It's Not Just for Trust: Designing for Emerging Uses of Explainable AI in Clinical Decision-Making

KATELYN MORRISON, Carnegie Mellon University, USA
ZEXUAN LI, University of Michigan, USA
MINSUK KIM, Carnegie Mellon University, USA
CHENGQI (MALIA) HONG, Carnegie Mellon University, USA
JIDAPA KRAISANGKA, Mahidol University, Thailand
PRISCILLA CORREA-JAQUE, Icahn School of Medicine at Mount Sinai, USA
CHARLES FAUVEL, Rouen University Hospital, France
SANDEEP SAHAY, Houston Methodist, USA
REBECCA R. VANDERPOOL, Ohio State University, USA
ALLEN EVERETT, Johns Hopkins University, USA
SHILI LIN, Ohio State University, USA
MANREET KANWAR, Allegheny Health Network, USA
RAYMOND BENZA, Icahn School of Medicine at Mount Sinai, USA
ADAM PERER, Carnegie Mellon University, USA

Explainable AI (XAI) is often viewed as a mechanism to promote the transparency and interpretability of AI recommendations in high-stakes domains, such as healthcare. This has led many studies to focus on designing and evaluating XAI to foster trust, calibrate reliance, enable algorithmic recourse, and support model understanding. However, this limited design scope restricts our understanding of how XAI can be used more broadly to support critical tasks in complex workflows, such as facilitating shared decision-making and supporting communication between stakeholders. Our work aims to address these critical gaps in the design for and understanding of XAI's emerging uses by iteratively prototyping XAI designs for an AI-powered clinical decision-support system with clinical stakeholders. We then created a high-fidelity prototype from those iterative sessions and used it as a design probe to uncover four emerging uses of XAI: collaboratively exploring treatment options, identifying and reflecting on treatment plans, communicating with stakeholders, and supporting health education. We reflect on the implications of designing XAI for emerging uses in healthcare.

Authors' Contact Information: Katelyn Morrison, kcmorris@andrew.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Zexuan Li, zexuanli@umich.edu, University of Michigan, Ann Arbor, Michigan, USA; Minsuk Kim, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Chengqi (Malia) Hong, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Jidapa Kraisangka, Mahidol University, Nakhon Pathom, Thailand; Priscilla Correa-Jaque, Icahn School of Medicine at Mount Sinai, New York, New York, USA; Charles Fauvel, Rouen University Hospital, Rouen, France; Sandeep Sahay, Houston Methodist, Houston, Texas, USA; Rebecca R. Vanderpool, Ohio State University, Columbus, Ohio, USA; Allen Everett, Johns Hopkins University, Baltimore, Maryland, USA; Shili Lin, Ohio State University, Columbus, Ohio, USA; Manreet Kanwar, Allegheny Health Network, Pittsburgh, Pennsylvania, USA; Raymond Benza, Icahn School of Medicine at Mount Sinai, New York, New York, USA; Adam Perer, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Manuscript submitted to ACM

CCS Concepts: • **Human-centered computing → HCI design and evaluation methods**;

Additional Key Words and Phrases: human-centered explainable AI, human–AI decision-making, iterative prototyping

ACM Reference Format:
Katelyn Morrison, Zexuan Li, Minsuk Kim, Chengqi (Malia) Hong, Jidapa Kraisangka, Priscilla Correa-Jaque, Charles Fauvel, Sandeep Sahay, Rebecca R. Vanderpool, Allen Everett, Shili Lin, Manreet Kanwar, Raymond Benza, and Adam Perer. 2026. It's Not Just for Trust: Designing for Emerging Uses of Explainable AI in Clinical Decision-Making. 1, 1 (July 2026), 27 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Explainable AI (XAI) has largely been designed and evaluated through the lens of four traditional uses: to foster trust, calibrate reliance, enable algorithmic recourse, and support understanding of model behavior [9, 26, 37, 49, 52, 72, 85, 95]. While these uses of XAI are core to developing safe, transparent, and responsible AI, they constrain the design space for how else AI explanations can be used in human–AI collaborations. Furthermore, mixed results and critiques on XAI's ability to actually achieve those four traditional uses [83, 93, 94] raise the question of whether we are designing XAI with the right uses in mind. For example, fostering trust and calibrating reliance on AI are proving to be more complex constructs that an AI explanation alone cannot influence [78]. Research has captured a variety of factors, such as task difficulty, domain expertise, and consistency of AI performance [12], that contribute to shaping trust and reliance behaviors during human–AI collaboration [24]. Within high-stakes decision-making contexts such as healthcare, factors beyond AI explanations, including clinical trial validation and institutional support, are key to gaining physicians' acceptance of AI-powered clinical decision-support systems (AI-CDSS) [8, 84].

Other critiques of XAI have highlighted XAI's lack of actionability during AI-assisted decision-making [83]. Singh et al. [83] state that an explanation is actionable if it provides information that helps people figure out what actions need to be taken in order to change the AI's prediction. Consequently, a subset of XAI research has focused on designing actionable AI explanations to be used for algorithmic recourse in human–AI decision-making contexts [5, 83, 96]. Beyond algorithmic recourse, emerging uses of XAI in empirical studies remain limited. Discussions in recent literature have alluded to XAI being used as a mediator in human–human interactions [19, 35, 43, 48] or as a tool to facilitate knowledge discovery [88]. Yet HCI work on XAI in human–AI decision-making has underexplored designs that focus on emerging uses.

Despite evidence that AI explanations often fall short of achieving those four traditional uses, research continues to pit different explanation configurations (*i.e.*, method, modality, content, and interactivity [4, 14, 23, 33, 53, 73]) against one another to identify design guidelines that will lead us closer to the "best" XAI configuration for fostering trust [95], enabling algorithmic recourse [93], supporting model understanding [95], or calibrating reliance [34, 49]. This highlights a gap in the literature where little research exists that seeks to design for and understand the emerging uses of XAI in human–AI collaboration. Although these traditional uses are important to the development of responsible AI, they do not reflect the complexity of real workflows, where many tasks are interdependent and extend beyond a single decision point. Furthermore, these uses do not account for the collaborative, communicative, and exploratory needs and challenges that arise within complex workflows, such as clinical workflows. Yet XAI methods offer mechanisms that can fundamentally support these broader uses when designed with them in mind. We argue that designing for emerging uses can help address longstanding challenges across human–AI collaboration and computer-supported cooperative work contexts. For example, in healthcare, AI explanations can support communication grounding in human–human interactions, such as when a physician uses AI to identify and discuss appropriate treatment options or justify clinical reasoning to a patient. However, with the current empirical XAI literature in clinical contexts focusing primarily on traditional uses of XAI and for a single task in isolation [72, 85], these broader needs presented by clinical workflows remain underexplored.

To investigate these critical gaps in the design for and understanding of XAI's emerging uses in a real context, we focus on an AI clinical decision-support tool for Pulmonary Arterial Hypertension (PAH). PAH is a rare and incurable disease that predominantly affects women [20]. This progressive disease demands intricate, shared decision-making across all stages of care, from medical history evaluation to long-term treatment planning and symptom monitoring. Although there are algorithmic tools for PAH that offer mortality risk calculations based on the patient's lab results and demographics to aid in treatment planning [18], these tools are often limited to risk prediction for the current visit without XAI, neglecting nuanced communication and reflective practices critical to patient care. These factors present PAH as a compelling context for examining emerging uses of XAI across the clinical workflow. Our investigation into these emerging uses is guided by the following questions: (*RQ1*) *How can existing XAI techniques be designed to better enable emerging uses by clinicians treating patients with PAH?* and (*RQ2*) *What emerging uses of XAI do clinicians treating PAH envision across their evolving workflow?*

Recognizing that the design for emerging uses of XAI in clinical workflows remains underexplored, we adopted an iterative prototyping approach with clinical stakeholders as a method of inquiry to narrow down and refine designs of existing XAI methods (Phase 1). We chose prototyping for Phase 1 because prototyping methods have been shown as appropriate ways to design for and understand XAI needs [35, 48]. These co-design sessions informed the design of our high-fidelity prototype in Phase 2, which we used as a design probe with seven PAH physicians to further explore the emerging uses of XAI in clinical decision-making. We use our final prototype as a design probe in Phase 2 because design probes have been previously shown to be useful in eliciting rich data within clinical decision-making [11]. To our knowledge, no prior work has used participatory design to intentionally design for or understand the emerging uses of XAI across the full clinical workflow.

Our study presents four emerging uses of XAI specific to clinical decision-making through a combination of interactive and exploratory XAI designs: (1) collaboratively exploring treatment options with patients, (2) generating hypotheses to identify or reflect on treatment plans, (3) communicating with other stakeholders, and (4) supporting education efforts about the disease. These findings emphasize that limiting XAI designs to traditional uses constrains its design space, limiting the opportunity to identify emerging uses of XAI. Our work contributes to the ongoing discourse on XAI by encouraging researchers to explore how existing XAI designs, metrics of success, and evaluations may limit how XAI is used in human–AI and human–human–AI decision-making.

2 Background & Related Works

To better motivate the need to focus on designing for and understanding the emerging uses of XAI in clinical decision-making, we first synthesize prior work on XAI techniques for clinical decision-making and what type of information they provide. We then identify how XAI has primarily been used in healthcare-related empirical studies and examine how existing evaluation practices have contributed to the narrow focus on traditional uses. Finally, we highlight gaps in the literature by reviewing studies that have explored the emerging uses of XAI in clinical contexts.

2.1 Explainable AI Techniques for Clinical Decision Support

According to numerous literature surveys (*e.g.*, [72, 90, 91]), explainable AI techniques have increasingly been developed and adapted to work with the complex models and data that are prevalent in the medical domain. Prentzas et al. [72] categorizes the development of XAI solutions for medical AI across three main categories: model-agnostic, interpretable models (model-specific), and explaining deep learning models (model-specific). Surveys have also grouped XAI solutions into local (instance-level) versus global (model-level) explanations [72, 90, 91]. A plethora of approaches exists within each of these categories. Across multiple surveys on XAI solutions for healthcare [72, 90, 91], the primary intended use of XAI has been to promote transparency and assist the end-user (*i.e.*, clinician or patient) in judging the model's accuracy. Grounded in the terminology used by human–computer interaction (HCI) researchers [51], we translate these categories to be either solutions that intend to "explain why" or solutions that intend to "explain why not/what if".

Explaining "Why". Feature attribution techniques have been used to help explain why the AI made a certain prediction [72]. For example, SHAP [55] was introduced in the medical community to help clinicians understand a risk prediction model [56], and has been extensively used across clinical applications [72]. This model-agnostic, local explanation works by attributing model predictions to individual input features and produces a one-time, static output [55]. Another approach, LIME [74], fits an interpretable model to the black-box model to explain a given prediction and similarly produces a one-time, static output. Although these approaches aim to support end-users in understanding why the AI predicted a particular outcome, numerous empirical studies have highlighted their limitations [15, 16, 38, 46]. For example, Kaur et al. [38] found that even data scientists misinterpret SHAP explanations and over-trust an imperfect AI. Chromik et al. [16] observed how people who see SHAP explanations over-estimate their understanding of the AI's behavior. Attribution-based XAI methods, like SHAP, can be difficult to interpret because the variables (*i.e.*, feature contributions) are presented as if they act independently, even though clinical features tend to be highly correlated and interact in complex ways. These limitations present a gap between what XAI designs currently provide and the intricacies of complex clinical decision-making.

Explaining "Why Not/What if". Numerous XAI techniques are designed to help users understand why a certain instance was not predicted to be a particular class or how the AI's prediction changes under different input conditions. Although some of these XAI techniques are automated, such as DICE [62], they may unintentionally leverage irrelevant or non-actionable variables [5]. In the context of clinical decision-making, this can make these explanations difficult to interpret or integrate into one's decision-making. Alternatively, researchers have developed interactive counterfactual explanations that gives users control over changing the input features themselves (*e.g.*, [5, 29, 96, 101]). For example, [29] introduces an interactive framework that allows clinicians to select anatomical segments in a source image and replace them with segments from a target image, enabling them to see how their changes affect the model's classification. Beyond local explanations, global explanations can face similar limitations. For example, partial dependence plots (PDPs) usually assume feature independence [72], which can make interpretation challenging in clinical decision-making. However, Krause et al. [44] mitigate this challenge with an interactive and dynamic system that shows all features updating in real time as users change them. These approaches along with the feature attribution approaches, have shaped how XAI has been used throughout clinical decision-making.

