AI Failure Cards: Understanding and Supporting Grassroots Efforts to Mitigate AI Failures in Homeless Services

Ningjing Tang
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Jiayin Zhi
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Tzu-Sheng Kuo
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

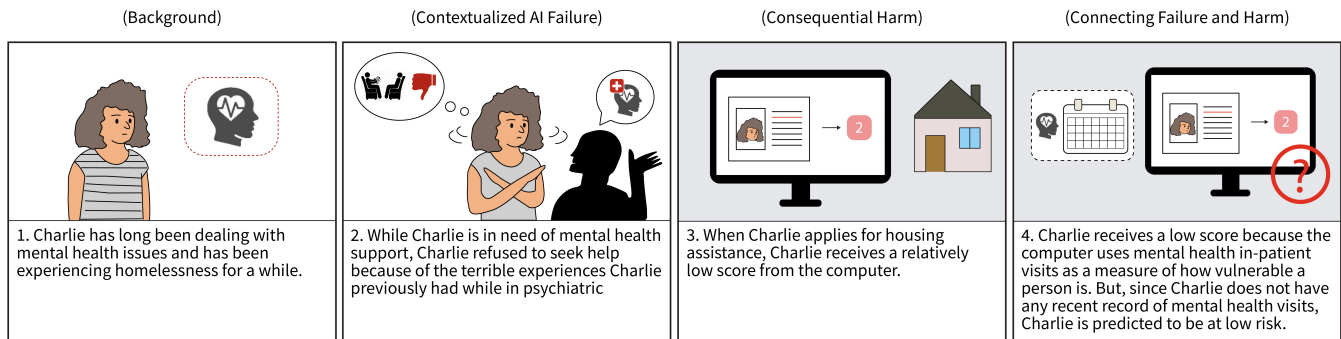Calla Kainaroi
Independant Researcher
USA

Jeremy J. Northup
Point Park University
Pittsburgh, Pennsylvania, USA

Kenneth Holstein
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Haiyi Zhu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Hoda Heidari
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Hong Shen
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

(Background) | (Contextualized AI Failure) | (Consequential Harm) | (Connecting Failure and Harm)

1. Charlie has long been dealing with mental health issues and has been experiencing homelessness for a while.

2. While Charlie is in need of mental health support, Charlie refused to seek help because of the terrible experiences Charlie previously had while in psychiatric

3. When Charlie applies for housing assistance, Charlie receives a relatively low score from the computer.

4. Charlie receives a low score because the computer uses mental health in-patient visits as a measure of how vulnerable a person is. But, since Charlie does not have any recent record of mental health visits, Charlie is predicted to be at low risk.

What do you think should be done to help people in a situation like Charlie?

Figure 1: We developed a novel method – the "AI Failure Cards" – to understand and support grassroots efforts to mitigate AI failures in the context of a locally deployed predictive housing allocation algorithm. We use three artifacts to scaffold the elicitation process: the *Onboarding Cards*, the *Failure Cards*, and the *Mitigation Cards*. Here, we present one example of the Failure Card on "target-construct mismatch (problematic proxies)," which captures and presents one of the common failures of predictive AI via comicboarding, accompanied by an elicitation question. We used an alternative layout in our workshops(see Appendix).

ABSTRACT

AI-based decision support tools have been used in a wide range of high-stakes settings. However, many of them have failed. Past literature in FAccT contributes important insights into how to detect and mitigate AI failures from a technical perspective. Recently, there are growing calls to understand AI failures as socio-technical and to support community-centered, grassroots-based mitigations to AI failures, in addition to top-down approaches. In this paper, we present *AI Failure Cards*, a novel method for both improving communities' understanding of AI failures and for eliciting their current practices and desired strategies for mitigation, with a goal to better support those efforts in the future. Through a series of workshops with unhoused individuals, frontline workers and service providers, as well as local policy advocates, we conducted an empirical investigation of our method in the context of a locally deployed predictive housing allocation algorithm. Our results suggest that the use of the method helped impacted communities better understand these AI failures. It also surfaced a wide range of existing grassroots practices and desired mitigation strategies. Finally, we discuss both the challenges and opportunities for supporting grassroots efforts in mitigating AI failures.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**; **HCI design and evaluation methods**.

KEYWORDS

homelessness, AI-based decision support, comicboarding, grassroots efforts

1 INTRODUCTION

AI-based decision support tools have been used in a wide range of high-stakes domains. For example, predictive algorithms have been used to help judges decide whether defendants should be detained or released while awaiting trial [1], assist child protection agencies in screening referral calls [31, 58, 66], help school districts decide student assignment [53, 55], and support allocation of housing resources [21, 34]. Despite widespread optimistic views on the capacity of AI to improve human decision making, many of them have failed in the real world and raised serious concerns [42]. Over the years, national debates and public outcry have erupted over biased and harmful outcomes caused by recidivism prediction [19], child welfare predictive analytics [13], and predictive policing [59]. These harmful outcomes are often the direct result of limitations or *failures* inherent to AI systems.

Past literature in FAccT contributes important insights into how to detect and mitigate AI failures from a technical perspective [2, 4, 76]. Recently, there are growing calls for understanding AI failures as *socio-technical* and to support grassroots mitigations to AI failures, in addition to top-down approaches [12, 38]. As Chancellor [12] notes, a socio-technical perspective on AI failures emphasizes not on "technical conceptions of performance through quantitative metrics, such as evaluating error rates and efficiency" but to attend to "who is impacted by failure, in what ways they are impacted by, how they cope with, and operate around failure".

However, understanding and supporting those grassroots efforts to mitigate AI failures presents a significant challenge. Mainstream media often foregrounds the advantages and capabilities of AI-based decision-support systems, generating much hype and excitement [75]. While previous research in AI literacy [39] has focused on the cultivation of understanding, skills, and competencies related to a wide range of AI systems [34, 62], few have aimed to make the common flaws, limitations, and failures of AI visible. In addition, while previous work has made an important contribution in broadening participation in AI design [34, 63], there is a lack of methods and tools for eliciting grassroots efforts and desired strategies to address AI failures. Developing a better understanding of these community-centered mitigation strategies can provide useful design guidelines to better support those efforts and achieve the desired changes. However, eliciting these strategies from impacted communities is challenging, as they often possess limited technical literacy and confront a variety of social challenges. These factors hinder their ability to recognize and connect their existing mitigation efforts to specific AI failures and brainstorm future mitigation strategies.

In this paper, we present the "*AI Failure Cards*," a novel method that aims to improve communities' understanding of AI failures and elicit their existing and desired mitigation strategies. We used three different artifacts – the Onboarding Cards, the Failure Cards, and the Mitigation Cards – to scaffold this elicitation process. The *Onboarding Cards* describe the basic decision-making process of the AI-based decision support tool. The *Failure Cards* capture a series of common socio-technical failures of the AI and present them via comicboarding, accompanied with an elicitation question. The *Mitigation Cards* enumerate a number of potential mitigation strategies, serving as the foundation to help participants brainstorm a wide range of actions they can take.

In this work, we document a case study of the use of our approach in the context of an AI-based predictive algorithm used in local homeless services. The Housing Allocation Algorithm (HAA) studied here is a type of predictive optimization algorithm [75], which prioritizes housing resources for people experiencing homelessness in a US county. It has been deployed for more than three years. Past work [34] has documented a series of existing failures of HAA, such as problematic proxies. Those failures generated widespread concerns among local homeless communities and motivated this work. We set out to understand how impacted community stakeholders respond to a recurring set of failures of HAA. In particular, to center the voices of those who are most directly affected by these failures, yet who lack the power to influence the mitigation process [78], we worked with frontline workers and service providers, current and former unhoused individuals in the region who are the direct decision subjects of HAA, as well as local policy advocates in homeless services.

Using the "AI Failure Cards," we conducted a series of workshops with unhoused individuals, frontline workers and service providers, as well as local policy advocates, to help them better understand AI failures and elicit their current mitigation practices and desired changes. Reflecting on their experiences during the study, our participants noted that our method has helped them better understand the root causes of these AI failures, identify shared patterns of AI failures across various sectors, and situate those failures within the broader social, institutional and structural contexts. They also shared a wide range of grassroots mitigation practices and desired changes, including trauma-informed practices, community-building efforts, contesting AI-informed decisions, and a set of proposals on preferred technical and social mitigation interventions.

Our contributions are three-fold:

- First, we introduce a novel method – the AI Failure Cards – to both help impacted community members better understand AI failures and to elicit grassroots efforts and desires surrounding mitigation;
- Second, we document a case study of the use of our approach in the context of a locally deployed predictive housing allocation algorithm. Through a series of workshops with unhoused individuals, workers and service providers, as well as local policy advocates, we conduct an empirical investigation of the initial effectiveness of our approach and collect a wide range of community-centered, grassroots-based mitigation practices, strategies and desired changes;

- Third, we discuss both the challenges and opportunities for supporting grassroots efforts in mitigating socio-technical failures of AI-based decision-support tools.

2 RELATED WORK

We outline the relevant work in two areas. First, we present an overview of current efforts aimed at cultivating AI literacy and describe how our work is positioned in this space. Next, we review the existing techniques, tools, and systems related to mitigating AI failures in the ML fairness literature, and describe how our work contributes to this area of research.

2.1 Critical AI Literacy

Researchers in HCI and FAccT has undertaken significant efforts to develop tools and methods to cultivate AI literacy. Defined as "a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" [39], AI literacy has become important topics in education and beyond.

