

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 fallouts from AI mistakes. While Explainable AI (XAI) has predominantly tackled algorithmic opaqueness, we propose that seamful design can foster Human-centered XAI by strategically revealing sociotechnical and infrastructural mismatches. We introduce the notion of *Seamful XAI* by (1) conceptually transferring “seams” to the AI context and (2) developing a design process that helps stakeholders design with seams, thereby augmenting explainability and user agency. We explore this process with 43 AI practitioners and users, using a scenario-based co-design activity informed by real-world use cases. We share empirical insights, implications, and critical 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, AI ethics, human-AI interaction, explanations

1 INTRODUCTION

“*We make it so easy to use that you won’t even have to think about it!*” exclaimed the VP of 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 [10, 39]—can be a double-edged sword. On the positive side, it promotes simplicity and ease of use [32, 39]. On the negative side, it conceals and abstracts important factors about the AI 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 uncritical AI acceptance or lack of agency for the end user. What can we do differently? How do we move beyond seamless AI? And what can 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. To that end, we can learn from the concept of *seamful design* [9, 10], a complement (not

Authors’ addresses: Upol Ehsan, Georgia Institute of Technology, Atlanta, GA, USA; Q. Vera Liao, Microsoft Research, Montreal, , Canada; Samir Passi, Microsoft, , USA; ; Mark O. Riedl, Georgia Institute of Technology, Atlanta, GA, USA, ; Hal Daumé III, University of Maryland & Microsoft Research, , USA, .

a rival) of seamless design [39]. With conceptual roots in Ubiquitous Computing, seamful design deals with seams—“mismatches”, “gaps”, “cracks”, “uncertainties” that arise when a technology is integrated into a physical world or a deployment context [8, 74]. 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 example, a classic example of seamful design is a “seamful map” revealing varying strength of WiFi coverage in the deployed 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.

In this paper, we introduce the notion of *Seamful Explainable AI* by transferring seamful design into AI and exploring its practical intersection with the field of Explainable AI (XAI). 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. While there has been commendable work on what seams can do in Ubiquitous Computing [8–11] and increased calls for seamful design in AI [21, 43], some key questions remain unexplored: what are seams in AI? How may we identify seams in AI? More importantly, how may we design with the seams to augment agency and explainability? To address these questions, we operationalize the notion of Seamful XAI by (1) conceptually transferring “seams” to the context of AI and (2) developing a design process that helps stakeholders identify and design with seams to enhance AI explainability and user agency.

Our conceptualization and operationalization is informed by the lens of Human-centered XAI [20, 23, 49], which prioritizes both people’s needs and end-goals for understanding AI, and sociotechnical and infrastructural perspectives. Similar to seamful design, HCXAI acknowledges that AI systems are 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—needs more than just algorithmic transparency.

We broadly conceptualize seams in AI as *mismatches and cracks between assumptions made in designing and developing the AI system and the reality of the deployment context*. Through this lens of revealing mismatches and with the goals of mitigating breakdowns and supporting user agency, seamfulness can offer an operational design space that facilitates explainability. The following example, inspired by real-world cases shared by our informants in the context of automated lending, can illustrate how revealing seamful information can help achieve these goals:

The loan application for Ahmed, a great customer, is rejected by the AI-powered system. When Nadia, the loan officer, checks the AI’s feature-level explanations, nothing seems amiss, making it difficult for her to decide whether to overrule the system. Behind the scenes, there is a mismatch (seams): when the system was last updated in 2020, applicants could only have three active loans, but regulatory changes in 2021 allow up to five. Ahmed currently has four active loans, which triggered a denial; however, if Nadia were aware of this mismatch (seam), she would be empowered to trigger a manual underwriting review.

The type of understanding afforded by seams 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 explanations. This holistic understanding equips Nadia to seek actions and contest the AI.

Furthermore, we take a *proactive* design stance and encourage practitioners to anticipate and design with seams to empowers users. This stance is in contrast with prior work that addresses seams after they appear in deployment [8, 10, 62]. Our focus on XAI demands a proactive outlook because XAI systems are put in place in anticipation of the need to reconcile incomplete user understanding of an AI system’s workings. Our stance also aligns with the emerging

Responsible AI (RAI) practices that emphasize proactively anticipating and mitigating individual and societal harms of AI [22, 31, 55, 58, 71]. Thus, our work also aims to introduce a new perspective and an actionable approach to this currently formidable RAI goal of anticipating and mitigating AI breakdowns.

In the rest of this paper, we situate the notion of Seamful XAI in the broader literature in seamful design, XAI, and RAI. Next, we share the what, how, and why of our proposed design process. Then, we share our methods for an empirical study to explore both the design process of, and the values afforded by, seamful XAI, using a scenario-based co-design activity. Finally, we share findings from the study with 43 users and practitioners of diverse roles. This paper presents a foundational step toward Seamful XAI through conceptual, methodological, and empirical advances:

- We conceptually transfer the notion of seams from Ubicomp to AI and introduce the notion of Seamful XAI.
- We operationalize seamful XAI by introducing a design process that aims to help designers and developers augment user agency and understanding by proactively identifying and designing with seams.
- We share empirical insights and implications from a formative scenario-based co-design study with 43 AI practitioners and end-users. We highlight how our process helped participants anticipate and design with seams, how designing seamful XAI can enhance explainability and user agency, foster multidisciplinary collaboration, and facilitate proactive harm mitigation. We also share challenges and tensions around the process.