2.2 Empirical Uses and Evaluation of Explainable AI in Clinical Decision-Support Systems

To understand why XAI remains limited to a narrow set of uses in clinical decision-making, we consider the XAI design resources available as well as how empirical and participatory design studies frame and evaluate XAI.

Toolkits. Srinivasu et al. [89] presents a literature survey on a series of toolkits and frameworks available to help designers, clinicians, and researchers develop XAI methods that align with user needs (*e.g.*, [35, 59, 67, 100]). Several of these toolkits, as analyzed by Srinivasu et al. [89], position XAI to be used for traditional uses.

Empirical Studies. Numerous empirical studies have been published on what information decision-makers need from an AI explanation (*e.g.*, [8, 19, 21, 50, 79, 99]). Several studies focus on understanding and/or comparing how XAI techniques impact user trust, reliance, decision confidence, and satisfaction (*e.g.*, [23, 28, 49, 70, 84]). For example, Lee and Chew [49] consider how AI explanations impact therapists' performance, decision agreement pre- and post-AI, and reliance on AI. Panigutti et al. [70] measure weight of evidence to capture how much the AI's prediction and explanation influenced their decision-making on a numerical estimation of patient's survival task. Both examples illustrate how XAI is being used and evaluated for one or more of the traditional uses throughout the HCI literature.

Participatory Design. Instead of conducting empirical studies or leveraging existing XAI toolkits, several other studies take participatory approaches to designing XAI solutions that address users' XAI needs and goals (*e.g.*, [8, 19, 69, 92, 99]). For example, Corti et al. [19] used examples of different domain-relevant XAI visuals to question clinicians about what information from the AI is necessary and how they may appropriate XAI throughout the evolution of their workflow. However, many of the XAI co-design studies in clinical contexts (*e.g.*, [69]) tend to narrow the design space to XAI being used to foster trust and calibrate reliance.

Across toolkits, empirical evaluations, and participatory design studies, prior work has largely focused on using XAI as a tool to foster trust, calibrate reliance, and support model understanding. These uses have shaped how we evaluate XAI's impact and reinforced its narrow use in human–AI clinical decision-making. In the next subsection, we synthesize the literature that has proposed or evaluated XAI for uses beyond trust, reliance, and understanding.

2.3 Emerging Uses of Explainable AI in Clinical Decision-Making

Beyond trust, reliance, and understanding, XAI has also been used to enable algorithmic recourse in mental health-care [47]and clinical decision-making [82], a use that was once emerging but is now relatively well-established. Literature has posed emerging uses of XAI, such as Lundberg et al. [56], who mention how SHAP could help improve clinical understanding of hypoxaemia during surgery, or Woensel et al. [97], who argue that AI explanations can help patients with a chronic disease better understand their disease and motivate long-term behavior change. However, neither of these works conduct an empirical study to ground these emerging uses in stakeholders' perspectives, workflows, or behaviors. Other work has commented on or alluded to the potential of using XAI to mediate multi-stakeholder communication [19, 35, 42]. Jin et al. [35]'s framework for designing user-centered XAI exposed how patients might use XAI to communicate with their caregivers and family members about their condition. In contrast, Corti et al. [19] observed clinicians imagining how they could use XAI to communicate with other stakeholders and justify their decisions. Our work builds XAI designs inspired by these allusions and discussions by taking a participatory design approach to designing for and understanding the emerging uses of XAI. By simulating diagnosis and treatment recommendations in our XAI prototype with clinical stakeholders, we surface concrete emerging uses for XAI in clinical decision-making.

3 Study Context: Pulmonary Arterial Hypertension

We recognize the importance of establishing the study context and mapping the journey of key stakeholders involved in PAH patient care. We present the clinical journey of managing a single patient with PAH, from primary care referral to monitoring treatment outcomes, as shown in Figure 1. Our journey map has been validated by our clinical collaborators. We introduce specific challenges with existing PAH risk assessment tools (examples shown in Figure 2). This journey

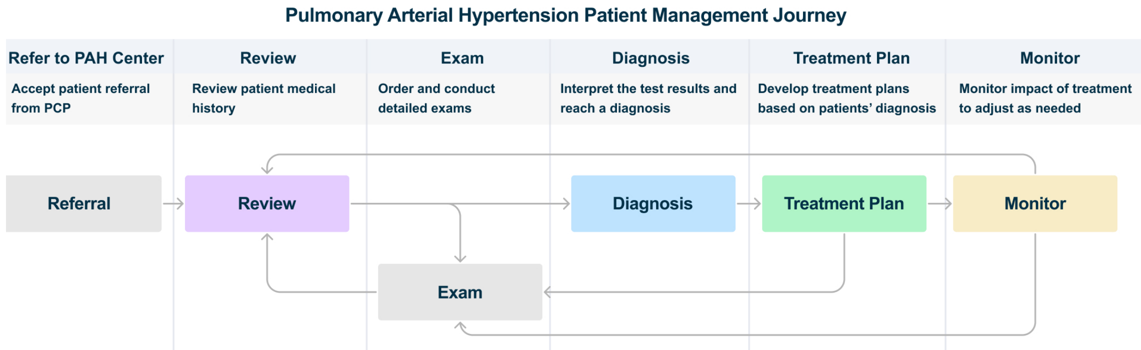

Fig. 1. **Single PAH patient management journey**. Journey validated by PAH physicians.

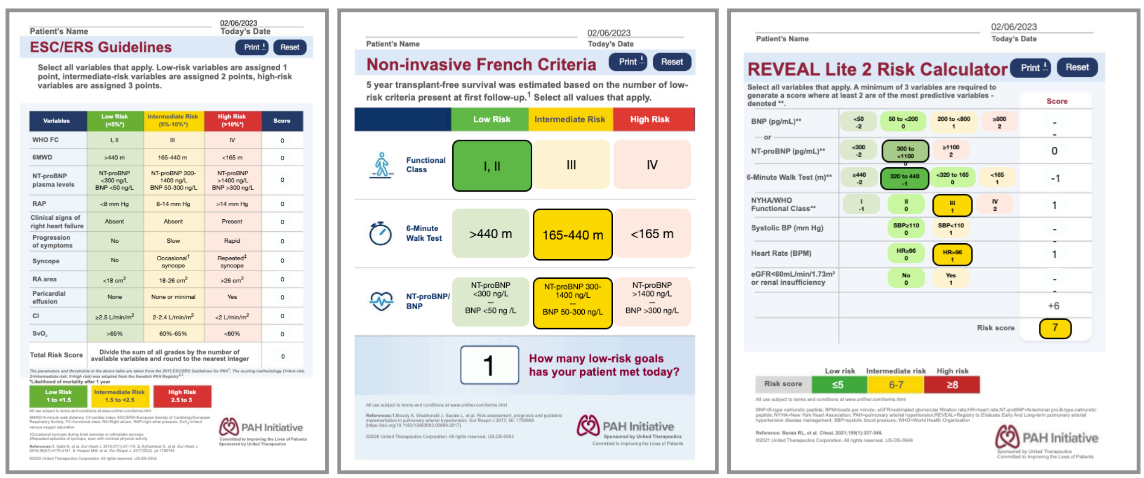

Fig. 2. **Examples of existing PAH risk calculators**. (**Left**) Interface for the ESC/ERS Guidelines risk assessment tool. (**Middle**) Non-Invasive French Criteria risk assessment tool. (**Right**) Reveal Lite 2 risk assessment tool. PAH physicians have referred us to these interfaces, explaining how they aid them in PAH diagnosis and treatment planning. These interfaces are offered by the PAH Initiative [18].

serves as a foundation for identifying and designing for emerging uses of XAI explored in our study. Each phase of the journey is detailed below.

Review Stage. After a patient is referred to a PAH care center by their primary care physician, the PAH physician will review the patient's existing medical history.

Exam Stage. After reviewing the patient's medical history, the physician requests exams if they need additional data before diagnosing the patient. For example, the patient may need an echocardiogram or right heart catheterization so the physician can determine any abnormalities in the heart's function. After the exams are completed, the physician will return to the review stage to consider any additional factors that may influence the diagnosis.

Diagnosis Stage. After the patient completes additional tests, the physician will calculate the patient's risk of mortality at 6 months, 1 year, or 5 years to help identify appropriate treatment interventions. This typically occurs just before the physician's visit with the patient. There are a variety of algorithm-based risk assessment calculators (*e.g.*, REVEAL [2], COMPERA [31], and ESC/ERS Guidelines [32]) to aid in diagnosis and treatment planning. For example, the calculators shown in Figure 2 operate on a green, yellow, and red color scale, which helps to easily communicate the patient's risk level. Some institutions have integrated these calculators into their electronic health record systems, while others have posters of them on the walls in the patient rooms or nothing at all, such as in community clinics.

Treatment Planning Stage. After calculating the patient's risk, the physician will identify an appropriate treatment plan, considering any existing treatment plans and co-morbidities the patient may have. Numerous PAH physicians leverage the ESC/ERS 2022 guidelines [32] when determining an appropriate treatment plan for PAH patients. However, selecting a treatment is a shared decision-making process between the patient and the physician, as certain treatments can require significant lifestyle modifications. This is where being able to visualize different scenarios of the patient's condition can help inform the next steps in treatment planning and encourage reflective practices.

Monitor Stage. After a treatment plan has been established, physicians monitor their patient's reactions to the treatment through follow-up visits, which typically occur every three to six months. The frequency of these visits depends on factors such as the patient's diagnosis, their response to the treatment plan, and external constraints, such as pharmacy or insurance bottlenecks. During these follow-ups, physicians reassess the patient's progress, often recalculating their risk based on new information — whether from exams, unexpected hospitalizations, or regular check-ups — to determine if modifications to the treatment plan are needed.

4 Phase 1: Iterative Prototyping of XAI Designs

Our iterative prototyping phase sets out to address our first research question: (*RQ1*) *How can existing XAI techniques be designed to better enable emerging uses by clinicians treating patients with PAH?* During this phase, we engage in low-fidelity prototyping of design concepts for AI explanations while gathering feedback from key stakeholders in PAH.

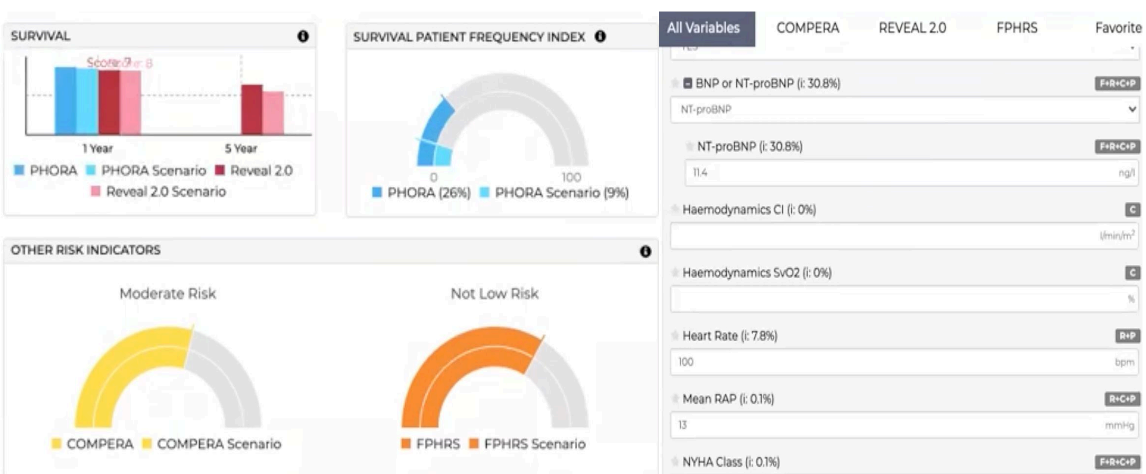

Fig. 3. **Inherited AI-CDSS design.** (**Left**) Visualizations designed by a subset of the clinical team to visualize how models would predict risk differently when values of the features were modified. (**Right**) A navigation tab for different risk prediction models and the features for each model. Each feature's value can be overwritten to create a new scenario.

