SoK: Safer Digital-Safety Research Involving At-Risk Users

Rosanna Bellini* Emily Tseng* Noel Warford† Alaa Daffalla*
Tara Matthews‡ Sunny Consolvo‡ Jill Palzkill Woelfer§ Patrick Gage Kelley‡
Michelle L. Mazurek† Dana Cuomo¶ Nicola Dell* Thomas Ristenpart*

*Cornell Tech †University of Maryland ‡Google §JumpCloud ¶Lafayette College

***Abstract*—Research involving at-risk users—that is, users who are more likely to experience a digital attack or to be disproportionately affected when harm from such an attack occurs—can pose significant safety challenges to both users and researchers. Nevertheless, pursuing research in computer security & privacy (S&P) is crucial to understanding how to meet the digital-safety needs of at-risk users and to design safer technology for all. To standardize and bolster safer research involving such users, we offer an analysis of 196 academic works to elicit 14 research risks and 36 safety practices used by a growing community of researchers. We pair this inconsistent set of reported safety practices with oral histories from 12 domain experts to contribute scaffolded and consolidated pragmatic guidance that researchers can use to plan, execute, and share safer digital-safety research involving at-risk users. We conclude by suggesting areas for future research regarding the reporting, study, and funding of at-risk user research.**

1. Introduction

A growing body of research in computer security & privacy (S&P) and human-computer interaction (HCI) is drawing attention to the digital security, privacy, and safety (i.e., *digital safety*) needs and experiences of at-risk users [23, 117, 148]. These works cover users facing a spectrum of risks, including those who face an immediate threat of experiencing a digital attack (survivors of intimate partner violence (IPV) [44], political activists [38]); an increased likelihood to be targeted (LGBTQIA+ people [39], political campaigners [34]); or disproportionate harm from an attack (children [108], people experiencing homelessness [124]).

It is crucial to understand and address the digital-safety needs and experiences of at-risk users, not only to develop effective mitigation approaches, but because improving their digital safety can benefit everyone. Nevertheless, conducting research that involves at-risk users can be daunting. When studying people who may be under a heightened threat of surveillance, harassment, or other digitally-mediated attacks, standard research activities—like recruiting, scheduling, providing participation incentives, and reporting results—may exacerbate risk to research participants and the group(s) they represent, as well as introduce risk to the researchers conducting the work. Thus any research involving at-risk users requires extra caution in order to ensure safety, which can slow research progress.

In this paper, we systematize knowledge from the S&P and HCI research communities to develop pragmatic guidance about reducing risk of harm in the planning, execution, and sharing of digital-safety research involving at-risk users (i.e., *at-risk research* hereafter). Our guidance reflects a systemization of "good" practices based on an analysis of 196 academic works and oral histories from an expert panel of S&P scholars, as guided by the following:

- **Q1:** What digital-safety risks are associated with research involving at-risk users?
- **Q2:** What practices do researchers report employing to help mitigate digital-safety risk in at-risk research?
- **Q3:** What pragmatic guidance might researchers follow to reduce the risk of harm in their digital-safety research involving at-risk users?

From our analysis of the 196 academic works, we identified 36 safety practices researchers reported using to address 14 explicitly articulated risks. We found a wide variety of practices, but only sparse and inconsistent reporting of them: the vast majority of works lacked sufficient detail to assess the safety of research procedures, and, by extension, to replicate safety practices. Furthermore, there were no widely recognized standards for reporting on digital-safety practices. The lack of consistency in reporting safety practices suggests that strategies for reducing the risks inherent with at-risk research may not be widely known or accepted.

To develop pragmatic guidance, we engaged S&P scholars (all co-authors) who have substantive experience in at-risk research. Building on written and oral histories [27, 98], we formulated six strategies for safer digital-safety research involving at-risk users. Such strategies propose that researchers assess and mitigate risks via threat modeling, select the lowest risk method that addresses the research goals, and handle data and publication with care, among others. The strategies complement well-documented guidance on more general ethical research approaches [53, 94, 109] by adding a strong focus on identifying and mitigating risk introduced by digital-safety research involving at-risk users. We hope this systematization of knowledge will help the community work toward more consistent use and reporting of safety strategies for methods used in at-risk research.

2. Background & Related work

In this section, we define key terms, then review methodological approaches to at-risk research from the S&P and HCI communities.

Digital-safety research. We use *digital-safety research* to refer to research about a person's or a group's state of security, privacy, safety, and autonomy, as it relates to their digital footprint. While there are multiple definitions of *safety* [13, 50], we use it to mean when technologies serve, enable, or empower activities rather than being sources of harm (e.g., vectors for harassment, surveillance).

Risk, harm, & research harm. *Risk* is the probability that a person—independently or as part of a group—will experience danger or harm [79]. *Harm* is a negative impact to a person's psychological/emotional, physical, financial/economic, or social/relational condition [42, 119], including injury to safety, rights, or welfare. Harm has been characterized across three dimensions—probability, severity, and duration [61]; severity has multiple dimensions, such as intent, scale, and urgency [119]. We use *research harm* to refer to harm caused to participants, the group(s) they represent, the public, or researchers themselves as a result of research activities [109]. For instance, negligent research approaches may destabilize communities on- and offline (e.g., via deanonymization) [28, 76], and research prototypes can lead to harmful, unintended consequences. In this paper, we focus on harm resulting from the planning, execution, or sharing of research involving at-risk users.

At-risk users. We define a user(s) as being *at-risk* if they face an elevated likelihood of an attack to their digital safety, have factors that influence or exacerbate their chances of being targeted, and/or experience heightened harm as a result of a digitally-mediated attack [148]. Influential factors may be due to *societal factors* (e.g., politics, marginalization), *relationship factors* (e.g., relying on a third party for digital support, having a relationship with an attacker) and *personal circumstances* (e.g., prominence in comparison to others, having access to a sensitive resource) [148]. Identifying differences in risk is crucial to not flatten at-risk users under one universal banner, which could result in inadequate harm mitigation [117, 148].

Recent scholarship in S&P and HCI has covered various topics related to at-risk users, such as the digital-safety needs of political activists [10, 38, 71, 80], survivors of IPV [29, 43, 45, 56, 82], people experiencing homelessness [124] or incarceration [104], and people who experience identity-based marginalization within society (e.g., queer and trans people [48], women [63, 131], and people of color [26]). From this growing volume of empirical research, works have synthesized contextual risk factors [148] and privacy risks for marginalized groups [117] but few directly address how to do digital-safety research safely.

Scholars have developed methodological guidance for research with specific at-risk groups and surveyed some research approaches to representing and evaluating risk and security [50, 119]. Slupska et al. [125] propose participatory threat modeling to elicit often overlooked risk factors of a single at-risk user group (survivors of IPV) for technology design, while Bhalerao et al. [23] call for work with at-risk users to employ confidentiality practices, be attuned to experiences of trauma, and adopt justice-oriented principles at specific (but not all) stages of research. In their review of marginalized users, Sannon and Forte [117] ask that researchers develop shared best practices, including suggestions for fairer participant reimbursement and author position statements. Collectively, these works reveal a need for theoretical and methodological convergence, to draw together HCI expertise in treating at-risk users sensitively with S&P expertise in mitigating digital-safety harm.

Digital-safety & ethical practices. Within the S&P and HCI research communities, it is generally considered common knowledge that research might increase harm to at-risk users. However, these communities lack agreed-upon best practices for measuring, reporting, and implementing approaches to combat the adverse effects of researcher involvement or published works. Some research ethics guidelines and practices have been established to limit or minimize risk of adverse or harmful outcomes in human subjects research broadly. For example, researchers must often secure approval from ethical review committees (e.g., institutional review boards) before conducting human subjects research. Many program committees and journal editors have also established ethical requirements—with varying formality—for publications in their venues; they may reject or retract submissions on ethical grounds (e.g., [62]).

