Seamful XAI: Operationalizing Seamful Design in Explainable AI

UPOL EHSAN, Georgia Institute of Technology, USA
Q. VERA LIAO, Microsoft Research, Canada
SAMIR PASSI, Microsoft, USA
MARK O. RIEDL, Georgia Institute of Technology, USA
HAL DAUMÉ III, University of Maryland & Microsoft Research, USA

Mistakes in AI systems are inevitable, arising from both technical limitations and sociotechnical gaps. While black-boxing AI systems can make the user experience seamless, hiding the seams risks disempowering users to mitigate fallout from AI mistakes. Instead of hiding these AI imperfections, can we leverage them to help the user? While Explainable AI (XAI) has predominantly tackled algorithmic opaqueness, we propose that seamful design can foster AI explainability by revealing and leveraging sociotechnical and infrastructural mismatches. We introduce the concept of Seamful XAI by (1) conceptually transferring “seams” to the AI context and (2) developing a design process that helps stakeholders anticipate and design with seams. We explore this process with 43 AI practitioners and real end-users, using a scenario-based co-design activity informed by real-world use cases. We found that the Seamful XAI design process helped users foresee AI harms, identify underlying reasons (seams), locate them in the AI’s lifecycle, learn how to leverage seamful information to improve XAI and user agency. We share empirical insights, implications, and reflections on how this process can help practitioners anticipate and craft seams in AI, how seamfulness can improve explainability, empower end-users, and facilitate Responsible AI.

CCS Concepts: • Human-centered computing → Scenario-based design; Empirical studies in HCI; HCI theory, concepts and models; Collaborative and social computing theory, concepts and paradigms; • Computing methodologies → Artificial intelligence.

Additional Key Words and Phrases: Seamful design, explainable AI, responsible AI, human-AI interaction

ACM Reference Format:
Upol Ehsan, Q. Vera Liao, Samir Passi, Mark O. Riedl, and Hal Daumé III. 2024. Seamful XAI: Operationalizing Seamful Design in Explainable AI. Proc. ACM Hum.-Comput. Interact. 8, CSCW1, Article 119 (April 2024), 29 pages. https://doi.org/10.1145/3637396

1 INTRODUCTION

"We make it so easy to use that you won’t even have to think about it!" exclaimed the VP of a fin-tech company while proudly pitching the latest AI-powered tool. "When you use our product, it’s seamless! One click, that’s it. Plug and play! It’s frictionless!"—firsthand account of an informant in our study.

Mistakes in AI systems are inevitable. Handling the mistakes is harder when the system’s decision-making is hidden or black-boxed. Seamless design—an often unquestioned design virtue [12, 42]—can be a double-edged sword. On the positive side, it promotes simplicity and ease of use [36, 42].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2024 Copyright held by the owner/author(s).
ACM 2573-0142/2024/4-ART119
https://doi.org/10.1145/3637396
On the negative side, it abstracts and conceals important factors about the AI system that are crucial for explainability. These factors can be both inside AI’s black box—such as its training and algorithms—or outside—such as sociotechnical infrastructures like data and integration practices. Breakdowns in AI systems often occur when the assumptions we make in design and development do not hold true when they are deployed in the real-world. For example, an AI system can fail when it’s trained on data from North America but deployed in South Asia, especially when the end user is unaware of this infrastructural mismatch. Concealing these mismatches through a seamless ideal can lead to downstream harms for end-users, such as unquestioned AI acceptance. What can we do differently? How do we move beyond seamless AI? And what do we gain by doing so?

One strategy would be to critically question the seamless ideal and, instead of hiding an AI system’s seams, leverage them in technology design. Here, the concept of ‘seamful design’ [11, 12], a complement (not a rival) of seamless design [42], can help. With conceptual roots in Ubiquitous Computing (Ubicomp), seamful design deals with seams—“mismatches”, “gaps”, “cracks”, and “uncertainties” that arise when a technology is deployed in the real world [10, 75]. Instead of hiding seams or treating them as problematic, seamful design argues for strategically revealing (and concealing) seams to support user agency, re-configuration, and appropriation. For instance, a classic example is a “seamful map” revealing varying strength of WiFi coverage in a space. Knowing exactly where the Wi-Fi is weakest will allow a user to best use it. Thus, awareness of the blind spots and uncertainties in the technology can help users leverage the technology better.

While there has been commendable work on seamful design in Ubicomp [10–13] and increased calls for it in AI [23, 46], key questions remain unexplored: how can we define seams in AI? How may we find them before deployment? Where do they appear in the AI’s lifecycle? How may we use them to augment agency and explainability? We tackle these questions in this paper.

We introduce the notion of ‘Seamful Explainable AI’ by transferring seamless design into AI and exploring its practical intersection with the field of Explainable AI (XAI). We operationalize the notion of Seamful XAI by (1) conceptually transferring “seams” to the context of AI systems and (2) developing a design process that helps stakeholders identify and use the seams to enhance AI explainability and user agency. We chose to bridge seamful design and XAI because the two areas have a natural connection: both challenge the notion of black-boxing. In fact, much of XAI’s proverbial “opening of the black box” can be viewed as undoing the seamlessness of AI systems.

In the context of AI, we define and conceptualize seams as mismatches and cracks between assumptions made when designing and developing the AI system and the reality of its deployment context. The following example, inspired by real-world automated lending cases from our informants, illustrates how revealing seamful information can facilitate explainability and support user agency:

The loan application for Ahmed, a great customer, is rejected by the AI-powered system. Nothing about the AI’s feature-level explanation seems off to Nadia, the loan officer. Behind the scenes, there was a mismatch (seam): during the last system update in 2020, applicants could only have three active loans, but regulatory changes in 2021 allow up to five. Ahmed has four active loans, which triggered the denial. Had Nadia been aware of this mismatch (seam), she would be empowered to trigger a manual underwriting review.

The type of understanding afforded by seamful information exceeds what is offered by algorithmic explanations. In the example, Nadia’s awareness of the AI’s blind spot, where model updates lag organizational policy shifts, can help her understand the “why not” question—why the AI could not take certain things into consideration, instead of just the “why” question answered by the feature-level explanation. This understanding equips Nadia to seek actions and contest the AI.

Our conceptualization of Seamful XAI is informed by and adds to two aspects: the sociotechnical perspectives of Human-centered XAI (HCXAI) [22, 26, 51] and the proactive or anticipatory design.
stance in Responsible AI (RAI) practices [25, 35, 56, 58, 71]. In a nutshell, we operationalize seams as a design element that could facilitate end-user explainability of the sociotechnical system—the human-AI assemblage [21]—rather than just the technical system or algorithmic black-box. In contrast to prior work in Ubicomp that addresses seams after they appear in deployment [10, 12, 61], our focus is proactive and holistic (pre- and post-deployment). While the end goal of many RAI approaches is anticipating harms, the end goal of our process extends the current paradigm holistically: beyond anticipating harms (breakdowns), it also articulates the reasons (seams) that lead to breakdowns, traces where in the AI lifecycle these seams may appear, and most importantly, aims to turn a negative into a positive—using the seam in an opportunistic manner to augment user agency and explainability. Stemming from the ethos of *Jugaad* (Frugal) Innovation [69], the aim is to be resourceful and strategically use the seams as part of the solution instead of trying to hide them. Thus, our work creates a new design space of “seams” and an actionable design process to facilitate XAI and RAI practices (elaborated further in Section 2).

Below we first situate the notion of Seamful XAI in the broader literature in seamful design, XAI, and RAI. Based on these perspectives, we introduce our proposed design process in Section 3. In Section 4, we share our methods for conducting an empirical study to apply the design process. Finally, we share findings from the study with 43 users and practitioners of diverse roles. This paper presents a foundational step toward Seamful XAI. Our research contributions have 3 layers:

- **Conceptual:** By introducing the notion of Seamful XAI, this paper breaks new ground and presents the pioneering effort to (a) conceptually translate and (b) explicitly define seams in the context of AI technologies—uncharted areas despite calls for seamful design in AI [23, 46].
- **Methodological:** To operationalize Seamful XAI, we offer a design process that allows stakeholders to proactively identify seams and use them to augment explainability and user agency.
- **Empirical:** The findings and implications of our co-design study with 43 AI practitioners and end-users showcase that Seamful XAI aids participants in anticipating and designing with seams, enhancing explainability and user agency, promoting collaboration, and facilitating harm mitigation.

## 2 BACKGROUND AND RELATED WORK

### 2.1 Seamless and seamful design

Our understanding of seamless and seamful design follows research in the field of Ubiquitous Computing that explores how users live with heterogeneous technological systems [10–13, 27]. Seamlessness is almost always considered a good, even commendable, design ideal [42]. Technology builders strive to provide users with seamless experiences—a click and you get what you want without having to think about how and why [36, 42]. Technological systems, however, are complex sociotechnical assemblages. They do not exist in isolation, and intertwine with other components, practices, and systems in consequential ways. Seamless experiences thus require black-boxed systems shielding users from the intricate logic and messy machinery of technologies [75, p. 8].

While seamless design emphasizes ease of use and concealment of messiness, seamful design prioritizes “configurability, user appropriation, and revelation of complexity, ambiguity, or inconsistency” [42, p. 1]. Seams—mismatches, gaps, and uncertainties—are the inevitable “cracks and bumps” between heterogeneous components of socio-technical systems [10, 75]. Seamful design focuses on leveraging them in ways that help users make better decisions around technology use [11, 13, 36, 42]. Central to seamful design is the notion of strategic revelations and concealment [42, 88]. It is not just about revealing seams but also about concealing them. Most importantly, we have to be strategic (intentional) about it. The goal is to support user agency, acknowledging that despite following the best processes, technology creators face uncertain downstream trajectories of use and must give
users the resources to act effectively [42]. The question, however, isn’t just about what seams to reveal (and what to hide), but also how and when to reveal them [42].

Our work builds on and contributes to these different strands of scholarship by situating seamful design in the context of AI. In doing so, we not only put forward a design process to anticipate, craft, and design with seams, but also make visible the values that seamful design provides for helping users work with and through the everyday failures and inevitable imperfections of AI systems.

2.2 Human-centered Explainable AI

Our work also aims to augment AI explainability. The need to support user understanding through explainability is paramount given the opaqueness of AI technologies and their ability to cause harms at scale without human scrutiny. This pressing need gives rise to the field of XAI [33], which has made commendable progress in producing a growing collection of techniques to enable algorithmic explanations. While the technical landscape of XAI is increasingly broad (for an overview of XAI techniques, see [5, 30, 32, 53]), the majority of XAI research typically aims to address user questions such as “how does the model make decisions” or “why does the model make a particular decision” [50] through means such as revealing how the model weighs different features and what rules it follows.