Total Participants	Role Breakdown	Experience Breakdown
18	16 PAH physicians 1 Nurse Practitioner 1 Behavioral Psychologist	1 with 1–5 years 5 with 10–19 years 12 with 20–40 years

Table 1. Participant demographics for Phase 1.

4.1 Methods

4.1.1 Inherited Prototype. Before we started iterative prototyping, we inherited an already existing AI-CDSS prototype built by a subset of the clinical team without human–computer interaction expertise (shown in Figure 3). We used this inherited prototype as a starting point for our visualizations because it emulates a core interaction and visualization that the clinical team wanted to further explore in the AI-CDSS: manipulating feature values to view and compare different scenarios. This scenario-creation interaction is facilitated by XAI methods, such as counterfactuals, which can be automated [62] or exploratory [96]. However, because this prototype was designed with minimal feedback from clinicians outside the team, we only treat this inherited prototype as a starting point for the possible types of interactions to offer. We describe how the low-fidelity prototype that we developed in Figma across two rounds build on this scenario-creation interaction while expanding to other XAI designs below.

4.1.2 Round 1 Prototype. In the first round, we developed an interactive Figma prototype showcasing an initial set of design concepts for explanations: subgroup insights, understanding driving factors of risk, and "what-if" hypothesis generation capabilities. We conducted 30-minute semi-structured interviews with five PAH clinicians (***P1–P5***) using the prototype as a probe to gather feedback on the design concepts and elicit potential emerging uses for AI explanations across the PAH patient care journey.

4.1.3 Round 2 Prototype. After talking to those five PAH clinicians, we updated the low-fidelity Figma prototype to reflect the feedback and conducted another set of 30-minute semi-structured interviews with 13 PAH clinicians (***P6–P18***) using the updated prototype as a probe. These sessions further refined our understanding of the roles that explanations could play throughout the PAH care journey, laying the groundwork for the final design of the high-fidelity prototype. As a result, we selected a subset of the design concepts to further develop in the high-fidelity prototype. Descriptions of the design concepts used throughout prototyping are provided in Section 4.2.

4.1.4 Participant Recruitment. The five participants for the first round of prototyping consisted of four physicians and one family nurse practitioner. The Round 1 interview sessions lasted for an average of 28.56 minutes. The second-round prototype was shown to 12 physicians and one behavioral psychologist for PAH patients (average interview time: 18.58 minutes). Participants were informed of all our IRB-approved study procedures and voluntarily participated in our study after signing a consent form. We aimed to recruit a group of stakeholders with diverse years of experience with PAH care, as shown in Table 1. However, we did not intentionally aim to recruit from a diverse set of roles, which is why we only have one nurse practitioner and one behavioral psychologist.

4.1.5 Analysis Approach. Our primary analysis method is a thematic analysis. Two thematic analyses were done: one was to capture the design feedback while prototyping the low-fidelity prototype, and the other was to capture any insights mentioned about emerging uses of XAI as our designs evolved. To do the thematic analyses, two authors identified relevant quotes from the interview transcripts and collaboratively grouped the quotes directly in Figma based

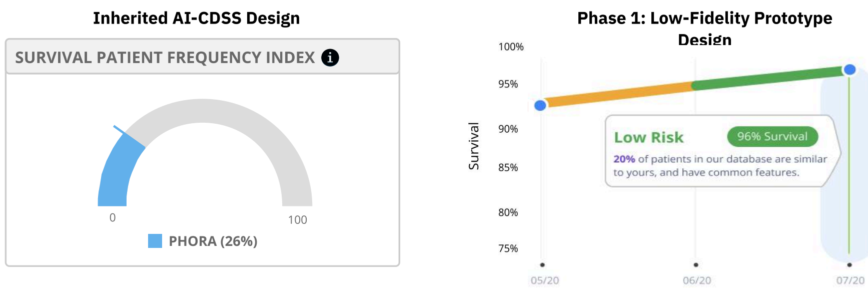

Fig. 4. **Evolution of the Subgroup Insights Design**. (**Left**) Inherited AI-CDSS design with donut chart visualizations showing the 'survival patient frequency index'. (**Right**) An iteration of Phase 1 designs showing text callouts for the instance frequency is integrated into the patient's historical risk trend line chart.

on recurring ideas and patterns observed. Each group of quotes was assigned a title to represent the theme. The findings from these thematic analyses are presented in Section 4.2.

4.2 Findings: XAI Design Concepts

Our initial design concepts were informed by existing features in the inherited early AI-CDSS prototype, insights from relevant literature on AI-CDSS systems, and informal discussions with PAH physicians. Based on this foundation, we developed a set of XAI design concepts that have been used for calibrating trust and reliance, but can also afford emerging purposes. These include providing insights into patient subgroups, understanding the factors driving risk predictions, and supporting effective communication with different stakeholders. Each design concept was iterated on after each round; we rationalize final designs with interview quotes from Phase 1 participants below.

4.2.1 ***Subgroup Insights***. Recent XAI work in human–AI collaboration has suggested that showing explanations based on model behaviors for different subgroups can improve the appropriate use of the model [9]. Similarly, having an understanding of the percentage of data similar to your instance that was present in the training set [30] or understanding the error boundaries of the AI model [63] can improve the users' appropriate use of the model.

Given that the inherited AI-CDSS designs already offered an instance frequency visualization, and recent literature has shown that it benefits human–AI teams to have a sense of when the model may perform poorly, we include a design communicating an aggregate statistic about the instance related to the training data in the low-fidelity prototype. For example, in Figure 4, when a risk prediction is made, the AI-CDSS shows an aggregate statistic for the percent of training data that is similar to the patient instance. However, many participants in Phase 1 struggled to see the value this design intended to provide. *P5* questioned the utility of the feature, asking, "*What is the value of knowing that 20% of patients in the database are similar to yours*?" and misunderstood its purpose, remarking, "*I actually thought that that was another level of risk*." Similarly, *P1*, *P2*, and *P3* could not envision how this explanation would add value to their workflow. *P4*, the nurse practitioner, critiqued the terminology, suggesting that while "frequency" might be misleading, and alternatives like "similarity" or "comparison" would not alleviate the underlying confusion. Given that participants found this design concept more confusing than valuable, we chose not to keep this feature in the Phase 2 high-fidelity prototype design.

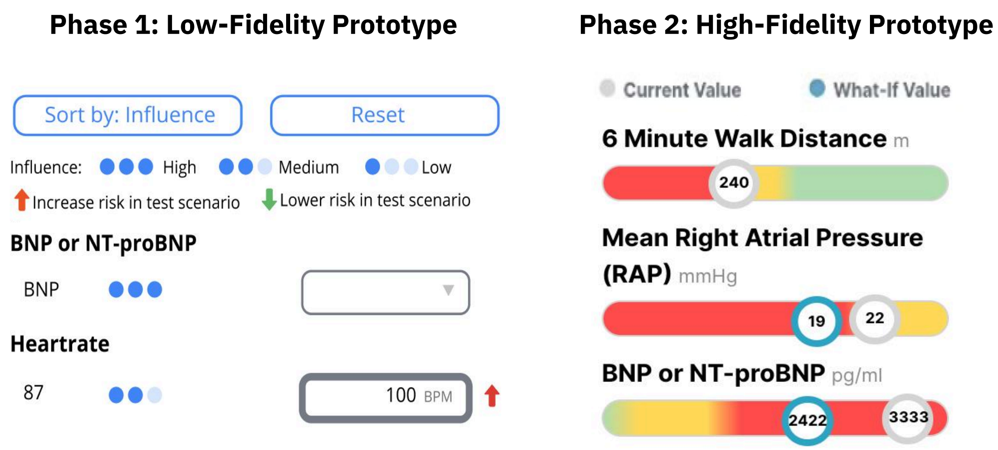

Fig. 5. **Evolution of Designs for Understanding Driving Factors of Risk Prediction**. (**Left**) One iteration of the Phase 1 low-fidelity prototype design used a three-dot system to represent the influence of a feature on the prediction. (**Right**) Phase 2 high-fidelity prototype uses partial dependence bars to visualize the impact of features on each other and the prediction simultaneously.

4.2.2 ***Understanding Driving Factors of Risk Prediction***. Previous work has found empirical evidence of understanding factors that drive predictions to be important in understanding AI and increasing actionability [5]. While the inherited AI-CDSS did not offer this design, we included a feature attribution style design in the low-fidelity prototype. The design is shown as blue dots increasing in quantity based on feature importance, as shown in Figure 5 (left) to represent important features driving the prediction. Although the participants from Phase 1 stressed the importance of understanding "*what's the driving variable behind all of this...*" (***P1***), the blue dot visual design in the Phase 1 prototype was confusing to some of the participants.

Initially, our Phase 1 low-fidelity prototype design enabled what-if explanations by overwriting the variables' values in a text box, shown in Figure **5** (left), as inspired by the inherited AI-CDSS. However, it was not integrated with feature importance. Interactive what-if explanations combined with feature importance visuals have been shown to help users understand driving risk factors and provide actionable insights [5]. During the interviews for the second round of the low-fidelity prototype in Phase 1, ***P6***, the behavioral psychologist, mentioned that "*It might be easier for people to drag a line than to like [input text in the variable field] 'let me try 550'... 'let me try 575'*" and that the current design "*...doesn't tell you how much of that [scenario risk score] is age, and how much of that is walk distance*" when you make complex "what-if" scenarios. As a behavioral psychologist, **P6** viewed the "what-if" explanation scenarios as a way to potentially assist patients in "*therapeutic transitions*" and help explain to them "*why we want you in rehab...*".

To combine information provided by feature importance visualizations with interactive "what-if" explanations, our final design (seen on the right side of Figure 5) in the Phase 2 high-fidelity prototype takes inspiration from Krause et al. [44] and Beretta et al. [3]. The XAI design for this concept draws on classic partial dependence plots (PDPs), but modifies them to be interactive and feature-dependent: the bars visualize the effect of different feature values on the predicted risk score as bar toggles are interacted with [44, 80]. This minimizes the feature-independence assumptions of classic PDPs and enables dynamic "what-if" reasoning that existing XAI algorithms do not support. To illustrate, consider a PAH patient presenting with an elevated heart rate (HR) and reduced six-minute walk distance (6MWD), whom the model classifies as high risk. A traditional XAI technique like SHAP would decompose this prediction into additive feature contributions (*e.g.*, indicating that HR contributed +15% and 6MWD contributed +10% to the elevated risk score relative to a baseline). While this answers why the model produced its current prediction, it does not address the question more relevant to clinical action: what would happen to this patient's risk if treatment successfully reduced their HR? Our interactive "what-if" explanations directly support this reasoning. A clinician can adjust the HR to a target value and immediately observe how the predicted risk changes for all other features simultaneously. This "what-if" interactivity reveals how changing one feature may affect the patient's other current values, according to the model. Our interviews with PAH clinicians suggested this intervention-oriented framing aligns more naturally with clinical decision-making than traditional XAI's feature attribution.

The grey circle (bar toggle) and value in Figure 5 (right) represent the patient's current value, while the blue circle (bar toggle) and value represent the user's selection for creating a "what-if" scenario. The color gradients show the categorization of risk predictions for every possible combination of values. Visually, the color gradients are red–yellow–green, as this aligns with the existing culture that PAH physicians have, noting that "*...a lot of us have just gotten used to seeing the risks that way*" (***P1***). These gradients are updated simultaneously for every feature as the what-if values are added and modified.

4.2.3 ***Hypothesis Testing Capabilities.*** Ludwig and Mullainathan [54] underscores the central role of hypothesis generation throughout the clinical workflow, while Zhang et al. [101] qualitatively observed clinicians' interest in using counterfactuals during treatment planning. However, the previous works are limited in their exploration of XAI designs that are adaptable to the evolving clinical workflow and enable emerging uses. Hypothesis testing was already incorporated in the inherited AI-CDSS design and has similarly been suggested by previous works [5]. ***P8*** motivates PAH clinicians' interest in hypothesis testing, saying, "*...it would be nice to be able to see and to share with patients like, you know, hey, if this happened, this is where your prognosis would go*". ***P6***, the behavioral psychologist, noted the importance of the variable included in making "what-if" scenarios, saying: "*I want to separate out the behavioral aspect of it from the part that they have no control over, which is their aging*". This aligns with related XAI research in AI-CDSS, which states that AI-CDSS should only offer clinical variables in what-if visualizations that can be realistically modified [5]. Based on this, we gathered feedback from PAH clinicians on which variables would make sense to include and in which order they should be presented.