For human subjects research in general, the S&P and HCI communities have institutionalized many ethical practices, such as obtaining informed consent from research participants [68, 109]. Similarly, the S&P community has long-standing norms about when and how to disclose new security vulnerabilities and attacks with the aim of limiting harm [62]. These practices and norms are critically important for promoting ethical research, but they are not necessarily sufficient for research involving at-risk users. As an example, research ethics discussions rarely encourage digital safety-enhancing behaviors by the research team when conducting indirect (i.e., no direct interaction) research about at-risk users [106] or enforce the use of practical safety plans [22]. While all digital-safety practices could be considered forms of ethical practice, not all ethical practices concern safety (e.g., conversations around standards of conduct and moral values). Adopting a safety-conscious approach helps raise awareness of the importance of safety as a specific area of focus in research ethics.

Open questions. While the aforementioned prior works are valuable for digital-safety research involving at-risk users, they have several limitations. Although Sannon and Forte reviewed recent privacy research about marginalized users, some at-risk users are not considered marginalized (e.g., journalists) and experience digital-safety risks beyond threats to privacy, including threats to their security or safety. Frameworks or meta-analyses that address the digital safety of at-risk users (e.g., [16, 50, 135, 148]) do not

discuss how research itself may further harms. Although most reviews [23, 135, 148] conclude with calls for clarity on best practice guidelines or recommendations for performing research safely, no prior works—to our knowledge—address the entire research process (as opposed to specific research stages [13, 23, 117])—despite calls to do so [13, 16, 23, 50, 117, 135, 148].

3. Methods: Analysis of existing works

To systematize existing approaches to safety in at-risk research, we start with an analysis of 196 academic works. We use the results of this analysis (presented in Section 4) and knowledge from eight experienced S&P scholars to generate pragmatic strategies to guide safer digital-safety research involving at-risk users (presented in Section 6). Four authors (first, second, third, and fifth) were involved in the analysis of existing works (the *analysis team*), while all 12 authors (referred to as an *expert panel*) contributed to the systematization of practices into strategies.

The analysis of existing works followed a rapid evidence review (RER) methodology [93, 142]. RERs are a rigorous approach that provide relevant and actionable evidence to strengthen policy and practice. They follow the same steps as a systematic review for identifying, selecting, critically evaluating and analyzing data, but some components are simplified (i.e., exclusion of grey literature) to ensure results are delivered quickly. In doing so, we followed established guidelines for conducting a research synthesis via an RER [93], as described below.

Corpus. To understand how research practices with at-risk users are being reported, we analyzed 196 peer-reviewed papers. We focused our analysis on digital-safety research (rather than technology research more broadly), because we expected that such works would be more likely to report their digital-safety considerations.

To begin, we retrieved an existing open dataset of 127 digital-safety papers from Warford et al. [148] (which include several authors of this work) who analysed 31 distinct at-risk population categories (e.g., journalists, sex workers). We used specialist databases including USENIX, IEEE Xplore, and the ACM Digital Library to identify additional works that were published in premier S&P and HCI venues after this initial dataset was collected: CCS, CHI, CSCW, IEEE S&P, NDSS, PETS, SOUPS, and USENIX Security. Using the same inclusion criteria[1], we considered 3,948 papers published between late 2020 and 2022. Only 65 of these works satisfied the criteria for inclusion. Four papers in our corpus each referenced a prior work containing further details about safety procedures. We therefore included those four additional papers, resulting in a total of 196 papers.[2]

Data extraction criteria and coding process. To develop a set of data extraction criteria, the first author read a randomly selected subset of 20 papers of the corpus (9.8%) to identify exemplar and outlier cases. The third author reviewed the coding criteria to ensure a broad coverage of risks and safety practices. Once this coding criteria was reviewed by two other authors, the first author then read each of the 196 works, and coded the reported methodological approaches to at-risk research (i.e., stated methods), ethical dimensions, and any safety practices performed by the works' authors. The coding process of the 196 papers was performed using an established protocol by one experienced coder, and results checked by a second author (similar to most RER approaches [93]). The results synthesis and clarification of evidence (described next) was performed by the first, second, and fifth authors in iterative steps [142].

Data analysis. To answer **Q1**, the analysis team developed, discussed, and refined our findings into 14 common risks that authors of the works reported as relevant to their research methods (see Table 1 in Section 4). To answer **Q2**, the process was repeated to identify 36 common, distinct safety practices; that is, strategies and actions that authors of the works reported using to support the safety of participants, the at-risk users they represented (i.e., non-participants), or research team members (see Table 2 in Section 4.3). The analysis team considered a practice *common* if it was reported more than three times among (a) research with the same at-risk group, or (b) research with at least three different at-risk groups. Open codes were then synthesized into two abstract categories (independent of practice employed or at-risk user group involved): *research risks* and *digital-safety practices*.

Further systematization with S&P scholars. In aggregate, existing works provided significant information about risks and safety practices involved in at-risk research, but did not provide pragmatic strategies for guiding such research. To answer **Q3** and to complement our RER, we engaged eight S&P scholars to elicit further knowledge on research practices and help formulate actionable safety strategies. These scholars, who may be considered as domain experts, are authors on this paper to represent their contributions to this work. We discuss the process used to develop the strategies in Section 6.

Expert engagement is known to increase the relevance, use, and dissemination of RERs [93, 142], and we hypothesized that expert accounts would provide invaluable information on digital safety that was absent from many works—such oral histories and traditions are often used to teach safe and ethical practices [98].

Limitations. Despite our approach, we cannot overcome bias [93, 142] that occurs when authors, reviewers, or publishers favor specific studies (e.g., promoting methodological orientations, studies of large sample sizes, etc.). Reviewers may request omissions of methodological information (which we have experienced ourselves in some of our prior work), or the authors may omit such information. We do not presume that absence of methodological descriptions on

1. To be included in the Warford et al. [148], dataset, the *primary* purpose of the work had to explore: (a) at-risk users *and* (b) digital-safety threats. Papers had to be full-length, peer-reviewed, and written in English (6,428 works). 127 papers were identified through a rigorous approach of determining relevancy; the dataset for this work is available at this link.

2. The complete list of 196 papers can be found at this link.

digital safety equates to unsafe research, but rather that such authors were either not encouraged or did not feel compelled to report their safety practices. Thus, our RER cannot draw conclusions about how safety practices were enacted, but we can comment on how such practices were accounted for, described, and justified in publications.

Our data was gathered systematically and is large enough to cover common digital-safety challenges in at-risk research as performed by the S&P and HCI communities. As a first systemization on this topic, we scoped our efforts to the research community's knowledge and felt it was appropriate to focus on a representative set of academic works by the privacy and security community. The works we analyzed were written in English and took place in predominantly Western contexts, most often in the U.S. Consequently, this systematization may skew toward Western audiences, which may not accommodate all digital-safety considerations in at-risk research.

4. Reported risks & Safety practices

From these 196 works, we identified 14 commonly anticipated risks incurred by digital-safety research to participants, the group(s) they represented, and researchers (Table 1), as well as the practices researchers used to mitigate those risks (Table 2). This compilation of risks is non-exhaustive as they are products of authors', reviewers', and venues' publication norms, but can provide a useful baseline for digital-safety research. We delineate these risks via *person affected* (participants and at-risk users, researchers) and by *research processes* (data collection, direct research encounters, publication of deliverables), which are illustrated by examples of *specific risks* (e.g., escalation of abuse).

4.1. Risks to participants and at-risk users

Research involves risks for all participants [68], but at-risk users can face greater dangers from data collection, direct research encounters, and publications. These risks apply most directly to participants themselves, but can also extend to other, non-participating members of the groups the participants represent.

From data collection. Empirically grounded research into at-risk users' digital safety involves collecting data about them, typically in a manner designed to minimize their discomfort and enhance their wellbeing. When either indirect methods (e.g., scraping data) or direct interaction are used to elicit at-risk user data, the risk of *unauthorized data access* increases. Parties gaining unauthorized access may be relatively benign, such as colleagues outside the research team, or actively malicious, such as nation-states. Adversaries may gain access to sensitive material or leak participants' research contributions to other harmful parties (e.g., journalists under surveillance [85]). Interception of recruitment materials also qualifies, as it could identify membership in a group of at-risk users (e.g., sex workers seeking anonymity [83]).

