbari, Steven Wu, and Himabindu Lakkaraju. 2022. The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective. arXiv:2202.01602 [cs.LG]

[99] Peter Gall Krogh and Ilpo Koskinen. 2022. How Constructive Design Researchers Drift: Four Epistemologies. *Design Issues* 38, 2 (2022), 33–46. https://doi.org/10.1162/desi_a_00680

[100] Todd Kulesza, Margaret Burnett, Weng-Keen Wong, and Simone Stumpf. 2015. Principles of Explanatory Debugging to Personalize Interactive Machine Learning. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (Atlanta, Georgia, USA) *(IUI '15)*. Association for Computing Machinery, New York, NY, USA, 126–137. https://doi.org/10.1145/2678025.2701399

[101] Florian Königstorfer and Stefan Thalmann. 2020. Applications of Artificial Intelligence in commercial banks – A research agenda for behavioral finance. *Journal of Behavioral and Experimental Finance* 27 (2020), 100352. https://doi.org/10.1016/j.jbef.2020.100352

[102] Vivian Lai, Yiming Zhang, Chacha Chen, Q. Vera Liao, and Chenhao Tan. 2023. Selective Explanations: Leveraging Human Input to Align Explainable AI. arXiv:2301.09656 [cs.AI]

[103] Markus Langer, Daniel Oster, Timo Speith, Holger Hermanns, Lena Kästner, Eva Schmidt, Andreas Sesing, and Kevin Baum. 2021. What do we want from

Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research. *Artificial Intelligence* 296 (2021), 103473. https://doi.org/10.1016/j.artint.2021.103473

[104] Matthias Carl Laupichler, Alexandra Aster, Jana Schirch, and Tobias Raupach. 2022. Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence* 3 (2022), 100101. https://doi.org/10.1016/j.caeai.2022.100101

[105] David J Lederer and Fernando J Martinez. 2018. Idiopathic pulmonary fibrosis. *New England Journal of Medicine* 378, 19 (2018), 1811–1823.

[106] Beth L Leech. 2002. Asking questions: Techniques for semistructured interviews. *PS: Political Science & Politics* 35, 4 (2002), 665–668.

[107] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

[108] Q. Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 10, 1 (Oct. 2022), 147–159. https://doi.org/10.1609/hcomp.v10i1.21995

[109] Shusen Liu, Bhavya Kailkhura, Donald Loveland, and Yong Han. 2019. Generative Counterfactual Introspection for Explainable Deep Learning. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. 1–5. https://doi.org/10.1109/GlobalSIP45357.2019.8969491

[110] Tania Lombrozo. 2010. Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology* 61, 4 (2010), 303–332. https://doi.org/10.1016/j.cogpsych.2010.05.002

[111] Tania Lombrozo. 2012. Explanation and abductive inference. *Oxford handbook of thinking and reasoning* (2012), 260–276.

[112] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376727

[113] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[114] Jörn Lötsch, Dario Kringel, and Alfred Ultsch. 2022. Explainable Artificial Intelligence (XAI) in Biomedicine: Making AI Decisions Trustworthy for Physicians and Patients. *BioMedInformatics* 2, 1 (2022), 1–17. https://doi.org/10.3390/biomedinformatics2010001

[115] Raju Maharjan, Darius Adam Rohani, Per Bækgaard, Jakob Bardram, and Kevin Doherty. 2021. Can We Talk? Design Implications for the Questionnaire-Driven Self-Report of Health and Wellbeing via Conversational Agent. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) *(CUI '21)*. Association for Computing Machinery, New York, NY, USA, Article 5, 11 pages. https://doi.org/10.1145/3469595.3469600

[116] Gary Marchionini. 2006. Exploratory Search: From Finding to Understanding. *Commun. ACM* 49, 4 (apr 2006), 41–46. https://doi.org/10.1145/1121949.1121979

[117] Aniek F. Markus, Jan A. Kors, and Peter R. Rijnbeek. 2021. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics* 113 (2021), 103655. https://doi.org/10.1016/j.jbi.2020.103655

[118] Claudia Mazo, Cathriona Kearns, Catherine Mooney, and William M. Gallagher. 2020. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers* 12, 2 (2020). https://doi.org/10.3390/cancers12020369

[119] Stephen McCarthy, Paidi O'Raghallaigh, Simon Woodworth, Yoke Lin Lim, Louise C. Kenny, and Frédéric Adam. 2016. An integrated patient journey mapping tool for embedding quality in healthcare service reform. *Journal of Decision Systems* 25, sup1 (2016), 354–368. https://doi.org/10.1080/12460125.2016.1187394 arXiv:https://doi.org/10.1080/12460125.2016.1187394

[120] Eric B. Meltzer and Paul W. Noble. 2008. Idiopathic pulmonary fibrosis. *Orphanet Journal of Rare Diseases* 3, 1 (26 Mar 2008), 8. https://doi.org/10.1186/1750-1172-3-8

[121] Matthew B Miles and A Michael Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. sage.