As shown in Figure 6, the inherited AI-CDSS featured hypothesis testing through the use of side-by-side bar charts (Figure 6, left). However, we chose to use line charts to better incorporate with a historical trends plot for side-by-side comparison in our low-fidelity prototype (Figure 6, middle). Interestingly, feedback on the low-fidelity prototype in Phase 1 highlighted more areas for improvement. For instance, ***P2*** found the scenario presentation designs confusing and suggested separating the plots entirely from the original patient data. Inspired by this, we refined the Phase 2 high-fidelity prototype to completely separate scenario generation and exploration from the data and visualizations of the current visit (Figure 6, right). This modification ensures clarity by isolating the current risk score from scenario-based risk comparisons, thereby supporting clinicians' needs for clear and actionable insights during hypothesis testing. Inspired by related work on interactive "what-if" explanations [96], we design for multiple comparisons, saving scenarios, and summarizing variables changed.

4.2.4 Summary of Phase 1 Insights. We provide a clear mapping of Phase 1 insights to Phase 2 design decisions in Table 2. The four key insights include removing subgroup insights, integrating feature importance into the "what-if" exploration, separating "what-if" exploration from real data, and designing to support human-to-human communication.

Phase 1 Insight	Source (P#)	Design Decision in Phase 2 Prototype
Subgroup insights visualization was confusing and unclear on how it supports workflows	P1–P5	Removed the subgroup insights entirely from the high-fidelity AI-CDSS prototype
Feature importance can enrich "what-if" explorations	P6	Mapped the feature importance on the clinical variables in the "what-if" exploration
Hypothetical and real data can be easily confused	P2, P4	Separated scenario exploration into its own pane, distinct from historical risk trends and replaced value overwriting with slider-based controls
XAI has the potential to support patient or trainee communication	P1, P3, P7, P8	Connected XAI component to a comparative graph and treatment guidelines

Table 2. Mapping Phase 1 insights to design decisions in the Phase 2 AI-CDSS.

Collectively, these insights help address *RQ1* by suggesting that clinicians gravitate more towards interactive and exploratory techniques that can more easily integrate with their decision-making processes rather than static, feature attribution methods.

5 Phase 2: Exploring Emerging Uses of XAI

Figure 7 summarizes the methods used for Phase 1 and how they inform the methods in Phase 2. After iterating on the XAI design concepts (Phase 1) described above, we used the resulting insights to inform the final high-fidelity prototype for Phase 2. We summarize how insights from Phase 1 inform the design of the high-fidelity prototype that we used as a design probe in the next subsection.

5.1 High-Fidelity AI-CDSS Design & Development

Inspired by designs from the inherited AI-CDSS, our prototype in Figure 8 presents the (1) patient's clinical variables, (1a) historical trends of each clinical variable, (2) a PAH risk prediction model called PHORA [36], (2a) a historical trend of predicted risk, (3) a "what-if" exploration feature (our final XAI design component), and (3d) a treatment

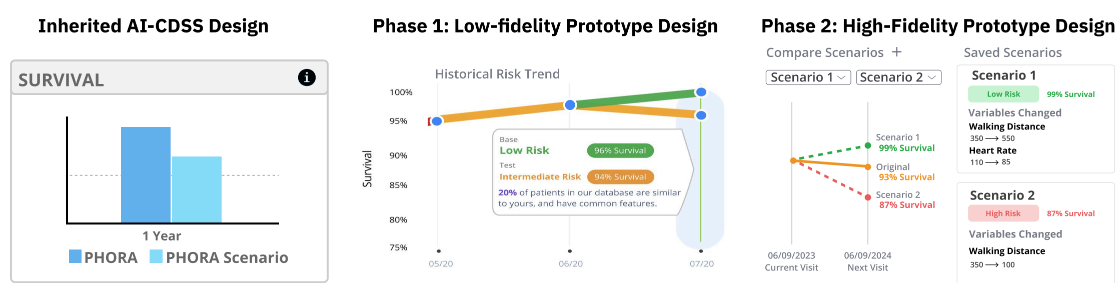

Fig. 6. **Evolution of the Hypothesis Testing Capabilities Design**. (**Left**) The inherited AI-CDSS design shows bar charts to compare against what-if explanations. (**Middle**) The Phase 1 low-fidelity prototype design adapted line charts as a historical trend to compare against what-if explanations. (**Right**) The Phase 2 high-fidelity prototype separates the what-if explanations from the patient's historical trend visualization, enabling emerging uses by offering a comparison of scenarios and a summary of variables modified.

guidelines component. This high-fidelity prototype presents an interactive, conditional explanation system that differs from classical partial dependence plots (PDPs) because it avoids the feature independence assumption. Our findings from Phase 1 confirmed that clinicians found the hypothesis testing functionality useful. They also imagined emerging uses for it, such as augmenting conversations with patients. Our findings also revealed the value that feature importance could have when integrated within "what-if" visualizations to guide comparative insights. Lastly, the findings motivate our choice to exclude the subgroup insights due to confusion and an unclear purpose.

We deployed the high-fidelity prototype online and used it as a design probe in 45-minute semi-structured interview sessions with seven PAH physicians to answer *RQ2*: *What emerging uses of XAI do clinicians treating PAH envision across their evolving workflow?* We detail the interview protocol, recruitment process, and analysis approach in the following subsections.

5.2 Method

We conduct semi-structured interviews with seven additional PAH physicians (**P19–P25**). It is important to note that the semi-structured interview was not intended to quantitatively evaluate the model's performance or physicians' appropriate reliance on the AI. These sessions were designed to be generative and help us better elicit the potential uses of the XAI design throughout the physician's journey. This simultaneously provides insight into whether our XAI designs can help enable those uses. We describe the structure of the interviews below.

5.2.1 Pre-probe. Before interacting with the dashboard, participants were presented with a hypothetical clinical scenario in which they diagnosed an anonymous patient from a physician-validated vignette, identified possible treatment plans, and communicated their decisions to the hypothetical patient (the interviewer). The physician-validated vignette was designed to be a complex intermediate-high risk patient case where the PHORA [36] risk assessment model similarly predicted high risk. The pre-probe was designed to get the physician thinking about the patient's case and how they would diagnose the patient before seeing an AI recommendation with explanations.

5.2.2 Probe. After participants communicated how they would diagnose and treat the patient, they were directed to the dashboard and provided a brief overview of the different features: model prediction, patient data, historical trends, "what-if" exploration, and treatment guidelines. Participants were then directed to finalize their diagnosis and treatment

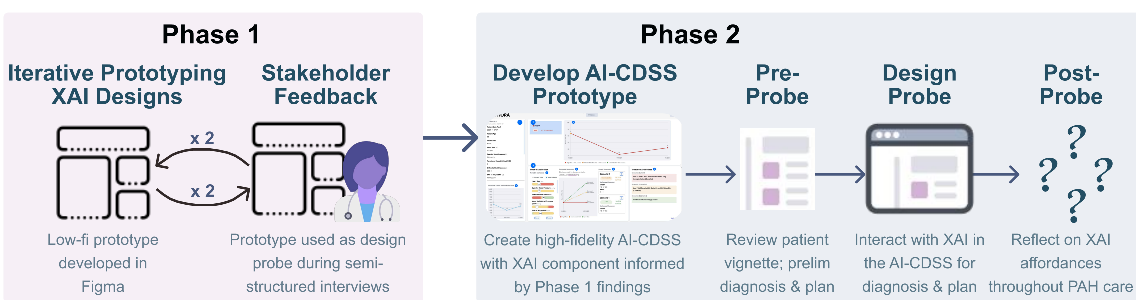

Fig. 7. **Summary of Phase 1 and Phase 2 methods.** Phase 1 consisted of iterative prototyping with feedback from key stakeholders. Informed by developed insights, we develop a high-fidelity AI-CDSS and use it as a design probe for Phase 2. Phase 2 consisted of semi-structured interviews that included a pre-probe, a probe, and a post-probe component.

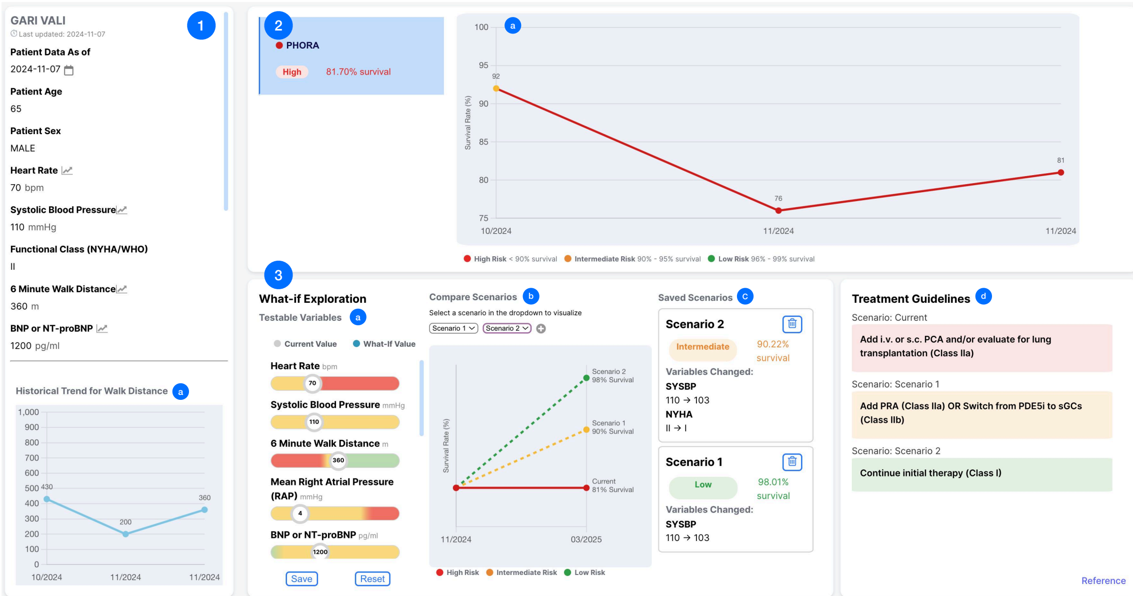

Fig. 8. **Final Phase 2 high-fidelity AI-CDSS prototype.** (**1**) List of clinical variables used in the risk model. (**1a**) Historical trends graph for each quantitative clinical variable. (**2**) Risk model output and (**2a**) historical trend of predicted risk. (**3**) Main XAI component informed by Phase 1 findings.

Total Participants	Role Breakdown	Experience Breakdown
7	7 PAH physicians	3 with 2–6 years 3 with 10–20 years 1 with >30 years

Table 3. Participant demographics for Phase 2.

plan using the dashboard and communicate the diagnosis and treatment plan to the interviewer as if they were the patient.

5.2.3 Post-probe. After participants interacted with the dashboard, they were given the opportunity to reflect on the XAI designs and how they might use them in practice. If not mentioned, they were asked explicitly how they could see XAI being used across the different phases of the PAH care journey identified in Figure 1.

5.3 Participant Recruitment & Analysis Approach

Representing five additional institutions and various experience levels, as shown in Table 3, seven PAH physicians (**P19–P25**) were recruited with the help of our clinical collaborators. We intentionally aimed to recruit only physicians for this phase as they are the ones reviewing patient history, calculating risk scores, and determining treatment options (the tasks simulated in this phase). Participants were informed of all our IRB-approved study procedures and voluntarily participated in our study after signing a consent form. Interview sessions lasted an average of 43.5 minutes.