For well-resourced adversaries, a breach of research data may occur via a *compelled disclosure*, where researchers could be forced to divulge data to third parties without participants' consent. For example, an adversary might use a subpoena from a legal authority (or even the threat of one) to compel a researcher to disclose what their participants told them [56, 58] or destroy information collected about an adversary. If participants disclose their participation in illegal activities [26], or the researcher identifies elder or child abuse [152], the researcher may be required to notify relevant authorities. These disclosures can exacerbate security threats for some at-risk users, like activists [26] or sex workers [83], who may take steps to avoid law enforcement.

From direct research encounters. Direct research encounters with at-risk participants could also create undue risk. Users at risk of oppression or stigmatization [148] may face coercion by adversaries, inhibiting them from freely participating in research. Data collected may be compromised if researchers do not consider the potential for participants to self-censor if, e.g., they are intimidated by adversaries or consider the research site unsafe [6]. Participants may even be compelled to attend research encounters with adversaries such as partners [44, 45], family members [88], or caregivers [90].

Participants who disclose information that adversaries disapprove of may face retaliation, harming their job prospects [116, 134], access to resources [140], or physical safety [139]. Similarly, participation in research that requires a user to change their protective actions, such as locking an adversary out of a compromised account [56, 85, 140], could incur an *escalation of abuse*. Such situations are highly dangerous, creating an acute risk of other harms, such as physical stalking or monitoring [80].

Digital-safety research may also *distress* or *re-traumatize* at-risk participants, by asking them to recount some of their most sensitive experiences [12, 44], or inadvertently triggering feelings of judgment or shame that they did not take "better steps" to protect themselves from their adversaries [56, 137] (c.f., [31]). Researchers may wish to help by sharing guidance on S&P best practices, but this must be done carefully, as improperly advising participants can *withhold benefit* from them or even cause additional harm. This could include omitting relevant advice about protective practices [113] or underplaying the severity of potential threats, leading participants to make unsafe choices about securing their safety [73]. Even well-informed advice, however, runs the risk of disrupting dedicated S&P support. For instance, participants may feel compelled to take part in research activities to get much-needed help [23, 43].

From publication of research deliverables. Research deliverables like papers and reports may pose risks to at-risk groups downstream of direct research encounters. *Misrepresentation* of participants' experiences of digital risk and harm could perpetuate myths about the causes of their vulnerability [83] or enforce unfair and negative stereotypes [118], as well as potentially leading people in power to create ineffective interventions [90]. Further, as many at-risk groups depend on confidentiality and anonymity as a protective practice, the risk that a reader may *de-anonymize*

Risks posed			Description	Example papers
to participants	...from data collection	Breach of confidentiality	Researchers may be compelled to disclose participant data to an authority without participants' consent, due to subpoena, duties to law enforcement, or parental rights.	[26, 56, 58, 152]
		Unauthorized access	Even when using best-practice data-security tools, adversaries may gain unauthorized access to sensitive participant data.	[83, 85]
	...from direct research, including primary interviews or when researchers offer digital-safety advice	Coercion of contributions	Adversaries may accompany participants to studies and provide or discourage responses, especially when the adversary is an intimate (e.g., a partner, family member, or caregiver).	[44, 56, 88, 90]
		Disruption to support	Researchers may disrupt the normal functioning of digital-safety services and place a participant's security in jeopardy. Participants may also conflate research activities with service provision and feel compelled to participate in research to receive support.	[23, 43]
		Distress and re-traumatization	At-risk participants may be prompted to recount moments where they experienced digital-safety harms, which may cause distress. This can extend to viewing the researcher as a physical threat to a participant's wellbeing.	[12, 31, 44, 56, 137]
		Escalation of abuse	Research activities may require or encourage participants to break routines or take protective actions like removing spyware, which may incite adversaries to escalate their abuse or retaliate against the participant.	[56, 80, 85, 140]
		Withhold benefit	If researchers do not inform participants about the viability of reported threats or available protective practices, participants may be at greater risk.	[73, 113]
	...from the publication of research products	Adversarial feedback	Research may publicize protective strategies in ways that inform adversaries, who then correspondingly adapt or escalate their attacks.	[21, 26, 40, 44, 82, 138]
		Deanonymization	Unsuccessfully paraphrased quotes or poor redaction of participant information might reveal the identities of at-risk participants, particularly those who are public figures.	[34, 44, 45]
		Misrepresentation	Research may inadvertently mischaracterize participants' digital-safety needs, which may disrupt their safety strategies or encourage risky or ineffective interventions.	[83, 90, 118]
to researchers		Burnout and vicarious trauma	Immersion in stories of hate, harassment, and abuse may incur vicarious trauma or secondhand traumatic stress, which may result in burnout or exhaustion.	[11, 31, 43, 91, 100, 139]
		Harassment and intimidation	Researchers may themselves experience hate and harassment due to public statements about their research. Scholars with marginalized identities are particularly susceptible.	[12, 40]
		Liability exposure	Researchers may be subject to criminal prosecution or civil litigation for failing to disclose observed vulnerabilities (of at-risk groups or technical systems) uncovered during their research.	[26, 88, 144]
		Surveillance	Adversaries who have strategies for digitally tracking and monitoring at-risk groups may extend these tactics to researchers.	[104, 114, 121]

TABLE 1: Risks posed by digital-safety research involving at-risk users, systematized from our analysis.

them in reports or papers could have severe consequences. Researchers may include idosyncratic details about participants' experiences [34], add demographics that facilitate jigsaw identification [43], or advertise the location of private communities which may be subsequently targeted.

S&P publications may motivate readers to eliminate barriers for at-risk groups [148], but may also be read by adversaries who then target at-risk users or incite others to do so [40]. These risks include providing instructions about how to attack a particular group of at-risk users, or naming specific software tools or online communities where adversaries can find further information [21, 44, 138]. Even activities that may initially seem beneficial could prove harmful to at-risk groups. For example, compiling disparate useful resources into reviews or best-practice guides could inadvertently reduce the burden of threat intelligence for adversaries [26, 82, 138].

4.2. Risks to researchers

Researchers can also incur emotional, physical, and legal harms that are under-reported in research deliverables [92].

Scholars can face *vicarious trauma* from repeated exposure to dark and distressing content like stories of digital-safety harms. Vicarious trauma, also known as 'second-hand' traumatic stress, is the emotional residue of exposure to traumatic stories and experiences [31, 91]. Research activities—like witnessing active discrimination against a group of at-risk users by other members of society [100], hearing accounts of technology-facilitated abuse [43, 44, 139], or reading online accounts of sexual violence in community groups [11]—can all evoke vicarious trauma. As digital-safety researchers must conduct these activities in the course of doing their jobs, this research puts them at risk of *burnout*, or exhaustion from work-related stress [91]. Experience of exhaustion may be exacerbated for researchers who are *exposed to legal liability* challenges. Researchers may face civil or criminal charges if appropriate parties or companies are not informed of relevant information [133], such as abuse of persons or systems [18, 22].

Researchers may become the target of *harassment and intimidation* in the process of publicizing their research findings, especially when their work revolves around ideas and concepts that may cause backlash, like the digital-safety needs of persecuted users [40]. Scholars with marginalized identities are known to be particularly vulnerable [12, 40]. Such actions may also extend to *surveillance* of researchers by adversaries who may take an interest in pursuing at-risk groups, thereby violating the privacy and physical safety of researchers involved in conducting this work.

4.3. Safety practices

Researchers may deploy *digital-safety practices* to mitigate the risks their research may pose to at-risk participants, the at-risk users they represent, and to the researchers themselves. Our analysis elicited 36 distinct digital-safety practices (SP1–SP36) summarized in Table 2. We discuss these according to six high-level categories (ordered roughly chronologically relative to research project timelines).

Researchers have commenced at-risk research by forming *professional partnerships*, either with practitioners (external experts) or researchers (outside of the initial research team)—for guidance, further training, or recruitment (SP1, SP2). This also included seeking *external review* outside of the bounds of a traditional research institution (i.e., local community, NGO) to address safety concerns (SP4), or incorporating at-risk users into the research team (SP3).