Past work has made an important contribution to improving AI literacy in education, both in K-12 and higher education. For example, Ottenbreit-Leftwich et al. investigated how K-12 students conceptualized and experienced AI to inform the design of AI curriculum. Researchers have evaluated existing educational frameworks [49, 71], and formulated theoretical guidelines for educators and policymakers about the methods and content to teach K-12 students about AI [43, 44, 69, 72, 80]. In addition, they also developed applications and toolkits to help teachers in machine learning education [52, 56, 70]. A growing line of work has also focused on addressing the need to integrate ethical topics into AI and ML coursework in college education [22, 57, 61].

With the increasing deployment of AI systems in everyday lives, a different line of work has explored how to enhance AI literacy among the general public and impacted communities. For example, researchers designed and evaluated a series of alternatives of confusion matrices to facilitate public understanding of the performance of algorithmic decision-making systems [62]. Others presented an interactive interface to improve the understanding of ML models for both experts and laypeople [29]. Some developed the AI lifecycle comicboarding method to provide impacted stakeholders with detailed knowledge about AI system design and deployment [34].

Despite their significant contribution, there is a noticeable gap in efforts aimed at making the common flaws, limitations, and failures of AI visible and understandable to impacted community stakeholders [20]. This gap is particularly pertinent as these stakeholders often possess limited technical literacy and confront a variety of social challenges [34], which could hinder their ability to comprehend the common limitations and risks of AI that will lead to real-world failures and harms. As a result, there have been increasing calls for centering critical perspectives in AI literacy. A critical approach to AI literacy aims at fostering stakeholders' ability towards a more "practical, problem-oriented analytical perspectives on the risks of AI" [67]. For example, Stefen Strauß has emphasized the importance of building critical AI literacy as a means to raise awareness and empower citizens to challenge the dominant narratives and

practices of AI, and to foster more democratic and inclusive forms of AI innovation [67].

2.2 Mitigating AI Failures in ML and FAccT Research

Past literature in ML has developed a variety of technical approaches regarding detecting and mitigating failures in AI systems, in particular, those concerning the issues on fairness, accountability, and transparency. A series of toolkits have been proposed to help developers and AI/ML practitioners identify and mitigate failures in AI systems. The bulk of these efforts are largely led by AI and ML practitioners and offer *technical* solutions to remedy AI failures, such as bias and unfairness – broadly defined as undue disparities in the outcomes produced by AI [3, 4, 77]. For example, the Fairlearn toolkit developed by Microsoft helped empower data scientists with measuring the AI model's disparate performance [4]. However, in their meta-analysis of the fairness literature, Black et al. observe that the ML community has given an outsized attention to the statistical modeling stage in the ML lifecyle – e.g., by imposing a variety of "fairness constraint" at the time of fitting a statistical model to the training data [6]. They point out that such remediation strategies, while seemingly effective in reducing disparity metrics, may be targeting the wrong stage of the lifecyle, and subsequently, hide (as opposed to address) the underlying cause of disparity in the model's outcomes.

Recognizing the limitations of merely relying on these top-down, technical approaches towards AI failures, there is a growing body of work in FAccT that centers on understanding and supporting bottom-up, grassroots mitigation efforts towards AI failures. This line of work aims to empower individuals and communities "from below" – people who lack adequate power and agency in the creation and evaluation of the technology but are significantly affected by its design and deployment. For example, Kulynych et al. developed the Protective Optimization Technologies (POTs) tookit, which aims providing means for impacted stakeholders to address the risk and harms of optimization systems [33]. Vincent et al. introduced the data leverage framework to highlight opportunities for grassroots actions to change technology company behavior related to a wide range of social concerns [74]. Li et al. discussed opportunities for data activism to empower everyday data producers in their interactions with tech companies [36]. This body of work also resonates with prior studies in HCI that emphasize the importance of including community stakeholders' perspectives in the design, development and evaluation of AI models [11, 30, 32, 35, 54, 81], as well as studies that support grassroots practices to the problems posed by public technology [24, 78].

Our work aims to bridge the existing gap in critical AI literacy and contribute to the understanding of grassroots mitigation efforts in ML and FAccT research. In this paper, we introduced the *AI Failure Cards*, a novel method to cultivate critical AI literacy among impacted communities by making a series of common AI failures visible and understandable to them. Our method improves their understanding of those failures and facilitate elicitation of a wide range of grassroots mitigation strategies.

2.3 Participatory Methods in AI/ML

In recent years, there has been increasing attention towards participatory approaches in the design of AI/ML systems [5, 18]. Advocates argue that involving diverse stakeholders throughout the AI development process can lead to more democratic and inclusive algorithmic systems, emphasizing empowerment through "giving power to the people" [35, 68]. However, there are also limitations and potential drawbacks of these participatory methods in AI/ML.

For example, Delgado et al. [17] note that while many practitioners support greater stakeholder participation, what constitutes meaningful participation remains ill-defined. Others, such as Robertson and Salehi [55] and Sloane et al. [65], further caution that participatory processes, if not carefully structured, risk inhibiting progress or inadvertently exploiting participants based on the choices available to them and the depth of their involvement. There are also critiques of participatory design approaches for often relying on unpaid or inadequately compensated labor from community members [82], raising ethical concerns about exploiting marginalized groups' time and knowledge [37].

In our study, participants expressed a strong desire to share their experiences with AI failures, yet often lacked the necessary methods, tools, and platforms to do so effectively. This gap motivated our research, leading us to adopt participatory methods that center community involvement in the AI mitigation process, promoting participation as a key component of justice [7, 8, 15]. While we recognize the potential limitations of participatory approaches in AI and ML, our goal is not to present community engagement as a panacea, but rather to contribute methodological insights that foster meaningful involvement of marginalized communities in AI design and development process.

3 STUDY DESIGN

Drawing upon prior HCI methods, we developed a novel method – the "AI Failure Cards" – to both help impacted communities better understand AI failures and to elicit their existing practices and desires for mitigation. Using this method, we conducted a series of workshops with unhoused individuals, frontline workers and service providers, as well as policy advocates, in the context of a locally deployed predictive housing allocation algorithm.

3.1 Study Context: AI-based predictive optimization in homeless services

We conducted our study in the context of an AI-based predictive algorithm used in local homeless services. The Housing Allocation Algorithm (HAA), studied here, is a type of predictive optimization algorithm [75] that prioritizes housing resources for people experiencing homelessness. It has been deployed in a US county for more than three years. HAA's assessment begins with county staff verifying applicant eligibility, followed by running HAA, which uses personal data from the county's warehouse unit to predict how likely the applicant will experience the following three situations if they remain unhoused over the next 12 months: more than four emergency room visits based on healthcare utilization data, at least one mental health inpatient funded by Medicaid, and at least one jail booking. Applicants are then scored and placed on housing wait-lists. In some cases, an alternative assessment (ALT HAA) based

on self-reported data is used, especially when applicants' data is less than 90 days old or when their vulnerability isn't accurately reflected by HAA's score, though its use is generally discouraged [48, 73]. Past work [34] has documented a series of existing failures of HAA. Those failures have generated widespread concerns among local homeless communities and motivated this work.

3.2 The AI Failure Cards

Taken as a group, the AI Failure Cards is designed to achieve two goals: (1) help non-technical community stakeholders better understand the socio-technical failures of an AI-based predictive algorithm; (2) elicit community-centered, grassroots-based practices and desires for mitigation, with a goal to support those efforts.

We iteratively co-designed the method with two service providers as representatives of the impacted community, who interact with both HAA and local homeless communities regularly. They serve as the community co-authors of this paper [14]. In order to scaffold the workshop discussion, we designed three kinds of artifacts, drawing upon previous HCI and FAccT research on card-based toolkits (e.g., [45, 61, 64]). Due to the unique reading and technology literacy challenges our participants are facing, following Kuo et al., we used a comicboarding approach to construct our cards, based on a set of gender-neutral personas. Comicboarding [46], a design technique that uses comic strips as a scaffolding tool, has been used to elicit perceptions of algorithmic systems in different contexts such as targeted advertisement[23]. It is particularly effective in our context, as low reading literacy presents a barrier to participation for many unhoused participants. Below we describe the three set of cards in details (all cards are included in the Appendix):

3.2.1 Artifact 1: The Onboarding Cards. The Onboarding Cards describe the basic workflow and key decision-making process of HAA and present them via comicboards [46]. We based information in our Onboarding Cards upon detailed technical reports published by the county about the HAA system [73] and validated the accuracy of our design with our community co-authors. As such, they served as a foundation to help study participants gain a quick and basic understanding of how HAA works, which serves as a foundation to surface socio-technical failures.

3.2.2 Artifact 2: The Failure Cards. The Failure Cards capture a series of community-centered socio-technical failures of HAA and present them via comicboards [46], accompanying with an elicitation question. We co-designed the failure cards with community stakeholders by leveraging real-world failure cases and utilizing existing theory-driven taxonomies. First, to avoid "reinventing the wheel," we collected a set of taxonomies and frameworks of AI failures and harms (e.g.,[60, 75]) from previous literature as our initial theoretical framing. In parallel, our community co-authors collected a set of real-world community-perceived failure cases of HAA via their personal network. A typical failure case contains a qualitative narrative describing how and why unhoused individual(s) was harmed by HAA. In total, 45 cases were collected. Second, using those on-the-ground failure cases as our initial dataset, we performed affinity diagramming [26] to cluster similar ideas and identify common themes. Much of the discussion with our community co-authors centered around the underlying causes of these

failures, with a number of key themes emerged from the process, such as problematic proxies and distribution shift. Third, through iterative discussion using deductive and inductive thinking, we examined whether and how the resulting clusters map back to the existing frameworks and eventually reached seven set of recurring failures, mapped well onto [75]:

(1) Intervention vs. prediction (good prediction may not result in good decisions),
(2) Target-construct mismatch (problematic proxies),
(3) Distribution shifts (training data rarely matches the deployment setting),
(4) Limits to prediction (social outcomes often defy meaningful prediction),
(5) Disparate performance (worse performance for some groups may be unavoidable),
(6) Lack of contestability (lack of viable mechanisms for contestability),
(7) Goodhart's law (decision-subjects may adapt in a way that defeats the goals of system).