2 BACKGROUND AND RELATED WORK

2.1 Seamless and seamful design

Our understanding of seamless and seamful design follows a long tradition of research in the field of Ubiquitous Computing that explores how users live with heterogeneous technological systems [8–11, 24]. Seamlessness is almost always considered a good, even commendable, design ideal [39]. 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* [32, 39]. Technological systems, however, are complex sociotechnical assemblages. They hardly ever exist in isolation, overlapping and intertwining with other components, practices, and systems in consequential ways. Seamless experiences thus require shielding users from the intricate logic and messy machinery driving today’s technologies [74, p. 8]. Unsurprisingly, seamless experiences and black-boxed systems go hand in hand.

While seamless design emphasizes ease of use and concealment of messiness, seamful design prioritizes “configurability, user appropriation, and revelation of complexity, ambiguity, or inconsistency” [39, p. 1]. Seams—mismatches, gaps, and uncertainties—are the inevitable “cracks and bumps” between heterogeneous components of socio-technical systems [8, 74]. Seamful design focuses on leveraging them in ways that help users make better decisions around technology use [9, 11, 32, 39]. Central to seamful design in the notion of *strategic revelations and concealment* [39, 85] thus, 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 [39]. The question, however, isn’t just about *what* seams to reveal (and what to hide), but also *how*, and *when* to reveal them [39].

Our work builds on and contributes to these different strands of scholarship, by situating seamful design values and practices within AI technologies. 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 system explanations is paramount given the complexity and opaqueness of AI technologies, and their risks in creating harms at scale without human scrutability. This pressing need gives rise to the field of explainable AI (XAI) [29], 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, 27, 28, 51]), they typically aim to address user questions such as “how does the model make decisions” or “why does the model make a particular decision” [48] through means such as revealing how the model weighs different features, and what rules it follows.

Despite its progress, the XAI field has received a fair amount of criticisms. The first is its current gaps from explanations that humans seek and produce, as long studied in the social sciences [46, 52, 57, 59] and philosophy [63]. Second, 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 technique can risk creating over-trust and over-reliance when model predictions are wrong [6, 68, 69, 88]. Technical explanations may in fact burden people who do not have the capability or motivation to engage [26, 80], and risk impairing their user experience. These pitfalls have been attributed to a techno-centric focus of the current XAI field [21, 49], pursuing technical advancement without a clear understanding of and aiming to support the end-goals that people seek explanations for.

In responses to these limits, an area of work that takes human-centered approaches to explainable AI has emerged within the HCI community [20, 23, 49, 84]. These works center the development of XAI technologies around people’s needs and end-goals, whether the end-goal is improving the model [61], making better decisions [40, 86], assessing model biases [17], seeking recourse [41], contesting the model [53], and so on [50, 79]. With this shift of focus, these works often expand the design space of XAI beyond algorithmic explanations, providing additional information as well as various interaction-level support such as interactive model interrogation and visualizations [33, 61, 86].

Our work is informed by emerging research in HCXAI that takes a sociotechnical lens [20]. This body of work challenges the algorithm-centrism in XAI where the predominant focus tends to be on revealing the algorithmic processes and not on the sociotechnical factors that govern the use of AI [23, 49]. Dhanorkar et al. [16] highlight how explainability needs evolve and depends on who needs what and when. Highlighting an XAI blind spot where the social factors are ignored, Ehsan et al. [19] 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.

Our work similarly adopts a sociotechnical perspective to expand the XAI design space and takes a goal-oriented stance. Focusing on AI supporting human decision-making (as opposed to complete automation), we aim to support the end-goal of actionability for decision-making [48] and contestability when AI is wrong [53].

2.3 Responsible AI in Practice

Our work is also informed by, and contributes to, the responsible AI (RAI) practices. RAI is concerned with putting theoretical principles of AI ethics into practice [7, 71, 76]. In principle, RAI works call for explicit 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, 65, 72]. However, recent work studying how practitioners deal with RAI issues on the ground, including AI fairness [15, 35, 54, 65], transparency [31, 37, 48], accountability [70], and overall harms mitigation [4, 71], report that practitioners grapple with tremendous challenges. Proactively anticipating potential harms for complex systems deployed in heterogeneous social contexts is inherently challenging. Prior

work called for inter-disciplinary approaches that involve technical experts, ethics experts, social scientists, and user experience (UX) professionals [42, 44, 58, 66, 67], and multi-stakeholder approaches by also involving end-users and impacted communities to align RAI with their values [14, 30, 77], as well as more actionable methods, processes, and tools to lower the barriers and empower individuals to engage in RAI practices [31, 48, 56]. 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 [64, 71].

We situate our proposed design process in the RAI ecosystem by taking a proactive, inter-disciplinary and multi-stakeholder approach that is intended to involve practitioners of diverse roles and end-users. To make the design process adoptable by practitioners, we draw valuable lessons from—and extend—recent work that supports RAI practices through structured but also flexible “scaffolding”, such as guiding frameworks [19, 25, 60], checklists [56], and processes [48, 70].

3 SEAMFUL XAI DESIGN PROCESS

At a high level, 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 both allows us to solidify the conceptual transfer and enable others, especially AI practitioners, 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. The proposed process is intended to be actionable, attending to the procedural aspect and providing various scaffolds for practitioners to follow.

Broadly, our design process has three steps (Fig. 1): envisioning breakdowns, anticipating the seams, and designing with seams. 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 the 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” [36] of user workflow.

Starting the process with breakdowns is motivated by the fundamental notion in seamful design literature that seams are revealed through breakdowns [8, 9, 39]. Articulating concrete breakdowns will offer anchoring points to trace back



Fig. 1. An overview of the Seamful XAI design process with key questions relevant to each step.

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) [13], which is a part of the TRIZ (Theory of Inventive Problem Solving) framework [38]. 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. [81] Once we have the causes for failures, we can then proactively search for solutions to prevent it. Thus, it provides a tractable structure for the challenging task of anticipating the 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.