Phase	**XAI Use**	Phase 1 (*P1–P18*)	Phase 2 (***P19–P25***)
Review	Educating physicians in-training	P2, P13	P19, P21
Diagnosis	Help patients understand the diagnosis	—	P20–22
	Iterative hypothesis exploration	P1, P8–P10, P12, P17	P20–25
Treatment	Celebrate & motivate patients	P2, P6, P8, P13, P16	P20, P21, P25
	Engage patients in treatment planning	P5, P6, P8	P19–P21
	Help the physicians gain patients' trust	P15	P20
Monitor	Discussions with care staff	P2	P20, P25
	Reflection on disease management	P10–P12	P19–21, P23

Table 4. Participants from both phases are mapped to which emerging use of XAI they imagined across the different phases in a PAH physician's journey of managing a single patient.

Following the same analysis process as described for Phase 1 in Section 4.1.5, we conduct a thematic analysis of the relevant quotes from the interviews to capture how participants imagined using the XAI designs to help them diagnose and treat the sample patient. A qualitative approach has been shown to be appropriate to elicit rich insights through the use of semi-structured interviews paired with a design probe [7]. Two authors collaboratively grouped quotes and assigned themes to each group. Any disagreements regarding groupings were discussed to reach a consensus. Themes were then mapped to the PAH journey in order to capture the emerging uses of XAI.

6 Findings on Emerging Uses of XAI

This section focuses on scenarios in which PAH clinicians thought they could use XAI, addressing *RQ2*. We include quotes from both the Phase 1 and Phase 2 because emerging uses were brought up in both phases. We group emerging uses of XAI based on a single PAH patient management journey in Table 4 with a list of which participants from both phases mentioned which uses.

6.1 XAI Use in the Diagnosis Workflow

While the diagnosis is not the start of the workflow, it is one of the most significant parts because it helps the physician determine what treatment options they will discuss with the patient. We observed two primary ways that physicians across Phase 1 and Phase 2 felt they would use the XAI features during their diagnosis workflow: (1) help patients understand the diagnosis and (2) iteratively explore hypotheses.

6.1.1 Help patients understand the diagnosis. Three of the seven physicians from Phase 2 mentioned that they could use the tool to help them communicate better with their patients about their diagnosis and risk score. For example, ***P21*** shared how it can be difficult to have, "*...a higher level abstract conversation about how BNP levels relate to you dying in a year...so maybe this [the XAI designs] would help bridge that gap a little bit*". Similarly, ***P20*** stated that the what-if explanation design can "*...**helps patients understand why I'm so concerned, even if they're unfamiliar with their disease states...***". ***P22*** further imagined how they could use the what-if explanations to help them communicate with their patient:

> "*If I tell them your mean right atrial pressure is 16, and we want to get it down to 9. They're gonna look at you like, what the heck are you talking about? But, if I show them how on that little graph thing ["what-if" comparison] if getting it down to 8 changes it from red to green, that, they understand.*"

However, ***P24***, specializing in pediatric PAH, where the patient usually is not old enough to participate in treatment discussions, raised the concern that the design of the "what-if" explanations component could distract caregivers (*e.g.*, parents) from focusing on the right information. Aside from patient communication, ***P20*** pointed out that the partial dependence bars design can bring "*...attention to those factors that are playing the biggest role in the ultimate risk category...*", which they imagined would be helpful.

6.1.2 Iterative Hypothesis Exploration. Six of the 18 clinicians from Phase 1 commented on how, with improvements, the "what-if" XAI design could be used for iterative hypothesis testing to identify the best treatment options. For example, ***P12*** said "*If a patient is going the wrong direction, you want to see where I can intervene...that's going to make the biggest difference*". Six of the seven physicians from Phase 2 used the "what-if" explanations as a way to validate their hypotheses or determine what needs to change in the patient's regimen and condition so that their risk decreases. For example, ***P24*** started off by experimenting with the 6-minute walk test variable to see how much the patient's risk would decrease. ***P22*** modified the testable variables until they were able to get the patient's risk lower. Their interaction led them to finalize a treatment plan by concluding, "*...now all she needs is a second drug*". ***P23*** specifically had used the XAI tool to help them: (1) "*...figure out when the risk becomes medium*" and (2) "*...[identify] which parameter affected the [risk score] the most...*" by observing how the partial dependence bar plots change. ***P21*** created what-if scenarios with the intention to validate how they expect a treatment would affect the patient's variables: "*making these [what-if scenarios] you know, this is what, realistically, I probably hope for by adding another agent.*"

6.2 XAI Use in the Treatment Planning Workflow

The treatment planning workflow follows the diagnosis, where the risk score calculation and medical history strongly influence how physicians determine which treatment path to pursue. One of the five participants from Phase 1 Round 1 and four of the 13 participants from Phase 1 Round 2 imagined how, with improvements, the what-if explanations could be used to have celebratory or motivational discussions with patients, while one participant from Phase 1 Round 1 and three from Round 2 mentioned its potential in being used to engage patients during treatment planning. One participant from Phase 1 also commented on its potential to help physicians gain patients' trust. Four of the seven physicians from Phase 2 imagined how the scenarios could augment their treatment planning workflow by better enabling them to (1) celebrate or motivate the patients, (2) engage the patients in treatment planning discussions, and (3) gain patients' trust.

6.2.1 Celebrate or motivate the patients. PAH treatments can be tough regimes, significantly altering the patient's lifestyle. These lifestyle changes can be difficult for patients to navigate and difficult for physicians to explain why and how the patient's condition could improve if the treatment plan is properly followed. For example, ***P8*** felt like with design improvements to the Phase 1 prototype, the what-if explanations could be used to help them rationalize difficult therapies:

> "*If we were going to introduce a new therapy and said, you know, your BNP has been climbing, it's looking worse. Let's say we started this new therapy. And our goal is that it'll do X, Y, and Z. Now we can look and see where your risk would be if we made our goals. I can see that meaning a lot to patients too. Just getting them to buy in on the therapy. Sometimes they're tough therapies.*"

Similarly, ***P16*** imagined how an improved version of the "what-if" explanations could help them frame their conversation with patients: "*I think [it] would hold a lot of weight with a patient like saying, you know, if we can change factor x by this, look what it does to your outcome, and again,* ***just helping frame that conversation***". ***P20*** validated the design updates to enable this emerging use by similarly imagining how the "what-if" explanations component could be used to help them "...***emphasize to them [patients] that we're unlikely to get where we need to go if we just stay where we are right now***".

As important as it is to motivate patients to adhere to difficult treatment plans, it is also important to celebrate their progress. ***P6*** (the behavioral psychologist) and ***P13*** imagined using "what-if" explanation visualizations to "*...reinforce, to encourage them that they're improving. Keep taking your meds...*" (***P13***). For example, ***P20*** imagined seeing patients that were doing really well and showing them the partial dependence bars' "*...trends going the other way and the colors moving more towards green [as] something they can celebrate*" instead of "*...telling them a whole bunch of mumbo jumbo about echoes and so forth*". ***P25*** similarly felt like this feature could be used by physicians to "*...help them [patients] be able to gauge how well they're doing and* ***reinforce positive behaviors***".

6.2.2 Engage patients in treatment planning. Traditionally, physicians would converse with patients without using visual aids when discussing treatment plans. However, integrating visual and data-driven tools into these discussions has the potential to better engage patients in the decision-making process [65, 66]. One participant from Phase 1 Round 1 and two from Round 2 mentioned this as a potential use. ***P6***, the behavioral psychologist, imagined how they could use the "what-if" explanations component "'*...to help people figure out ways to be more physically active*". ***P8*** further suggests how the what-if explanation design could be used to "*...shape a discussion about treatment...we can dive into the details of how this would come about. And then this* ***may help explain why I'm not likely to add more medication onto your regimen at this time...***". Three of the seven physicians from Phase 2 validated that the modified designs for the "what-if" explanation functionality can indeed empower them to engage patients in the treatment planning process. For example, ***P20*** said "***We need ways to include patients in their own decision-making using data. And I think that this does that very, very well.***" ***P19*** imagined how if they could "*...pull this up in the exam room, this is more information to show the patient or talk to the patient about how, what therapies we feel, at the current time are beneficial, unnecessary*". ***P21*** imagined how this could aid in difficult conversations: "*I often run into like, you know, patients not doing well on like oral therapy. And we're having the conversation about this. I think this could help there...*".

6.2.3 Help the physicians gain patients' trust. Physicians need to build a strong relationship with their patients so that the patients trust and value their impressions and treatment suggestions. P15 felt like the "what-if" explanations component "***...gives a visual justification***" to their treatment decisions. However, ***P20*** brought up how it can be difficult initially to engage with new patients because "*They don't know my experience level*", but having "*...concrete examples of where we are and where we would like to be...could engender confidence so early on in a relationship...so that they don't think I'm just pulling this stuff out of nowhere.*"

6.3 XAI Uses in other parts of the workflow

While physicians primarily calculate risk scores during the diagnosis and treatment planning phases of their workflow, there are instances where calculating a risk score during the review and monitoring phases is necessary or beneficial. Only four participants across all phases mentioned using the "what-if" explanations component to support the education of physicians-in-training during the review phase, and nine participants mentioned emerging uses of the "what-if" explanations component during the monitoring phase.

6.3.1 Educating Physicians in-training. Medicine strongly depends on its apprenticeship model throughout education and training. Four of the 25 participants across the two phases mentioned using the "what-if" explanations component to help them educate and train their residents and fellows. This can occur when physicians review a patient's case with their interns and residents. For example, ***P2*** imagined how they could use this "*...for teaching and sort of education purposes, modifying the variables and...looking to see how that changes the risk*". ***P13*** similarly imagined a scenario where they'd, "*...play with it when I teach others about risk stratification. And this is a good teaching tool even outside of clinical care*". Beyond using it to help train others about PAH risk score calculations, ***P21*** imagined using the "what-if" explanations while teaching their residents how to select and modify treatment plans:

> "*[I would] take them through like, okay, this is the risk score right now before we're even starting therapy. We're gonna start X, Y plus Z. This is what we're hoping for when we see them in the clinic in 2 months. And then I'll do that What-if tool where you could see this is where we realistically could see that in a few months...*".

6.3.2 Reflection on Disease Management. An important process of PAH care is the monitoring phase, where physicians reflect on their patients' treatment plans and decide if they need to make any changes. Three participants from Phase 1 Round 2 and four of the seven physicians from Phase 2 saw the "what-if" explanations component's potential to encourage and support reflection on disease management. For example, ***P22*** said while they are reviewing a patient's chart, the CDSS could bring to their attention that, "*...maybe my gestalt is not as good as I thought...maybe I need to get more aggressive with this patient's treatment*". ***P19*** said:

> "*I don't always have the luxury of time when I'm in the exam room, and sometimes data is missing, right? If there's an echo done outside the institution, that may take time to get that report back. Or maybe there's a functional class adjustment...So, you know, during my admin time, I could potentially see using this calculator when I get additional data*".

6.3.3 Discussions with Care Staff. Physicians frequently interact with clinical staff, who also interact with the patient. Three participants out of the 25 imagined how the "what-if" explanations component could aid in discussions with other stakeholders. For example, ***P2*** brought up how, with improvement, they could use the XAI feature to present to their colleagues how they can make a patient's condition better: "*...we all work with our palliative care colleagues a lot and to be able to graphically show our palliative care colleagues like this patient is getting sicker, and I can make them better...*". ***P25*** confirms that the new designs can enable such an application, imagining that "*...it would probably be good for them [nurses] to see how their [patients] risk is changed and why it's changed...*". ***P20*** said they could imagine how this could be valuable in situations when they talk with their nurses or get a call from a referring provider about their patient's condition.