Despite its progress, XAI as a technical field has had its fair share of critique. Empirical studies examining the effectiveness of AI explanations in user interactions have found mixed results and even pitfalls of XAI. For examples, recent HCI studies repeatedly found that popular types of XAI techniques can risk creating over-trust and over-reliance when model predictions are wrong [7, 66, 68, 91]. Technical explanations may in fact burden people who do not have the capability or motivation to engage [29, 82] and risk impairing their user experience. These pitfalls have been attributed to the techno-centric focus of the current XAI field [23, 51], pursuing technical advancements without a clear understanding of and aiming to support the end-goals that people seek explanations for. In response to these limits, the area of HCXAI has emerged within the HCI community [22, 26, 51, 87]. These works center the development of XAI technologies around people’s needs and end-goals, whether the end-goal is improving the model [60], making better decisions [43, 89], assessing model biases [19], seeking recourse [44], contesting the model [54], and so on [52, 81].

Inherent in HCXAI is also a sociotechnical perspective [22]: AI systems are often situated in socio-organizational environments, and people’s understanding must span factors both inside and outside the algorithmic black-box; thus, explainability of the socio-technical elements—the human-AI assemblage [21]—needs more than just algorithmic transparency. This sociotechnical perspective has brought prior work to “expand explainability”—defining new design spaces concerning the sociotechnical factors that govern the use of AI [6, 26, 31, 51]. For example, Dhanorkar et al. [18] highlight how explainability needs evolve and depends on who needs what and when. Ehsan et al. [21] challenge the algorithm-centrism and propose a design framework for social transparency of AI, by leveraging information about how other people interact and reason with the AI system.

In this work, we introduce seams as a new design space for HCXAI—facilitating end-user understanding of possible mismatches in and breakdowns of the sociotechnical system. In doing so, seams can complement existing explainability efforts by helping users understand not just why the AI did something, but also why the AI may not or cannot do something. Focusing on AI supporting human decision-making (vs. complete automation), we aim to support the end-goal of actionability for decision-making [50] and algorithmic contestability [54].
2.3 Responsible AI in Practice

Our work is also informed by, and contributes to, RAI practices. RAI is concerned with putting theoretical principles of AI ethics into practice [8, 71, 78]. In principle, RAI works call for proactive considerations of “how things can go wrong”—risks, harms, and ethical issues in general—and addressing these issues proactively during development, rather than reactively after deployment [4, 63, 72]. However, recent work studying how practitioners deal with RAI issues on the ground, including AI fairness [17, 38, 55, 63], transparency [35, 40, 50], accountability [70], and overall harms mitigation [4, 71], report that practitioners grapple with tremendous challenges [67]. Proactively anticipating potential harms for complex systems deployed in heterogeneous social contexts is inherently challenging [77].

To tackle this challenge, prior work called for inter-disciplinary approaches that involve technical experts, ethics experts, social scientists, and user experience (UX) professionals [45, 47, 58, 64, 65], and multi-stakeholder approaches by also involving end-users and impacted communities to align RAI with their values [16, 34, 79]. Research also aims to provide structured yet flexible “scaffolding”, such as guiding frameworks [21, 28, 59], checklists [57], processes [50, 70] and tools to lower the barriers and empower individuals to engage in RAI practices [35, 50, 57]. Recent work emphasizes the need to strategically introduce friction in RAI practices to promote critical reflection [73]. Meanwhile, the challenges for RAI practices must also be addressed at the organizational level, by creating incentive structures and organizational frameworks, strategies and workflows that prioritize RAI practices [62, 71, 86].

Aligning with the goal of providing actionable guidance for practitioners to take a proactive and multi-stakeholder approach to RAI, our proposed design process for seamful XAI tackles the challenging work of “anticipation” [9] by offering two complementary perspectives to existing RAI work. First, seamful design encourages practitioners to systematically examine the sources of potential harms in the development lifecycle rather than requiring them to fully anticipate all possible harms in the deployment context. This offers a more tractable and actionable approach to RAI. Second, seamful design prioritizes user understanding and their agency to take actions, rather than deterministic mitigation in development, acknowledging that despite the best intention, technology developers may not be able to anticipate all user needs and take actions on behalf of the users. Thus, our work also aims to introduce a new perspective and an actionable approach to this currently formidable RAI goals of anticipating and mitigating AI breakdowns.

3 SEAMFUL XAI DESIGN PROCESS

Our goal is to adapt the concept of seams to the context of XAI systems to augment explainability and support user agency. We do so by developing a design process of Seamful XAI, which allows us to solidify the conceptual transfer and enable stakeholders to engage in seamful XAI design.

In the context of AI, we conceptualize seams as mismatches and cracks between assumptions made in designing and developing AI systems and the reality of their deployment contexts. To develop the design process, we draw on key notions from the seamful design literature (reviewed in Section 2.1): 1) seams are inevitable, arising from the integration of heterogeneous sociotechnical components during technology deployments. Seams are revealed through system breakdowns. 2) Instead of treating seams as problematic negatives to be erased, they can be leveraged by strategically revealing (and concealing) seams to users. 3) The goal of this strategic revelation (and concealment) is to support user agency (actionability, contestability, and appropriation). Our process encourages AI practitioners to take proactive and multi-stakeholder approaches to anticipate seams and design with them to mitigate potential AI breakdowns in deployment.
Our process offers a few key benefits to complement existing RAI approaches. First, what makes it seamful lies around empowering users to turn seemingly negatives into positives—i.e., leverage the seams in an opportunistic manner to achieve user goals in an informed manner. Second, the process not only empowers generation but also filtering (many RAI techniques only focus on the generative part). Through the process, participants can not only anticipate seams (mismatches) from certain breakdowns but they can also figure out what to do with them—strategically filter which seams to show and which to hide, making the design space tractable. This addresses a common weakness of anticipatory methods that are excellent at generating risks but can create situations that overwhelm practitioners when it comes to knowing what to do with the anticipated harms. Finally, a key feature of this process is its interconnected and tightly coupled nature: not a single seam is untethered. Every seam can be traced back to a breakdown, its source in the AI lifecycle, and what user goal it serves if revealed.

Broadly, our design process has three steps (Fig. 1): envisioning breakdowns, anticipating the seams, and designing with seams. We grounded the development of the process using theoretical and empirical means. Conceptually, the diverging and converging elements of the “diamonds” in human-centered design served as a guiding design lens. Step 1-2 could be thought of as the generative or diverging (“<” of the diamond) part whereas 2-3 is the filtering or converging (“>” half of the diamond). Another theoretical inspiration was TRIZ (Theory of Inventive Problem Solving) framework [41](elaborated in 3.2) whose adversarial thinking substantially helped participants get traction in the generative task without getting overwhelmed. Stemming from the ethos of Jugaad (Frugal) Innovation [69], the aim is to be resourceful and strategically use the seams as part of the solution instead of trying to hide them. Beyond the conceptual grounding, we ran 4 focus group discussions (FGDs) with 11 AI practitioners and researchers from 4 technology companies and 2 universities. In total, we had 17 total discussion sessions within the research team punctuated by the 4 FGDs that helped us strike a balance between depth and breadth. For instance, we worked with the participants to iterate on the number of lifecycles (e.g., since seams predominantly appear in deployment, 3 slots are reserved for post-deployment stages in step 2). Once we reached a stable version of the process, we ran 4 pilot studies to further refine the process.

Below we propose how each step should be performed and share rationales behind our proposal. In Fig. 2, we present the whiteboard used in our study to facilitate the process to provide a visual illustration for these steps. These steps should be interpreted as a suggested starting point to engage in designing seamful XAI instead of a prescriptive guidance. In this section, we use the term “stakeholders” to broadly refer to practitioners and end-users. Even though in the empirical study we mainly focused on practitioners’ perspective, this process can also involve end-users in a participatory manner. For example, they can be involved in the same session with the practitioners to engage in a dialogue to co-anticipate and design with seams.
3.1 Envisioning Breakdowns (Step 1)

The design process begins by asking the stakeholder “what could go wrong?” in their use case, drawing on their knowledge about the target users and the deployment context. The goal is to articulate multiple types of breakdown (example breakdowns are shown in Part 1 of Fig. 2). To facilitate brainstorming, stakeholders should consider variations in user characteristics and usage contexts instead of a “hero persona” or the ideal “golden path” [39] of user workflow.

Starting the process with breakdowns is motivated by the fundamental notion in seamful design literature that seams are revealed through breakdowns [10, 11, 42]. Articulating concrete breakdowns will offer anchoring points to trace back to what types of seams can be revealed in the next step. Starting with this step will also ensure the identified seams are intentionally grounded in potential breakdowns that the seamful XAI design should aim to mitigate.

3.2 Anticipating & Crafting Seams from Breakdowns (Step 2)

This is a generative step to produce rich seams that can be filtered and designed with in the next refining step. To effectively generate seams, we propose three scaffolds: using the Breakdowns to anticipate the seams, tracing the seams through Stages in the AI Lifecycle, and Prompts for Crafting (writing out) the seam for system users (Part 2 in Fig. 2).

3.2.1 Using breakdowns to anticipate seams.

The stakeholder will be asked to engage in an adversarial role playing—given the breakdown anticipated in Step 1, “what might we (as developers, designers, researchers, etc.) do to make the breakdown happen?” In other words, what type of mismatches in assumptions between design and use can lead to these breakdowns? The answer to this question will produce a list of seams.

The adversarial thinking is motivated by the lenses of Anticipatory Failure Determination (AFD) [15], which is a part of the TRIZ (Theory of Inventive Problem Solving) framework [41]. Used in high-stakes domains like nuclear engineering to proactively mitigate failures, AFD ‘inverts’ the problem to make the failure and the search for its underlying reasons a goal directed task [83]. Once we have the causes for failures, we can then proactively search for solutions to prevent them. Thus, it provides a tractable structure for the challenging task of anticipating seams.

3.2.2 Anticipating Seams along stages of the AI’s lifecycle.

To further guide the anticipation of seams with actionable steps, we ask stakeholders to think through stages of an AI lifecycle (shown in part 2 of Fig. 2): data collection and creation, model training, model testing, data inputs in deployment, use of outputs, and updates and maintenance. “Where in the AI’s lifecycle can these mismatches or seams arise and what are they?” In each stage, if applicable, practitioners should pinpoint assumptions that they made in the development or design that can cause the given breakdown.