Motivated by an infrastructural perspective [47, 83], 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 infrastructure, we went beyond the traditional AI lifecycle stages (data collection, model testing, and training) and expands into the situated use of AI (e.g., input into the AI, use of the AI’s output, and system maintenance). The formation of these stages are informed by human-centered AI research studying AI development practices [16, 23, 79] 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 consume effectively to support their decision-making. To help with this

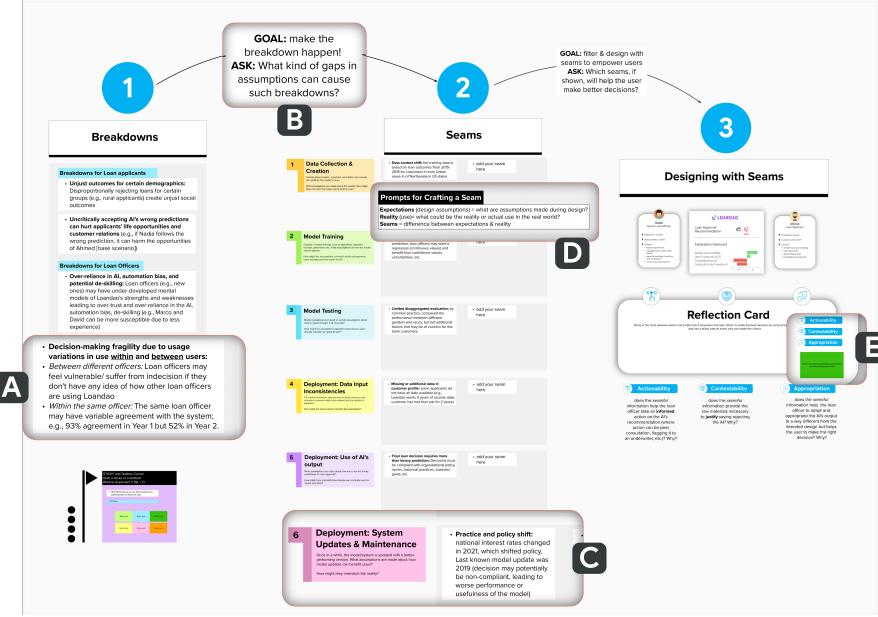


Fig. 2. A screenshot of the virtual whiteboard used for the seamful XAI design activity in the study, with zoomed-in examples. **Area 1:** Envisioning breakdown (Step 1). In the study, we provided sample breakdowns (A), which participants could either use directly or get inspiration for their own envisioning. **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.)

part, 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 and reality. Formative pilots showed that these prompts not only simplify the generation process but also make the generated seams more actionable and consistent across participants. The incentive for this scaffold also stems from the goal of seamful design—merely revealing seams is not enough; the revelation should 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 step aligns our process with goals 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. Below we share the guiding questions that should 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 helping the user to make the right decision? 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).

We also offer some practical suggestions for the design process based on what we learned from pilots:

When generating the seams: Instead of doing it collectively, the process can be performed individually by different stakeholders. We recommend to have one facilitator who serves as the junction point for these conversations and diverse perspective. This alleviates the challenges of getting multiple teams in the room at the same time. It can also mitigate conformity pressure and group thinking that can happen when brainstorming. Stakeholders can also engage in brainwriting [82] 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 the business goals or organizational values.

4 METHODS

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 [73].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 seamful 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

To put our design process into practice, we need a scenario of an AI use case, which should include both user personas¹ 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 by envisioning their own breakdowns, seams, and the impact of seamful designs.

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,

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.



Fig. 3. A visual presentation of the lending scenario used in our study showcasing the “backstory” of each persona. The different loan officer personas were used by participants to think of variations in the scenario to generate new breakdowns and seams.

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 experiences 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.. 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 Loandao, an 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, Loandao 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 varied in both organizational and technological experience from Nadia (Fig. 3). For instance, some participants in the study considered Marco’s use of Loandao 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 global pandemic using video conferencing and online whiteboarding tools to conduct the activities virtually. 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

Table 1. Details on study participants: Organizational roles and amount of experience.

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

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 the brainstorming activity, we provided a list of four realistic breakdowns in the lending scenario as well as exemplar seams, all informed by our 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 by discussing the 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 explainable AI, their role, domain, geographic location, and number of years of experience, along with questions where they can provide examples. Out of 258 entries, we recruited 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.