If utilizing direct research approaches, researchers reported being responsive to their own *positionality*, or how the personal characteristics of the researcher may influence participants' safety and the data participants are willing to provide (SP5, SP8, SP9). These practices included justifying their methods of *participant engagement*, such as conducting pilot studies or studies with proxies for at-risk participants (SP6, SP7), as well as the need for additional training or therapeutic support for at-risk users and researchers on emotive topics (SP10, SP11).

Whether conducting direct or indirect research, works often described using *privacy-preserving data collection* approaches, which involved minimizing the amount of data collected (SP12–SP15) or securing the safety of data collection sites using encryption or access control (SP16, SP17, SP20). At-risk participants were also commonly encouraged to use protective practices in contributing to research, such as choosing safer communication modalities or using pseudonyms (SP18, SP19). Works that used *secure data storage and processing* aimed to significantly limit access to data pertaining to at-risk users through access control (SP21), redaction (SP22), encryption (SP23), or secure processing in transit and at rest (SP24). These practices could also extend to preserving the privacy of the research team in contexts where the research team identified a risk of being targeted by adversaries (as described in Section 4.2).

In addition, researchers also practiced *researcher accountability* to at-risk users, including adapting to the particular needs of at-risk users (SP25, SP26), ensuring transparency in what data are collected (SP27, SP29), and minimizing the risk of exploitative approaches that cause further harm (SP28). Finally, several practices identified the need for researchers to critically analyze *sharing and evaluating deliverables*, such as additional steps to ensure anonymity of those involved in research (SP30–SP33, SP35, SP36) and analysing the potential for adversaries to learn more about at-risk users (SP34).

Challenges. The numerous risks posed to at-risk participants and the users they represent (Section 4.1), to researchers (Section 4.2), and the digital-safety practices used in response (Section 4.3) may be a daunting array of considerations for digital-safety research. In most cases, papers in our dataset did not provide sufficient detail, if any, about why particular risks were considered and safety practices used: 27.0% of works (n=53) did not report any digital-safety practices, while 9.6% of works (n=19) contained only statements clarifying approval by an ethics body (such as an IRB), and another 28.5% of papers (n=56) contained one or two sentences pertaining to safety (excluding the 14 works that only reported approval by an ethics body). We determined that only 32 works (16.3%) in our corpus contained clear justifications on why safety practices were used, who performed such practices, and actionable descriptions of the practices they employed. While we organised these safety practices into high-level descriptive categories (Table 2), these labels are not practical, actionable, or comprehensive enough to cover research from start to finish.

5. Methods: Development of strategies

Our analysis of existing works did not provide an answer to **Q3**, as pragmatic safety guidance for planning, executing, and sharing at-risk research was neither offered by nor the focus of those works. We had to go beyond the contents of those papers to gain an understanding of current practices.

Expert panel. To do so, we engaged eight S&P scholars (all co-authors on this paper): established researchers with extensive experience across a broad cross-section of at-risk research. We recruited these scholars by word-of-mouth. The resulting 12 members of our expert panel (i.e., eight experienced S&P scholars, four from our analysis team) together have over 60 years of experience in computer security and digital-safety research involving at-risk users across industry, academia, and non-profit sectors. Our expert panel has, in aggregate, worked on many styles of research projects, ranging from short-term (less than one year) exploratory studies to long-term engagements (5 years or more) involving survivors of IPV [21, 43, 44, 45, 56, 140], political campaign workers [34], refugees [122], survivors of human trafficking [30], political activists [38], journalists [149], low-income communities [154], returning citizens (post-incarcerated individuals) [18, 19], sex workers [83], people experiencing homelessness [153], targets of occupational bullying and harassment [20], online content creators [136], and more. We have therefore worked with a substantial cross-section—18 or 58.1%—of the 31 at-risk groups identified in Warford et al.'s [148] corpus. Our fields span S&P, HCI, and criminology.

Process. The analysis team worked with the S&P scholars to develop a set of strategies that would provide pragmatic guidance for the planning, execution, and sharing of digital-safety research involving at-risk users. The nine-month process of strategy development involved three phases.

In the first phase, we conducted a series of informal discussions about safety challenges and practices in at-risk research. In the second phase, we performed a structured oral history elicitation, known for providing a rich image of past work [15]. The first author implemented a protocol

Category	ID	Digital-safety practices	Example papers
Professional partnerships & Ethical review	SP1	Elicit expert (academic) opinion on topic area	[17, 31, 67, 70, 82, 83, 112, 132, 136]
	SP2	Form professional partnerships (e.g., support services for at-risk users)	[44, 52, 72, 80, 82, 99, 105, 124, 134, 145]
	SP3	Invite and include an at-risk user to join research team	[17, 83, 97, 112]
	SP4	Seek external (non-institutional) ethical review approval or monitoring	[30, 43, 44, 78]
Positionality & Participant engagement	SP5	Build rapport with participants for understanding digital-safety needs	[1, 33, 34, 38, 73, 91, 97, 113, 137]
	SP6	Conduct pilot studies with general (non-at-risk) users	[5, 30, 33, 64, 67, 95, 101]
	SP7	Conduct studies with proxies for at-risk users (e.g., advocacy groups)	[2, 24, 33, 70, 74, 104, 132]
	SP8	Include researchers whose identities affirm participants'	[2, 6, 38, 64, 97, 110, 112, 113, 132, 134]
	SP9	Practice responsiveness in data collection sessions to potential threats	[3, 38, 49, 89, 100, 101, 124, 127, 128, 132]
	SP10	Provide professional therapeutic support for emotive topics	[7, 11, 30, 48, 95, 100, 101, 115, 144]
	SP11	Train team members in working with digital-safety risks	[7, 38, 115, 121]
Privacy-preserving data collection	SP12	Discourage participant self-disclosure (e.g., personal histories)	[1, 7, 25, 52, 70, 75, 118, 123, 137, 144]
	SP13	Focus data collection on supporting participant safety needs	[24, 34, 38, 66, 81, 97, 120, 121, 123, 129]
	SP14	Do not collect or ask for participant demographic data	[17, 26, 64, 83, 84, 104, 120, 124, 136, 145]
	SP15	Do not collect personally identifiable information on participants	[30, 43, 44, 52, 54, 58, 73, 85, 95, 143]
	SP16	Implement protocols for researchers to prevent stalking by adversaries	[30, 60, 80]
	SP17	Separate potential threats from at-risk users during data collection	[6, 72, 88, 96, 97, 100, 110, 115]
	SP18	Permit participants to contribute false information (e.g., pseudonyms)	[17, 54, 58, 78, 83, 100]
	SP19	Offer participants many modalities to contribute (e.g., audio, notes)	[4, 7, 24, 34, 57, 67, 90, 107, 117, 130]
	SP20	Secure confidentiality and privacy of online and in-person research sites	[6, 24, 30, 43, 44, 77, 100, 113, 134, 139]
Secure data storage & processing	SP21	Implement strict data access control measures for research data	[1, 7, 34, 51, 80, 112, 134, 136, 139, 147]
	SP22	Redact participant information prior to analysis by research team	[59, 86, 95, 107, 114, 128, 130, 140, 143, 156]
	SP23	Use encryption for research data in-transit and at-rest	[52, 60, 75, 85, 86, 87, 101]
	SP24	Use non-encrypted safe storage for research data in-transit and at-rest	[7, 30, 34, 90, 97, 114, 130, 132]
Researcher accountability	SP25	Conduct data collection sessions around participant schedules	[1, 35, 54, 65, 97, 111, 120, 128, 139]
	SP26	Offer formal proof of identity as professional researchers	[70, 82, 97, 112, 114, 115]
	SP27	Only use data from publicly accessible sites (e.g., no authorization)	[11, 32, 40, 97, 103, 138, 147, 155]
	SP28	Provide proportional incentives to participants for contributions	[54, 64, 72, 73, 82, 110, 134, 139, 145, 151]
	SP29	Be transparent with participants about risks incurred by research	[24, 26, 38, 54, 57, 69, 95, 110, 113, 128]
Sharing & evaluating deliverables	SP30	Do not attribute reported data contributions with participant identifiers	[7, 8, 9, 34, 55, 84, 114, 117, 134]
	SP31	Do not report participant demographics in research deliverables	[17, 24, 43, 77, 78, 83, 117, 120, 144, 145]
	SP32	Do not report participant names, pseudonyms, or identifiers	[9, 48, 71, 78, 101, 114, 121, 143, 145, 155]
	SP33	Paraphrase or withhold sources of data (e.g., websites they use)	[2, 9, 17, 40, 59, 69, 78, 123, 136, 155]
	SP34	Evaluate research deliverables for adversarial feedback or education	[34, 38, 44, 59, 82, 113]
	SP35	Selectively edit participant data in research deliverables	[7, 9, 11, 40, 55, 124, 139, 140, 150, 151]
	SP36	Provide participants control of their contributions (e.g., permit redaction)	[7, 47, 54, 75, 91, 113, 114, 117, 136]