Finally, instead of presenting these high-level categories directly to participants, we translated the seven failures of HAA into seven synthetic cases (i.e., personal stories) with our community co-authors and presented those cases via comicboarding, to both protect the privacy of the unhoused individuals in our initial dataset and to better communicate those failures to participants. In our final design, each card contains four panels, centering around how a specific type of AI failure has harmed unhoused individual(s), including providing background information, contextualizing AI failures, discussing consequential harms, and connecting failures and harms. The card is then followed by an elicitation question for mitigation strategies, such as "what do you think can be done to help people in situations like Alex and Kai?"

3.2.3 Artifact 3: The Mitigation Cards. To better facilitate the brainstorming process, we also developed a final set of artifact – the Mitigation Cards. Those cards enumerate a set of potential mitigation strategies via short descriptions. We derived our Mitigation Cards from prior research that gathered the perspectives of communities impacted by AI-based predictive algorithms [34, 66]. Using these community feedback as our initial dataset, we performed affinity diagramming [26] iteratively in our group to cluster similar ideas and synthesize various options. We intentionally developed the cards across a wide range of potential actions (e.g., across technical, social and structural level of changes), with a goal to enable the elicitation of a diverse range of community-based mitigation practices. To avoid promoting "bluesky thinking," which might lead to infeasible solutions that ultimately frustrate our participants [27], we used those mitigation cards to prepopulate the solution space to help better scaffold discussion. Finally, to support asynchronous deliberation between different stakeholders while ensuring the comfort of the participants, we also iteratively selected and presented the participants with additional "Mitigation Cards" generated from other stakeholder groups in previous sessions [28, 40]. In total, nine mitigation cards were iteratively developed and used in our study.

3.3 Study Protocol

We chose to conduct workshops for our study because workshops facilitates collective exploration of shared problems and the development of solutions [79], well suited for our study purposes. Each workshop was conducted exclusively with participants belonging to the same stakeholder group, which ensures that they felt comfortable and open to sharing [79]. In each workshop, we started by walking through the workflow of HAA using the Onboarding Cards. Then, they were shown seven Failure Cards one by one. We asked participants to think aloud and share their experiences related with the presented Failure Card, and used an elicitation question to probe their current practices and desires for mitigation. We also presented them with Mitigation Cards to facilitate brainstorming. In addition, we included selected responses from participants in the previous sessions as added Mitigation Cards, to provide an opportunity of deliberation between stakeholders in a safe space, considering the potential unequal power relations [28, 83]. Demographic information was collected at the end of the workshop. The study was approved by our institutional IRB.

3.4 Recruitment

We adapted a purposive sampling approach [51] to recruit participants from three stakeholder groups: frontline workers and service providers, unhoused individuals, and local policy advocates in homeless services. 1) We recruited workers and service providers due to their hands-on experience with HAA and direct interaction with the local unhoused population. Those were recruited with a list of contacts provided by our community co-authors; 2) we also recruit individuals with lived experiences of homelessness to elicit on-the-ground insights into the challenges and impacts HAA introduced to homelessness. Considering the sensitivity of this matter, we recruited through local churches who have established trust relationships with the homeless community, by leveraging the their "walk-in ministries" to distribute flyers via church staff; 3) We expanded our recruitment to include local policy advocates in homeless services, via personal network, to engage with individuals who have expertise and experience in policy advocacy. All the policy advocates we recruited were non-institutional; they all had direct experiences working with local communities and are passionate about advocating on behalf of the unhoused population.

In total, we recruited 22 participants: 11 individuals with experience of homelessness, 8 frontline workers and service providers, and 3 policy advocates. We conducted six workshop sessions with two to six participants in each, and two individual interview sessions due to participant no-shows. For the unhoused individuals we recruited through each church, we conducted the study session in a room provided to us by the respective church; we spoke with all frontline workers and service providers, and policy advocates via Zoom. The study sessions lasted 96 minutes on average, and each participant was compensated $60 for their participation. Participants' self-reported demographics are in Appendix.

3.5 Data Analysis

We utilized a reflexive thematic analysis approach to analyze our study data [9, 10]. Open coding was carried out by at least two of the authors on transcriptions from 768 minutes of audio, resulting

in a total of 345 codes. We held weekly meetings to engage in discussions to resolve disagreements. Following the standard practice of reflexive thematic analysis, we did not calculate inter-coder reliability, as reaching consensus are integral to theme development [41]. Our analysis was conducted across all Failure Cards, given our goal of understanding overarching themes in participants' responses across the full set of Failure Cards, which aligns with common practices in HCI storyboarding studies [16, 28]. After coding, we derived higher-level themes through affinity diagramming. This process yielded 113 first-level themes, nine second-level themes, and three third-level themes. We detail our findings in Section 4, with the section descriptors correlating with the second and third-level themes.

3.6 Limited Scope of the Case Study

The goal of our study is not to capture all failures in AI-based decision-support tools, but to study the initial effectiveness of the AI Failure Cards as a method to both improve communities' understanding of AI failures and to elicit grassroots mitigation strategies. In our work, we choose to focus on a series of community-perceived failures of a predictive housing allocation algorithm and work closely with local homeless communities and their representatives. Future work should expand the scope of this method to other AI applications and communities, to further validate its effectiveness and adaptability across different contexts and AI systems. This would not only help in refining the AI Failure Cards method but also contribute to a broader understanding of AI impacts and community-led mitigation strategies.

4 FINDINGS

In this section, we present our findings around three high-level themes identified through our analysis: (1) understanding AI failures, (2) the mitigation of AI failures in participants' grassroots practices, and (3) the proposals for mitigation strategies. Overall, we found that our method has improved participants' understanding of the root causes, shared patterns, and broader contexts of AI failures (Section 4.1). Building upon their understanding of AI failures and day-to-day experiences, our participants shared the grassroots practices they currently employ to mitigate these failures (Section 4.2). Furthermore, our participants proposed detailed actionable mitigation strategies that they believe could be implemented to address AI failures more effectively (Section 4.3). Throughout this section, participants with experiences of homelessness are identified with an identifier beginning with a "P," frontline workers and service providers are identified with a "W," and local policy advocates are identified with an "A".

4.1 Understanding AI Failures

Participants from varied backgrounds enriched their understandings of HAA's failures in different dimensions using the AI Failure Cards in our workshops. Our findings demonstrated that our method helped them uncover the root causes of individual failures, recognize shared patterns across various sectors, and situate those AI failures in broader contexts. Our approach enhanced both individual insights and collective comprehension of the complex landscape of AI failures.

4.1.1 Understanding the root causes of perceived individual failures. Throughout the study, our participants **enhanced their understanding of the AI failures through connecting the root causes to the individual cases they encounter.**

For example, when presented with the Failure Card about "Distribution shifts (training data rarely matches the deployment setting)" (Figure 6 in Appendix), P5 resonated with the case and reasoned his experience with the same underlying cause. P5 talked about how they traveled from another county and was misjudged by HAA, as the local county did not have any data records of him. Given the same card, W2 shared that many unhoused individuals who came from outside the county, were impacted by the same flaw, as the HAA's training on local resident data led to their lower scores: *"[For travellers,] they need to start everything over."*

Similarly, when presented with the Failure Card on "Target-construct mismatch (problematic proxies)" (Figure 5 in Appendix), P2 pointed out that the cause behind the story presented in the card could be that HAA uses mental health inpatient and emergency room visits as two of the proxies to predict vulnerability. They shared that many unhoused individuals, including himself, avoid seeking emergency room care due to its high cost: *"[It costs] a lot of money. How do you go to ER, [to be considered by HAA]?"* W5, who interacted with multiple unhoused individuals echoed this sentiment, and shared there are similar cases for the same root cause: *"There are lots of Charlie[s]."* Charlie is the character featured on our card who declines to seek mental health services due to previous encounters with institutional violence. *"They don't trust psychiatric hospitals because of their traumatizing experiences,"* W2 said. Consequently, their score was adversely affected because HAA used mental health inpatient as one of its indicators for vulnerability.

*4.1.2 Understanding shared patterns of AI failures across various sectors.* Our participants also **identified and recognized the shared patterns of AI failures across different contexts**.