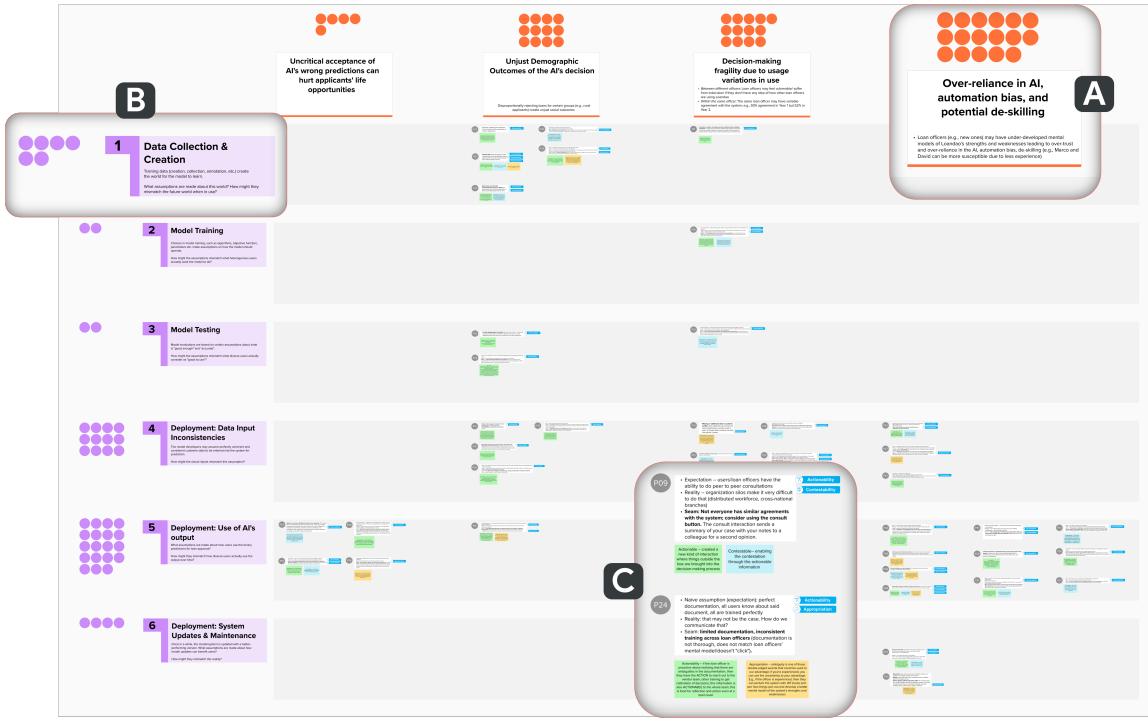
4.4 Qualitative Analysis

Overall, we analyzed roughly 2795 minutes of transcribed videos from the 43 study sessions as informed by the grounded theory approach [12, 78]. 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, e.g., participants saying the process is ‘fun and engaging’). 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.

Beyond the analysis of the videos and their transcripts, 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 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.



5.1.2 Anticipating and crafting seams. Participants' reflection comments suggested 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 was "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). Initially participants found the task of writing out the seams "an intimidating task but writing down the [ideal design] expectations and the reality [real use]" made the process "reproducible and doable even for first-timers" (P17, P32):

I like this framing because nowadays people design for the ideal, the golden path...but technology *never* works that way. 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, emphasis added)

Having the process broken down "into bite sized pieces" made the "learning curve [of writing the seam] less steep, promot[ing] self-reliance" (P06, P07). This participant encapsulates the notion of a "beautiful seam" [39] in AI:

Writing out the expectations and reality like this makes me feel like I'm gather raw materials for the seam. Now, the job is to stitch the information in the right way...Framing the seam in a way that helps the AI user makes for a *beautiful seam*. (P34, emphasis added)

5.1.3 Designing with Seams. **Every participant was also able to design with the seams they crafted:** to strategically select from the available seams those that would be worth revealing to augment a users' 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). This participant articulates the value added by the filtering activities at a process level:

Most frameworks I've used *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 (step 3) and figure out what to *do* with them *systematically*. This reminds me of one of the diamonds of the double diamonds of user-centered design...[steps] 1-2 being the generative part and 2-3 as the filtering part. (P06, UXR, emphasis added)

5.1.4 Transferring and teaching about seams. In addition to being able to craft and design with 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:

The mark of a good design process is how transferable it is. I’m surprised how easily I was able to 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 as well. (P40, a senior Data Scientist)

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. Below we share one example of a transfer case, where P03 provided the usage context first, followed by the conceptualization of seams and how seamful XAI would bring benefit into this context. This transfer example covers an AI-powered international aid allocation algorithm that did not perform as expected:

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, the algorithm ended up being biased anyway [which is a breakdown]...Turns out that there was model leakage [in reality] where the algorithm latched on to neighborhood data like ZIP codes, found some weird correlations, and was discriminating based on that...the seam here was the mismatch between what we expected around how blinding the algorithm would make it fairer and how it actually worked in use... if we were aware of the mismatch, then it’d have helped us not just explain why the bias existed but also help our users address the AI’s flaws. (P03)

This example illustrate 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 provide “peripheral vision that can highlight AI blind spots otherwise 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. This process makes what’s around the box less opaque...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, highlights its fallibility, and showcases the strengths and weaknesses of the system**, which can facilitate **calibration of trust and reliance in 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). If a user is “aware of the seam, then [they] have a better understanding, a clearer picture of the contours of the system’s performance—what it can do vs. what it can’t” (P06). 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 what 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:

By showing the seam, you bust the mythology that AI is God-like...it reduces mindless AI acceptance. Awareness of the limitations helps me understand its strengths and weaknesses, making it more explainable. The improvement in explainability seems to be an implicit outcome of seamful design. (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**. This participant captures the point effectively:

The explainability we get with seams is *very different*..For example, if I know there’s annotator bias in the data, I can do forward causal inference and see if the path was the reason for the biased AI outcome. It helps me understand 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, which is for explainability. (P28, an XAI researcher, emphasis added)

Having information of both why and why-not can promote “holistic understanding and counterfactual reasoning mechanisms” (P24). Improvement in counterfactual reasoning can lead to gains in explainability [19, 45].

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 open spaces for how humans interact with systems” (P32). The new interactions can be 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). Equating perfectionist tendencies with seamlessness, participants reflected on how the

“western obsession with perfection often leads to an unachievable seamless ideal” (P13) and how seamfulness can be empowering, further encapsulated by this participant:

Seamless AI design is a take it or leave it paradigm...user has no voice. By identifying seams along the AI’s lifecycle, users evaluate what to do with it—having that choice, intrinsically, is empowering. (P18)

Information in seams can **convert “unknown unknowns” to “known unknowns”**, which can empower users to know “where to start an investigation; for instance, 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 my 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 addressing at least one of the three dimensions. Below we synthesize participants’ justification.