[122] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[123] Tim Miller. 2023. Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 333–342. https://doi.org/10.1145/3593013.3594001

[124] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[125] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Trans. Interact. Intell. Syst.* 11, 3–4, Article 24 (sep 2021), 45 pages. https://doi.org/10.1145/3387166

[126] Meike Nauta, Ron van Bree, and Christin Seifert. 2021. Neural Prototype Trees for Interpretable Fine-Grained Image Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14933–14943.

[127] Anne-Marie Nussberger, Lan Luo, L. Elisa Celis, and M. J. Crockett. 2022. Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence. *Nature Communications* 13, 1 (03 Oct 2022), 5821. https://doi.org/10.1038/s41467-022-33417-3

[128] Hanno N. Olinger, Johannes J. Britz, and Martin S. Olivier. 2007. Western privacy and/or Ubuntu? Some critical comments on the influences in the forthcoming data privacy bill in South Africa. *International Information & Library Review* 39, 1 (2007), 31–43. https://doi.org/10.1080/10572317.2007.10762729 arXiv:https://doi.org/10.1080/10572317.2007.10762729

[129] Jerome A. Osheroff, Jonathan M. Teich, Blackford Middleton, Elaine B. Steen, Adam Wright, and Don E. Detmer. 2007. A Roadmap for National Action on Clinical Decision Support. *Journal of the American Medical Informatics Association* 14, 2 (03 2007), 141–145. https://doi.org/10.1197/jamia.M2334 arXiv:https://academic.oup.com/jamia/article-pdf/14/2/141/2563507/14-2-141.pdf

[130] Michael Quinn Patton. 2014. *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

[131] Simona Pekarek Doehler and Evelyne Berger. 2019. *On the Reflexive Relation Between Developing L2 Interactional Competence and Evolving Social Relationships: A Longitudinal Study of Word-Searches in the 'Wild'*. Springer International Publishing, Cham, 51–75. https://doi.org/10.1007/978-3-030-22165-2_3

[132] Milda Pocevičiūtė, Gabriel Eilertsen, and Claes Lundström. 2020. *Survey of XAI in Digital Pathology*. Springer International Publishing, Cham, 56–88. https://doi.org/10.1007/978-3-030-50402-1_4

[133] Christina Popescu, Grace Golden, David Benrimoh, Myriam Tanguay-Sela, Dominique Slowey, Eryn Lundrigan, Jérôme Williams, Bennet Desormeau, Divyesh Kardani, Tamara Perez, Colleen Rollins, Sonia Israel, Kelly Perlman, Caitrin Armstrong, Jacob Baxter, Kate Whitmore, Marie-Jeanne Fradette, Kaelan Felcarek-Hope, Ghassen Soufi, Robert Fratila, Joseph Mehltretter, Karl Looper, Warren Steiner, Soham Rej, Jordan F Karp, Katherine Heller, Sagar V Parikh, Rebecca McGuire-Snieckus, Manuela Ferrari, Howard Margolese, and Gustavo Turecki. 2021. Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study. *JMIR Form Res* 5, 10 (25 Oct 2021), e31862. https://doi.org/10.2196/31862

[134] Alisha Pradhan, Amanda Lazar, and Leah Findlater. 2020. Use of Intelligent Voice Assistants by Older Adults with Low Technology Use. *ACM Trans. Comput.-Hum. Interact.* 27, 4, Article 31 (sep 2020), 27 pages. https://doi.org/10.1145/3373759

[135] Alun Preece, Dan Harborne, Dave Braines, Richard Tomsett, and Supriyo Chakraborty. 2018. Stakeholders in Explainable AI. arXiv:1810.00184 [cs.AI]