For instance, when discussing the Failure Card on "Lack of contestability (lack of viable mechanisms for contestability)" (Figure 9 in Appendix), A3 noted that this issue is widespread in public-sector AI applications, including another local AI-based decision-making tool used for child maltreatment prediction: *"They never disclose the predictive risk [either] ... because they are concerned about the potential disparity in negative impacts [on decision-subjects]."* In addition to the local government that deployed HAA and the child maltreatment prediction tool, A2 noted that other government agencies also hastily adopted AI without adequately considering the legitimacy of introducing AI-based predictive optimization systems in critical public services: *"We shouldn't be coming at it from a starting point [that] all processes can be improved upon by introducing algorithms. That should be a question, not an assumption."*

4.1.3 Understanding AI failures in broader contexts. Our participants also pointed that the failures highlighted in the cards were not merely due to the technical limitations of HAA. Instead, they pointed out that **these issues stemmed from fundamental social and structural problems that extend beyond HAA.** For example, P6 believed that instead of developing AI tools, *"they need to open up more shelters, hospitals and recruit more people."* Similarly, P2 argued that: *"the problem that I'm having ... with this tool is [that it is] not addressing employment, [which I believe is] the real reason*
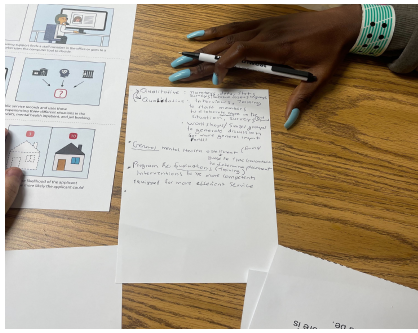
(a) Participants shared with us the mitigation strategies they wrote.
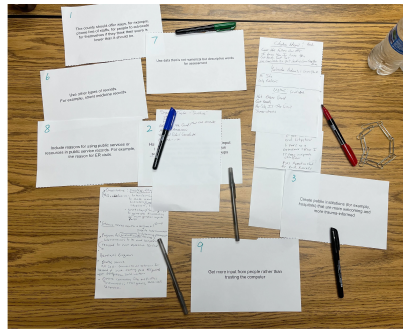
(b) A collection of mitigation strategies generated by participants.
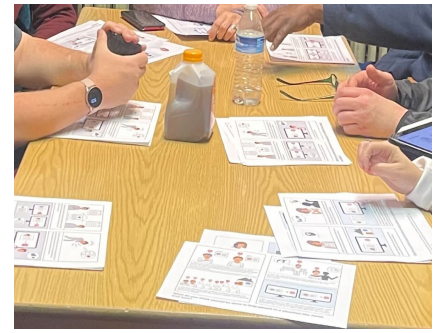
(c) Participants shared their thoughts while reading the Failure Cards.

Figure 2: Photos taken during the in-person workshops, capturing community members' discussion and engagement in generating mitigation strategies for HAA's failures.

for homelessness." These comments indicate the need for broader changes to address the homelessness crisis. However, given the limited capacity of individual participants to enact such changes, they have developed grassroots practices to mitigate the impact of these AI failures on the ground, as detailed in the next section.

4.2 Mitigating AI Failures via Grassroots Practices

Given AI Failure Cards, participants shared various grassroots practices they currently implement or are aware of to mitigate the failures of HAA and improve the assessment process. These practices include: (1) integrating more trauma-informed care practices into the pre- and post-assessment stages, (2) building a stronger community support system for the unhoused population, and (3) developing strategies to contest HAA's decisions.

4.2.1 Mitigation through trauma-informed care practices. When HAA fails to recognize the vulnerabilities of unhoused individuals, workers and service providers incorporate additional care practices both before and after the assessment.

During the pre-assessment process, when frontline workers document an unhoused individual's situation for input into HAA, **they formulate care-based questions to elicit vulnerabilities that can positively impact an individual's HAA score**. For example, when reading the Failure Card about "Target-construct mismatch (problematic proxies)" (Figure 5 in Appendix), explained how they rephrased the question with care instead of straightforwardly asking: "Are you a victim of domestic violence." W6 said: *"I feel like it's more so a journey, and explaining what the questions mean, and kind of digging through their life, because some people just say no, but then you get to know them better. They explained some stories. They may realize later it was abuse."*

Following the assessment, as frontline workers were unable to contest or explain HAA's decision, **they adopted a more trauma-informed approach to communicate HAA's decision and explored workarounds to connect unhoused individuals with alternative resources**. For example, when discussing the Failure Card on "Lack of contestability (lack of viable mechanisms for contestability)" (Figure 9 in Appendix), W2 mentioned that instead

of stating "you're not qualified" without providing reasons, they opted for a more empathetic approach, saying, *"Sorry, we just have too many people waiting."* Another service provider, W4, shared that when they believes the AI system has made a problematic assessment: *"I don't really care what the assessment says [...] this is a good opportunity to connect this person to other services."*

4.2.2 Mitigation through community network building. Participants also organized various grassroots community-building initiatives to alleviate the harm stemming from HAA's failures. Both the unhoused community and local service providers took active roles in driving these initiatives.

Among the community with experience of homelessness, **participants emphasized the importance of information-sharing, recognizing the substantial disparities in accessing information**. For example, when discussing the Failure Card related on "Target-construct mismatch (problematic proxies)" (Figure 5 in Appendix), P5 pointed out that unhoused individuals had varied inclination to seek information, which can impact their score assigned by the AI system: *"It's about everything from the willingness to get up out of bed and do something to find the information. [...] They don't go to the library. They don't go to the mental health institute, all those sorts of things, which are all helpful."* P5 noted that individuals might be reluctant to actively seek help due to past traumatizing experiences with existing institutions. Therefore, they believed that curating and sharing information related to resources within the community could be beneficial. This belief was affirmed in one of our workshops, where unhoused participants actively exchanged information about local resources, including shelters and food.

Service providers also shared their efforts to enhance and broaden the support system when HAA fell short in allocating sufficient housing resources. For example, when discussing the Failure Card about "Lack of contestability (lack of viable mechanisms for contestability)" (Figure 9 in Appendix), W1 shared their attempts to develop a system that reported bed availability from all local shelters: *"There could be a central collection point where people knew where there were beds available. [...] We had tried to develop a [system for] bed availability. Just call it and we'd have bed counts from smaller shelters and larger shelters all around the county."*

These efforts were appreciated by unhoused individuals, like P7, who found centralized information provided by service providers useful: *"Talk to the counselor. [...] She's helped me with a lot of stuff. She has information."*

4.2.3 Mitigation through contesting the AI Failures. As mentioned in Study Context, the existing decision-making system around HAA offers a number of override mechanisms, such the ALT HAA. However, workers in general are discouraged from using these mechanisms. Due to those constraints, during the workshop participants shared a number of other contesting mechanisms they developed.

Participants shared that they have actively **leveraged the existing county-based systems as means to collectively raise their concerns on problematic assessments by HAA**. For example, some participants mentioned the case conferences that are held regularly by the local county and viewed it as an opportunity to voice their concerns. In a typical case conference, frontline workers and service providers will gather and discuss a few specific cases that they think led to problematic HAA scores. Participants perceived this as a democratic practice to contest the failures of HAA, where the decision is made by collective human deliberation and discussion instead of being made by human individual or the AI. According to W5, *"It's not just one person making a decision about it on their own. It's people getting together they discuss it, they talk about well what are the vulnerabilities versus experiencing and what is the computer system missing? And can we all agree that?"*

Participants also shared **grassroots approaches to developing low-tech or no-tech alternatives to HAA**. For example, a small group of frontline workers and service providers developed an alternative assessment tool with carefully phrased questions for understanding an individual's vulnerability. These questions assessed an individual's vulnerability, even if they might not seek emergency room assistance but could be facing other challenges, such as an overdose. The participants who created the tool envisioned its use for contesting AI decisions and referring them to the county. They also shared how the development process assisted them in reflecting on their practices related to assessing vulnerability and in rephrasing questions to more effectively gather relevant information about an individual. W4 noted that, *"I do feel like just for me personally, it was helpful even just to go through the process of making up that sheet and going through thinking different questions we could be asking. if I'm working with someone maybe I can think of ways to rephrase things or to maybe get them prioritize."*

4.3 Re-Imagining Mitigation via Technical, Human-centered and Institutional Changes

In addition to the existing grassroots practices, our participants suggested additional mitigation strategies, spanning technical improvements, human-centered interactions, and institutional changes.

4.3.1 Technical improvement on the algorithmic system. Participants suggested rethinking HAA's system design, including integrating qualitative data, expanding the types of quantitative data used, and incorporating a holistic service allocation mechanism into the system.

Many participants suggested **integrating qualitative information into HAA:** *"Words instead of numbers can be useful,"* W1

said, when reading the Failure Card about "Target-construct mismatch (problematic proxies)" (Figure 5 in Appendix), *"Hospitalization [count] isn't the only way to tell the story about vulnerability."* Participants believed that qualitative narratives could assist in capturing the real situations of unhoused individuals behind the quantitative measurements used by the system. For example, in addition to mental health visit counts, A3 suggested incorporating a space for frontline workers to document unhoused individuals' responses to questions, such as *"How much do you think you need mental health help,"* which might reveal more information about an individual.

Other participants emphasized the need to **diversify the types of data used by HAA**. For instance, W4 argued that, in addition to hospital visits, data from street-level medical services could be valuable because unhoused individuals often turn to street medicine due to a lack of trust in institutions like hospitals. They believed this improvement is highly feasible because *"[The street medical services already] have documentation of their clients. I think that should be viewed the same as the hospital visit."*

Participants also proposed **a more holistic system that extends beyond housing resources to include other support services.** They found it particularly helpful in cases where housing resources are not available: *"As opposed to just telling them approved, denied, if the resources are in front of me... I can say here's some useful resources I could pass along that might help defuse the situation,"* A2 said. However, they also acknowledged the challenges of integrating data and services from multiple sources: *"I remembered the child welfare had data sharing agreement with some hospitals and that type of data sharing is extremely uncommon and was probably very challenging to work out."*

4.3.2 Enhance human-centered interaction experience. Besides technical improvements, participants also proposed ways of improving the interaction experience with the system from a human-centered perspective.