Actionability comes to the forefront when we focus on what people can *do* with the information in the seams. A major way in which seam addressed actionability was when *the seamful 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” or “flagging for a peer consultation to double check [the user’s] intuition” or “recommend customer 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 AI “decisions are surprising” (P36, P21). Contesting a decision is a critical path of exercising user voice in algorithmic decision-making [3, 53, 75]. One of the key themes around contestability was when the seamful 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 “most experienced users reject the AI”, it provides them the “right ammunition to say no to the AI” (P12, P30, P27).

Appropriation supports the “how” when users are dissatisfied with the current system, that is, “*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). Participants underscored that there is a “need to have the right expertise and skill set to exercise the appropriation properly...[because]...with great power comes great responsibility” (P26). 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, which encompass all parts of the AI lifecycle in our study. 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**. There is an intrinsic value of generating seams. The generated list of 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 or harm assessments in a resourceful way" (P21). The proactive anticipation affords us a unique value:

There's a dual value proposition—once you've the seams, you can neutralize many of them simply because you identified them ahead of time... There'll be some unfixable seams like regulatory policy changes making the AI non-compliant. In those cases, you can leverage these seams resourcefully to help with better decision-making. So basically you get two for the price of one. (P33)

Typical "AI Ethics and Responsible AI frameworks call for envisioning harms" (P03); what is unique about seamful design in XAI is that when seams are revealed, "they aren't shown for the sake of showing them...they are strategically used in opportunistic ways to help the user" (P14). Thus, the seamful design process empowers users "not just discover the problem but *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

In participants' reflection on the seamful XAI design process and how it may be applied in their own work, several themes emerged about how they view the 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 than high-level AI Ethics guidelines" (P38), elaborated by this participant:

Why the heck didn't we think like this before? I've personally used tons of AI Ethics tools. Everything's so high level there, which my clients hate. This [process]is specific and customers love that...I'd go as far as to say that 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:

Risk and harm can be nebulous words...When we make things concrete, it's actionable. I love how we started with a specific breakdown, worked through the [AI] lifecycle, and then used the seams...What's unique here is that the lifecycle allows you to pinpoint where things could go wrong...the concreteness is immensely helpful because we've no clue how to enact the high level AI Ethics guidelines. (P25)

Moreover, our process has "a nice balance between top-down and bottom-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 in contrast to "the traditional impact assessments that are too top down" (P03).

5.3.2 Fostering cross-disciplinary collaboration in a critically reflective manner. Participants believed that, when done right, the design process can facilitate **multi-stakeholder and interdisciplinary discussions**, provides a **inclusive and engaging way to lower the barriers for participation of critical reflective practices**.

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:

Typically, large tech teams work in silo... They lack visibility... Going through this deliberative process where we have end-to-end perspective can really us out. For example, if I know my product manager foresees a mismatch in deployment, as a data scientist, I can try to mitigate that at an upstream stage. It also makes it clear what is whose responsibility, which is important for accountability. (P41)

Second, participants appreciated how the activity can empower multidisciplinary stakeholders to communicate effectively “by providing a shared medium and language to communicate” (P14). They found the “vocabulary of seams helpful and terms like ‘breakdowns’ accessible” (P13). Many highlighted the potential to have inclusive interactions through the process. 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 the process “can level the playing field...on the board, everything is flat and everyone is important; a client-facing PM is the expert for post-deployment seams...the process shows how we’re all important” (P26), which can foster an inclusive discourse.

Third, participants appreciated the inclusive “co-design nature of the activity” (P30). Envisioning it as “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). This participant elaborates further:

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

Last, 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 the adversarial thinking participants did to anticipate seams by making the breakdowns happen:

I had so much fun doing the super villain thinking! As technologists, we’re the good guys trying to prevent bad outcomes...Getting a temporary license to be unshackled from that *hero expectation* is *freeing*—it gives us permission to not hold back ...That’s what makes this process so different. It’s not as dull and serious as the other stuff out there. This achieves the same goal without that heaviness...that’s effective. I can see this as a group activity people want to join! (P24, an AI Researcher, 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 impact of harms from “reckless and irresponsible AI innovations” (P33). Participants strongly resonated with “the paradigm shift in AI proposed by the seamful XAI process” (P21). Citing recent scandals around AI-powered technology, they highlighted how “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). This approach is not just proactive, but 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 envisioning seams from breakdowns and designing with them, the process “transforms the information into actionable insights to make things right” (P30). Participants highlighted how it was a “tractable proactive design method that is a sharp turn from the reactive AI design” (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). 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 what type of XAI models we need, understand where it 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 a “surefire way to stop all breakdowns” (P34):

Breakdowns will always happen...If we use this process, we can lessen the fallout from breakdowns. Think of this like thinking of designing a car’s crumple zone that saves lives during a crash. Seamful information can provide the tools users need to handle the breakdowns, which is the critical point. (P19)

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, highlight 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,... team mandates, and the customer’s interest” (P18) for the process to have the highest impact.

Second, at a process level, participants *initially struggled to craft or frame the seam’s information in a manner that would be useful to the end user*. “One thing that threw [some participants] off initially was how to talk about the gap in a way that would be constructive” (P10). 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, we were able to address this challenge through the scaffolding provided by the “crafting equation of expectations vs. reality” (P07). Participants felt “the equation was bang on... [and] boiling it down to a recipe really 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?... Something like this seam impacts 35% of the final AI decision vs. that seam impacts 10% would be a killer feature because then I can prioritize seams using both the impact and the stickers. (P04, a senior XAI focused UX researcher)

Second, there was a high 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). Moreover, accompanying the “so what?” elements with seamful information can “boost adoption because clients want answers and, more importantly, hand holding to get those answers” (P03). Adding action items, they felt, would “get users to the answer faster” (P08).