Using an infrastructural perspective [49, 85], our proposed AI lifecycle stages consider common technical infrastructures in model development and ask stakeholders to trace their “seams” with the infrastructures in deployment. To capture the sociotechnical elements, we went beyond the traditional AI lifecycle stages (data collection, model testing, and training) and expanded into the situated use of AI (e.g., input into the AI, use of the AI’s output, and maintenance). The formation of these stages are informed by HCAI research studying AI development practices [18, 26, 81] and four in-depth discussions with industry and academic experts on seamful design and XAI.

3.2.3 Using prompts to craft the seam.

Once stakeholders locate a seam with the above two scaffolds, they will need to craft the seam in a manner that end-users can use to support their decision-making. To craft the seam, we suggest the following prompts for scaffolding: stakeholders should ask “What are the expectations during design?”; next, they should consider: “What could be the reality or actual use in the real world?”. The seam can be written down as the difference between the expectation
Fig. 2. A screenshot of the virtual whiteboard used for the seamful XAI design activity in the study, with zoomed-in examples. **Area 1:** Envisioning breakdown (Step 1). In the study, we provided sample breakdowns \((A)\), which participants could either use directly or get inspiration for their own envisioning. **Area 2:** Anticipating & crafting seams (Step 2). We provided guiding prompts \((B)\) for effectively crafting the seams. We also shared exemplary seams \((C)\) for each stage of the AI lifecycle framework. **Area 3:** Designing with seams (Step 3). We asked participants to articulate their reasoning for choosing a seam and tag which user goals the selected seam \((E)\) can support for augmenting user agency. (Appendix A.1 has a non-annotated higher resolution version of this picture.)

and reality (the expectation vs. reality equation). Formative pilots showed that these prompts not only simplify the generation process but also make the generated seams more actionable and consistent across participants. This scaffold is motivated by the goal of seamful design—merely revealing seams is not enough; it has to be intentional in supporting user agency.

### 3.3 Designing with Seams (Step 3)

After the generative step of anticipating and crafting seams, the last step is to filter and design with them—to allow “strategic revelation and concealment” by selecting the most relevant seams that can empower end users. This filtering part can be thought of as the converging aspect (the “>” of the diamond of user-centered design). This step aligns our process with the goal of seamful design: supporting user agency. We ask stakeholders to consider: to empower the user to make better decisions and mitigate fallouts from the potential breakdowns identified in Step 1, “which seams are relevant to show? Which seams are acceptable to hide?”

Here we suggest another scaffold to help stakeholders to effectively filter which seams can support user agency—a Reflection Card that “houses” the seams (Fig. 2). While it’s not the only way to present seamful designs, this card serves as a design artifact that can be shown directly to users and weave seams into the system interface.

To guide the filtering process in a deliberative manner, we unpack the concept of “user agency” into three user goals based on seamful design and XAI literature. The guiding questions below can help stakeholders decide how the seams can address agency through supporting these goals.

- **Actionability:** Does the seamful information help the user take informed actions on the AI’s recommendation? Why?
• Contestability: Does the seamful information provide the resources necessary to justify saying no to the AI? Why?
• Appropriation: Does the seamful information help the user to adapt and appropriate the AI’s output in a way different from the provided design but help the user? Why?

Stakeholders are asked to articulate how a selected seam can address at least one of the three aforementioned user goals. If the seam does not address the goals, it can be concealed. To promote traceability and facilitate downstream design decisions, they will tag the selected seams with the targeted user goals (stickers in the Reflection Card in Fig. 2).

**Practical suggestions for the design process:** When generating the seams: Instead of doing it collectively, the process can be performed individually by different stakeholders. We recommend to have one facilitator serving as the junction point for these conversations. This alleviates the challenges of getting multiple teams synchronously, mitigate conformity pressure, and group thinking. Stakeholders can also engage in brainwriting [84] where seams are generated separately followed by a discussion to refine things.

When filtering seams: When many seams are considered relevant for supporting user agency, some prioritization strategies can be utilized, such as ranking the the seams based on how many dimensions of user goals they address (e.g., by the number of stickers they had). Stakeholder can also engage in an additional prioritization step, such as by evaluating the alignment of a seam with business goals or organizational values.

Post-deployment: The process is designed to be a “living” one, used both pre- and post-deployment. Given not all downstream harms might be proactively anticipated (e.g., in large-scale AI systems), teams can take multiple passes (e.g., quarterly) throughout the product’s lifetime.

4 METHODS: EMPIRICAL STUDY TO APPLY THE DESIGN PROCESS

Using our proposed design process described above, we conduct an empirical study with AI practitioners. The study is intended to be both a formative validation of the proposed method and to gain empirical insights about how seamful XAI design can help AI product development practices, especially in terms of augmenting user agency and explainability. This study also enables us to reflect on—and improve—the proposed process.

To learn from a wide range of perspectives, we chose to conduct an interview study that centers around a scenario-based design activity. A scenario-based design (SBD) allows simultaneously suspending the need to specify high-fidelity technical details while “envisioning future use possibilities” through immersive storytelling [74]. Participants went through the design process outlined above (in Sec. 3) with a given AI product scenario. Following the design activity, we interviewed them as they critically reflected on the process including: (a) the transferability of the design process to product contexts that they were familiar with, (b) how seamless design can impact AI explainability and user agency in their contexts, and (c) the effectiveness of the process. Below we begin by introducing the scenario we used and how we developed it, then describe the procedure, recruitment and qualitative data analysis.

4.1 Developing the Scenario for the Design Process

We need a scenario of an AI use case, which should include both user personas and usage contexts. It should balance two critical aspects: first, it should be concrete enough to generate design ideas, and second, it should afford enough malleability for participants to co-opt it.

---

1In our study, when we refer to the ‘user’, we mean the loan officer (not the applicant) who is the user of Loandao. The seamful information thus should be crafted such that it helps the officer make their decision.
We chose lending as the target domain for the scenario because of its broader relatability and consequential nature. This choice was informed by 7 consultation sessions with 18 diverse stakeholders (AI and HCI researchers, practitioners, and users) where we explored multiple possible scenarios including cybersecurity, radiation oncology, lending, and hiring. We found that even for people who did not work in the lending domain, everyone nonetheless had some level of lived experience with loans. Participant feedback also suggested that seams revealed in lending situations had the potential to be impactful for both applicants and loan officers.

We developed our scenario by basing it on real-world use cases as informed by domain experts’ insights and feedback. We did so by having iterative conversations with six loan officers who had prior experiences with automated lending support. All participation was voluntary and through personal rapport. Specifically, we constructed the narrative in our lending scenario through nine conversation sessions, where each session involved at least three loan officers. We asked the loan officers to share stories about real-world mismatches between idealized AI design and its use, which helped us generate example seams to show to the SBD participants; a list of realistic breakdowns is shown in Fig. 2. We co-constructed not only the narrative but also developed the different personas involved in the story with the loan officers. The loan officers also shared a realistic list of feature-level explanations provided by the systems they currently use, which we used to ground the AI system introduced in the SBD scenario as well. Finally, the loan officers validated the usefulness of the exemplar seams we generated to seed the brainstorming with our study participants later.

Below we share the backstory where all names are pseudonyms (captured visually in Fig. 3):

Nadia, a senior loan officer, is using Lonadao, a lending decision-support software, to evaluate if Ahmed, the applicant, should get a loan or not. Despite Ahmed being one of the bank’s best customers, Lonadao rejects his loan application. When Nadia checks Loandao’s feature-level explanations, she cannot find anything that leads her to question Loandao’s rejection. She is in a tough spot: she knows Ahmed has a good track record with the bank, yet the AI has rejected him. Should she accept the AI’s decision? What else does Nadia need to know to make an informed decision?

To encourage stakeholders to consider variations in user characteristics and usage contexts, as suggested in Step 1 of our process, we add two additional loan officer personas, Marco and David, who vary in both organizational and technological experience from Nadia (Fig. 3). For instance, some participants in the study considered Marco’s use of Lonadao and found new seams by asking: what type of seam would be more relevant to a less experienced officer?

4.2 Study Procedure

We conducted the study during the Covid-19 pandemic using video conferencing and online whiteboarding tools. Lasting 65 minutes on average, our study activity consisted of three phases.
The first phase is the introduction. After providing informed consent, participants watched an orientation video that covered the lending scenario and the personas, provided a high-level idea of the concept of seams as mismatches between design and use, and why seams can help augment agency and explainability. We then answered clarifying questions from participants. Next, we showed participants the Reflection Cards (where seams will be presented to users) and primed participants to think of the card as expensive real-estate: a seam had to earn its place to be there. This priming helped participants craft seams intentionally later on. We also primed them to think of examples in their own domain where such a process could apply (a point we returned to in the third phase of the study). Then, we provided an end-to-end example walkthrough of the steps of design activity (Fig. 2): selecting one of the breakdowns, identifying and crafting the seams for it, and then transferring the seams to the Reflection Card and arguing for why that seam addressed at least one of three dimensions of user agency: actionability, contestability, or appropriation.

The second phase is the design activity, following the design process outlined in Sec. 3 and applying it to the lending scenario. Here, the participant led the activity and we (as facilitators) guided them. To seed brainstorming, we provided four examples of realistic breakdowns in AI-powered lending and exemplar seams, all informed by discussions with the loan officers as described in the last section. Participants started with breakdowns (Step 1), followed by anticipating and crafting seams (Step 2), and then going through a filtering process to design with the seams (Step 3).

The third phase is an interview reflecting on the design activity and transferring it to their own product contexts. We started with transfer cases: we asked if participants could share an example in their domain where they could apply seamful design, for instance as a reframing of a prior project, or a speculative example in their domain. Once participants found an example, we asked them to highlight the breakdowns, the seams, and how awareness of the seams would impact the user experiences of their products. Next, participants shared their overall thoughts on the effectiveness of the design process: what worked well and why. Finally, participants shared their final thoughts on how thinking seamfully affects the explainability of the system, how seamful information can empower users, and what challenges they faced as well as suggestions for improvements.