Participants suggested **integrating more human explanation in the loop during the HAA assessment.** For example, P7 expressed a desire for the AI system to interpret human natural expressions as input: *"Talk to real people as your input."* Besides, there were also suggestions to strengthen the role of human mediators in the interactions between applicants and HAA. For example, A1 recommended incorporating more human explanation in the loop to provide applicants with a greater sense of control and trust. According to them, having human workers to explain the context that *"the score is not the end of all, it is just a risk score and there are something you can do"*, *"can provide the population with [not only] a true reality, but also a sense of control and trust."*

Our participants also shared the desires for HAA to **proactively conduct more check-ins with applicants, rather than relying on applicants to initiate contact with frontline workers and the system.** In particular, when reading the Failure Card on Limits to prediction (social outcomes often defy meaningful prediction) (Figure 7 in Appendix), which described that HAA failed to predict applicants' vulnerability when their situation changed overnight, A2 recognized the importance of proactively checking in with applicants and updating their data regularly when people's risks are *"not a static, unchanging thing"*. They further underscored

the significance of this practice, especially for vulnerable populations who may face challenges in actively reaching out to frontline workers and providing updates. A1agreed with their proposal and commented, *"it puts a lot of the onus on the county ... rather than on the individual to initiate contact, which I'm assuming that like for this population, is both challenging and not exactly a pleasure"*.

However, we observed that enhancing certain aspects of interaction processes in HAA could pose its own challenges, particularly due to the fact that human caseworkers are already overwhelmed with substantial workloads, including significant amounts of invisible emotional labor. P10 shared that the caseworkers for HAA they had met are all overburdened by too many cases they had to deal with: *"My caseworker made it very clear to me. She's overloaded with people. She's like, I have way too many people on my thing. So they don't have enough staff to cover everybody. I think that's what they're doing. They don't have enough staff to deal with [applicants]."*

4.3.3 Institutional Changes. Finally, participants advocated for broader institutional changes to improve the overall sociotechnical ecosystem surrounding HAA. These ideas include alternative mechanism to demonstrate individual's vulnerability beyond HAA's score, shifting the housing model from individual-based to community-based, and introducing regular evaluation processes across the entire AI development pipeline.

Some participants mentioned the need of **alternative mechanism beyond HAA score to demonstrate an individual's vulnerability**. When reading the story of Charlie in the Failure Card on "Target-construct mismatch (problematic proxies)" (Figure 5 in Appendix), W3 pointed out that there should be *"clear pathways for providers that are working with people, or potentially the people themselves, to be able to demonstrate those vulnerabilities beyond the predictive risk score."* They noted the existing natural support system in deed provided rich information which can be further leveraged to demonstrate an unhoused individual's circumstances: *"that be an outreach worker or a shelter staff person, a service coordinator in the behavioral health system, it could be police, the natural support has a lot of information of a person. [The frontline worker] can say like, yep, I know the score is low, but here are all the things that we've observed, they're at risk of these types of exploitation in their current situation, we would like them to be prioritized for a housing program, despite the score."*

Some participants shared desires to **shift the housing model used by HAA from focusing on individuals to a community.** This change aims to provide a stronger safety net for all who are experiencing homelessness. For example, W1 spoke from their experience that individuals who *were* housed could sometimes face increased risks of overdoses, as *"they are no longer within a community that can look out for them"*. They suggested that more holistic support is essential beyond just providing housing for individuals: *"I think that sometimes, housing an individual solo, doesn't reduce the harm when it's somebody who is very dangerously compromised with their substance use disorder. It doesn't mean that they ought not to be housed. It just means that we should be looking at some different models of housing."*

Other participants proposed **a more frequent and systematic evaluation program for HAA to increase its accountability, across the entire AI development pipeline.** For example, A2

suggested regular audits based on a checklist to evaluate HAA's use of proxies, data, and training methods: *"What are the proxy variables that are being used? What's not being used? Is that accurate? Should we revise that? [We should] require regular internal processes to make sure that the results being outputted are as accurate as possible."* A1 echoed the idea and asked for ways to increase HAA's accountability at the early stage:*"Imagine at the procurement stage, for some of these tools, there could be some additional accountability. You could imagine requiring impact assessments or audits."*

5 DISCUSSION: SUPPORTING GRASSROOTS EFFORTS IN MITIGATING AI FAILURES

Drawing upon previous HCI and FAccT research on card-based toolkits (e.g., [45, 61, 64]), in this work, we presented a novel method - *AI Failure Cards* – to both improve communities' understanding of common failures of a predictive housing allocation algorithm, and to elicit their current practices and desired strategies for harm mitigation. Through a series of workshops with unhoused individuals, workers and service providers, as well as local policy advocates, we found that the method is promising in supporting community members to better understand the AI failure cases they encountered in their everyday lives and facilitating the discussion of a wide range of existing mitigation efforts. In addition, the cards were effective in helping participants propose a set of feasible and actionable directions to further mitigate these AI failures.

Next, informed by our findings, we discuss several design opportunities, limitations and directions for future work.

5.1 Implications for Practice and Policy

We call for ML/AI practitioners and policymakers' attention to and support of these grassroots efforts, following Green and Viljeon's call on "algorithmic realism" [25]. As Green and Viljeon noted, when facing the negative societal impact of AI systems, practitioners often adopt a top-down, formalist method, focusing on directly making technical repairs.In contrast, grassroots mitigation engage with the complexities of sociotechnical systems in a more bottom-up and organic manner, adapting dynamically to the intricacies of local contexts [25]. By integrating these top-down and bottom-up approaches, we can gain a holistic understanding of how to better mitigate AI failures. Recognizing the efforts taken by community members, in what follows, we discuss how ML/AI practitioners and policymakers can learn from these grassroots efforts to better support and empower impacted communities from below.

First, our findings demonstrate that when provided with adequate and easy to understand information of AI failures, participants are able to bring up feasible and actionable mitigation strategies, rooted in their contextual knowledge and lived experiences. These strategies can be adopted by AI/ML practitioners and further integrated into the system design. For example, when reading the Failure Card of Target-construct mismatch, some participants proposed to diversify the data sources used in HAA, in particular, from emergency room visits to street medicine visits. Based on their lived experiences, using records of street medicine visits can better capture the situation of unhoused individuals, as many of them are not willing to visit emergency rooms due to past experiences of institutional violence. It is worth to note, however, even these seemingly

technical mitigations can involve and often require complex social and institutional negotiations and accommodations, which further reveals the socio-technical nature of AI failures. For example, after recognizing the need to diversify the data sources, some of our participants acknowledged the complexities and difficulties of sharing data across different government agencies and organizations.

Second, we also observed that many of the existing and proposed strategies from our participants involve developing a series of "workarounds" for the HAA assessment process. This is partially because as "people from below" [78], they lack sufficient power and technical capacity to directly intervene on AI-based decision-making systems. That said, some of those workaround strategies can be very effective, suggesting opportunities to design formal processes and mechanisms to systematically incorporate those workarounds into the broader system. For instance, recognizing their inability to directly override HAA decisions and acknowledging existing support systems outside of HAA, frontline workers and service providers, actively connect unhoused individuals to various relevant resources. These practices, as discussed in section 4.2.2, are not direct interventions in HAA but rather strategies to repair and navigate its failures and harms. Such practices could be shared and incorporated into training programs for frontline workers and service providers, helping them to better manage the system's limitations in allocating scarce housing resources.

Thirdly, the grassroots initiatives shared by our participants, as described in section 4.2.3, also involve developing a set of low-tech and non-tech alternative assessment tools, as a way to contest the flawed decisions made by HAA. These efforts present a valuable foundation for practitioners who want to develop alternative assessment systems and for policymakers who aim to develop mechanisms for contesting AI-driven decisions. Although our findings discuss the challenges in establishing such mechanisms, these grassroots tools provide a practical starting point. Practitioners and policymakers can leverage these existing low-tech and non-tech approaches as a basis to initiate and progressively refine methods for contesting the decisions made by AI-based decision tools.

Finally, in addition to a series of mitigation strategies, some of our participants questioned the legitimacy of introducing AI-based predictive tools into critical public services (e.g., housing allocation) and emphasized the need for additional deliberation and community engagement before system procurement, as discussed in section 4.1.2. We echo their suggestions and urge practitioners and policymakers to proactively consider and consult with impacted communities not only on *how* to improve AI-based decision-support tools and mitigate AI failures but also, perhaps more importantly, on *whether* we need such an AI system in the first place. Involving impacted communities before system procurement ensures that the voices of those most affected are heard and their concerns are addressed, leading to more equitable and effective implementations of AI in public services.

5.2 Limitations and Future Work

As a "proof-of-concept" case study, there are a number of limitations that we have reflected on.

First, in this study, we chose to present the same comicboards to all participant groups to simplify the process and lower barriers to participation. Future research could adapt our method by customizing the content according to participants' expertise and backgrounds. This approach may uncover unique mitigation strategies specific to different stakeholder groups.

Moreover, while our methods facilitated discussions around common AI failures, we acknowledge that there are other "unknown unknown" AI failures that might not be identified in our study. Complementary methods could be used to enhance our understanding of real-world AI failures that our study materials did not cover.

Furthermore, our failure taxonomy and the depicted cases in the cards were grounded in the specific context of HAA. Although we believe our approach can extend to various domains, adaptation requires dedicated collaboration with community partners. This entails actively seeking input from local collaborators, identifying context-specific failure instances, and developing customized sets of failure cards aligned with each unique setting.

Lastly, the use of our method with impacted communities underscores opportunities for practitioners and policymakers. Integrating both top-down and bottom-up approaches to mitigate AI failures is crucial. However, translating grassroots practices and feedback into broader decision-making systems necessitates new tools, processes, and methods. We aim to contribute to this through future research.