Third, there was a popular suggestion to *create a collection 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 a “low hanging yet rich source of seams...the limitations documentations that already exist in organizations” (P07). They argued that “turning the latent information in the limitations or risk documentation in tech companies into activated knowledge can be a resourceful way to go about seamful design” (P40).

6 DISCUSSION

The empirical study we conducted provides formative validation for the value of transferring the concept of seamful design into XAI, as well as our proposed design process to operationalize the concept. In this section, based on our empirical findings, we reflect on the roles of seamful XAI can play in expanding the design space of XAI and enabling RAI practices, as well 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 Human-centered XAI (HCXAI) by explicitly showing us how to use infrastructural mismatches (seams) in a resourceful way to augment traditional notions of 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. This understanding is crucial in complex decision-making scenarios. In 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 so in another, we can calibrate when to rely on it. This participant underscores the utility of understanding the ‘why-not’ part:

I really like the seam where the AI’s updates lagged bank policy changes...how it considered 4 loans to be a bad thing because of its training [that limits to 3 loans] when in reality the bank had relaxed its policies [for up to 5 loans]...Two important things happen if I know this [seamful] information. [First,] by giving the broader context, it explains why the AI couldn’t make the right call...[Second], it makes my trust on the AI nuanced [where] I know scenarios it will fail but are also where it may succeed...It’s no longer an all-or-nothing play...The AI’s 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: not only does this outlook provide a lens of analysis when it comes to explainability, but it also fosters “thinking of new interaction paradigms” (P28) in XAI systems that include seams.

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

As we learned in our findings, the seamful XAI design process enables us to enact RAI activities in a tractable and resourceful manner. We found that the combination of the focused and structured nature of the process along with the on-the-ground situated nature of seams enables participants to transform abstract concepts like harms into specific instances. There are three main ways in which our design process extends the reach and impact of RAI activities.

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 (causes for breakdowns), 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 seam 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 [10, 11, 39]. 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 [24].

6.3 Implications for Research & Practice

We share three implications the seamful outlook in XAI has on both research and practice. First, taking an infrastructural view, a seamful approach allows us to leverage the imperfections in AI systems in resourceful ways 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 inform how we think about trust calibration in AI, especially when addressing over-inflated expectations on AI performance. An acknowledgement of AI imperfections does not, however, entail that we give up and do not work on problems [62]. 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 pragmatic and resourceful in designing Human-AI interactions that leverage the imperfections to empower users such that they have the tools to deal with the fallout from breakdowns.

Second, a seamful XAI outlook appropriately calibrates the expectations of our agency as creators by helping us recognize that we cannot always be in control of our deployed systems. Given that AI is a tricky design material [18, 26, 34, 87] due to its stochasticity, changing, and non-deterministic nature, it is difficult to envision exactly how events unfold in use [74]. 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 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 design stance is a differentiating factor from the majority of prior work in seamful design where inventions dealt with seams in-situ [8, 10, 62]. The core implication of the unbridgeable nature of the “sociotechnical gap” [1]—the difference between what the system can technically afford and what users socially 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.

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.

For future iterations, we are inspired by Phil Agre’s notion of adopting a split identity, one where we simultaneously plan “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 specifically, and AI more broadly. We developed and validated a design process to help stakeholders anticipate, identify, filter, and design with seams. We made visible how the strategic revelation and concealment of sociotechnical and infrastructural seams enhances explainability, augments user agency, and, in fact, helps 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 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 Tabik, Alberto Barbado, Salvador García, 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, 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.
- [7] Richard Benjamins, Alberto Barbado, and Daniel Sierra. 2019. Responsible AI by design in practice. *arXiv preprint arXiv:1909.12838* (2019).
- [8] Gregor Broll and Steve Benford. 2005. Seamful 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.
- [9] Matthew Chalmers. 2003. Seamful Design and Ubicomp Infrastructure.
- [10] Matthew Chalmers, Andreas Dieberger, Kristina Höök, and Åsa Rudström. 2004. Social Navigation and Seamful Design. In *In Japanese Journal of Cognitive Science, Special Issue on Social Navigation*. 171–181.
- [11] 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>
- [12] Kathy Charmaz. 2014. *Constructing Grounded Theory (Introducing Qualitative Methods series) 2nd Edition*. Sage, London.
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 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.
- [17] Jonathan Dodge, Q Vera Liao, Yunfeng Zhang, Rachel KE Bellamy, and Casey 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.
- [18] 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>
- [19] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.
- [20] 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.
- [21] Upol Ehsan and Mark O Riedl. 2021. Explainability pitfalls: Beyond dark patterns in explainable AI. *arXiv preprint arXiv:2109.12480* (2021).
- [22] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The Algorithmic Imprint. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. 1305–1317.
- [23] Upol Ehsan, Philipp Wintersberger, Q Vera Liao, Martina Mara, Marc Streit, Sandra Wachter, Andreas Riener, and Mark O Riedl. 2021. Operationalizing human-centered perspectives in explainable AI. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, 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.
- [28] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51, 5 (2018), 1–42.
- [29] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science Robotics* 4, 37 (2019).
- [30] 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.