6.4 Perspectives on Trusting & Relying on the AI-CDSS

As previous research has already uncovered, calibrating clinicians' trust in AI is not as simple as just showing an explanation to validate a prediction [8, 84]. While our study did not have trust and reliance as a core research goal, we still probed the physicians during Phase 2 for their perspectives on trusting and relying on this AI-CDSS for diagnosis and treatment planning. While we did not get noteworthy reactions from five of the seven physicians, two physicians had very interesting perspectives. ***P21*** based their perceived trust on the fact that the tool was in line with what they are used to doing. Aside from ***P20*** mentioning that trust for them comes over time, they also mentioned how they are basing their trust on the AI-CDSS due to the fact that they knew and admired some of the physicians who developed the model that was in the AI-CDSS:

> "*trust for me comes over time...this is the first time I'm playing around with this, so I don't know If I can say I trust it. It seems to make a lot of sense, and I know the people behind the creation of this...so I have maybe an insight that many people don't...but I know [Physician 1] and [Physician 2] well, and you know I've admired their work. So, yeah, there's trust based on kind of me knowing who all is behind this...*"

7 Discussion & Limitations

We use iterative prototyping to co-design XAI capabilities that align with clinicians' collaborative, communicative, and exploratory needs. We implement those designs in a high-fidelity prototype and use it to probe for emerging uses of XAI while simultaneously identifying designs that best enable those emerging uses throughout the evolving Pulmonary Arterial Hypertension (PAH) care journey of a single patient. Our findings demonstrate that by combining interactive and dynamic partial dependence bar plots with "what-if" explanation generation and exploration capabilities, the uses of XAI in an AI-CDSS can extend well beyond fostering trust, calibrating reliance, enabling algorithmic recourse, and supporting model understanding. Throughout the PAH care journey, PAH clinicians envisioned how XAI can enhance clinical conversations, facilitate planning and reflective practices, and support educational efforts. We discuss the implications of designing XAI for emerging uses in clinical decision-making by drawing from theories and literature on sensemaking [71, 77], grounding [17], clinical reasoning [68], and design [22]. We further situate our findings within related work to ground these implications, and conclude by outlining the limitations of our study design and suggesting future directions to explore.

7.1 Designing Process-Oriented XAI to Support Emerging Uses of XAI in Clinical Decision-Making

While many empirical studies tend to pit two XAI methods against one another to find which is the "best" to use (*e.g.*, [34, 95]), our iterative prototyping led us to design for emerging uses of XAI without limits. With our design, we capture how PAH clinicians felt subgroup insights were confusing and how feature importance visualizations integrated into "what-if" explorations with a separate scenario-comparison feature could be used to collaboratively explore treatment options with patients, identify and reflect on treatment plans, communicate with other stakeholders, and support health education of trainees, nurses, and patients. These findings can result from designing process-oriented [102] XAI rather than designing XAI that focuses on explaining "why" an AI behaved a certain way (output-oriented). We discuss these emerging uses in detail and their implications across a series of dimensions below: communication grounding (Section 7.1.1); planning, reflecting, and sensemaking (Section 7.1.2), and educational interactions (Section 7.1.3). We further connect the emerging uses of XAI to relevant theories and present example outcomes of using XAI in different ways in Table 5.

7.1.1 ***Designing XAI to Facilitate Grounding in Conversations.*** Our research highlights how clinicians imagined "what-if" explanations' role in their clinical conversations: providing them additional insights and visuals to better communicate with their patients, which is fundamental to patient care [1, 25] and a key determinant of a physician's credibility [6, 81]. We also observed how our "what-if" generation and exploration design in our high-fidelity prototype can give physicians the freedom to generate visual aids to use in celebratory conversations, which is something that Szymanski et al. [92] found to be valued by patients. Beyond celebratory conversations, we observed clinicians' interest in using explanations as a way to motivate patients, corroborating Kim et al. [42]'s findings. The behavioral psychologist who participated in our iterative prototyping sessions approached these designs as a way to connect with the patient, helping them understand on another level why their physician is advising them to pursue a different

Emerging Use	Theory	Example Outcome	Section #
Support multi-stakeholder collaboration and coordination	Grounding [17], Design [22]	Stakeholders reach shared understanding and alignment on next steps through grounding processes	6.2.2, 6.3.3
Enable hypothesis exploration, sensemaking, and reflection	Sensemaking [71, 77], Design [22], Reasoning [68]	User reasons forward by exploring scenarios, identifying patterns and reflecting on decisions rather than trying to judge the correctness of a recommendation by reasoning backward	6.1.2, 6.3.2
Build trust and motivation between stakeholders	Grounding [17]	Expert builds credibility; stakeholders remain engaged and committed	6.2.1, 6.2.3
Bridge knowledge gaps across stakeholders	Reasoning [68]	Stakeholders understand complex information appropriate to their expertise level	6.1.1, 6.3.1

Table 5. Four emerging uses of XAI that expand traditional uses, grounded in theory: multi-stakeholder collaboration, hypothesis exploration and reflection, trust and motivation building, and bridging stakeholder knowledge gaps.

treatment route, such as rehabilitation therapy. This is similar to what we heard from physicians in phase 2. The nurse practitioner is underrepresented in our discussion, as their comments primarily focused on interface-level design feedback during iterative prototyping rather than on emerging uses. Physicians also expressed interest in using XAI to communicate with other stakeholders, such as nurses and palliative care staff, hinting at XAI's critical role in providing physicians with powerful visuals to better facilitate grounding. Ultimately, these emerging uses, overlapping with findings from previous works [5, 19, 35], represent broader uses that either support multi-stakeholder collaboration and coordination or build trust and motivation among stakeholders (Table 5).

As argued by Miller [57], XAI methods should be grounded in social and cognitive psychology. Rohlfing et al. [76] expand upon this with a conceptual framework for explanations as an interactive social process that involves an explainer and explainee. Recent studies have similarly emphasized the importance of drawing from communication theories to design human-centered XAI [40, 58]. Through the lens of social science and the concept of grounding in communication [17], AI explanations can be seen as a communication guide to help all parties reach a point of mutual understanding or common ground regarding an outcome or concept. In this sense, our findings suggest how our XAI designs can support these grounding processes within clinical decision-making. Corti et al. [19] similarly conclude that XAI may have value in multi-stakeholder conversations across clinical workflows. In contrast to our interactive design to enable this grounding, Spillner et al. [86] propose an XAI method that focuses on explaining the points of uncommon ground between the human and AI. However, the intentional design of XAI to better facilitate grounding introduces interesting design challenges, as it will be necessary to align XAI with human explanation strategies [57, 60].

7.1.2 ***Designing Interactions to Facilitate Planning, Reflection, & Sensemaking.*** Another use that we captured throughout our prototyping and interview sessions was being able to explore hypotheses, engage in sensemaking, and reflect on treatment decisions (Table 5). We observed how our XAI designs can encourage reflective practices by enabling clinicians to revisit and evaluate their initial line of reasoning through interacting with the scenarios. Similarly, physicians' interactions with the XAI component suggest how it can enable hypothesis exploration and sensemaking behaviors [71, 77]. Related work has also acknowledged that XAI has properties that can help facilitate the sensemaking process users undergo when understanding AI systems [27]. A series of studies have found how XAI can be used to help people plan tasks by providing actionable insights [48], enable hypothesis generation during decision-making [101], and support hypothesis generation and testing during the model development lifecycle [10]. Zhang and Reicherts [102] comment that AI systems can enable these emerging uses when designed with a process-oriented approach or by designing for the system to be appropriated by its users [22]. Expanding from this line of work, we suggest that our process-oriented XAI designs may be able to help clinicians reason forward by enabling them to consider the decision context, rather than trying to make sense of why an AI recommendation fits a given input. Alternatively, Naiseh [64] discuss how the concept of social XAI can encourage users to reflect on their own decisions, potentially reducing biases within human–AI teams.

Ultimately, designing interactions alongside XAI insights that facilitate planning and reflection should involve holistically understanding the information needs and stakeholder conversations that occur before and after a diagnosis is made. For example, our "what-if" explanation design is tied to risk prediction and simultaneously displays recommended treatments from the treatment guidelines [32]. By doing this, the physicians recognized the value our XAI designs could bring during their administrative time, as they review patients' conditions and treatments. These findings underscore the value of explanations that not only support immediate decision-making but also enhance ongoing reflection and reasoning within clinical workflows.

7.1.3 ***Designing Mechanisms to Enable Clinical Educational Interactions.*** The clinicians in our study highlighted the value that our interactive "what-if" explanations design could bring to educating others about PAH. For example, they recognized how they could leverage the XAI visuals while teaching trainees about treatment impact, nurses about a patient's PAH condition, and patients about the disease itself. This captures the broader use of XAI for bridging the knowledge gaps between stakeholders, whether that be physician and nurse, physician and trainee, or physician and patient (Table 5). As we observed from the feedback and interactions with the design probe, as well as seen in prior literature, XAI has the potential to support skill and knowledge development [48, 64, 88]. On a related thread, Kim et al. [43] similarly discovered end-users' interest in using XAI to help them improve their skills. However, it is important to understand that not all XAI designs may successfully support educational interactions. XAI designs that intend to support the education of medical trainees will need to adapt to pedagogies used within medicine, such as learning through differentials and variations [68]. For example, designs could include mechanisms that allow users to easily compare multiple scenarios to enable clinical reasoning, as shown in our designs. By designing XAI systems that integrate learning and educational capabilities tailored to the reasoning skills needed for decision-making, we can create tools that not only support clinicians in their current workflows but also contribute to the apprenticeship model seen throughout clinical contexts.

7.2 XAI for AI-CDSS Ethical Implications & Considerations

As we observed physicians simulate diagnosing and treating a hypothetical patient with our high-fidelity prototype, we observed how the participants were overlooking the accuracy and confidence of the underlying model. This behavior made it apparent how physicians can fall to confirmation bias or be misled by imperfect XAI systems, especially if they are anchoring on an incorrect piece of information or lacking enough expertise to spot inaccuracies. Therefore, it is imperative to focus on mechanisms and designs that minimize the misuse of "what-if" explanations in AI-CDSS. Giving the end-user so much agency over crafting the explanation to use in clinical conversations can be consequential, as explanations have been known to disagree with each other [45] and present misleading or confusing information [39, 61]. This could misinform junior physicians about the severity of their patients' conditions or lead them to misjudge the impact of treatments on their patients' conditions. Similarly, when being used for educational purposes, unidentified

XAI inaccuracies can unintentionally de-skill users [87]. When using XAI in conversations with other stakeholders to justify decisions or identify potential next steps, XAI designs must consider the potential inaccuracies and disagreements that may arise. As designers of XAI for high-stakes contexts where users appropriate XAI beyond calibrating trust and reliance, it is becoming an important challenge for the field to not only design for appropriately trusting and relying on the AI predictions but also on the information and insights being derived by the XAI and the context where those insights are shared.

7.3 Limitations & Future Work

We initially narrowed our recruitment to clinicians for this study to explore the clinical validity of the designs and emerging uses, as well as to ensure that the designs were rooted in the clinical workflow. Although our iterative prototyping method actively engaged the PAH clinicians in understanding how XAI can bring value to their workflow, our study raises new questions that need to be addressed in future research. Specifically, we address the lack of including patients and caregivers in our interviews.

7.3.1 Lack of Patient Perspective. Our study lacks insight into the evolving needs for and uses of explanations from the patients' and caregivers' perspective. Given that many emerging uses of XAI were situated in contexts involving patients, it would be ideal to continue future design iterations with patients and physicians simultaneously. As such, our ongoing current research is focused on the patients' and caregivers' perspectives to validate the clinicians' perspectives and design a patient-facing tool. Patients' and caregivers' perspectives can help validate the potential of XAI to augment patient-physician communication. It can also help uncover emerging uses of XAI in the context of at-home disease management. For example, Wong et al. [98] found that exploring counterfactuals helps patients better understand their diagnosis, evaluate trade-offs between different treatment options, and actively participate in discussions. Conducting studies with patients can also help identify contexts in which patients are uncomfortable using XAI for an emerging use. For example, Kim et al. [41] found that patients may only value such functionality in low-stakes situations as they fear that they would miss testing the right scenarios in high-risk situations. Carmichael [13] also allude to different perceptions of explanations as the associated risk of the decision changes. Most importantly, future work with patients on XAI designs in AI-CDSS can help ensure that the designs are not "*increasing the distance between clinicians and the patient*" [75]. Such a metric could be used in future studies to validate the designs of XAI in AI-CDSS.

7.3.2 Focus on Physicians' Perspectives. A majority of the perspectives we capture for design feedback are from physicians. As such, the emerging themes that we identified reflect a PAH physician's workflow and perspectives. Not only does this limit the generalizability of our findings, but it also limits the potential additional XAI designs and emerging uses that we could have identified. We encourage future works to explore how designs may differ depending on the context and target stakeholder.