6 CONCLUSION

In this paper, we present the AI Failure Cards, a novel method that improves impacted communities' understanding of AI failures and elicits their current practices and desired strategies for mitigation. We documented an early use of our method through a series of workshops with unhoused individuals, frontline workers and service providers, as well as local policy advocates, in the context of a locally deployed predictive housing allocation algorithm. Our results suggested that the use of method helps improve their understanding of the AI failures and facilitates the elicitation of a wide range of community-centered, grassroots-based mitigation strategies. Finally, we discussed design opportunities to better support those grassroots efforts and called for combing both "top-down" and "bottom-up" approaches in mitigating socio-technical failures of AI-based decision support tools.

7 RESEARCH ETHICS AND SOCIAL IMPACT

7.1 Ethical Consideration

The study was approved by our institutional IRB. In addition, due to the high level of vulnerability of our study participants, we followed best practices from previous research and sought additional help from domain experts. Throughout the study, following best practice [14], we worked closely with two community co-authors to collect real-world failure cases of HAA, design all the three sets of cards, and ensure that our study material is accurate and trauma-informed. In addition to the two community co-authors, we also connected and actively collaborated with local churches for hosting in-person workshops with unhoused individuals. We actively consulted with church staff to create a workshop environment that fosters participant safety and avoids sense of objectification. We also openly shared that the workshop is for research purpose and

we are working independently from the county and agency that deploy HAA, following best practices in prior work [47].

7.2 Positionality

As researchers, we recognize that our work is influenced by our own identity, experiences and values. Indeed, this project started when some of the representatives from the local homeless community reached out to two of the authors of this paper regarding the existing failures of HAA, given our past work in this domain. At that time, HAA has been deployed locally for more than three years. Past work have documented a series of failures and harms associated with HAA [34]. Those failures and harms have manifested in the real world and generated widespread concerns among the local homeless communities, which motivated this work.

We are researchers who receive research training in the United States from fields including Human-Computer Interaction, Communication, and Social Work. All authors reside in the county where HAA is deployed. Throughout the study, we worked closely with two community co-authors as our co-researchers. One co-author possesses an MSW degree and serves as a frontline worker directly engaged with the unhoused community. The other co-author, holding a Psy.D. degree, brings experiences from working in counseling centers and maintaining close connections with homeless service providers in the county where HAA is deployed.

7.3 Adverse and Unintended Impact

We acknowledge that, as academic researchers and community co-authors, our ability to directly intervene in the closed-door processes involving the design and deployment of HAA is limited. We position our research within a broader call that aims at understanding and supporting grassroots efforts in mitigating AI failures. We discuss in section 5 about how our research findings can contribute valuable insights for policymakers and technical practitioners to combine both "top-down" and "bottom-up" approaches in mitigating socio-technical failures of AI-based decision support tools.

ACKNOWLEDGMENTS

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2022. Machine bias. In *Ethics of data and analytics*. Auerbach Publications, 254–264.

[2] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, A Mojsilović, et al. 2019. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development* 63, 4/5 (2019), 4–1.

[3] Rachel KE Bellamy, Kuntal Dey, Michael Hind, Samuel C Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Jacquelyn Martino, Sameep Mehta, Aleksandra Mojsilovic, and others. 2018. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943* (2018).

[4] Sarah Bird, Miro Dudík, Richard Edgar, Brandon Horn, Roman Lutz, Vanessa Milan, Mehrnoosh Sameki, Hanna Wallach, and Kathleen Walker. 2020. Fairlearn: A toolkit for assessing and improving fairness in AI. *Microsoft, Tech. Rep. MSR-TR-2020-32* (2020).

[5] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–8.

[6] Emily Black, Rakshit Naidu, Rayid Ghani, Kit Rodolfa, Daniel Ho, and Hoda Heidari. 2023. Toward Operationalizing Pipeline-aware ML Fairness: A Research Agenda for Developing Practical Guidelines and Tools. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*. ACM. https://doi.org/10.1145/3617694.3623258

[7] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. 2021. Envisioning Communities: A Participatory Approach Towards AI for Social Good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (Virtual Event, USA) *(AIES '21)*. Association for Computing Machinery, New York, NY, USA, 425–436. https://doi.org/10.1145/3461702.3462612

[8] Alan Borning and Michael Muller. 2012. Next steps for value sensitive design. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1125–1134.

[9] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.

[10] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[11] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300271 event-place: Glasgow, Scotland Uk.

[12] Stevie Chancellor. 2023. Toward Practices for Human-Centered Machine Learning. *Commun. ACM* 66, 3 (2023), 78–85.

[13] Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. How child welfare workers reduce racial disparities in algorithmic decisions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–22.

[14] Cynthia E Coburn and William R Penuel. 2016. Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational researcher* 45, 1 (2016), 48–54.

[15] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

[16] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly exploring application design through speed dating. In *UbiComp 2007: Ubiquitous Computing: 9th International Conference, UbiComp 2007, Innsbruck, Austria, September 16-19, 2007. Proceedings 9*. Springer, 429–446.

[17] 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).

[18] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–23.

[19] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science advances* 4, 1 (2018), eaao5580.

[20] Lilian Edwards and Michael Veale. 2017. Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke L. & Tech. Rev.* 16 (2017), 18.

[21] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[22] Casey Fiesler, Natalie Garrett, and Nathan Beard. 2020. What do we teach when we teach tech ethics? A syllabi analysis. In *Proceedings of the 51st ACM technical symposium on computer science education*. 289–295.

[23] Liza Gak, Seyi Olojo, and Niloufar Salehi. 2022. The distressing ads that persist: Uncovering the harms of targeted weight-loss ads among users with histories of disordered eating. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.

[24] Patricia Garcia, Tonia Sutherland, Niloufar Salehi, Marika Cifor, and Anubha Singh. 2022. No! Re-Imagining Data Practices Through the Lens of Critical Refusal. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (Nov. 2022). https://

doi.org/10.1145/3557997 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[25] Ben Green and Salomé Viljoen. 2020. Algorithmic realism: expanding the boundaries of algorithmic thought. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 19–31. https://doi.org/10.1145/3351095.3372840 event-place: Barcelona, Spain.

[26] Bruce Hanington and Bella Martin. 2012. Universal methods of design: 100 ways to research complex problems. *Develop innovative ideas, and design effective solutions: Rockport Publishers* (2012).

[27] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[28] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2019. Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20*. Springer, 157–171.

[29] Md Naimul Hoque and Klaus Mueller. 2022. Outcome-Explorer: A Causality Guided Interactive Visual Interface for Interpretable Algorithmic Decision Making. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2022), 4728–4740. https://doi.org/10.1109/TVCG.2021.3102051

[30] Michael Katell, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and P. M. Krafft. 2020. Toward situated interventions for algorithmic equity: lessons from the field. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 45–55. https://doi.org/10.1145/3351095.3372874 event-place: Barcelona, Spain.

[31] Anna Kawakami, Venkatesh Sivaraman, Hao-Fei Cheng, Logan Stapleton, Yanghuidi Cheng, Diana Qing, Adam Perer, Zhiwei Steven Wu, Haiyi Zhu, and Kenneth Holstein. 2022. Improving human-AI partnerships in child welfare: understanding worker practices, challenges, and desires for algorithmic decision support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18.

[32] P. M. Krafft, Meg Young, Michael Katell, Jennifer E. Lee, Shankar Narayan, Micah Epstein, Dharma Dailey, Bernease Herman, Aaron Tam, Vivian Guetler, Corinne Bintz, Daniella Raz, Pa Ousman Jobe, Franziska Putz, Brian Robick, and Bissan Barghouti. 2021. An Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 772–781. https://doi.org/10.1145/3442188.3445938 event-place: Virtual Event, Canada.

[33] Bogdan Kulynych, Rebekah Overdorf, Carmela Troncoso, and Seda Gürses. 2020. POTs: Protective Optimization Technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 177–188. https://doi.org/10.1145/3351095.3372853 event-place: Barcelona, Spain.

[34] Tzu-Sheng Kuo, Hong Shen, Jisoo Geum, Nev Jones, Jason I Hong, Haiyi Zhu, and Kenneth Holstein. 2023. Understanding Frontline Workers' and Unhoused Individuals' Perspectives on AI Used in Homeless Services. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[35] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, and Ariel D. Procaccia. 2019. WeBuildAI: Participatory Framework for Algorithmic Governance. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019). https://doi.org/10.1145/3359283 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[36] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1151–1161. https://doi.org/10.1145/3593013.3594070 event-place: Chicago, IL, USA.

[37] Calvin A Liang, Sean A Munson, and Julie A Kientz. 2021. Embracing four tensions in human-computer interaction research with marginalized people. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 2 (2021), 1–47.

[38] Cindy Kaiying Lin and Steven J Jackson. 2023. From Bias to Repair: Error as a Site of Collaboration and Negotiation in Applied Data Science Work. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32.

[39] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–16.

[40] 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/

[41] 3313831.3376445

[41] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.

[42] Sean McGregor. 2021. Preventing repeated real world AI failures by cataloging incidents: The AI incident database. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 15458–15463.

[43] Fengchun Miao, Wayne Holmes, Ronghuai Huang, Hui Zhang, et al. 2021. *AI and education: A guidance for policymakers*. UNESCO Publishing.

[44] Punya Mishra and Matthew J Koehler. 2006. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers college record* 108, 6 (2006), 1017–1054.

[45] 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.

[46] Neema Moraveji, Jason Li, Jiarong Ding, Patrick O'Kelley, and Suze Woolf. 2007. Comicboarding: using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1371–1374.