4.3 Study Recruitment

Our aim was to gather multi-disciplinary perspectives from AI stakeholders, including researchers, practitioners, and end-users of AI-powered applications. We recruited through online ads with a sign-up form survey (advertised via email and social media), which gathered 258 total submissions. In this survey, we asked for the following: experience with AI systems, experience with XAI, their role, domain, geographic location, and number of years of experience, along with questions where they can provide examples. For screening, we balanced the following factors: range of work experience (from early career to experienced), XAI experience, variety of domains, geographic diversity, project timelines and budget, and most importantly, scheduling constraints across multiple time zones. We prioritized participants who provided detailed work examples, which showcased their motivation to engage. Our final pool comprised of a diverse group of 43 participants comprising of Product Managers, Data Scientists, UX Researchers, Designers, AI Ethics/Responsible AI experts, Researchers (both Academic and Industry-based), and Loan Officers. Table 1 provides a breakdown of their roles, domains, and experience. Each participant was provided US $50 as an appreciation for their time. 24 out of 43 participants self-identified as female while the rest identified as male. While the majority of participants were located in the US (22 out of 43), we had participants from Australia, Belgium, Canada, Denmark, Germany, India, Netherlands, and the United Kingdom. Not only do we have Loan Officers representing their voices as potential end users of the system, we also recruited a few Data Scientists who specialized in developing automated lending systems.
Table 1. Details on study participants: Organizational roles and experience levels.

| ID | Organizational Role                                      | Experience |
|----|----------------------------------------------------------|------------|
| P28 | Academic Researcher (XAI)                              | >3 ys.     |
| P29 | Academic Researcher (XAI)                              | >4 ys.     |
| P31 | Academic Researcher (XAI)                              | >10 ys.    |
| P32 | Academic Researcher                                    | >7 ys.     |
| P33 | Academic Researcher (XAI)                              | >4 ys.     |
| P07 | Industry Researcher (XAI)                              | >5 ys.     |
| P21 | Industry Researcher                                    | >4 ys.     |
| P24 | Industry Researcher                                    | >10 ys.    |
| P34 | Industry Researcher (XAI)                              | >4 ys.     |
| P10 | AI Ethics/ Responsible AI Expert                        | >3 ys.     |
| P11 | AI Ethics/ Responsible AI Expert                        | >5 ys.     |
| P19 | AI Ethics/ Responsible AI Expert                        | >10 ys.    |
| P20 | AI Ethics/ Responsible AI Expert                        | >10 ys.    |
| P13 | Data Scientist                                          | >5 ys.     |
| P16 | Data Scientist                                          | >10 ys.    |
| P17 | Data Scientist                                          | >4 ys.     |
| P18 | Data Scientist                                          | >5 ys.     |
| P40 | Data Scientist (Lending focus)                          | >3 ys.     |
| P41 | Data Scientist (Lending focus)                          | >3 ys.     |
| P42 | Data Scientist (Lending focus)                          | >4 ys.     |
| P43 | Data Scientist (Lending focus)                          | >7 ys.     |
| P02 | Product Manager                                         | >5 ys.     |
| P03 | Product Manager                                         | >4 ys.     |
| P11 | Product Manager                                         | >10 ys.    |
| P14 | Product Manager                                         | >5 ys.     |
| P19 | Product Manager                                         | >8 ys.     |
| P22 | Product Manager                                         | >15 ys.    |
| P08 | AI-UX Designer                                         | >15 ys.    |
| P12 | AI-UX Designer                                         | >10 ys.    |
| P15 | AI-UX Designer                                         | >3 ys.     |
| P26 | AI-UX Designer                                         | >5 ys.     |
| P27 | AI-UX Designer                                         | >4 ys.     |
| P01 | UX Researcher                                           | >5 ys.     |
| P04 | UX Researcher                                           | >5 ys.     |
| P05 | UX Researcher                                           | >5 ys.     |
| P06 | UX Researcher                                           | >5 ys.     |
| P20 | UX Researcher                                           | >5 ys.     |
| P23 | UX Researcher                                           | >5 ys.     |
| P35 | Loan Officer                                            | >6 ys.     |
| P36 | Loan Officer                                            | >8 ys.     |
| P37 | Loan Officer                                            | >3 ys.     |
| P38 | Loan Officer                                            | >10 ys.    |
| P39 | Loan Officer                                            | >7 ys.     |

4.4 Qualitative Analysis

Overall, we analyzed 2796 minutes of transcribed videos from the 43 study sessions as informed by the grounded theory approach [14, 80]. Taking an inductive approach, we started the process with an open-coding scheme and iteratively produced in-vivo codes (generating codes directly from the data). One researcher performed the iterative coding, punctuated by frequent discussions with the research team where we constantly compared and contrasted the codes, refining and reducing the variations in each round. We referred to the videos and whiteboards as needed. Next, we analyzed the data using axial codes, which involves finding relationships between the open codes and clustering them into different categories (e.g., ‘revealing the AI’s blind spots’). Finally, we unified the axial codes and consolidated them to selective codes (e.g., ‘enhancing explainability’). We discuss the selective codes as the main themes in our findings section where the axial codes are bolded (ones that add to the theme), wherever applicable.

We also collated all seams from all participants and clustered them along their respective breakdowns and AI lifecycle stages, creating a single snapshot of the entire study to get a sense of the distribution of breakdowns chosen and seams generated (visually depicted in Fig. 4).

5 FINDINGS

In this section, we report on our key empirical findings, organized in four parts. First, we describe how our design process helped participants to not only identify, craft, and design with seams in the context of the study scenario, but also transfer learning to their own product contexts. Second, we describe the value of introducing and operationalizing the notion of seamful XAI in terms of augmenting explainability and agency goals. Third, we describe the efficacy of our design process more broadly regarding how it helps participants proactively design for failures, make concrete
Fig. 4. A bird’s eye view of all seams from all participants for all breakdowns along all AI lifecycle. A: the dots above this breakdown provide the number of times it was chosen (e.g. 17 times). All the seams appearing below this column were crafted in connection to this particular breakdown. B: the dots showcase the number of seams crafted for a given AI lifecycle stage (6 seams were crafted for this stage). C: a zoomed-in view to provide a sample of seams crafted by our participants along with the justifications for how the seams enhances agency. (Appendix A.2 has a non-annotated higher resolution version of this picture.)

abstract RAI goals, and address often amorphously-defined AI harms. Finally, we touch upon challenges and areas of opportunities concerning our process as described by our participants.

5.1 Design Activity Effectiveness: Anticipating, Crafting, and Designing with Seams

We first provide an overview of our study activities, including information about the seams generated. Then we share the context for how participants identified and designed with seams.

5.1.1 Overview. Our 43 participants generated a total of 85 seams. By following our process, every participant was able to identify and craft at least one seam. This is encouraging because only one participant had some knowledge about seamful design before the study; everyone else started with no knowledge. As depicted in Fig. 4, overall, participants filtered and chose to reveal 47 relevant seams. For example, around dataset and model training, seams like “skewed data” (P03), “data and model drift” (P07), and “the AI system can overly focus on one parameter such as credit score and ignore other parameters” (P16) were identified. On deployment side, participants crafted seams that looked at gaps in usage from data input as well as updating and maintenance; e.g., seams like “reports from multiple credit bureau’s may have conflicts which impacts AI’s decision” (P38), “model updates does not guarantee compliance with policies” (P37), and “variable peer-to-peer agreement with AI over time” (P39) were crafted. Participants placed these seams on the Reflection Card and justified why the seamful information augments user agency.
Each breakdown had at least 5 participants working on it. All AI lifecycle stages had seams crafted along them. The most popular breakdown was “Over-reliance on AI, automation bias, and potential de-skilling” (N=17) and the most popular AI lifecycle stage was “Deployment: Use of AI’s prediction” (N=23). While participants added seams to all AI lifecycle stages, the majority (35 out of 47) were anticipated in deployment stages (4-6). No obvious differences in the number and type of seams addressed were observed for participants of different job roles. Below we share how our process helped participants anticipate, craft, and design with seams. We then underscore how participants found the process effective in transferring the concepts to their own domains.

5.1.2 Anticipating and crafting seams. Participants highlighted three main reasons why the process was effective in empowering them to engage in seamful design, all of which are tied to the scaffolds our process provided. First, the adversarial thinking during breakdown was helpful because “the first impulse of when we think of breakdowns is to stop them from happening” (P19). “Making breakdown a goal-centered task is super effective” because as “technologist[s] [we are] often too solution oriented and don’t pause enough to think through what the exact problems are” (P07, P24). The “provocative” nature of asking ‘what could we do to make the breakdown happen’ helped participants to “not go to the solution first” but instead “have the freedom to really think about bad outcomes without being burdened by the hero expectations” (P03, P01, P24).

Second, having the AI lifecycle stages listed out helped participants “find a home for the seams because the gaps could come from anywhere” (P25). The different stages also made the “thinking process deliberative and structured” and “showcased the sociotechnical reality of AI systems” (P29, P23). Third, prompts for crafting were “instrumental” to frame and write out the seam “in a productive way” (P15, P20, P43). “Coming up with the seam is one thing but figuring out how to write it in a way that is helpful to someone else is a different ball game” (P06). Writing down the [ideal design] expectations and the reality [real use]” made the process “reproducible and doable even for first-timers” (P17, P32). Having the process broken down “into bite sized pieces” made the ”learning curve less steep, promot[ing] self-reliance” (P06, P07):

No real-world technology follows the golden path. Having this expectations vs. reality equation has boiled the art of crafting a seam into a recipe that anyone can easily follow...that’s powerful. (P08, a senior AI UX designer)

5.1.3 Designing with Seams. Every participant succeeded to design with the seams they crafted: to strategically select seams that would be worth revealing to augment user agency.

This step was supported by the use of reflection cards and the breakdown of agency into three dimensions: actionability, contestability, and appropriation. Participants appreciated how the deliberative exercise helped them design seamfully. First, the “decomposition of agency into the three dimensions helped them “grapple with the abstract concept of agency” (P07, P10). They underscored how the three dimensions were “more accessible, operational, and on the ground” (P15). Second, the deliberation empowered participants to be “decisive about filtering the seams and ensuring that every seam in the reflection card had earned its place” (P33). This way, there was no “seam bloat” problem where “practitioners would jam too many seams and overwhelm users” (P39, P28). Third, the deliberative process helped participants “embed their design voices because the seams revealed are just as intentional as the ones hidden” (P18).

Most RAI frameworks stop at the generation phase, leaving us with a list of harms and no idea of what to do with them. This is the first one that lets me filter and figure out what to do with them systematically. It’s like the double diamonds of user-centered design...[steps] 1-2 being the generative part and 2-3 as the filtering part. (P06)
5.1.4 Transferring and teaching about seams. Participants reported that the process was teachable and helped them transfer the concepts to their own use cases. As such, every participant succeeded in the transfer task with varying levels of depth by reframing their past projects through a seamful lens. These comments demonstrate the effectiveness of our process as all participants, except one, was introduced to the concept of seams for the first time with our 45 minutes co-design activity.