- [31] 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.
- [32] Nicole Hengesbach. 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>
- [33] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M Drucker. 2019. Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
- [34] 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>
- [35] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudík, 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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>
- [40] 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.
- [41] 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.
- [42] 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>
- [43] 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).
- [44] 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>
- [45] 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 Rosina Weber (Eds.). Springer International Publishing, Cham, 163–178. https://doi.org/10.1007/978-3-030-58342-2_11
- [46] Frank C Keil. 2006. Explanation and understanding. *Annu. Rev. Psychol.* 57 (2006), 227–254.
- [47] Paul M Leonardi. 2013. Theoretical foundations for the study of sociomateriality. *Information and organization* 23, 2 (2013), 59–76. Publisher: Elsevier.
- [48] 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.
- [49] Q Vera Liao and Kush R Varshney. 2021. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. *arXiv preprint arXiv:2110.10790* (2021).
- [50] 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).
- [51] Zachary C Lipton. 2016. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490* (2016).
- [52] Tania Lombrozo. 2012. Explanation and abductive inference. (2012).
- [53] Henrietta Lyons, Eduardo Veloso, and Tim Miller. 2021. Conceptualising contestability: Perspectives on contesting algorithmic decisions. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–25.
- [54] 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).
- [55] 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>

- [56] 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.
- [57] Bertram F Malle. 2006. *How the mind explains behavior: Folk explanations, meaning, and social interaction*. Mit Press.
- [58] 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>
- [59] Tim Miller. 2018. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* (2018).
- [60] 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.
- [61] Shweta Narkar, Yunfeng Zhang, Q Vera Liao, Dakun 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.
- [62] Tommy Nilsson, Carl Hogsden, Charith Perera, Saeed Aghaei, David J Scruton, Andreas Lund, and Alan F Blackwell. 2016. Applying seamful design in location-based mobile museum applications. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 12, 4 (2016), 1–23.
- [63] Andrés Páez. 2019. The pragmatic turn in explainable artificial intelligence (XAI). *Minds and Machines* 29, 3 (2019), 441–459.
- [64] Samir Passi. 2021. *Making Data Work: The Human and Organizational Lifeworlds of Data Science Practices*. Ph. D. Dissertation.
- [65] Samir Passi and Solon Barocas. 2019. Problem Formulation and Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (FAT* '19). Association for Computing Machinery, New York, NY, USA, 39–48. <https://doi.org/10.1145/3287560.3287567>
- [66] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science: Collaboration, Translation, and Accountability in Corporate Data Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 136 (nov 2018), 28 pages. <https://doi.org/10.1145/3274405>
- [67] Samir Passi and Phoebe Sengers. 2020. Making data science systems work. *Big Data & Society* 7, 2 (2020), 2053951720939605. <https://doi.org/10.1177/2053951720939605>
- [68] Samir Passi and Mihaela Vorvoreanu. 2022. *Overreliance on AI: Literature Review*. Technical Report MSR-TR-2022-12. Microsoft. <https://www.microsoft.com/en-us/research/publication/overreliance-on-ai-literature-review/>
- [69] Forough Poursabzi-Sangdeh, Daniel G Goldstein, Jake M Hofman, Jennifer Wortman Vaughan, and Hanna Wallach. 2021. Manipulating and measuring model interpretability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–52.
- [70] Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 33–44.
- [71] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 7 (apr 2021), 23 pages. <https://doi.org/10.1145/3449081>
- [72] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23.
- [73] Mary Beth Rosson and John M Carroll. 2009. Scenario based design. *Human-computer interaction. boca raton, FL* (2009), 145–162.
- [74] Gloire Rubambiza, Phoebe Sengers, and Hakim Weatherspoon. 2022. Seamless Visions, Seamful Realities: Anticipating Rural Infrastructural Fragility in Early Design of Digital Agriculture. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 451, 15 pages. <https://doi.org/10.1145/3491102.3517579>
- [75] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, Atlanta GA USA, 59–68. <https://doi.org/10.1145/3287560.3287598>
- [76] Ben Shneiderman. 2021. Responsible AI: Bridging from ethics to practice. *Commun. ACM* 64, 8 (2021), 32–35.
- [77] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. Participation is not a design fix for machine learning. *arXiv preprint arXiv:2007.02423* (2020).
- [78] Anselm Strauss and Juliet M. Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Techniques and Procedures*. Sage, New York.
- [79] Harini Suresh, Steven R Gomez, Kevin K Nam, and Arvind Satyanarayan. 2021. Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
- [80] Maxwell Szymanski, Martijn Millecamp, and Katrien Verbert. 2021. Visual, textual or hybrid: the effect of user expertise on different explanations. In *26th International Conference on Intelligent User Interfaces*. 109–119.
- [81] Christian M Thurnes, Frank Zeihsel, Svetlana Visnepolschi, and Frank Halfell. 2015. Using TRIZ to invent failures-concept and application to go beyond traditional FMEA. *Procedia engineering* 131 (2015), 426–450.
- [82] Arthur B VanGundy. 1984. Brain writing for new product ideas: an alternative to brainstorming. *Journal of Consumer Marketing* (1984).
- [83] Janet Vertesi. 2014. Seamful spaces: Heterogeneous infrastructures in interaction. *Science, Technology, & Human Values* 39, 2 (2014), 264–284.
- [84] 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.

- [85] Mark Weiser. 1994. Creating the invisible interface: (invited talk). In *Proceedings of the 7th annual ACM symposium on User interface software and technology*. 1.
- [86] 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.
- [87] 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.
- [88] 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