7.3.3 Lack of Comparison to other XAI Techniques. Another limitation of this work is the lack of a larger qualitative or quantitative comparison to existing XAI techniques. This would further provide evidence that existing feature attribution methods may limit how XAI is used in emerging ways. Future work could consider running a within- or between-subjects study that compares an existing XAI technique, such as SHAP, to a more process-oriented, interactive technique, such as the one we presented. Such a study design would further validate the designs that enable emerging uses and identify the contexts in which they are most effective. Beyond lacking comparison to other XAI techniques, we do not explore the implications of imperfect XAI. It is necessary to design safeguards to ensure users apply insights derived from XAI appropriately, especially when XAI presents inaccurate or misleading information. Future work is encouraged to further explore XAI designs that investigate safeguarding users from inappropriate reliance on derived insights.

8 Conclusion

Inspired by the numerous potential emerging uses of XAI within and outside of clinical decision-making, our work uses iterative prototyping and the use of a design probe to identify designs that can enable emerging uses. Across the two phases, with feedback and insights from 25 PAH clinicians representing 23 different institutions, we present detailed scenarios on how PAH clinicians would utilize interactive XAI visuals that combine information about feature importance and facilitate interactive, forward-looking explorations. Our study highlights numerous emerging uses of XAI in clinical decision-making: from collaboratively exploring treatment options and identifying/reflecting on treatment plans to communicating with stakeholders and supporting health education. Despite still knowing little about how clinicians will utilize XAI in human–human–AI decision-making, we hope this work encourages discussions about new designs for XAI systems and evaluation metrics for XAI's impact on human–human–AI collaborations, extending beyond trust, reliance, understanding, and algorithmic recourse.

Acknowledgments

Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL164906.

Use of Generative AI Disclosure. ChatGPT, Copilot, Writeful, and Grammarly were utilized throughout the sections of this paper to improve writing, spelling, grammar, and the phrasing and flow of information, or going from detailed outlines to paragraph format. Any generated content used was reviewed and further modified by the authors to accurately represent the work, ensure originality, and align with their writing style.

References

[1] Temi A Adekunle, Joy M Knowles, Sarah V Hantzmon, Maya N DasGupta, Kathryn I Pollak, and Sarah E Gaither. 2023. A qualitative analysis of trust and distrust within patient-clinician interactions. *PEC innovation* 3 (2023), 100187.

[2] Raymond L Benza, Mardi Gomberg-Maitland, C Greg Elliott, Harrison W Farber, Aimee J Foreman, Adaani E Frost, Michael D McGoon, David J Pasta, Mona Selej, Charles D Burger, et al. 2019. Predicting survival in patients with pulmonary arterial hypertension: the REVEAL risk score calculator 2.0 and comparison with ESC/ERS-based risk assessment strategies. *Chest* 156, 2 (2019), 323–337.

[3] Isacco Beretta, Eleonora Cappuccio, and Marta Marchiori Manerba. 2023. User-Driven Counterfactual Generator: A Human Centered Exploration.. In *xAI (Late-breaking Work, Demos, Doctoral Consortium)*. 83–88.

[4] 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*. 1–21. https://dl.acm.org/doi/10.1145/3544548.3581314

[5] Aditya Bhattacharya, Jeroen Ooge, Gregor Stiglic, and Katrien Verbert. 2023. Directive Explanations for Monitoring the Risk of Diabetes Onset: Introducing Directive Data-Centric Explanations and Combinations to Support What-If Explorations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 204–219.

[6] Johanna Birkhäuer, Jens Gaab, Joe Kossowsky, Sebastian Hasler, Peter Krummenacher, Christoph Werner, and Heike Gerger. 2017. Trust in the health care professional and health outcome: A meta-analysis. *PloS one* 12, 2 (2017), e0170988.

[7] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.

[8] Eleanor R Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuścińska, Suzanne Currie, J Marc Overhage, Erika S Poole, and Jofish Kaye. 2023. Healthcare AI treatment decision support: Design principles to enhance clinician adoption and trust. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–19.

[9] Ángel Alexander Cabrera, Adam Perer, and Jason I Hong. 2023. Improving human-AI collaboration with descriptions of AI behavior. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–21.

[10] Ángel Alexander Cabrera, Marco Tulio Ribeiro, Bongshin Lee, Robert Deline, Adam Perer, and Steven M Drucker. 2023. What did my AI learn? How data scientists make sense of model behavior. *ACM Transactions on Computer-Human Interaction* 30, 1 (2023), 1–27.

[11] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. " Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proceedings of the ACM on Human-computer Interaction* 3, CSCW (2019), 1–24.

[12] Shiye Cao and Chien-Ming Huang. 2022. Understanding user reliance on AI in assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.

[13] Josie Carmichael. 2024. Translating Human-Centred Artificial Intelligence for Clinical Decision Support Systems into Practice: A Medical Retina Case Study.. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–5.

[14] Federico Maria Cau and Lucio Davide Spano. 2026. Exploring the impact of explainable AI and cognitive capabilities on users' decisions. *User Modeling and User-Adapted Interaction* 36, 1 (2026), 3. https://link.springer.com/article/10.1007/s11257-025-09438-0

[15] Michael Chromik and Andreas Butz. 2021. Human-XAI interaction: a review and design principles for explanation user interfaces. In *IFIP Conference on Human-Computer Interaction*. Springer, 619–640.

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

[17] HH Clark. 1991. Grounding in communication. *Perspectives on socially shared cognition/American Psychological Association* (1991).

[18] United Therapeutics Corporation. 2024. *PAH Initiative Risk Assessment Calculators*. Retrieved 12-Sep-2024 from https://www.pahinitiative.com/hcp/risk-assessment/calculators

[19] 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*. 1–21.

[20] Karla Cruz Morel, Vinicio De Jesus Perez, and Arunabh Talwar. 2020. Gender and Race Disparities in Pulmonary Hypertension Diagnosis and Treatment. *Pulmonary Hypertension: Controversial and Emerging Topics* (2020), 195–202.

[21] Shipi Dhanorkar, Christine T Wolf, Kun Qian, Anbang Xu, Lucian Popa, and Yunyao Li. 2021. Who needs to know what, when?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 1591–1602.

[22] Alan Dix. 2007. Designing for appropriation. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK*. BCS Learning & Development.

[23] Yuhan Du, Anna Markella Antoniadi, Catherine McNestry, Fionnuala M McAuliffe, and Catherine Mooney. 2022. The role of xai in advice-taking from a clinical decision support system: A comparative user study of feature contribution-based and example-based explanations. *Applied Sciences* 12, 20 (2022), 10323.

[24] Sven Eckhardt, Niklas Kühl, Mateusz Dolata, and Gerhard Schwabe. 2025. A survey of AI reliance. *Comput. Surveys* 58, 6 (2025), 1–37.

[25] Pouyan Esmaeilzadeh, Tala Mirzaei, and Spurthy Dharanikota. 2021. Patients' perceptions toward human–artificial intelligence interaction in health care: experimental study. *Journal of medical Internet research* 23, 11 (2021), e25856.

[26] Timo Freiesleben and Gunnar König. 2023. Dear XAI community, we need to talk! Fundamental misconceptions in current XAI research. In *World Conference on Explainable Artificial Intelligence*. Springer, 48–65.

[27] Elisa Gagnon, Anouk de Regt, and Liette LaPointe. 2025. Clarity in Complexity: Advancing AI Explainability through Sensemaking. In *Proceedings of the 58th Hawaii International Conference on System Sciences, HICSS 2025*. 1400–1409.

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

[29] Grace Guo, Lifu Deng, Animesh Tandon, Alex Endert, and Bum Chul Kwon. 2024. Mimicri: Towards domain-centered counterfactual explanations of cardiovascular image classification models. (2024), 1861–1874.

[30] Kazuaki Hanawa, Sho Yokoi, Satoshi Hara, and Kentaro Inui. 2020. Evaluation of similarity-based explanations. *arXiv preprint arXiv:2006.04528* (2020).

[31] Marius M Hoeper, Christine Pausch, Karen M Olsson, Doerte Huscher, David Pittrow, Ekkehard Grünig, Gerd Staehler, Carmine Dario Vizza, Henning Gall, Oliver Distler, et al. 2022. COMPERA 2.0: a refined four-stratum risk assessment model for pulmonary arterial hypertension. *European Respiratory Journal* 60, 1 (2022).

[32] Marc Humbert, Gabor Kovacs, Marius M Hoeper, Roberto Badagliacca, Rolf MF Berger, Margarita Brida, Jørn Carlsen, Andrew JS Coats, Pilar Escribano-Subias, Pisana Ferrari, et al. 2022. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: Developed by the task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS). Endorsed by the International Society for Heart and Lung Transplantation (ISHLT) and the European Reference Network on rare respiratory diseases (ERN-LUNG). *European heart journal* 43, 38 (2022), 3618–3731.

[33] Christina Humer, Andreas Hinterreiter, Benedikt Leichtmann, Martina Mara, and Marc Streit. 2022. Comparing effects of attribution-based, example-based, and feature-based explanation methods on ai-assisted decision-making. *OSF Preprints* 2 (2022). https://www.researchgate.net/publication/364581372_Comparing_Effects_of_Attribution-based_Example-based_and_Feature-

based_Explanation_Methods_on_AI-Assisted_Decision-Making

[34] Christina Humer, Andreas Hinterreiter, Benedikt Leichtmann, Martina Mara, and Marc Streit. 2024. Reassuring, misleading, debunking: Comparing effects of XAI methods on human decisions. *ACM Transactions on Interactive Intelligent Systems* 14, 3 (2024), 1–36.

[35] Weina Jin, Jianyu Fan, Diane Gromala, Philippe Pasquier, and Ghassan Hamarneh. 2021. EUCA: The end-user-centered explainable AI framework. *arXiv preprint arXiv:2102.02437* (2021).

[36] Manreet K Kanwar, Mardi Gomberg-Maitland, Marius Hoeper, Christine Pausch, David Pittrow, Geoff Strange, James J Anderson, Carol Zhao, Jacqueline V Scott, Marek J Druzdzel, et al. 2020. Risk stratification in pulmonary arterial hypertension using Bayesian analysis. *European Respiratory Journal* 56, 2 (2020).

[37] Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. 2022. A survey of algorithmic recourse: contrastive explanations and consequential recommendations. *Comput. Surveys* 55, 5 (2022), 1–29.

[38] Harmanpreet Kaur, Harsha Nori, Samuel Jenkins, Rich Caruana, Hanna Wallach, and Jennifer Wortman Vaughan. 2020. Interpreting Interpretability: Understanding Data Scientists’ Use of Interpretability Tools for Machine Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14. doi:10.1145/3313831.3376219

[39] Maxime Kayser, Bayar Menzat, Cornelius Emde, Bogdan Bercean, Alex Novak, Abdala Espinosa, Bartlomiej W Papiez, Susanne Gaube, Thomas Lukasiewicz, and Oana-Maria Camburu. 2024. Fool Me Once? Contrasting Textual and Visual Explanations in a Clinical Decision-Support Setting. *arXiv preprint arXiv:2410.12284* (2024).

[40] Bernard Keenan and Kacper Sokol. 2023. Mind the gap! Bridging explainable artificial intelligence and human understanding with Luhmann’s Functional Theory of Communication. *arXiv preprint arXiv:2302.03460* (2023).

[41] Dajung Kim, Niko Vegt, Valentijn Visch, and Marina Bos-De Vos. 2024. How Much Decision Power Should (A) I Have?: Investigating Patients’ Preferences Towards AI Autonomy in Healthcare Decision Making. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[42] Minjung Kim, Saebyeol Kim, Jinwoo Kim, Tae-Jin Song, and Yuyoung Kim. 2024. Do stakeholder needs differ?-Designing stakeholder-tailored Explainable Artificial Intelligence (XAI) interfaces. *International Journal of Human-Computer Studies* 181 (2024), 103160.

[43] Sunnie SY 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*. 1–17.