TABLE 2: The 36 digital-safety practices we identified in our analysis of 196 existing works. The practices are organized into broad categories to aid readability and paired with a random sample of (up to 10) example works.

where each expert was paired with another expert to share oral histories. Each expert had extensive experience and training in discussing challenging topics, and were paired with experts of similar professional experience to minimize potential shame or embarrassment. These pairings collaboratively recorded relevant experiences across their work with at-risk users in a shared, access-controlled document. In this way, we elicited new knowledge, not previously identified in our analysis of existing works: research experiences about digital-safety risks, which we present in this work as *anecdotes*. In total, the 12 members of the expert panel contributed 57 individual accounts of oral histories (per author M: 6.3, SD: 2), totalling 6,600 words (word count per item ranged from 71 to 504, M: 186, SD: 86).

We then used these oral histories in a third phase: consensus building. The first author conducted a content analysis to identify salient themes. The expert panel regrouped to discuss these oral histories in a series of focused meetings to iterate on designing strategies for safer at-risk group research. The panel grounded the strategies in specific, actual research practice from the oral histories, making iterative refinements until we reached consensus. By discussing and identifying areas of disagreement, we were able to concretize six strategies for safer digital-safety research involving at-risk users.

Strategies were chosen as a viable deliverable as they necessitate a plan of action that is designed and (ideally) implemented to achieve a goal (thereby answering **Q3**). We intend for these strategies to help researchers prioritize and customize practices appropriate to their particular research context. The strategies, labeled S1–S6, appear in Table 3. We subsequently linked these strategies back to the findings from our analysis of existing works (Table 2), identifying how reported safety practices could be applied to support the strategies. Note that our linking between the strategies in Table 3 and practices in Table 2 are examples and not exhaustive. The specific practices researchers should apply for each strategy will vary depending on their research goals, context, and the people involved.

Limitations. In sharing oral histories and developing strategies, the 12 members of the expert panel provide personal accounts, which were susceptible to cognitive biases—where a researcher's expectations, opinions, prejudices, or memory may affect their ability to accurately report. During the elicitation phase, all accounts were collected via a standardized, structured procedure and corroborated by multiple co-researchers (all co-authors on this paper) to triangulate these accounts.

Also, the six proposed strategies may not cover all instances that could affect the digital safety of participants, the group(s) they represent, or researchers during the planning, execution, and reporting of at-risk research. There

will undoubtedly be instances in which a strategy does not apply or would not be appropriate. As a result, we envision researchers using these strategies in conjunction with careful reflection on their particular context.

6. Strategies for safer at-risk research

Here we present six strategies that guide researchers to think through which safety practices may apply (including but not limited to the 36 from our review) and how to enact safer research in their context. Rather than sorting through previously applied practices—not all of which may apply to a given research context—the six strategies raise issues that apply across many at-risk research contexts.

Our six strategies begin by highlighting a foundational frame that was omnipresent throughout our discussions, and was agreed upon by all 12 members of the expert panel to provide an orientating mindset for the six strategies:

Research should be treated as an intervention.

This means that one should assume a priori that research will have an impact on the at-risk participants and potentially other members of the at-risk group(s) they represent. Explicitly interventionist research traditions (e.g., clinical trials) are well-understood as sites for potential positive and negative effects, and have established procedures for handling impacts. In digital-safety research, it is tempting to assume many types of studies—such as observational studies or online measurement studies—can be executed without affecting participants or the populations they represent.

But, as reflected in our anecdotes, even digital-safety research with observational or descriptive aims (e.g., surveys) can result in harm. For example, asking participants to recount difficult experiences may trigger trauma responses in participants (*re-traumatization*, Section 4.1), or burnout in researchers over time (*burnout and vicarious trauma*, Section 4.2). Even online measurement studies can have impact if, for example, adversaries learn new tactics from the results, or target researchers they may who disagree with the study findings (*harassment/intimidation*, Section 4.2).

After realizing that research should be treated as an intervention, the importance of employing a strategic approach should become apparent, as should the need to identify areas where strategies could be applied to mitigate risk. We posit that planning research to minimize potential harm and maximize potential benefit is especially important when it involves at-risk users, because harm in this space has the very real potential to be outsized.

6.1. Engage experts early

Digital-safety research involving at-risk users often benefits from a wide range of expertise. We suggest engaging experts as early as possible in research planning. They can make critical contributions to identifying appropriate safety practices and helping to ensure that potential problems are caught and corrected before they lead to harm. Early engagement also helps to avoid putting the expert(s) in the awkward situation of pointing out problems after a research protocol has been fully developed or worse, deployed.

We use "experts" to mean professionals or advocates who have worked with members of the at-risk group, as well as people with expertise relevant to the research more broadly. This might include lawyers, psychotherapists, security engineers, or others well-versed in specific domains. Moreover, users who were formerly at elevated risk of digital-safety threats (i.e., no longer under immediate threat) may be able to offer important expertise. This strategy aims to mitigate the risk that well-intentioned—but unprepared—researchers could inadvertently conduct research that causes harm.

Anecdotes. Our work has greatly benefited from expert engagement. In an ethnographic project with people who had experienced incarceration [19], an early partnership with a frontline service organization gave us access to experts who reviewed our research protocol and advised on appropriate language to use with participants. As researching incarceration can incur emotional responses, the organization also provided participants and the research team with therapeutic support throughout the study to help process exposure to upsetting accounts of trauma and discrimination.

In other research involving people experiencing homelessness [153], our research plans were reviewed by professionals from partner support organizations. These experts helped provide us with an overview of participants' technology use and digital-safety concerns that informed threat modeling (introduced in strategy A2, Section 6.2). With their help, we adapted our recruitment procedures to minimize coercion, provided safe and comfortable locations for interview sessions, and ensured ethical incentive amounts were provided.

Applying the strategy. The inclusion of experts can be useful in overseeing most, if not all, security practices (Table 2), yet we focus on those most relevant to experts. Many at-risk groups have advocates and other support professionals who might be *potential research partners* (SP2) or provide *external review* (SP4) of the safety of research engagements (e.g., review research protocols for safety practices). Researchers should consider *engaging with domain experts* (SP1, SP3) from the beginning and structure the engagement to be mutually beneficial (discussed further in S5 Section 6.5). Doing so helps to ensure that the research plan explicitly considers predictable effects of the research on participants (like those covered in Section 4), and plans for making outcomes beneficial, rather than harmful.