[136] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W. Crandall, Nicholas A. Christakis, Iain D. Couzin, Matthew O. Jackson, Nicholas R. Jennings, Ece Kamar, Isabel M. Kloumann, Hugo Larochelle, David Lazer, Richard McElreath, Alan Mislove, David C. Parkes, Alex 'Sandy' Pentland, Margaret E. Roberts, Azim Shariff, Joshua B. Tenenbaum, and Michael Wellman. 2019. Machine behaviour. *Nature* 568, 7753 (01 Apr 2019), 477–486. https://doi.org/10.1038/s41586-019-1138-y

[137] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[138] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2018. Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence* 32, 1 (Apr. 2018). https://doi.org/10.1609/aaai.v32i1.11491

[139] Jane Ritchie and Liz Spencer. 2002. Qualitative data analysis for applied policy research. In *Analyzing qualitative data*. Routledge, 173–194.

[140] Negar Rostamzadeh, Diana Mincu, Subhrajit Roy, Andrew Smart, Lauren Wilcox, Mahima Pushkarna, Jessica Schrouff, Razvan Amironesei, Nyalleng Moorosi, and Katherine Heller. 2022. Healthsheet: Development of a Transparency Artifact for Health Datasets. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) *(FAccT '22)*. Association for Computing Machinery, New York, NY, USA, 1943–1961. https://doi.org/10.1145/3531146.3533239

[141] Jannik Schaaf, Martin Sedlmayr, Johanna Schaefer, and Holger Storf. 2020. Diagnosis of Rare Diseases: a scoping review of clinical decision support systems. *Orphanet Journal of Rare Diseases* 15, 1 (24 Sep 2020), 263. https://doi.org/10.1186/s13023-020-01536-z

[142] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

[143] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan. 2020. Explainable Deep Learning Models in Medical Image Analysis. *Journal of Imaging* 6, 6 (2020). https://doi.org/10.3390/jimaging6060052

[144] Venkatesh Sivaraman, Leigh A Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. 2023. Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 754, 18 pages. https://doi.org/10.1145/3544548.3581075

[145] Dylan Slack, Sophie Hilgard, Emily Jia, Sameer Singh, and Himabindu Lakkaraju. 2020. Fooling LIME and SHAP: Adversarial Attacks on Post Hoc Explanation Methods. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (New York, NY, USA) *(AIES '20)*. Association for Computing Machinery, New York, NY, USA, 180–186. https://doi.org/10.1145/3375627.3375830

[146] Helen Smith. 2021. Clinical AI: opacity, accountability, responsibility and liability. *AI & SOCIETY* 36, 2 (01 Jun 2021), 535–545. https://doi.org/10.1007/s00146-020-01019-6

[147] Kacper Sokol and Peter Flach. 2020. Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Barcelona, Spain) *(FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 56–67. https://doi.org/10.1145/3351095.3372870

[148] Nasim Sonboli, Robin Burke, Michael Ekstrand, and Rishabh Mehrotra. 2022. The Multisided Complexity of Fairness in Recommender Systems. *AI Magazine* 43, 2 (Jun. 2022), 164–176. https://doi.org/10.1002/aaai.12054

[149] Lisa Sparks, H Dan O'Hair, and Kevin B Wright. 2012. *Health communication in the 21st century*. John Wiley & Sons.

[150] Irene Steenbruggen and Meredith C. McCormack. 2023. Artificial intelligence: do we really need it in pulmonary function interpretation? *European Respiratory Journal* 61, 5 (2023). https://doi.org/10.1183/13993003.00625-2023 arXiv:https://erj.ersjournals.com/content/61/5/2300625.full.pdf

[151] David F. Steiner, Kunal Nagpal, Rory Sayres, Davis J. Foote, Benjamin D. Wedin, Adam Pearce, Carrie J. Cai, Samantha R. Winter, Matthew Symonds, Liron Yatziv, Andrei Kapishnikov, Trissia Brown, Isabelle Flament-Auvigne, Fraser Tan, Martin C. Stumpe, Pan-Pan Jiang, Yun Liu, Po-Hsuan Cameron Chen, Greg S. Corrado, Michael Terry, and Craig H. Mermel. 2020. Evaluation of the Use of Combined Artificial Intelligence and Pathologist Assessment to Review and Grade Prostate Biopsies. *JAMA Network Open* 3, 11 (11 2020), e2023267–e2023267. https://doi.org/10.1001/jamanetworkopen.2020.23267

[152] R Street. 2001. Active patients as powerful communicators: the linguistic foundation of participation in health care. In *Hand of Language and Social Psychology*, W.P. Robinson (Ed.). Wiley, 541–560.