[44] Josua Krause, Adam Perer, and Kenney Ng. 2016. Interacting with Predictions: Visual Inspection of Black-Box Machine Learning Models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI ’16)*. Association for Computing Machinery, New York, NY, USA, 5686–5697. doi:10.1145/2858036.2858529

[45] Satyapriya Krishna, Tessa Han, Alex Gu, Steven Wu, Shahin Jabbari, and Himabindu Lakkaraju. 2022. The disagreement problem in explainable machine learning: A practitioner’s perspective. *arXiv preprint arXiv:2202.01602* (2022).

[46] I Elizabeth Kumar, Suresh Venkatasubramanian, Carlos Scheidegger, and Sorelle Friedler. 2020. Problems with Shapley-value-based explanations as feature importance measures. In *International conference on machine learning*. PMLR, 5491–5500.

[47] Anisio Lacerda, Claudio Almeida, Leonardo Ferreira, Adriano Pereira, Gisele L Pappa, Wagner Meira, Debora Miranda, Marco A Romano-Silva, and Leandro Malloy Diniz. 2023. Algorithmic Recourse in Mental Healthcare. In *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[48] Christine P Lee, Min Kyung Lee, and Bilge Mutlu. 2024. The AI-DEC: A Card-based Design Method for User-centered AI Explanations. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 1010–1028.

[49] Min Hun Lee and Chong Jun Chew. 2023. Understanding the effect of counterfactual explanations on trust and reliance on ai for human-ai collaborative clinical decision making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–22.

[50] Min Hun Lee, Silvana Xin Yi Choo, and Shamala D/O Thilarajah. 2024. Improving Health Professionals’ Onboarding with AI and XAI for Trustworthy Human-AI Collaborative Decision Making. arXiv:2405.16424 [cs.HC] https://arxiv.org/abs/2405.16424

[51] 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. doi:10.1145/3313831.3376590

[52] Zachary C Lipton. 2018. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* 16, 3 (2018), 31–57.

[53] Han Liu, Vivian Lai, and Chenhao Tan. 2021. Understanding the Effect of Out-of-distribution Examples and Interactive Explanations on Human-AI Decision Making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 408 (Oct. 2021), 45 pages. doi:10.1145/3479552

[54] Jens Ludwig and Sendhil Mullainathan. 2024. Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics* 139, 2 (2024), 751–827.

[55] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[56] Scott M Lundberg, Bala Nair, Monica S Vavilala, Mayumi Horibe, Michael J Eisses, Trevor Adams, David E Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature biomedical engineering* 2, 10 (2018), 749–760.

[57] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.

[58] Dimitry Mindlin, Fabian Beer, Leonie Nora Sieger, Stefan Heindorf, Elena Esposito, Axel-Cyrille Ngonga Ngomo, and Philipp Cimiano. 2025. Beyond one-shot explanations: a systematic literature review of dialogue-based xAI approaches. *Artificial Intelligence Review* 58, 3 (2025), 81.

[59] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–45.

[60] Katelyn Morrison, Donghoon Shin, Kenneth Holstein, and Adam Perer. 2023. Evaluating the impact of human explanation strategies on human-AI visual decision-making. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–37.

[61] Katelyn Morrison, Philipp Spitzer, Violet Turri, Michelle Feng, Niklas Kühl, and Adam Perer. 2024. The Impact of Imperfect XAI on Human-AI Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–39.

[62] Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. 2020. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 607–617.

[63] Hussein Mozannar, Arvind Satyanarayan, and David Sontag. 2022. Teaching humans when to defer to a classifier via exemplars. In *Proceedings of the aaai conference on artificial intelligence*, Vol. 36. 5323–5331.

[64] Mohammad Naiseh. 2024. Social eXplainable AI (Social XAI): Towards Expanding the Social Benefits of XAI. In *The Impact of Artificial Intelligence on Societies: Understanding Attitude Formation Towards AI*. Springer, 169–178.

[65] Catherine Nasrallah, Cherish Wilson, Alicia Hamblin, Christine Hariz, Cammie Young, Jing Li, Jinoos Yazdany, and Gabriela Schmajuk. 2024. Patient perceptions of an electronic-health-record-based rheumatoid arthritis outcomes dashboard: a mixed-methods study. *BMC Medical Informatics and Decision Making* 24, 1 (2024), 1–15.

[66] Catherine Nasrallah, Cherish Wilson, Alicia Hamblin, Cammie Young, Lindsay Jacobsohn, Mary C Nakamura, Andrew Gross, Mehrdad Matloubian, Judith Ashouri, Jinoos Yazdany, et al. 2024. Using the technology acceptance model to assess clinician perceptions and experiences with a rheumatoid arthritis outcomes dashboard: qualitative study. *BMC Medical Informatics and Decision Making* 24, 1 (2024), 140.

[67] Harsha Nori, Samuel Jenkins, Paul Koch, and Rich Caruana. 2019. Interpretml: A unified framework for machine learning interpretability. *arXiv preprint arXiv:1909.09223* (2019). https://arxiv.org/abs/1909.09223

[68] Geoffrey Norman. 2005. Research in clinical reasoning: past history and current trends. *Medical education* 39, 4 (2005), 418–427.

[69] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. 2023. Co-design of human-centered, explainable AI for clinical decision support. *ACM Transactions on Interactive Intelligent Systems* 13, 4 (2023), 1–35.

[70] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–9.

[71] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.

[72] Nicoletta Prentzas, Antonis Kakas, and Constantinos S Pattichis. 2023. Explainable AI applications in the medical domain: A systematic review. *arXiv preprint arXiv:2308.05411* (2023).

[73] Marissa Radensky, Doug Downey, Kyle Lo, Zoran Popovic, and Daniel S Weld. 2022. Exploring the Role of Local and Global Explanations in Recommender Systems. 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 290, 7 pages. doi:10.1145/3491101.3519795

[74] 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. doi:10.1145/2939672.2939778

[75] Tabea E Röber, Rob Goedhart, and Sİ Birbil. 2024. Clinicians' Voice: Fundamental Considerations for XAI in Healthcare. *arXiv preprint arXiv:2411.04855* (2024).

[76] Katharina J Rohlfing, Philipp Cimiano, Ingrid Scharlau, Tobias Matzner, Heike M Buhl, Hendrik Buschmeier, Elena Esposito, Angela Grimminger, Barbara Hammer, Reinhold Häb-Umbach, et al. 2020. Explanation as a social practice: Toward a conceptual framework for the social design of AI systems. *IEEE Transactions on Cognitive and Developmental Systems* 13, 3 (2020), 717–728.

[77] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 269–276. doi:10.1145/169059.169209

[78] Tracy L. Sanders, Alexandra D. Kaplan, Keith MacArthur, William G. Volante, and P. A. Hancock. 2021. *Trust and Human Factors: Foundations of Trust in Automation*. Cambridge University Press, 77–98.

[79] Tjeerd AJ Schoonderwoerd, Wiard Jorritsma, Mark A Neerincx, and Karel Van Den Bosch. 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies* 154 (2021), 102684.

[80] Julian Senoner, Simon Schallmoser, Bernhard Kratzwald, Stefan Feuerriegel, and Torbjørn Netland. 2024. Explainable AI improves task performance in human-AI collaboration. *arXiv preprint arXiv:2406.08271* (2024).

[81] Samer H Sharkiya. 2023. Quality communication can improve patient-centred health outcomes among older patients: a rapid review. *BMC Health Services Research* 23, 1 (2023), 886.

[82] Harsh Sharmam, Rajeshwari Pradhan, Harsh Mogha, Shivpratap Singh, and Rupesh Kumar. 2025. Counterfactual AI in Healthcare: Enhancing Decision-Making and Outcome Prediction. In *2025 Seventh International Conference on Computational Intelligence and Communication Technologies (CCICT)*. IEEE, 608–615.

[83] Ronal Singh, Tim Miller, Liz Sonenberg, Eduardo Velloso, Frank Vetere, Piers Howe, and Paul Dourish. 2024. An Actionability Assessment Tool for Explainable AI. *arXiv preprint arXiv:2407.09516* (2024).

[84] 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*. 1–18.

[85] Venkatesh Sivaraman, Katelyn Morrison, Will Epperson, and Adam Perer. 2025. Over-Relying on Reliance: Towards Realistic Evaluations of AI-Based Clinical Decision Support. *arXiv preprint arXiv:2504.07423* (2025). Presented at CHI '25 Workshop on Envisioning the Future of Interactive Health.

[86] Laura Spillner, Nima Zargham, Mihai Pomarlan, Robert Porzel, and Rainer Malaka. 2025. Finding Uncommon Ground: A Human-Centered Model for Extrospective Explanations. *arXiv preprint arXiv:2507.21571* (2025).

[87] Philipp Spitzer, Joshua Holstein, Katelyn Morrison, Kenneth Holstein, Gerhard Satzger, and Niklas Kühl. 2024. Don't be Fooled: The Misinformation Effect of Explanations in Human-AI Collaboration. *arXiv preprint arXiv:2409.12809* (2024).

[88] Philipp Spitzer, Niklas Kühl, Marc Goutier, Manuel Kaschura, and Gerhard Satzger. 2024. Transferring Domain Knowledge with (X) AI-Based Learning Systems. *arXiv preprint arXiv:2406.01329* (2024).

[89] Parvathaneni Naga Srinivasu, N Sandhya, Rutvij H Jhaveri, and Roshani Raut. 2022. From blackbox to explainable AI in healthcare: existing tools and case studies. *Mobile Information Systems* 2022, 1 (2022), 8167821.

[90] Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. 2020. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10, 5 (2020), e1379.

[91] Qiyang Sun, Alican Akman, and Björn W Schuller. 2025. Explainable artificial intelligence for medical applications: A review. *ACM Transactions on Computing for Healthcare* 6, 2 (2025), 1–31.

[92] Maxwell Szymanski, Vero Vanden Abeele, and Katrien Verbert. 2024. Designing and Evaluating Explanations for a Predictive Health Dashboard: A User-Centred Case Study. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[93] Sohini Upadhyay, Himabindu Lakkaraju, and Krzysztof Z Gajos. 2025. Counterfactual Explanations May Not Be the Best Algorithmic Recourse Approach. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*. 446–462.

[94] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour* (2024), 1–11.

[95] Xinru Wang and Ming Yin. 2021. Are explanations helpful? a comparative study of the effects of explanations in ai-assisted decision-making. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 318–328.

[96] Zijie J. Wang, Jennifer Wortman Vaughan, Rich Caruana, and Duen Horng Chau. 2023. GAM Coach: Towards Interactive and User-Centered Algorithmic Recourse. 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 835, 20 pages. doi:10.1145/3544548.3580816

[97] William Van Woensel, Floriano Scioscia, Giuseppe Loseto, Oshani Seneviratne, Evan Patton, Samina Abidi, and Lalana Kagal. 2022. Explainable clinical decision support: towards patient-facing explanations for education and long-term behavior change. In *International Conference on Artificial Intelligence in Medicine*. Springer, 57–62.

[98] Melisa L Wong, Francesca M Nicosia, Alexander K Smith, Louise C Walter, Vivian Lam, Harvey Jay Cohen, Kah Poh Loh, Supriya G Mohile, Carling J Ursem, and Margaret L Schwarze. 2022. "You have to be sure that the patient has the full picture": Adaptation of the Best Case/Worst Case communication tool for geriatric oncology. *Journal of geriatric oncology* 13, 5 (2022), 606–613.

[99] 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*. 1–13.

[100] Jiehuang Zhang and Han Yu. 2022. A methodological framework for facilitating explainable AI design. In *International Conference on Human-Computer Interaction*. Springer, 437–446.

[101] Shao Zhang, Jianing Yu, Xuhai Xu, Changchang Yin, Yuxuan Lu, Bingsheng Yao, Melanie Tory, Lace M Padilla, Jeffrey Caterino, Ping Zhang, et al. 2024. Rethinking human-ai collaboration in complex medical decision making: A case study in sepsis diagnosis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[102] Zelun Tony Zhang and Leon Reicherts. 2025. Augmenting human cognition with generative ai: Lessons from ai-assisted decision-making. *arXiv preprint arXiv:2504.03207* (2025). https://arxiv.org/abs/2504.03207 Presented at CHI '25 Workshop on Tools for Thought.

Received June 2025; revised December 2025; accepted April 2026