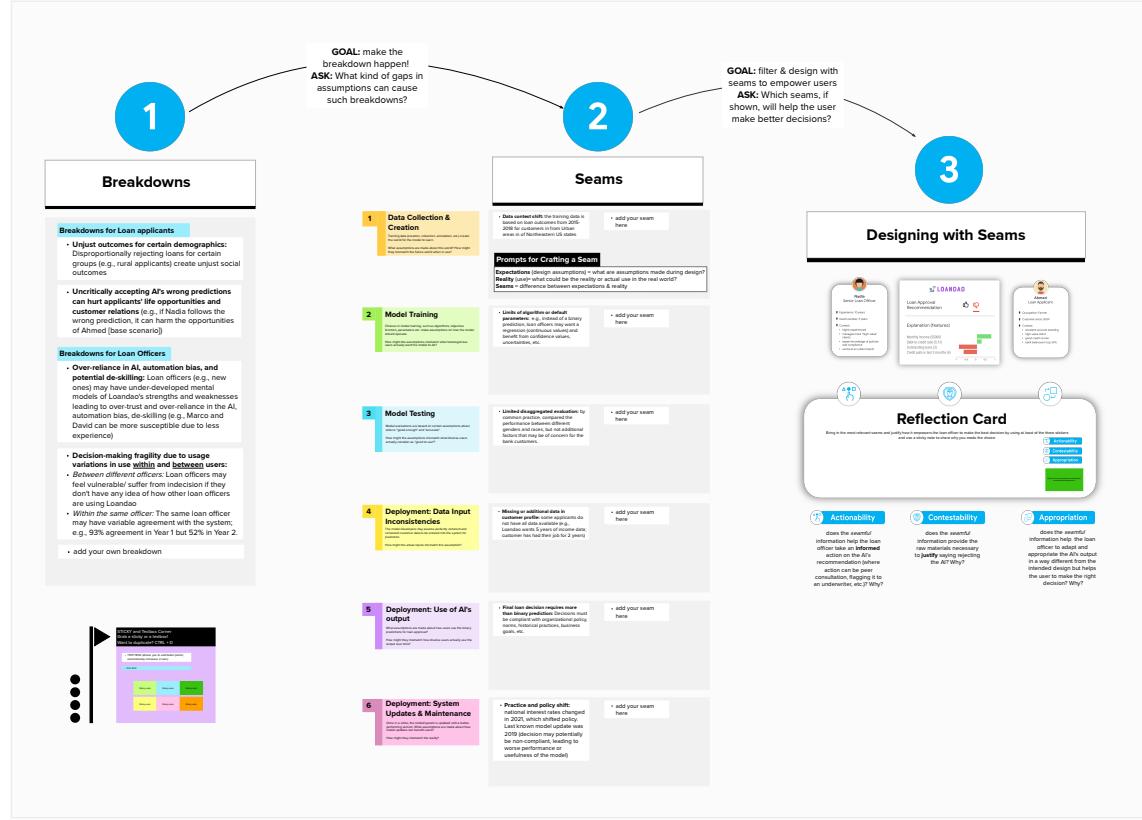


Fig. 5. A higher resolution picture of Figure 2 showing a screenshot of the virtual whiteboard used for the seamful XAI design activity in the study, with zoomed-in examples. **Area 1:** Envisioning breakdown (Step 1). In the study, we provided sample breakdowns, which participants could either use directly 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

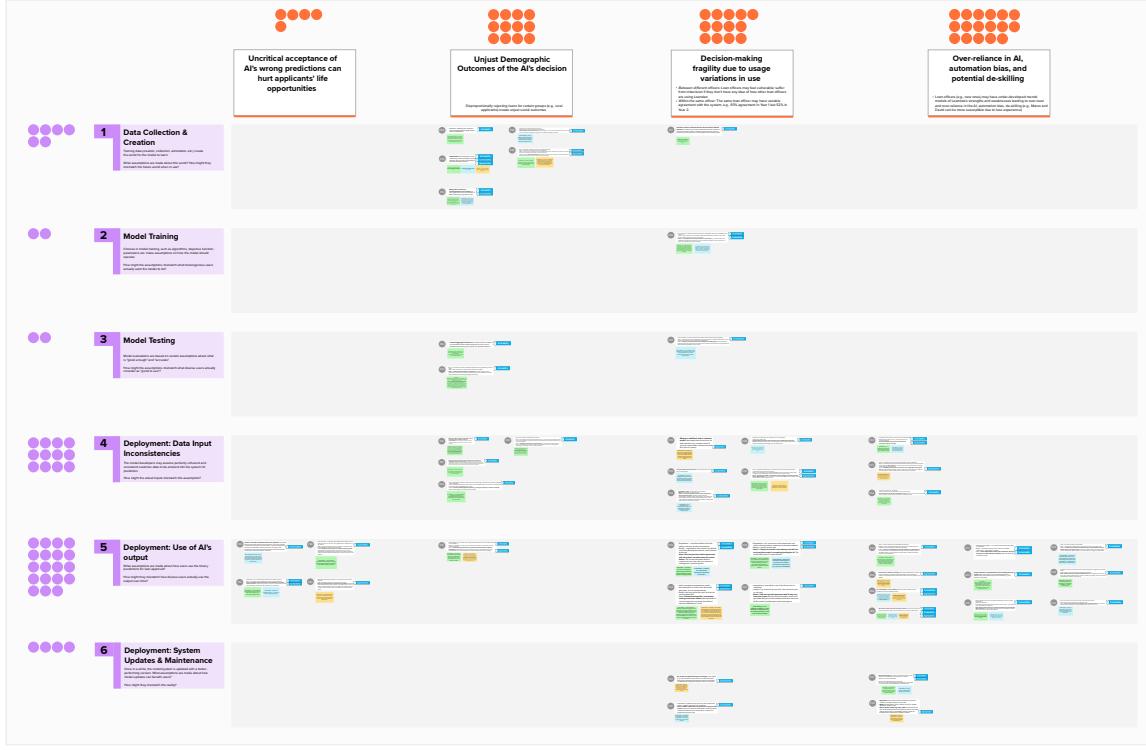


Fig. 6. A higher resolution picture of Figure 4 showing a bird's eye view of all seams from all participants for all breakdowns along all AI lifecycle. Breakdowns are the columns while the AI lifecycle stages are on the rows. Te dots above breakdowns provide the number of times a particular one was. 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.