Overall, participants found the process to be a “teachable exercise, [one that’s]...generalizable to a lot of AI use cases” (P14). In particular, they appreciated “how the process converted abstract concepts like seams into specific mismatches” (P27). They also enjoyed the “simplicity of the 1-2-3 process of going from breakdowns to seams to designing with seams on the [reflection] cards” (P05). The following participant sheds light on what worked for successful transfer:

I’m surprised how easily I could map this new lens onto my old projects...What worked really well for me was how the process was broken down to bite-sized pieces and transitions were clear...loved the adversarial thinking we did for the breakdowns. (P40)

The transfer use cases covered a diverse array of AI-powered applications in domains such as finance, healthcare (radiology, oncology, mental health), agriculture tech, self-driving cars, humanitarian efforts, and air travel. One transfer example was around an airline algorithmic pricing gone awry during Covid-19. “If travel agents knew [the seam] of how the model was built on pre-Covid data and it was not capturing real-time market dynamics, then they wouldn’t either trust it blindly or reject it completely” (P20). This awareness could have helped calibrate their reliance appropriately, allowing them to contest and appropriate. Another example focused on AI-powered treatment planning in Radiation Oncology (RadOnc). If RadOncs in South Asia knew the seamful information of the model being trained on North American data with large variance annotations, it would have helped them contest the AI recommendations and fight organizational pressure to “accept the made in USA AI model” (P07). It would have helped them also appropriate the recommendations because treatment standards vary regionally; for example, in South Asia, “hypofractinations [higher doses per fraction] are the norm due to machine and electricity black outs, things you don’t have to worry in the US” (P24). Below is an in-depth walk-through of a transfer case about an AI-powered international aid allocation algorithm gone wrong:

Our [ideal] assumption was that if we redacted the protected categories such as race, the algorithm would be blind to these and thus be fair. However, there was model leakage [breakdown in reality] where the algorithm latched on to ZIP codes, made weird correlations, and was discriminating based on that...if we’re aware of the seam of model leakage, then it’d have explained why the bias existed. (P03)

This example illustrates the depth of transfer. The fact that all participants “were able to do the transfer exercise in a limited time is a testament to ease of adoption of the process” (P25).

5.2 The Value Propositions of Seamful XAI

In this section, we share empirical insights on the value propositions for Seamful XAI. We focus on the two goals that our work set out to approach with seamful XAI: enhancing explainability and user agency. Beyond this, we also share how our process can benefit AI product practices.

5.2.1 Enhancing Explainability. A core motivation for this work was to enhance explainability through seamful information. Participants’ responses validated this position and further demonstrated how seams can enhance explainability. Understanding is a key part of explainability. One of the core value propositions of seams is that it exposes the user to actionable information that augments one’s understanding. There is a “natural fit between explainability and seamful design...
because the places where the seams occur have the highest chance of creating negative downstream impact...[which] when revealed can explain the AI’s actions in a situated manner” (P30). Below we share three ways in which seamful information helps with explainability.

First, seamful information provides **peripheral vision by situating AI systems in the broader sociotechnical context**. With seamful information, “our blinders are taken away and we can see more, [which] helps us to understand the AI’s working, especially its mistakes” (P43). Seams make it “clear that an AI doesn’t live in a vacuum” (P04). A “model-centered, seamless view can only explain what’s going on inside the [black-] box” (P21). Revealing seams along the “entire AI lifecycle provides “peripheral vision that can highlight AI blind spots [that] otherwise [will] be missed if we take a seamless view” (P09), further elaborated by this participant:

For us data scientists, the AI isn’t the black box because we build it, but everything around it is...like how the data collection conditions impact my model’s performance...Having that awareness is extremely important for me to explain the AI. (P16)

Second, seamful XAI can **reveal the AI’s blind spots, highlight its fallibility, and showcase the strengths and weaknesses of the system**, which can facilitate **calibration of trust and reliance on the AI**. Participants felt that awareness of the blind spots can “improve explainability because if I know what the system cannot do, I can better explain why it’s unable to make certain decisions” (P25). Seams also convey the information about the blind spots in “a lay friendly manner, making things more accessible to non-AI experts” (P02). A clearer picture of the AI’s strengths and weaknesses is essential “to know when to trust the AI vs. not” (P07), which is one of the “most challenging things with AI because performance is never uniformly good or bad” (P23). Trust can be calibrated “from knowing the AI’s blind spots, which the seams can reveal” (P33). Having that understanding is empowering because “it can help people make better sense of AI’s output when things go wrong” (P34). This participant further encapsulates this point: “The improvement in explainability seems to be an implicit outcome of seamful design. By showing the seam, you bust the mythology that AI is all knowing...it reduces mindless AI acceptance” (P34).

Third, seamful information affords a unique type of understanding and explainability by helping users understand why the AI may not or cannot do certain things:

The explainability from seams is very different...If I know there’s annotator bias, I can do forward causal inference. It explains why the AI cannot make fair decisions even if everything looks good at a feature level... The usual XAI techniques tells me the ‘why-part’, but seams can help me understand the ‘why-not’ part. (P28, emphasis added)

Having information of both why and why-not can promote “holistic understanding and counterfactual reasoning mechanisms” (P24), which can improve explainability [21, 48].

5.2.2 Augmenting Agency. Participants’ comments also validated the end-goal of seamful XAI in supporting user agency, and delineated its paths to do so.

Broadly, by providing peripheral vision of the AI’s blind spots, **seamful information expands the action space of what users can do**, augmenting agency. At its core, “seams are about opening up new kinds of interactions with the system; knowing the limitations opens spaces for how humans interact with systems” (P32). Examples of user actions include new interactions in the form of “peer-to-peer consultations... contesting the AI...or appropriating the AI’s output in a way different from its intended use” (P21). Participants reflected on how the “western obsession with perfection often leads to an unachievable seamless ideal” (P13) and how seamfulness can be empowering: “seamless design is a take it or leave it paradigm...user has no voice. After identifying seams, users evaluate what to do with it—having that choice is empowering.” (P18)
Information in seams can convert “unknown unknowns” to “known unknowns” that can empower users to know “where to start an investigation; e.g., if [they] know that the model doesn’t capture a regulatory change, while [they] can’t fix the model, [they] have two things—understand the failure and know where to begin [the] investigation” (P03). “Without seams, people may not even know what to ask for” (P07). Often AI users “automatically assume that they are at fault and AI is right”, knowing the “limitations through the seams” can convert “unknown unknowns to known unknowns, which can empower [users] to contest the AI.” (P05, P07, P34)

In our study, we scaffolded and operationalized user agency along three dimensions of user goals with a decision-support AI system: actionability, contestability, and appropriation. When participants designed with seams (step 3 in Fig. 2), they also articulated how the chosen seams augment user agency. Below we synthesize participants’ justification along with concrete examples of user actions.

**Actionability** comes to the forefront when we focus on what people can do with the information in the seams. A major way in which seams addressed actionability was when the seamless information empowered the user to “act on it in an informed manner” (P08). For instance, the action could be “accepting the AI’s decision, clearly knowing why it’s appropriate,” “flagging [it] for a peer consultation to double check [the user’s] intuition,” or “recommend[ing] customers to correct their applications or credit reports if mismatches around data inconsistencies show up” (P19, P40, P39).

**Contestability** becomes relevant when the AI’s decisions “go against users’ intuition” or are “surprising” (P36, P21). Contesting a decision is a critical path of exercising user voice in algorithmic decision-making [3, 54, 76]. One of the key themes around contestability was when the seamless information provided justifiable information to say no to the AI, allowing the user to exercise their domain expertise instead of blindly trusting the AI. For instance, if stakeholders “are aware of data drifts that produce biased decisions” or that “prior users rejected the AI”, it gives the “right ammunition to contest the AI despite organizational pressures to use it” (P12, P30, P27).

**Appropriation** supports the “how”: when users are dissatisfied with the current system, “how to reinterpret or override the system” (P10, emphasis added). This dimension has an explicit “shift in the focus towards the human as the main driver” (P22). Seams facilitated appropriation when they afforded information that empowers the user to “co-opt”, “re-interpret”, or “re-work” the system to “aligns with the user’s goal” (P03, P29, P41, P28). For instance, if participants knew of the “mismatch where masking of protected categories (such as race) negatively impacted the AI’s decision, [they] could fiddle with the input to produce fairer decisions” (P36). Appropriation can be thought of as “feedback to the system, a way of sharing their voice back” (P40).

Notably, every justification touches on a combination of technology, infrastructure, and work practices that encompasses the AI lifecycle. Thus, participants were able to “view” the AI in its situated environment, addressing agency by explicitly considering the human-AI assemblage.

5.2.3 The dual value proposition: revealing seams while proactively mitigating harms. Our design process can be a resourceful way to not just reveal seams but also anticipate and mitigate harms. Generating seams has an intrinsic value. The generated seams “practically become a ready-made task list of potentially harmful things to work on” (P09). Working through the list of seams can “serve as a map to help companies conduct risk assessments in a resourceful way” (P21):

There’s a dual value proposition: once you’ve generated the seams, you can neutralize many of them because of proactive identification. Even if some seams are unfixable (e.g., regulation changes making the AI non-compliant), you can leverage them resourcefully to help with better decision-making. Thus, you get two for the price of one. (P33)

Typical “Responsible AI frameworks call for envisioning harms, which is a great start” (P03); what is unique about Seamful XAI is that “seams aren’t just shown for the sake of showing them...they are used in opportunistic ways to help the user” (P14). Thus, it empowers users to “discover the
problem *and* do something about it” (P34), going deeper than practices that stop at anticipating harms.

5.3 Benefits of the Design Process for AI Product Development Practices

Several themes emerged about how participants viewed the Seamful XAI process can bring benefits to AI product development practices.