[47] Ann Oakley. 2013. Interviewing women: A contradiction in terms. In *Doing feminist research*. Routledge, 30–61.

[48] Allegheny County Department of Human Services. 2020. *Allegheny Housing Assessment (AHA) Frequently Asked Questions (FAQs)*. Retrieved January 20, 2023 from https://www.alleghenycounty.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=6442472819

[49] Viktoriya Olari and Ralf Romeike. 2021. Addressing ai and data literacy in teacher education: A review of existing educational frameworks. In *The 16th Workshop in Primary and Secondary Computing Education*. 1–2.

[50] Anne Ottenbreit-Leftwich, Krista Glazewski, Minji Jeon, Cindy Hmelo-Silver, Bradford Mott, Seung Lee, and James Lester. 2021. How do elementary students conceptualize artificial intelligence?. In *Proceedings of the 52nd ACM technical Symposium on computer science education*. 1261–1261.

[51] Lawrence Palinkas, Sarah Horwitz, Carla Green, Jennifer Wisdom, Naihua Duan, and Kimberly Hoagwood. 2013. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health* 42 (11 2013). https://doi.org/10.1007/s10488-013-0528-y

[52] Abel A Reyes, Colin Elkin, Quamar Niyaz, Xiaoli Yang, Sidike Paheding, and Vijay K Devabhaktuni. 2020. A preliminary work on visualization-based education tool for high school machine learning education. In *2020 IEEE Integrated STEM Education Conference (ISEC)*. IEEE, 1–5.

[53] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. 2021. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[54] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. 2021. Modeling Assumptions Clash with the Real World: Transparency, Equity, and Community Challenges for Student Assignment Algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445748 event-place: <conf-loc>, <city>Yokohama</city>, <country>Japan</country>, </conf-loc>.

[55] Samantha Robertson and Niloufar Salehi. 2020. What If I Don't Like Any Of The Choices? The Limits of Preference Elicitation for Participatory Algorithm Design. *arXiv preprint arXiv:2007.06718* (2020).

[56] Juan David Rodríguez-García, Jesús Moreno-León, Marcos Román-González, and Gregorio Robles. 2020. Introducing artificial intelligence fundamentals with LearningML: Artificial intelligence made easy. In *Eighth international conference on technological ecosystems for enhancing multiculturality*. 18–20.

[57] Jeffrey Saltz, Michael Skirpan, Casey Fiesler, Micha Gorelick, Tom Yeh, Robert Heckman, Neil Dewar, and Nathan Beard. 2019. Integrating ethics within machine learning courses. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–26.

[58] Devansh Saxena, Karla Badillo-Urquiola, Pamela J Wisniewski, and Shion Guha. 2020. A human-centered review of algorithms used within the US child welfare system. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–15.

[59] Aaron Shapiro. 2017. Reform predictive policing. *Nature* 541, 7638 (2017), 458–460.

[60] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla, Jess Gallegos, Andrew Smart, Emilio Garcia, et al. 2022. Sociotechnical harms: scoping a taxonomy for harm reduction. *arXiv preprint arXiv:2210.05791* (2022).

[61] Hong Shen, Wesley H Deng, Aditi Chattopadhyay, Zhiwei Steven Wu, Xu Wang, and Haiyi Zhu. 2021. Value cards: An educational toolkit for teaching social impacts of machine learning through deliberation. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 850–861.

[62] Hong Shen, Haojian Jin, Ángel Alexander Cabrera, Adam Perer, Haiyi Zhu, and Jason I Hong. 2020. Designing alternative representations of confusion matrices to support non-expert public understanding of algorithm performance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.

[63] Hong Shen, Leijie Wang, Wesley H. Deng, Ciell Brusse, Ronald Velgersdijk, and Haiyi Zhu. 2022. The Model Card Authoring Toolkit: Toward Community-centered, Deliberation-driven AI Design. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 440–451. https://doi.org/10.1145/3531146.3533110

[64] Hong Shen, Leijie Wang, Wesley H Deng, Ciell Brusse, Ronald Velgersdijk, and Haiyi Zhu. 2022. The model card authoring toolkit: Toward community-centered, deliberation-driven AI design. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 440–451.

[65] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–6.

[66] Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1162–1177.

[67] Stefan Strauß. 2021. " Don't let me be misunderstood": Critical AI literacy for the constructive use of AI technology. *TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis/Journal for Technology Assessment in Theory and Practice* 30, 3 (2021), 44–49.

[68] Harini Suresh, Rajiv Movva, Amelia Lee Dogan, Rahul Bhargava, Isadora Cruxên, Ángeles Martinez Cuba, Guilia Taurino, Wonyoung So, and Catherine D'Ignazio. 2022. Towards intersectional feminist and participatory ML: A case study in supporting Feminicide Counterdata Collection. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 667–678.

[69] Matti Tedre, Tapani Toivonen, Juho Kahila, Henriikka Vartiainen, Teemu Valtonen, Ilkka Jormanainen, and Arnold Pears. 2021. Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. *IEEE access* 9 (2021), 110558–110572.

[70] Tapani Toivonen, Ilkka Jormanainen, Juho Kahila, Matti Tedre, Teemu Valtonen, and Henriikka Vartiainen. 2020. Co-designing machine learning apps in K–12 with primary school children. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 308–310.

[71] David Touretzky, Christina Gardner-McCune, Cynthia Breazeal, Fred Martin, and Deborah Seehorn. 2019. A year in K–12 AI education. *AI magazine* 40, 4 (2019), 88–90.

[72] David Touretzky, Christina Gardner-McCune, Fred Martin, and Deborah Seehorn. 2019. Envisioning AI for K-12: What should every child know about AI?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 9795–9799.

[73] Rhema Vaithianathan and Chamari I Kithulgoda. 2020. *Using Predictive Risk Modeling to Prioritize Services for People Experiencing Homelessness in Allegheny County*. Technical Report. Centre for Social Data Analytics, Auckland, New Zealand. https://www.alleghenycounty.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=6442473749

[74] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. Data Leverage: A Framework for Empowering the Public in Its Relationship with Technology Companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 215–227. https://doi.org/10.1145/3442188.3445885 event-place: Virtual Event, Canada.

[75] Angelina Wang, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. 2022. Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. *Available at SSRN* (2022).

[76] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. 2019. The What-If tool: Interactive probing of machine learning models. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 56–65.

[77] James Wexler, Mahima Pushkarna, Tolga Bolukbasi, Martin Wattenberg, Fernanda Viégas, and Jimbo Wilson. 2020. The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2020), 56–65. https://doi.org/10.1109/TVCG.2019.2934619

[78] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R Rick, and Lilly Irani. 2021. HCI tactics for politics from below: meeting the challenges of smart cities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[79] Sue Wilkinson. 1998. Focus group methodology: a review. *International journal of social research methodology* 1, 3 (1998), 181–203.

[80] Gary KW Wong, Xiaojuan Ma, Pierre Dillenbourg, and John Huan. 2020. Broadening artificial intelligence education in K-12: where to start? *ACM Inroads* 11, 1 (2020), 20–29.

[81] Pak-Hang Wong. 2020. Democratizing Algorithmic Fairness. *Philosophy and Technology* 33, 2 (2020), 225–244. https://doi.org/10.1007/s13347-019-00355-w Publisher: Springer Verlag.

[82] Peter Wright and John McCarthy. 2015. The politics and aesthetics of participatory HCI. *Interactions* 22, 6 (2015), 26–31.

[83] John Zimmerman and Jodi Forlizzi. 2017. Speed dating: providing a menu of possible futures. *She Ji: The Journal of Design, Economics, and Innovation* 3, 1 (2017), 30–50.

APPENDIX

A PARTICIPANTS DEMOGRAPHICS

| Demographic Information	Participant Counts or Statistics
Race	Black or African American (9), White (12), Hispanic (1)
Age	mean: 45.8 , maximum: 71, minimum: 33
Gender	Female (14), Male (8)
Education Level	Highschool degree or equivalent (14), Some college, no degree (1), Bachelor's degree (2), Master's degree (3), Professional degree (1), Doctorate (1)
Homelessness Status (Unhoused Individuals Only)	previously unhoused (7), currently unhoused (4)
Homelessness Chronicity (Unhoused Individuals Only)	decades (1), years (6), months (4)
Services Provided (Frontline Workers and Service Providers only)	Health Service (1), Shelter (3), Informing Design (1), Street Outreach (3)

Table 1: Aggregated participants' self-reported demographics

B   THE AI FAILURE CARDS

B.1   The Onboarding Cards
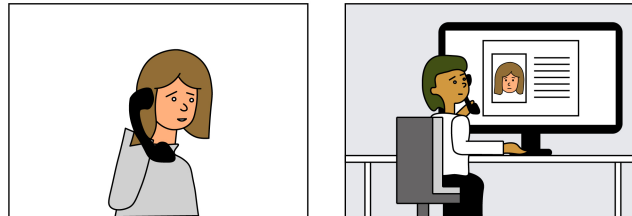
An applicant who wants to apply for housing support finds a staff member in the office or goes to a frontline worker. Then the staff or the worker uses the computer tool to decide.
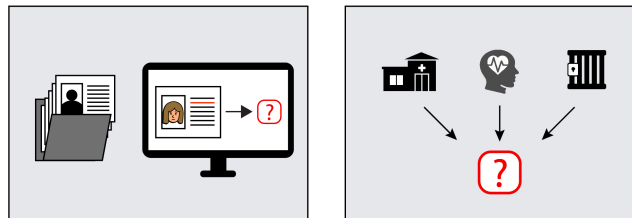
The computer first finds out the application's public service records and uses those records to decide the likelihood of the applicant experiencing three different situations in the future. The three situations are: emergency room visits, mental health inpatient, and jail booking.
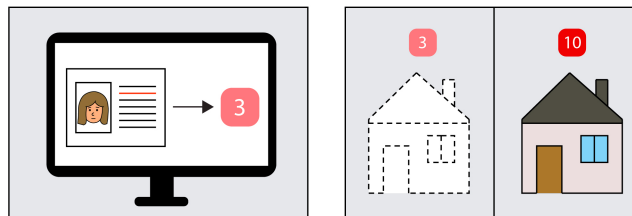
The computer generates a risk score between 1 and 10 based on the likelihood of the applicant experiencing the three kinds of situations. The higher the score, the more likely the applicant could get housing assistance.