Domain experts might be able to help the research team throughout the research process. For example, they might help think through threat models and risk mitigation, review research protocols, recruit participants or contribute to other logistics, review manuscripts for information that might identify a participant or educate an adversary (see also S6, Section 6.6), perform or assist with direct data collection, be on call to help address unexpected situations, and more. Domain experts can help prepare researchers for emotional reactions to the discussion of sensitive or

ID	Strategy title	Description	Example digital-safety practices
S1	Engage experts early	Consult or partner with domain experts from the beginning to inform and help facilitate safe research plans.	SP1, SP2, SP3, SP4, SP10
S2	Assess and mitigate risks by threat modeling	Apply the S&P practice of threat modeling to research protocols, and continuously update threat models to guide ongoing safety mitigations.	SP11, SP16, SP17, SP20
S3	Select the lowest risk method that addresses the research goals	Before soliciting at-risk users for high-touch methods like interviews, consider proxies (e.g., advocates), or indirect methods (e.g., online measurement).	SP6, SP7, SP12, SP14, SP15, SP27
S4	Respect that at-risk users self-manage risk	At-risk users are often experts in managing their safety risks. Give them choice in how they engage with research safety protocols, and respect the choices they make.	SP9, SP18, SP19, SP25, SP26, SP29
S5	Be an advocate for at-risk users' needs	Research, by its nature, can be extractive. Build reciprocity with at-risk users, and work to help them achieve their goals.	SP5, SP8, SP13, SP28, SP36
S6	Handle data and publications carefully	Data collection and analysis should follow security best-practice, and publications should avoid revealing identities or informing adversaries.	SP21, SP22, SP23, SP24, SP30, SP31, SP32, SP33, SP34, SP35

TABLE 3: Six strategies (S1–S6) for safer digital-safety research involving at-risk users. Each strategy represents a way of thinking about at-risk research. We list example safety practices for each strategy (SP1-SP36, see Table 2), but note that the practices researchers should apply will depend on their research goals, context, and people involved.

traumatizing experiences and help plan for or even provide trauma-informed care [31]. This is known as care planning, which may involve, for example, ensuring *professional therapeutic support* (SP10) is available before, during, or after the research.

All this can involve substantial time and energy from domain experts, and so researchers should consider how to provide proportional reimbursement—which may not necessarily be financial (see S5, Section 6.5). Example approaches include paying *domain experts* for their work (SP1), providing assistance to organizations via *professional partnerships* in return for expert time (SP2), offering co-authorship of academic papers (SP3), or creating and disseminating research reports that would be useful for the experts and their communities.

6.2. Assess and mitigate risks by threat modeling

As the name implies, digital-safety research works in an environment that can contain or attract potential adversaries. This strategy suggests applying threat modeling to the research process itself. In S&P, threat modeling is the practice of identifying relevant adversaries and enumerating their capabilities and goals (c.f., [141]). Identifying threat models is often the first step that engineers take to incorporate security into system design. We argue that threat modeling can help researchers identify and mitigate potential digital-safety risk to their participants, the group(s) they represent, or the researchers themselves.

Anecdotes. Threat modeling improved the digital safety of our research involving groups concerned about surveillance by nation-state actors [38]. As part of that work, we studied the privacy practices of political activists who campaigned against their government. We began our research planning by building an understanding of threats the activists faced, in close consultation with an expert (i.e., a researcher who had personal experience in the country; see also S1, Section 6.1). We determined that the activists' primary adversaries were nation-state actors who had purview over the country's entire telecommunications infrastructure, including activists' access to the global Internet. These adversaries could arrest activists, confiscate their devices, surveil them, and even cause physical harm. As the political climate in the country evolved, so did activists' threat models.

As a result, the interview protocol was designed to allow participants to reveal or not reveal the specific tactics they had used to evade arrests or surveillance. The researchers in this work anticipated that some participants might need to use some of the tactics again in the future (e.g., they do not want their adversaries to discover those tactics from the research). In an effort to create a safe data collection environment for participants, we ensured that the researcher who conducted the interviews had a shared cultural context and spoken language with participants. Recruitment protocols were not revealed, even in the eventual publication. These safeguards helped alleviate participant concerns and mitigate potential harm.

Applying the strategy. Early on, we suggest that researchers perform a threat modeling exercise. This can be a brief, five-part description of (1) the adversary(ies) and their capabilities, (2) the adversary's target and the target's defensive capabilities, (3) the adversary's goal(s), (4) the impact of a successful attack (on the target or others), and (5) the likelihood of an attack. Even coarse estimates (e.g., likely or unlikely) can be helpful. Most of the safety practices in Table 2 could be used to help assess or mitigate research harm; therefore, we discuss the most salient practices as examples to illustrate what this could encompass.

For researchers who are not very familiar with the at-risk group or context, threat modeling can be performed with the help of domain experts (see S1, Section 6.1) and literature reviews (see S3, Section 6.3). Exploring at-risk frameworks [148] and systematic literature reviews [117] may also be helpful in identifying common contextual risks to consider.

Crucially, researchers should consider if and how the

research might change the threats. For example, if the threat modeling suggests that the research might facilitate attacks or exacerbate harms, we would recommend that researchers revise their plan to include appropriate mitigations. For instance, the threat modeling might reveal that researchers consider *isolating potential threats* from at-risk groups through careful method design (e.g., separating family members to elicit data on S&P practices [90]) (SP20), implementing *an anti-stalking protocol* for researchers who suspect they may be followed [80] (SP16), or outlining approaches to *strict data access control measures* if the threat of malicious insiders becomes apparent (SP17).

Depending on the circumstances, researchers may find that the threat model does not materially change in light of the research. Independent of what the threat modeling reveals, it offers a structured rationale to help ensure reasonable safety practices have been considered for the participants, those they represent, and the researchers. The ability to model threats can still be continuously improved through further digital-safety training (SP11).

6.3. Select the lowest risk method that addresses the research goals

To give researchers the space to do the preparatory work to improve research safety, we suggest they take extra time to learn more about the at-risk group and their relevant threat models (see S2, Section 6.2). Typically the highest-risk method (aside from entirely ignoring at-risk user needs) is to directly engage at-risk users as part of the research. However, it may also be possible to learn what is needed in lower-risk ways that avoid direct engagement (e.g., indirect measurement or proxy studies)

Taking extra time and considering whether direct engagement may be avoided can reduce the burden imposed on those who are already vulnerable [41]. This may help avoid the re-traumatization of participants or vicarious trauma for the researchers (see Table 1). Any such decision should balance safety with the risk of further marginalizing an at-risk group. It can be harmful to de-prioritize giving marginalized groups the opportunity to speak for themselves [36], particularly if they often have others speak on their behalf (e.g., disabled people, children [90, 105, 126]). For instance, at-risk users may actively distrust the intentions of other people claiming to speak on their behalf (e.g., sex workers [24, 129]), especially if their perspectives have previously been misrepresented in published works (see misrepresentation, Section 4.1).

Anecdotes. We sought to study the perspectives of abusers in technology-mediated IPV to complement work that explored survivors' perspectives. However, we were unsure how to safely engage abusers. We delayed direct data collection, and instead, conducted measurement studies of online communities that discuss technology abuse tactics [21, 138]. Our measurement studies shed light on abuser perspectives, providing us with valuable expertise.

In our study of political activists [38], we recruited "diaspora activists," proxies for the activism movement who were from the larger at-risk group, but physically more remote from the threat. This helped us answer our research questions, while reducing the risk of research harm.

Applying the strategy. Deciding upon the lowest-risk method to address the research goals is challenging and involves thinking through various factors, such as: (a) Do the researchers have the experience to proceed with the method safely? (b) What is this population's history of exclusion, and can the research be structured to not further their marginalization? (c) Can the research risks for this method be mitigated? and (d) Do the benefits of using this method for the research substantially outweigh the risks?

We suggest that before researchers determine that they must involve at-risk participants in direct engagements like interviews or focus groups, that they first explore safer, alternative methods for answering their research questions. For example, researchers might identify *proxies* (e.g., advocates who work with at-risk groups [45, 113], people who were previously part of the at-risk group), perform pilot studies with proxies or general users, or leverage indirect data sources (e.g., public datasets, online forums, data from prior research, or academic literature) (SP6, SP7, SP27).

When marginalization is a concern, researchers can look for proxies who are closer to the population of interest, or explore alternate ways to include at-risk groups in the research that offer greater benefit to them (such as employment on the research team, SP4). Another approach could be to use measurement studies or sources of indirect data, such as examining online records [155]).

Importantly, working with proxies or indirect methods does not eliminate the need for safety mitigations or avoiding unnecessary burdens on participants. Proxies may themselves be at-risk (as identified by [81]) and their time is valuable (see also S5, in Section 6.5); researchers using indirect methods should consider that some scholars have highlighted [106, 146] that public data are not necessarily expected to be used in research.