5.3.1 Turning abstractions into specifics. Participants expressed that the seamful XAI design process is focused, specific, and structured, **empowering them to transform abstract concepts in AI harms mitigation to specific instances.** Highlighting that the “process is tangible and generative” (P21), participants emphasized that it can be “effective in turning high-level guidelines into practice because seams are on-the-ground” (P36). This, in turn, makes the process “more practically implementable with customers” (P38), elaborated by this participant:

> Why the heck didn’t we think like this before? I’ve used tons of AI Ethics tools. My clients hate how abstract they are. This [process] is specific and customers would love it...this is the next leap from impact or risk assessments. (P03, a Product Manager)

In short, our process “takes the story multiple levels deeper into the solution space than just brainstorming about potential problems” (P39). The specificity, especially with breakdowns, can also help “kick-start the design process in the right way, which is very hard to do in practice” (P24). Participants appreciated the “hands-on nature of the process”, especially the “creative antagonistic thinking [they] did around breakdowns, and then getting [their] hands dirty to filter the seams” (P06, P18). The hands-on nature can make abstract concepts tractable, elaborated by this participant:

> Harm can be a nebulous word...This [process] grounds everything– anchor on a specific breakdown, work through the [AI] lifecycle, and then use the seams...Unlike other RAI tools, the lifecycle stage allows you to pinpoint where things could go wrong (P25)

Moreover, our process has “a nice balance between top-down and bottom-up aspects: it has top-down ideas like going from breakdowns to seams with bottom-up aspects such as users generating the specific seams for a specific breakdown at a specific AI lifecycle” (P16). This is in contrast to “traditional impact assessments that are too top down” (P03).

5.3.2 Fostering cross-disciplinary collaboration in a critically reflective manner. Participants believed that the design process can facilitate **multi-stakeholder and interdisciplinary discussions**, providing an inclusive and engaging way to lower the participation barrier.

> First, participants underscored how the process is an integrative approach that “stitches together all the layers of AI development in a way that can benefit all stakeholders” (P28). Even though some participants were aware of certain seams, they highlight a “lack of a framework to think through these issues” (P17), further elaborated by this participant:

> This process gives end-to-end visibility, breaking silos rampant in tech companies. For example, if my PM foresees a seam in deployment, as a data scientist, I can address it upstream. It also helps with accountability; we know who is responsible for what. (P41)

Second, “by providing a shared medium and language to communicate” (P14), participants appreciated how the activity can empower inclusive multidisciplinary stakeholders to communicate effectively. They found the “vocabulary of seams helpful and terms like ‘breakdowns’ accessible” (P13). By “crafting the seam in plain words that anybody can understand, the process becomes very powerful. No one need[ed] a fancy machine learning degree to participate” (P40). This, in turn, can lessen feelings of “being a second-class citizen of a tech team” because “on the [design] board, it’s a level playing field. Be it a PM, a UXR, or a Designer...the process shows we’re all important” (P26).
Third, participants appreciated the inclusive “co-design nature of the activity” (P30). Envisioning it as “a reflective tool for enacting Responsible AI in a step-by-step manner” (P06), participants described how the “process can challenge and break down [their] own assumptions” (P08):

I really like the co-design element. It made me feel included. I can see doing this with my team. The flow was clear... Having pre-filled examples of breakdowns and seams made it less intimidating and lowered the activation energy. (P34, a UX Researcher)

Lastly, participants strongly resonated with “how fun and engaging” the process was (P06) to lower the barriers for participation. Many “loved the simplicity and clarity”, emphasizing that it was “easier, more fun and hands-on than anticipated” (P12, P15). A key driver of engagement was the adversarial thinking participants did to anticipate seams by making the breakdowns happen:

I had so much fun doing the super villain thinking! We’re always expected to be the good guys... Getting a temporary license to be unshackled from that hero expectation is freeing—I don’t have to hold back ... This is the X-factor of this process. This achieves the same goal without the heaviness or dullness of other RAI stuff. (P24, emphasis added)

5.3.3 Taking a proactive design approach. This design process promotes a shift towards a proactive design approach in (X)AI, which can mitigate the harms stemming from “the current reactive AI design–where we wait for things to fall apart from doing anything–slows down progress” and “is expensive in the long run” (P30, P22). Participants found the process a “tractable proactive method that is a sharp departure from the reactive AI design” (P19). Beyond proactiveness, Seamful XAI carries the process till the end—“it goes beyond just asking ‘what could go wrong?’, a point where most harms envisioning frameworks stop nowadays” (P30). By empowering people to leverage a seemingly negative seam into a positive, the process “transforms the information into actionable insights to make things right” (P30). This participant summarizes the value of a proactive process:

This proactive process lets us ask more meaningful questions... knowing the AI’s blind spots can foster better feature engineering. It can inform where the AI may fall short and how we need to use seams to stop-gap the understanding. (P29, an XAI researcher)

A proactive approach, however, is not “the panacea of preventing AI harms” (P34):

Breakdowns are inevitable... This process can lessen the fallout. The proactive stance is like designing a car with a crumple zone that’s lifesaving during crashes. Seamful information empowers users to handle AI mistakes without catastrophic failure. (P19)

When asked if the process was too resource intensive, participants stressed that the “investment was worth it because you can pay a cost now or pay a higher cost later” (P30).

5.4 Challenges and Opportunities

Every process has its challenges as well as areas of improvements. Below we touch on both.

Participants speculated two main challenges for practitioners to provide seamful XAI around value tensions between teams and crafting a seam in a productive manner. First, at an organizational level, competing incentives and value tensions between teams can impact which seams are shown or hidden. “Each team has its own agenda; the product [team] has a go-to market strategy that may conflict with AI ethics division’s– who wins out?” (P34). To solve this problem, participants suggested that “before teams even get started, they should put their cards on the table and get some alignment of goals—we don’t want a fight when people are transferring seams to the reflection [card]” (P30). Participants, especially policy advisors, highlighted that “if we don’t get alignment ahead of time, the process won’t be as effective” (P28). They felt it’s “imperative to harmonize business goals and the customer’s interest” (P18) for the process to have the highest impact.
Second, at a process level, in the beginning, participants struggled to craft or frame the seam’s information in an actionable manner to help the user albeit they were successful by the end of it. While participants “got the concept of seams fairly easily”, “the challenge was how to phrase it such that it was actionable or helped the user” (P08, P12). Thankfully, the scaffolding provided by the “crafting equation of expectations vs. reality” (P07) addressed this challenge. “The equation was bang on; boiling it down to a recipe helped [them] devise the seams effectively” (P06).

Participants proposed three ways to improve the process through (a) knowing the impact of a seam, (b) adding action items, and (c) creating an organizationally accessible collection of seams. First, they wished to know the impact of a particular seam on the outcome. Knowing the impact would make the seamful information more actionable because it could “evaluate which seams to tackle first vs. later” (P23), a perspective effectively expressed here:

Show me what’s the impact of this seam vs that seam: seam X impacts 35% of the AI decision vs. seam Y impacts 10% would be a killer feature because then I can prioritize using the impact data and the stickers on the board. (P04, an XAI focused UX researcher)

Second, there was a demand for adding action items and recommendations along with the seamful information. Participants suggested that the process would be “more powerful if [they] knew what to do with the seam” (P33). They recommended that adding the action item could “help less experienced end users to take the right action and make things consistent across the team” (P19).

Third, there was a popular suggestion to create a corpus of seams accessible across the organization. Having a corpus of seams categorized across “domains, problems, and AI life cycles stages would be a good starter pack” (P20). Participants had a few recommendations: one way could be “a design feature where users can flag useful seams, which are then added to the collection” (P01). Some highlighted an “inexpensive yet rich source of seams—the limitations documents” (P07). They argued that “turning the latent information in the limitations documentation in tech companies into activated knowledge can be a resourceful way to do seamful XAI” (P40).

6 DISCUSSION & IMPLICATIONS

Our study provides formative validation for the value of transferring and operationalizing the concept of seamful design into XAI. In this section, based on our empirical findings, we reflect on the roles Seamful XAI can play in expanding the design space of XAI, enabling RAI practices, and broader implications for research and practices around AI technologies.

6.1 Expanding the Design Space of XAI

A seamful outlook expands the XAI design space beyond the bounds of the algorithm. It extends the prior work in HCXAI by explicitly showing us how to use infrastructural mismatches (seams) in resourceful ways to augment explainability. Our findings showcase how seams can provide the peripheral vision of things outside the black-box that are required to make complex decisions, can reveal the AI’s blind spots, and can calibrate user trust in AI.

Seams bring to the forefront infrastructural mismatches that would otherwise likely be hidden or not considered. By doing so, as shown in Sec. 5.2.1, seamful information affords a unique type of understanding that helps with explainability – knowing why the AI may not or cannot take certain things into account. In complex real-world cases where the “ground truth is not black and white, we need to know when we can trust the AI” (P30, emphasis added). If we know why the AI does well in one situation and not in another, we can calibrate when to rely on it:

Two important things happen: [first,] awareness of the broader context explains why the AI couldn’t make the right call...[Second,] it makes my trust on the AI nuanced. I
know where it might fail but also where it’d succeed...It’s no longer an all-or-nothing play. The AI is neither idiotic nor God-like (P32, an XAI researcher)

*A seamful outlook expands the epistemic canvas of XAI.* It "broadens the domain of things considered for XAI" (P33). This expansion has implications: 1) it provides a unique lens of analysis for XAI, 2) it fosters "thinking of new interaction paradigms" (P28) that opportunistically leverage seams.

### 6.2 Enabling resourceful, tractable, and actionable ways to do Responsible AI (RAI)

Our findings showcased how the seamful XAI design process enables us to enact RAI activities in a tractable and resourceful manner. The structured nature coupled with the on-the-ground situated nature of seams enable participants to transform abstract concepts like harms into specific instances. The Seamful XAI process extends the reach and impact of RAI activities in three ways.

First, the process serves as a *resourceful way of proactively mitigating harms*. There is a "two for the price of one" (P33) angle here: through the exercise of anticipating and crafting seams, we have a ready-made list of actionable mismatches to work on, many of which can lead to harms. Thus, we have a proactive chance to neutralize the seams (rendering them irrelevant).

Second, our process *not only generates a potential list of harms, but it also provides guidance on “what to do with them”* (P06). This guidance is facilitated by the filtering and designing with seams (going from step 2 to 3 in Fig. 2). Participants found the exercise of filtering seams to be one of “the most innovative parts [because] it continued beyond merely generating harms, which is where most RAI approaches stop” (P04). Stakeholders can feel “overwhelmed and lost [when] they are left with a huge list and nothing to do about it” (P22). The filtering process not only forces us to critically think about how to leverage the seams but also prioritize the most important ones.