[153] Hariharan Subramonyam, Jane Im, Colleen Seifert, and Eytan Adar. 2022. Solving Separation-of-Concerns Problems in Collaborative Design of Human-AI Systems through Leaky Abstractions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 481, 21 pages. https://doi.org/10.1145/3491102.3517537

[154] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, 3319–3328. https://proceedings.mlr.press/v70/sundararajan17a.html

[155] Aurélien Tabard, Wendy E. Mackay, and Evelyn Eastmond. 2008. From Individual to Collaborative: The Evolution of Prism, a Hybrid Laboratory Notebook. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, CA, USA) *(CSCW '08)*. Association for Computing Machinery, New York, NY, USA, 569–578. https://doi.org/10.1145/1460563.1460653

[156] Jaime Teevan, Susan T. Dumais, and Daniel J. Liebling. 2010. A Longitudinal Study of How Highlighting Web Content Change Affects People's Web Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1353–1356. https://doi.org/10.1145/1753326.1753530

[157] Andrea Tocchetti, Lorenzo Corti, Agathe Balayn, Mireia Yurrita, Philip Lippmann, Marco Brambilla, and Jie Yang. 2022. A.I. Robustness: a Human-Centered Perspective on Technological Challenges and Opportunities. (2022). arXiv:2210.08906 [cs.AI]

[158] Richard Tomsett, Dave Braines, Dan Harborne, Alun Preece, and Supriyo Chakraborty. 2018. Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems. arXiv:1806.07552 [cs.AI]

[159] Sana Tonekaboni, Shalmali Joshi, Melissa D. McCradden, and Anna Goldenberg. 2019. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. In *Proceedings of the 4th Machine Learning for Healthcare Conference (Proceedings of Machine Learning Research, Vol. 106)*, Finale Doshi-Velez, Jim Fackler, Ken Jung, David Kale, Rajesh Ranganath, Byron Wallace, and Jenna Wiens (Eds.). PMLR, 359–380. https://proceedings.mlr.press/v106/tonekaboni19a.html

[160] Marko Topalovic, Nilakash Das, Pierre-Régis Burgel, Marc Daenen, Eric Derom, Christel Haenebalcke, Rob Janssen, Huib A.M. Kerstjens, Giuseppe Liistro, Renaud Louis, Vincent Ninane, Christophe Pison, Marc Schlesser, Piet Vercauter, Claus F. Vogelmeier, Emiel Wouters, Jokke Wynants, and Wim Janssens. 2019. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *European Respiratory Journal* 53, 4 (2019). https://doi.org/10.1183/13993003.01660-2018 arXiv:https://erj.ersjournals.com/content/53/4/1801660.full.pdf

[161] Eric J. Topol. 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* 25, 1 (01 Jan 2019), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[162] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. https://doi.org/10.1126/science.185.4157.1124 arXiv:https://www.science.org/doi/pdf/10.1126/science.185.4157.1124

[163] Niels van Berkel, Maura Bellio, Mikael B. Skov, and Ann Blandford. 2023. Measurements, Algorithms, and Presentations of Reality: Framing Interactions with AI-Enabled Decision Support. *ACM Trans. Comput.-Hum. Interact.* 30, 2, Article 32 (mar 2023), 33 pages. https://doi.org/10.1145/3571815

[164] Misha Vaughan and Catherine Courage. 2007. SIG: Capturing Longitudinal Usability: What Really Affects User Performance over Time?. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI EA '07)*. Association for Computing Machinery, New York, NY, USA, 2149–2152. https://doi.org/10.1145/1240866.1240970

[165] Misha Vaughan, Catherine Courage, Stephanie Rosenbaum, Jhilmil Jain, Monty Hammontree, Russell Beale, and Dan Welsh. 2008. Longitudinal Usability Data Collection: Art versus Science?. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (Florence, Italy) *(CHI EA '08)*. Association for Computing Machinery, New York, NY, USA, 2261–2264. https://doi.org/10.1145/1358628.1358664

[166] Filip Velickovski, Luigi Ceccaroni, Josep Roca, Felip Burgos, Juan B. Galdiz, Nuria Marina, and Magí Lluch-Ariet. 2014. Clinical Decision Support Systems (CDSS) for preventive management of COPD patients. *Journal of Translational Medicine* 12, 2 (28 Nov 2014), S9. https://doi.org/10.1186/1479-5876-12-S2-S9

[167] 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* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 17, 19 pages. https://doi.org/10.1145/3544548.3581506

[168] Giulia Vilone and Luca Longo. 2021. Classification of explainable artificial intelligence methods through their output formats. *Machine Learning and Knowledge Extraction* 3, 3 (2021), 615–661.