If research with proxies or indirect methods is inappropriate, researchers could moderate the amount of data collected from at-risk groups. This can mean *discouraging participants from disclosing* sensitive information (about themselves or others [144]), recognizing that some participants may benefit from sharing (SP12). A commonly safety practice is to collect little-to-no *identifiable information* such as participant *demographics* (SP14, SP15).

6.4. Respect that participants self-manage risk

At-risk users can be well aware of the risks they face and active in managing these risks themselves. To be safer *and* respectful, at-risk participants should be offered the information, and authority to make decisions regarding how they will engage with safety measures planned by researchers, as part of maintaining their own safety. Researchers should plan safe options, provide information that could impact participant decisions, and guide participants in cases where they are unsure or ask for help.

Anecdotes. Many of our studies embed choice in how

participants engage with research across the data collection pipeline. For example, a key concern in our research has been ensuring participants have choice over communication modalities, since at-risk groups may be at heightened risk of surveillance. In one example, when interviewing marginalized groups (LGBTQIA+ people, women, racial/ethnic minorities) who work in computer security, to solicit sensitive anecdotes about their experiences with vulnerability discovery, we provided participants a choice of using phone calls, video chat, or voice chat on a range of platforms [46]. In this and other studies, we have offered participants choice around how they are represented in our data, such as allowing them to opt into audio and/or video recording, or note-taking. In nearly all cases, we offer participants the option of remaining anonymous throughout their interactions with researchers and/or in eventual publications.

This strategy extends to giving at-risk participants respect and decision-making power in their interpersonal interactions with researchers. In our research regarding the privacy of low-income women in deeply patriarchal contexts, we learned that some women who wanted to participate feared repercussions from their husband or mother-in-law. Thus, we offered participants the choice of speaking with us alone, in a group with other women participants, or with their husband or mother-in-law present (to mitigate the risk of harm were they to be excluded).

Applying the strategy. We suggest researchers consider how to provide options that enable participants to manage their own risks. This typically involves providing (a) choices in how participants can engage in the research, and (b) information to inform such choices. Providing participant choice could take the form of offering multiple informed consent options or *communication modalities* such as text-based or voice-based contributions [115], as detailed above (SP19). This may increase the complexity of studies as it relies on being *responsive to emergent threats* (SP9); for example, adding remote data collection procedures may introduce threats not relevant to in-person data collection at workplaces [69] or public locations [80]. Complexity can also be problematic if the options overwhelm participants.

Nevertheless, we have found that even simple options and information can be beneficial—such as permitting participants to *contribute false information* (e.g., a pseudonym as in [17, 83]) or for researchers to *provide formal proof of identity* to potential participants (SP18, SP26). Participants enacting choices around risk can be subtle, for example, a participant proposing *data collection sessions around their schedule*, when they feel safe and able to participate (SP25). Researchers can also be *transparent about the risks of the research* (e.g., Tseng et al.'s approach to understanding escalation [139]), ensuring participants have good information upon which to base their risk assessments (SP29).

This strategy does not override the need for researchers to maintain up-to-date threat models (see S2, Section 6.2): researchers should not burden participants with all threat modeling and risk assessment. Instead, this strategy suggests that once a threat model has been identified and mitigations have been created, participants should be informed and given choice about how to engage with the mitigations—and should be trusted to make that choice.

6.5. Be an advocate for participants

Research involving at-risk users will have an impact on them. This strategy encourages researchers to be advocates for participants, shaping beneficial impacts in the near- and long-term. While ethical practice guides assert that the benefits of a study should outweigh the risks [68, 109], this strategy goes further, encouraging researchers to think about how to structure research to have direct, near- and long-term digital-safety or other benefits for participants.

Anecdotes. We have engaged in a range of activities to advocate for and support at-risk participants in our research. In some of our more mature work (i.e., after several years of working with the at-risk community), we have developed programs in which team members are trained to provide direct, individual assistance to at-risk individuals regarding their digital-safety [43, 56, 139, 140]. This service enables data collection for research, but it runs as a service first and foremost. At-risk individuals receive assistance regardless of whether they consent to participate in research, and research data is only collected if they consent.

Of course, it may not be feasible or advisable to provide such high-touch assistance. Other ways we have advocated for or supported at-risk groups in our research include direct financial incentives (as is often suggested for work with at-risk groups [23]); and collating curated lists of well-established digital-safety tools (e.g., encrypted messaging apps) or practices (e.g., using a password manager; providing pointers to advice guides that have been vetted by security experts) [43]. We have also advocated for systemic change by supporting others who advocate for change, by writing and making public reports for community members, by and pushing for regulation in legislative efforts [37].

Applying the strategy. To achieve immediate positive impact for participants and the group(s) they represent, researchers should adopt a mindset of being advocates through *proportional incentives* (SP28). In some cases, researchers may not know what participants want or need, especially if they are new to working with the at-risk group. Nonetheless, we propose that by *ensuring that participants engage researchers they identify with* (e.g., they are similar in race, ethnicity, age, gender expression) during data collection, the benefits of disclosing challenging accounts of risk (like racism to someone who "gets it" [137]) may be more immediately evident to the participant. (SP8) By *building rapport with participants and experts* (SP5), researchers can better understand community priorities and goals for participating in digital-safety research, and plan and implement activities that meet those needs. This could be done by focusing *data collection on supporting participants' needs* over 'interesting' content from a research perspective (e.g., sensationalist content [104, 118]) that may not effectively advocate for their wellbeing (SP13).

Similarly, researchers could consider keeping a regular dialogue with those who work with at-risk groups to ensure they do not overpromise and underdeliver. While doing so is undoubtedly challenging, providing *participants control of their contributions to any research deliverables* (e.g., through redaction) can help to ensure these efforts for advocacy are reflective of participant need (SP36).

This approach can be applied to shorter- and longer-term research projects; researchers should not need to approach every project involving at-risk users as though they are entering a long-term commitment. By centering the needs of at-risk participants and the group(s) the represent, and taking reasonable steps to address them, researchers can adopt a "scientist-advocate" viewpoint that can improve research *and* benefits for at-risk user(s).

We discourage researchers from promising or implying assurances that they cannot guarantee. For example, we suggest that researchers do not assure participants that the research will result in systemic change, or that such change will be swift. Effective methods to address complicated digital-safety issues are often slow to materialize, can involve unexpected road blocks, and may even depend on fundamental changes to society [131].

6.6. Handle data and publications carefully

Our final strategy is to handle the data collected from at-risk research and the resulting publications, talks, and other outputs with care. While care is always recommended for human-subjects research, the sensitivity of the data (including audio or video recordings, and intermediary analysis documents) and results generated from digital-safety research involving at-risk users warrants special protections.

Anecdotes. When we interviewed people involved with U.S. political campaigns [34], the audio recordings we collected contained stories that, if made public or accessed by adversaries, could potentially harm the participant's career, the campaign for which they worked, or even the political party their candidate represented. To mitigate this risk, two research team members transcribed the audio recordings themselves to avoid sharing data outside of the immediate research team (e.g., with a professional transcription company). Security experts reviewed our reports to ensure they did not include attacks or vulnerabilities that might inform adversaries (in this case sophisticated nation-state actors).

Across other published works, we used many other protections, including omitting details about research procedures (e.g., recruitment methods, see S2 Section 6.2) [38], editing quotations [21, 44, 138], and excluding demographic information to prevent re-identification [34, 82, 83].

Our experiences have also demonstrated a role for caution in handling media attention after publication. After publishing our work understanding abuser tactics in IPV, we received inquiries from reporters seeking our expertise. Some reporters may be incentivized to seek sensationalist headlines. For example, several have contacted us wanting to explicitly write stories on stalkerware, despite the fact that our research suggested that it is a less prevalent attack vector in IPV than everyday privacy violations like account compromise. Our statements to reporters explicitly state this, and we always tell them to contact the communications staff of our partner organizations.