Third, seamful design encourages us to use the mismatches and leverage them in an opportunistic manner that empowers users [12, 13, 42]. In most “harm envisioning tasks, there is no impetus to leverage the harms” (P18). Thus, there may be missed opportunities where the user could have been empowered to act on the risks instead of just being made aware of them. For the seams that can be neutralized, our process allows that proactively. For the ones that we cannot do anything (e.g., model lags and organizational policy shifts), a seamful outlook prompts us to use uncertainties and limitations, as Gaver would put it, in a resourceful way [27].

### 6.3 Implications for Research & Practice

We share three implications of Seamful XAI on research and practice. *First*, from an infrastructural view, a seamful approach allows us to leverage the imperfections in AI systems without needing to over-hype their capabilities or impose unachievable ideals. Understanding infrastructural mismatches can help users, especially non-AI experts, to be aware of the fallibility of AI systems. This awareness can shape how we think about trust calibration in AI, especially when addressing over-inflated expectations. Acknowledging AI is imperfect does not entail that we give up and do not work on problems [61]. It simply means we are realistic about what we can solve and opportunistic about the rest. We leverage the gaps and use uncertainty as a resource instead of thinking of them as inherent flaws. An infrastructural view also changes how we design: instead of obsessing over perfection, we can be resourceful about Human-AI interaction designs that leverage inevitable imperfections to empower users with the tools to deal with the fallout from breakdowns.

*Second*, a seamful XAI outlook can set the right expectations about the limits of our powers as technologists. Given that AI is a tricky design material [20, 24, 37, 90] due to its stochasticity, changing, and non-deterministic nature, it is difficult to envision exactly how events unfold in use [75]. Given that breakdowns emerge in use, where developers or designers are unlikely to be around, it is in our best interests to empower end users with tools to address the fallout from breakdowns. Moreover, a seamful outlook acknowledges that creators don’t always have the control.
to dynamically change the system in real time (e.g., model updates may not be possible due to regulatory reasons and thus may lag organizational policy shifts). This realization can prompt changes in how we study AI failures and design fallback plans.

Third, there is intrinsic value in adopting a proactive stance towards seamful XAI design even if we cannot anticipate or prevent all breakdowns. Our proactive stance is a differentiating factor from the majority of prior work in seamful design where inventions exclusively dealt with seams post-deployment [10, 12, 61]. The core implication of the unbridgeable nature of the “sociotechnical gap” [1]—the difference between what the system can technically afford and what users need—is that it may be impossible to exhaustively anticipate all seams. Despite that impossibility, there is value in a proactive stance because the very act of anticipatory exercises can (a) convert many “unknown unknowns” to “known unknowns” to promote proactive mitigation (discussed above) and (b) help us better understand and address the sociotechnical gap.

No AI system is perfect. None will be. Seams are inevitable. A seamful view acknowledges the world as is. A seamless ideal projects a vision of the future, yet unrealized and arguably unrealizable. Seamfulness embraces the imperfect reality of spaces we inhabit and makes the most out of it.

7 LIMITATIONS & FUTURE WORK
We have taken a formative step on how to incorporate seamful design in the context of (X)AI. Given that this is a first step, the resulting insights should be scoped accordingly. We acknowledge limitations from using a scenario-based design, where there is data dependency on the particular context. Our findings should be interpreted as formative rather than evaluative. We hope that future work could evaluate how the process works longitudinally, explore how organizational resources impact engagement in it, and investigate the consequences of a seamful design approach in how it impacts the user experience. There might some cases (e.g., large-scale AI models) where proactive identification of all downstream breakdowns may not be possible before deployment. In such cases, as shared in 3.3, the process should be used iteratively post-deployment. For future iterations, we are inspired by Phil Agre’s notion of adopting a split identity, one where we simultaneously plant “one foot...in the craft work of design and the other foot in the reflexive work of critique” [2]. We have planted one foot in design by introducing the notion of Seamful XAI. Now, we seek to learn from and with the broader HCI and XAI communities through critically constructive reflections.

8 CONCLUSION
In this paper, we demonstrated how seamful design advances Human-centered XAI. Through introducing and operationalizing the concept of Seamful XAI, we showed what seams look like in the context of XAI, and more broadly, AI. We developed and validated a design process to help stakeholders anticipate, filter, and design with seams. Our study showcased how the strategic revelation and concealment of sociotechnical and infrastructural seams enhance explainability, augment user agency, and, in fact, help identify and mitigate downstream AI harms. Recognizing that failures are a normal part of AI systems, seamful XAI embraces the incomplete and partial nature of AI systems, and instead of sweeping breakdowns under the rug, attempts to leverage seams as useful decision-making aids for end users. Through helping users live with systems that are inevitably imperfect, Seamful XAI aims to empower users to work with, through, and sometimes around such systems.

ACKNOWLEDGMENTS
With our deepest gratitude, we appreciate the time our participants generously invested in this project. This project would not have been possible without their involvement. Special thanks to Matthew Chalmers, one of the pioneers of Seamful Design, for engaging with us throughout this
project and sharing constructive and kind feedback, which substantially helped with the conceptual development. We are grateful to members of the Fairness, Accountability, Transparency, and Ethics (FATE) group alongside other researchers and practitioners at Microsoft for their feedback that informed the empirical work. At multiple stages of the project over the years, we are appreciative of the generative discussions with Michael Muller, Sauvik Das, Sashank Varma, Munmun De Choudhury, Jennifer Wortman Vaughan, Kate Crawford, Mehrnoosh Sameki, Ben Noah, Gonzalo Ramos, and Mary Czerwinski. This project was partially supported by Microsoft Research and by the National Science Foundation under Grant No. 1928586.

REFERENCES

[1] Mark S. Ackerman. 2000. The Intellectual Challenge of CSCW: The Gap between Social Requirements and Technical Feasibility. Hum.-Comput. Interact. 15, 2 (sep 2000), 179–203. https://doi.org/10.1207/S15327051HCI1523_5
[2] P Agre. 1997. Toward a critical technical practice: Lessons learned in trying to reform AI in Bowker, G., Star, S., Turner, W., and Gasser, L., eds, Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide, Erlbaum (1997).
[3] Evgeni Aizenberg and Jeroen Van Den Hoven. 2020. Designing for human rights in AI. Big Data & Society 7, 2 (2020), 2053951720949566.
[4] Saleema Amershi. 2020. Toward Responsible AI by Planning to Fail. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Virtual Event, CA, USA) (KDD ’20). Association for Computing Machinery, New York, NY, USA, 3607. https://doi.org/10.1145/3394486.3409557
[5] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabih, Alberto Barbado, Salvador Garcia, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion 58 (2020), 82–115.
[6] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing 7, 1 (Oct. 2019), 2–11. https://doi.org/10.1609/hcomp.v7i1.5285
[7] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–16.
[8] Richard Benjamins, Alberto Barbado, and Daniel Sierra. 2019. Responsible AI in design practice. arXiv preprint arXiv:1909.12838 (2019).
[9] Philip AE Brey. 2012. Anticipatory ethics for emerging technologies. NanoEthics 6, 1 (2012), 1–13.
[10] Gregor Broll and Steve Benford. 2005. Seamless Design for Location-Based Mobile Games. In Entertainment Computing - ICEC 2005, Fumio Kishino, Yoshifumi Kitamura, Hirokazu Kato, and Noriko Nagata (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 155–166.
[11] Matthew Chalmers. 2003. Seamless Design and Ubicomp Infrastructure.
[12] Matthew Chalmers, Andreas Dieberger, Kristina Höök, and Åsa Rudström. 2004. Social Navigation and Seamless Design. In In Japanese Journal of Cognitive Science, Special Issue on Social Navigation. 171–181.
[13] M. Chalmers, I. MacColl, and M. Bell. 2003. Seamful design: showing the seams in wearable computing. In 2003 IEE Eurowearable. 11–16. https://doi.org/10.1049/ic:20030140
[14] Kathy Charmaz. 2014. Constructing Grounded Theory (Introducing Qualitative Methods series) 2nd Edition. Sage, London.
[15] Leszek Chybowski, Katarzyna Gawdzińska, and Valeri Souchkov. 2018. Applying the anticipatory failure determination at a very early stage of a system’s development: overview and case study. Multidisciplinary Aspects of Production Engineering 1 (2018).
[16] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2021. Stakeholder Participation in AI: Beyond” Add Diverse Stakeholders and Stir”. arXiv preprint arXiv:2111.01122 (2021).
[17] Wesley Hanwen Deng, Manish Nagireddy, Michelle Seng Ah Lee, Jatinder Singh, Zhiwei Steven Wu, Kenneth Holstein, and Haiyi Zhu. 2022. Exploring How Machine Learning Practitioners (Try To) Use Fairness Toolkits. arXiv preprint arXiv:2205.06922 (2022).
[18] 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 Designing Interactive Systems Conference 2021. 1591–1602.
[19] Jonathan Dodge, Q Vera Liao, Yunfeng Zhang, Rachel KE Bellamy, and Case Dugan. 2019. Explaining models: an empirical study of how explanations impact fairness judgment. In Proceedings of the 24th international conference on intelligent user interfaces. 275–285.
[20] Graham Dove, Kim Halskov, Jodi Forlizzi, and John Zimmerman. 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI ’17). Association for Computing Machinery, New York, NY, USA, 278–288. https://doi.org/10.1145/3025453.3025739

[21] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–19.

[22] Upol Ehsan and Mark O Riedl. 2020. Human-centered explainable ai: Towards a reflective sociotechnical approach. In International Conference on Human-Computer Interaction. Springer, 449–466.

[23] Upol Ehsan and Mark O Riedl. 2021. Explainability pitfalls: Beyond dark patterns in explainable AI. arXiv preprint arXiv:2109.12480 (2021).

[24] Upol Ehsan, Koustuv Saha, Munmun De Choudhury, and Mark O Riedl. 2023. Charting the Sociotechnical Gap in Explainable AI: A Framework to Address the Gap in XAI. Proceedings of the ACM on Human-Computer Interaction 7, CSCW1 (2023), 1–32.

[25] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The Algorithmic Imprint. In 2022 ACM Conference on Fairness, Accountability, and Transparency. 1303–1317.

[26] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Martina Mara, Marc Streit, Sandra Wachtler, Andreas Rienzer, and Mark O Riedl. 2021. Operationalizing human-centered perspectives in explainable AI. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. 1–6.

[27] William W Gaver, Jacob Beaver, and Steve Benford. 2003. Ambiguity as a resource for design. In Proceedings of the SIGCHI conference on human factors in computing systems. 233–240.

[28] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. Commun. ACM 64, 12 (2021), 86–92.