[169] Lev S Vygotsky. 2012. *Thought and language*. MIT press.

[170] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology (Harvard JOLT)* 31, 2 (2018 2017), 841–888. https://heinonline.org/HOL/P?h=hein.journals/hjlt31&i=859

[171] Seán Walsh, Evelyn E.C. de Jong, Janna E. van Timmeren, Abdalla Ibrahim, Inge Compter, Jurgen Peerlings, Sebastian Sanduleanu, Turkey Refaee, Simon Keek, Ruben T.H.M. Larue, Yvonka van Wijk, Aniek J.G. Even, Arthur Jochems, Mohamed S. Barakat, Ralph T.H. Leijenaar, and Philippe Lambin. 2019. Decision Support Systems in Oncology. *JCO Clinical Cancer Informatics* 3 (2019), 1–9. https://doi.org/10.1200/CCI.18.00001 arXiv:https://doi.org/10.1200/CCI.18.00001 PMID: 30730766.

[172] Elena A Wood, Brittany L Ange, and D Douglas Miller. 2021. Are We Ready to Integrate Artificial Intelligence Literacy into Medical School Curriculum: Students and Faculty Survey. *Journal of Medical Education and Curricular Development* 8 (2021), 23821205211024078. https://doi.org/10.1177/23821205211024078 arXiv:https://doi.org/10.1177/23821205211024078 PMID: 34250242.

[173] Yao Xie, Melody Chen, David Kao, Ge Gao, and Xiang 'Anthony' Chen. 2020. CheXplain: Enabling Physicians to Explore and Understand Data-Driven, AI-Enabled Medical Imaging Analysis. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376807

[174] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen

Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 202, 30 pages. https://doi.org/10.1145/3544548.3581500

[175] Nur Yildirim, Alex Kass, Teresa Tung, Connor Upton, Donnacha Costello, Robert Giusti, Sinem Lacin, Sara Lovic, James M O'Neill, Rudi O'Reilly Meehan, Eoin Ó Loideáin, Azzurra Pini, Medb Corcoran, Jeremiah Hayes, Diarmuid J Cahalane, Gaurav Shivhare, Luigi Castoro, Giovanni Caruso, Changhoon Oh, James McCann, Jodi Forlizzi, and John Zimmerman. 2022. How Experienced Designers of Enterprise Applications Engage AI as a Design Material. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 483, 13 pages. https://doi.org/10.1145/3491102.3517491

[176] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 134, 21 pages. https://doi.org/10.1145/3544548.3581161

[177] Haoran Zhang, Amy X. Lu, Mohamed Abdalla, Matthew McDermott, and Marzyeh Ghassemi. 2020. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. In *Proceedings of the ACM Conference on Health, Inference, and Learning* (Toronto, Ontario, Canada) *(CHIL '20)*. Association for Computing Machinery, New York, NY, USA, 110–120. https://doi.org/10.1145/3368555.3384448

[178] Zijian Zhang, Koustav Rudra, and Avishek Anand. 2021. Explain and Predict, and Then Predict Again. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 418–426. https://doi.org/10.1145/3437963.3441758

[179] Zijian Zhang, Jaspreet Singh, Ujwal Gadiraju, and Avishek Anand. 2019. Dissonance Between Human and Machine Understanding. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 56 (nov 2019), 23 pages. https://doi.org/10.1145/3359158

[180] Leonieke G. Zomerdijk and Christopher A. Voss. 2010. Service Design for Experience-Centric Services. *Journal of Service Research* 13, 1 (2010), 67–82. https://doi.org/10.1177/1094670509351960 arXiv:https://doi.org/10.1177/1094670509351960

[181] Erdal İn, Ayşegül A. Geçkil, Gürkan Kavuran, Mahmut Şahin, Nurcan K. Berber, and Mutlu Kuluöztürk. 2022. Using artificial intelligence to improve the diagnostic efficiency of pulmonologists in differentiating COVID-19 pneumonia from community-acquired pneumonia. *Journal of Medical Virology* 94, 8 (2022), 3698–3705. https://doi.org/10.1002/jmv.27777 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.27777