Figure 3: The Onboarding Card capturing the workflow and decision making process of HAA.

B.2   The Failure Cards
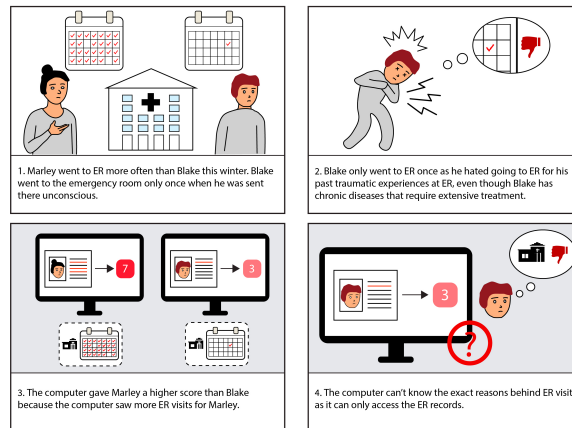
What do you think should be done to help people in a situation like Marley and Blake?

Figure 4: The Failure Card capturing the recurring flaw of "Intervention vs. Prediction".
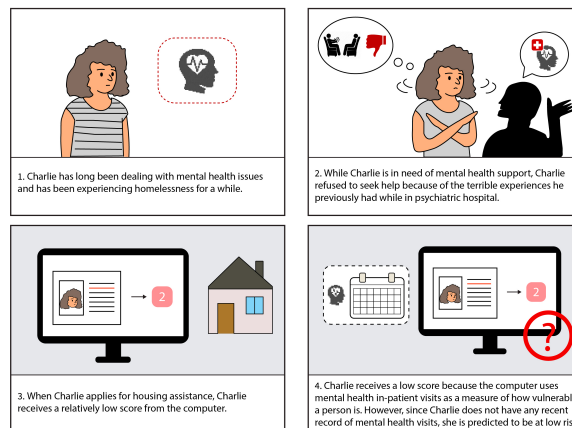
What do you think should be done to help people in a situation like Charlie?

Figure 5: The Failure Card capturing the recurring flaw of "Target-Construct Mismatch".
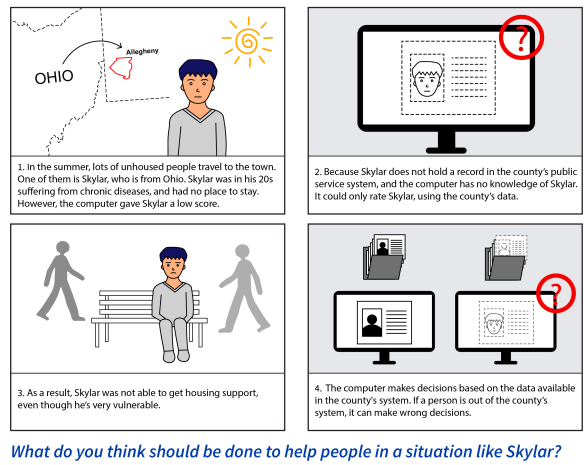
Figure 6: The Failure Card capturing the recurring flaw of "Distribution Shift".

Figure 7: The Failure Card capturing the recurring flaw of "Limits of Prediction".

1. Casey was new to the street: He is 18, kicked out of his home because he was transgender. He did not know how to protect himself from harassment and violence on the street. However, the computer gave Casey a low score.

2. The computer ranks people with chronic homeless records higher, overlooking newcomers like Casey, who are often younger, who also desperately require immediate support.

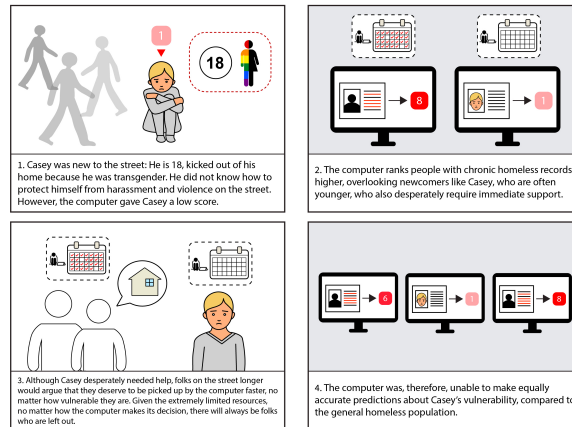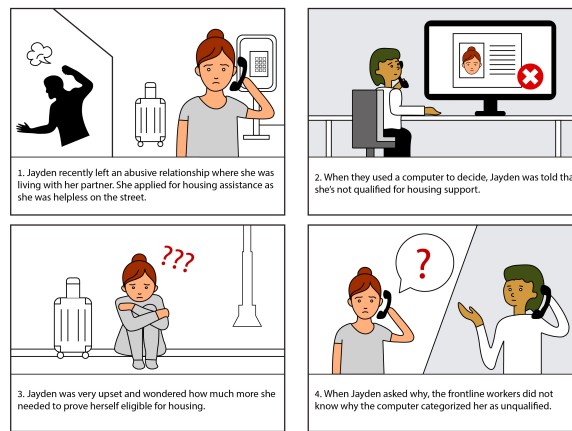3. Although Casey desperately needed help, folks on the street longer would argue that they deserve to be picked up by the computer faster, no matter how vulnerable they are. Given the extremely limited resources, no matter how the computer makes its decision, there will always be folks who are left out.

4. The computer was, therefore, unable to make equally accurate predictions about Casey's vulnerability, compared to the general homeless population.

What do you think should be done to help people in a situation like Casey?

Figure 8: The Failure Card capturing the recurring flaw of "Disparate Performance".

1. Jayden recently left an abusive relationship where she was living with her partner. She applied for housing assistance as she was helpless on the street.

2. When they used a computer to decide, Jayden was told that she's not qualified for housing support.

3. Jayden was very upset and wondered how much more she needed to prove herself eligible for housing.

4. When Jayden asked why, the frontline workers did not know why the computer categorized her as unqualified.

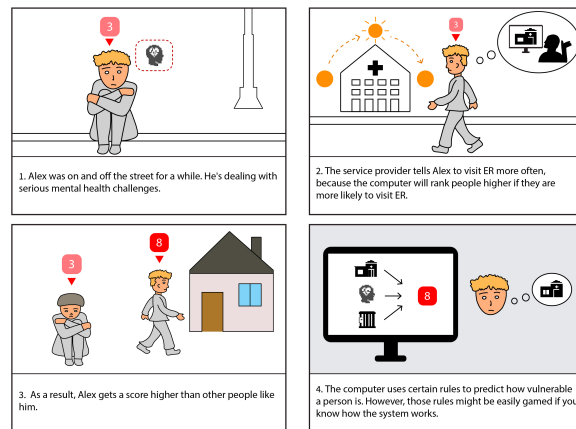*What do you think should be done to help people in a situation like Jordan?*

Figure 9: The Failure Card capturing the recurring flaw of "Lack of contestability".

Figure 10: The Failure Card capturing the recurring flaw of "Goodhart's Law".

B.3 The Mitigation Cards

1. Use other types of records. For example, street medicine records

2. Use data that is not numerical but descriptive words for assessment

3. Include reasons for using public services or resources in public service records. For example, the reason for ER visits.

Your own ones ...

4. Get more input from people rather than trusting the computer

5. Better training program to educate service providers about trauma-informed care and harm reduction

6. The county should offer ways, for example, phone line of staffs, for people to advocate for themselves if they think their score is lower than it should be

7. Host community workshop to encourage input from community members within various social groups, for example, LGBTQ groups and homeless youth.

8. Create public institutions (for example, hospitals) that are more welcoming and more trauma-informed
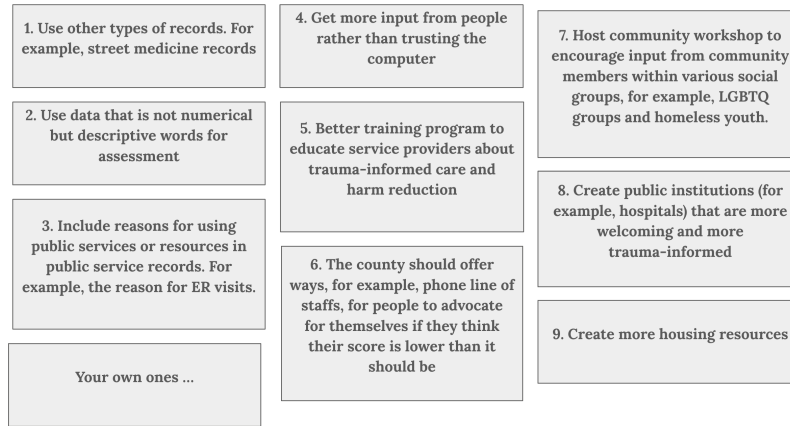
9. Create more housing resources

Figure 11: The Mitigation Cards used in the study.