[29] Bhavya Ghai, Q Vera Liao, Yunfeng Zhang, Rachel Bellamy, and Klaus Mueller. 2021. Explainable active learning (xal) toward ai explanations as interfaces for machine teachers. Proceedings of the ACM on Human-Computer Interaction 4, CSCW3 (2021), 1–28.

[30] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Spector, and Lalana Kagal. 2018. Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). IEEE, 80–89.

[31] Ana Valeria González, Gagan Bansal, Angela Fan, Yashar Mehdad, Robin Jia, and Srinivasan Iyer. 2021. Do Explanations Help Users Detect Errors in Open-Domain QA? An Evaluation of Spoken vs. Visual Explanations. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, 1103–1116. https://doi.org/10.18653/v1/2021.findings-acl.95

[32] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Forse Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. ACM computing surveys (CSUR) 51, 5 (2018), 1–42.

[33] David Gunning, Mark Stefić, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. Science Robotics 4, 37 (2019).

[34] Aaron Halfaker and R Stuart Geiger. 2020. Ores: Lowering barriers with participatory machine learning in wikipedia. Proceedings of the ACM on Human-Computer Interaction 4, CSCW2 (2020), 1–37.

[35] Amy Heger, Liz B. Marquis, Mihaela Vorvoreanu, Hanna Wallach, and Jennifer W. Vaughan. 2022. Understanding Machine Learning Practitioners’ Data Documentation Perceptions, Needs, Challenges, and Desiderata. Proceedings of the ACM on Human-Computer Interaction, 6, CSCW2 (2022), 1–30.

[36] Nicole Henselbach. 2022. Undoing Seamlessness: Exploring Seams for Critical Visualization. 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 364, 7 pages. https://doi.org/10.1145/3491101.3519703

[37] Lars Erik Holmquist. 2017. Intelligence on Tap: Artificial Intelligence as a New Design Material. Interactions 24, 4 (jun 2017), 28–33. https://doi.org/10.1145/3085571

[38] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–16.

[39] Matthew K Hong, Adam Fournier, Derek DeBellis, and Saleema Amershi. 2021. Planning for natural language failures with the ai playbook. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–11.

[40] Sungsoo Ray Hong, Jessica Hullman, and Enrico Bertini. 2020. Human factors in model interpretability: Industry practices, challenges, and needs. Proceedings of the ACM on Human-Computer Interaction 4, CSCW1 (2020), 1–26.

[41] Imoh M Ilevbare, David Probert, and Robert Phaal. 2013. A review of TRIZ, and its benefits and challenges in practice. Technovation 33, 2-3 (2013), 30–37.
[42] Sarah Inman and David Ribes. 2019. “Beautiful Seams”: Strategic Revelations and Concealments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI ’19). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300508

[43] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C Ahn, Thomas H McCoy, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–14.

[44] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 353–362.

[45] Pariya Kashfi, Agneta Nilsson, and Robert Feldt. 2017. Integrating User experience practices into software development processes: implications of the UX characteristics. PeerJ Computer Science 3 (10 2017), e130. https://doi.org/10.7717/peerj-cs.130

[46] Harmanpreet Kaur, Eytan Adar, Eric Gilbert, and Cliff Lampe. 2022. Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory. arXiv preprint arXiv:2205.05057 (2022).

[47] Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and Douglas Eck. 2019. Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI EA ’19). Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299059

[48] Mark T. Keane and Barry Smyth. 2020. Good Counterfactuals and Where to Find Them: A Case-Based Technique for Generating Counterfactuals for Explainable AI (XAI). In Case-Based Reasoning Research and Development (Lecture Notes in Computer Science), Ian Watson and Rossina Weber (Eds.). Springer International Publishing, Cham, 163–178. https://doi.org/10.1007/978-3-030-58342-2_11

[49] Paul M Leonardt. 2013. Theoretical foundations for the study of sociomateriality. Information and organization 23, 2 (2013), 59–76. Publisher: Elsevier.

[50] 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. 1–15.

[51] Q Vera Liao and Kush R Varshney. 2021. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. arXiv preprint arXiv:2110.10790 (2021).

[52] Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI. arXiv preprint arXiv:2206.10847 (2022).

[53] Zachary C Lipton. 2016. The myths of model interpretability. arXiv preprint arXiv:1606.03490 (2016).

[54] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising contestability: Perspectives on contesting algorithmic decisions. Proceedings of the ACM on Human-Computer Interaction 5, CSCW1 (2021), 1–25.

[55] Sarah Inman and David Ribes. 2019. “Beautiful Seams”: Strategic Revelations and Concealments. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI ’19). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300508

[56] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C Ahn, Thomas H McCoy, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. 2021. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 1–14.

[57] Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. 2021. Algorithmic recourse: from counterfactual explanations to interventions. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 353–362.

[58] Pariya Kashfi, Agneta Nilsson, and Robert Feldt. 2017. Integrating User experience practices into software development processes: implications of the UX characteristics. PeerJ Computer Science 3 (10 2017), e130. https://doi.org/10.7717/peerj-cs.130

[59] Harmanpreet Kaur, Eytan Adar, Eric Gilbert, and Cliff Lampe. 2022. Sensible AI: Re-imagining Interpretability and Explainability using Sensemaking Theory. arXiv preprint arXiv:2205.05057 (2022).

[60] Claire Kayacik, Sherol Chen, Signe Noerly, Jess Holbrook, Adam Roberts, and Douglas Eck. 2019. Identifying the Intersections: User Experience + Research Scientist Collaboration in a Generative Machine Learning Interface. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland Uk) (CHI EA ’19). Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290607.3299059

[61] Mark T. Keane and Barry Smyth. 2020. Good Counterfactuals and Where to Find Them: A Case-Based Technique for Generating Counterfactuals for Explainable AI (XAI). In Case-Based Reasoning Research and Development (Lecture Notes in Computer Science), Ian Watson and Rossina Weber (Eds.). Springer International Publishing, Cham, 163–178. https://doi.org/10.1007/978-3-030-58342-2_11

[62] Paul M Leonardt. 2013. Theoretical foundations for the study of sociomateriality. Information and organization 23, 2 (2013), 59–76. Publisher: Elsevier.

[63] 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. 1–15.

[64] Q Vera Liao and Kush R Varshney. 2021. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. arXiv preprint arXiv:2110.10790 (2021).

[65] Q Vera Liao, Yunfeng Zhang, Ronny Luss, Finale Doshi-Velez, and Amit Dhurandhar. 2022. Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI. arXiv preprint arXiv:2206.10847 (2022).

[66] Zachary C Lipton. 2016. The myths of model interpretability. arXiv preprint arXiv:1606.03490 (2016).

[67] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising contestability: Perspectives on contesting algorithmic decisions. Proceedings of the ACM on Human-Computer Interaction 5, CSCW1 (2021), 1–25.

[68] Michael Madaio, Lisa Egede, Hariharan Subramonyam, Jennifer Wortman Vaughan, and Hanna Wallach. 2022. Assessing the Fairness of AI Systems: AI Practitioners’ Processes, Challenges, and Needs for Support. Proceedings of the ACM on Human-Computer Interaction 6, CSCW1 (2022).

[69] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. 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, 1–14. https://doi.org/10.1145/3313831.3376445

[70] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 1–14.

[71] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish. 2021. Algorithmic Impact Assessments and Accountability: The Co-Construction of Impacts. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (Virtual Event, Canada) (FAccT ’21). Association for Computing Machinery, New York, NY, USA, 735–746. https://doi.org/10.1145/3442188.3445935

[72] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency. 220–229.

[73] Shweta Narkar, Yunfeng Zhang, Q Vera Liao, Dakuo Wang, and Justin D Weisz. 2021. Model LineUpper: Supporting Interactive Model Comparison at Multiple Levels for AutoML. In 26th International Conference on Intelligent User Interfaces. 170–174.

[74] Tommy Nilsson, Carl Hogsden, Charith Perera, Saeed Aghaei, David J Scruton, Andreas Lund, and Alan F Blackwell. 2016. Applying seamless design in location-based mobile museum applications. ACM Transactions on Multimedia
[84] Arthur B VanGundy. 1984. Brain writing for new product ideas: an alternative to brainstorming. *Journal of Consumer Marketing* (1984).

[85] Janet Vertesi. 2014. Seamful spaces: Heterogeneous infrastructures in interaction. *Science, Technology, & Human Values* 39, 2 (2014), 264–284.

[86] Mihaela Vorvoreanu, Amy Heger, Samir Passi, Shipi Dhanorkar, Zoe Kahn, and Ruotong Wang. 2023. *Responsible AI Maturity Model*. Technical Report MSR-TR-2023-26. Microsoft. https://www.microsoft.com/en-us/research/publication/responsible-ai-maturity-model/

[87] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[88] Mark Weiser. 1994. Creating the invisible interface: (invited talk). In *Proceedings of the 7th annual ACM symposium on User interface software and technology*. 1.

[89] 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.

[90] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13.

[91] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 295–305.
A APPENDIX

A.1 High Resolution Picture of the Whiteboard

Fig. 5. A higher resolution picture of Figure 2 showing a screenshot of the virtual whiteboard used for the seamful XAI design activity in the study, with zoomed-in examples. **Area 1**: Envisioning breakdown (Step 1). In the study, we provided sample breakdowns, which participants could either use directly or get inspiration for their own envisioning. **Area 2**: Anticipating & crafting seams (Step 2). We provided guiding prompts for crafting the seams can procedurally guide the process of writing down the seam in an effective manner. We also shared exemplary seams for each stage of the AI lifecycle framework. **Area 3**: Designing with seams. We asked participants to articulate their reasoning for choosing a seam and tag which user goals the selected seam can support for augmenting user agency. If viewed on a PDF viewer like Adobe Acrobat, the reader can zoom in as necessary to read the text on this board.
### A.2 High Resolution Picture of the Bird’s-eye View of All Seams

Fig. 6. A higher resolution picture of Figure 4 showing a bird’s eye view of all seams from all participants for all breakdowns along all AI lifecycle. Breakdowns are the columns while the AI lifecycle stages are on the rows. The dots above breakdowns provide the number of times a particular breakdown was used. All the seams appearing below this column were crafted in connection to that particular breakdown. The dots next to each lifecycle stage showcase the number of seams crafted for that stage. Each row has the seams along with the justifications for how the seams enhances agency in the sticky notes. If viewed on a PDF viewer like Adobe Acrobat, the reader can zoom in as necessary to read the text on this board.

Received January 2023; revised July 2023; accepted November 2023