Applying the strategy. We suggest that researchers map out the expected lifecycle of data collected, think through ways in which the data may be exploitable, and define policies regarding sensitivity levels based on the threat models developed with experts (see Sections 6.1 and 6.2). This can inform mitigations, such as *implementing strict data access control measures* (SP21), and *deletion or redaction schedules* (SP22). Authentication and data protection should follow the state-of-the-art in computer security, such as using *secure cloud or dedicated infrastructure*, and employing *encryption* (SP23, SP24).

Before publication, we suggest that *papers and other research artifacts be reviewed for their potential to inform adversaries* (SP34). Researchers could *reduce the granularity of demographic information* they report by only using aggregated summary statistics, paraphrasing participant quotes to *remove or change potentially identifying word choices*, or not *associating any contributions with identifiers* (SP30, SP31, SP32, SP35). In some cases, researchers may need to *withhold the origins of data*, such as websites or physical research sites (SP33). Researchers might also ask *domain and privacy experts* to assess the possibility of participant identification (see S1, Section 6.1); in doing so, it may be helpful to provide the experts with a per-participant breakdown of all data in the publication.

A researcher's concern that publications might inform future adversaries may also cause them difficulty deciding whether and when to disclose attack details. We suggest that researchers consider disclosing attack details when the gain from disclosure is high (e.g., to push forward future mitigations) or if adversaries can discover these tactics easily (SP34). In these situations, reporting them is unlikely to impact future attacks.

7. Future directions

Our analysis of research risks (Section 4) and systemization of strategies and relevant safety practices (Section 6) can help researchers plan, execute, and report on safer digital-safety research involving at-risk users, a need highlighted by other works [13, 16, 23, 50, 117, 135, 148]. But some of our strategies stand in contrast to, or generate friction with, research norms. For example, delaying research efforts (Section 6.3) may appear to conflict with the research community's drive for progress, action, and publication [14]. Our suggestion to consider the impact that disclosing vulnerabilities might have on at-risk participants and the group(s) they represent (Section 6.6) requires updating the disclosure processes used by the security community [35].

Here, we discuss how these friction points are both practical career challenges for researchers and opportunities for collective improvement in the S&P and HCI research communities. To achieve this, we revisit the broader digital-safety research ecosystem (see Section 2), and discuss po-

tential alterations to research *deliverables*, study *protocols*, research *community-building*, and scientific *funding*.

Reporting safety strategies in deliverables. Across the venues where digital-safety research is growing, publications rarely mention how digital safety was addressed during the research (Section 4). The works that *do* report safety strategies are valuable for research, practice, and training as the field evolves. Toward enabling future at-risk research, we suggest that the S&P community normalize including a "Safety strategies" section in publications, potentially under a methods section. It may describe the research's high-level approach to safety, as well as what safety practices were employed to support participants, the group(s) they represent, and the researchers involved (Table 2). The research community can encourage including these details in calls for papers, as well as in guidelines for peer review. A focus on safety strategies could draw more attention to safety in the field, and allow researchers who are doing excellent work in this area to demonstrate their approach — potentially leading to the discovery of unreported safety practices.

Incentivizing safer research protocols. The research community might also work towards establishing standards for safety practices. Importantly, these standards must be flexible and context-sensitive, to account for the wide array of contingencies and study designs possible in at-risk research. One way to achieve this context-specificity is to incentivize researchers to think critically and systematically about their safety procedures early, via concrete plans for safety within research proposals. Proposal reviewers could evaluate research plans for safety plans and make sure they are consistent with our strategies. As a start, reviewers could consider checking whether suitable expert partners will be consulted (S1, Section 6.1). Ethics review boards might similarly lean on our strategies to help in evaluating research protocols. Such early consideration and dialogue with protocol evaluators will also smooth the way to downstream publication of safety strategies.

Considering research safety early is particularly important in a funding climate where researchers are increasingly being asked to share anonymous data or make it accessible for secondary use. Data sharing can play an important role in at-risk research, because it is difficult to acquire knowledge of digital-safety harms experienced by hard-to-reach populations, but it may also inadvertently enable their adversaries to worsen their attacks (Section 6.6). To reconcile these potential harms and benefits, we encourage further research into the possibility of sharing data in at-risk research via improved computational tools, research procedures, and data-sharing paradigms.

Increasing the complexity of digital-safety research protocols may place additional time demands on both researchers and at-risk users; for example, longer timelines to ethics board approvals (S5, Section 6.5). As such, more work is needed to cultivate best practices throughout the research community, to ensure this time and effort is used effectively and properly rewarded [4, 117, 148].

Funding programs. One mechanism for cultivating and rewarding best-practice at-risk research is to improve research funding. Targeted grants could allow researchers to budget for the time and expertise to do this work, particularly for academic researchers who primarily rely on federal or national grant funding (e.g., NSF, NIH, EPSRC, ARC). Programs might allow support for non-traditional roles, such as community coordinators who can manage relationships with partners or dedicated mental health and well-being support. In other fields, funders have programs specifically designed to train students in both practice and research, such as the U.S. NIH Medical Scientist Training Program [102], which provides funding to train students as clinician-scientists qualified in medical practice and research. Similar programs could be considered in computer science.

Building a digital-safety research community. Digital safety is becoming more critical as online hate and harassment threatens at-risk users [135], as tools for surveillance are increasingly normalized in consumer technologies [21, 29, 138], and as computing pervades critical domains like health, education, and finance. To rise to these challenges, we need to increase the amount of research about technology and digital-safety problems involving at-risk users. Yet safe versions of this research can take significant time and energy—to do threat modeling (S2, Section 6.2), develop partnerships with suitable experts (S1, Section 6.1), prepare for emotional labor (S3, Section 6.3), and more.

Research communities can promote safe work with at-risk users by acknowledging this invisible labor and building support infrastructure (e.g., training resources and mentorship programs) [15, 92]. To make safety-focused labor visible, institutions should give researchers time and incentives to do it [81], and by training new researchers through workshop and conference development. Ultimately, this requires a push for research communities to value quality over quantity of publications and research deliverables: a tradeoff we see as a chance to improve the work we do, in service of the at-risk communities we are inspired to support.

8. Acknowledgements

We are deeply grateful to our shepherd and anonymous reviewers for their efforts to help improve this manuscript. We would also like to thank Elissa Redmiles, Rebecca Umbach, Georgina Powell, and the other participants of the Protecting At-Risk Users Workshop for their insightful contributions. This work was funded in part by NSF Award #1916096, as well as gifts from JPMorgan Chase. This material is based in part upon work supported by DARPA under grant HR00112010011. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the United States Government or DARPA.

References

[1] Daniel A. Adler, Emily Tseng, Khatiya C. Moon, John Q. Young, John M. Kane, Emanuel Moss, David C. Mohr, and Tanzeem Choudhury. Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[2] Tanisha Afnan, Yixin Zou, Maryam Mustafa, and Florian Schaub. Aunties, Strangers, and the FBI: Online Privacy Concerns and Experiences of Muslim-American Women. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, 2023.

[3] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. Privacy Concerns and Behaviors of People with Visual Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 2015.

[4] Tousif Ahmed, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. Addressing Physical Safety, Security, and Privacy for People with Visual Impairments. In *Proceedings of the Twelth USENIX Conference on Usable Privacy and Security*, 2016.

[5] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. "I am uncomfortable sharing what I can't see": Privacy Concerns of the Visually Impaired with Camera Based Assistive Applications. In *29th USENIX Security Symposium (USENIX Security)*, 2020.

[6] Deena Alghamdi, Ivan Flechais, and Marina Jirotka. Security practices for households bank customers in the kingdom of Saudi Arabia. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, 2015.

[7] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Joshua Gracie, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[8] Noura Alomar and Serge Egelman. Developers Say the Darnedest Things: Privacy Compliance Processes Followed by Developers of Child-Directed Apps. *Proceedings on Privacy Enhancing Technologies*, 2022(4), 2022.

[9] Ashwaq Alsoubai, Jihye Song, Afsaneh Razi, Nurun Naher, Munmun De Choudhury, and Pamela J. Wisniewski. From 'Friends with Benefits' to 'Sextortion:' A Nuanced Investigation of Adolescents' Online Sexual Risk Experiences. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[10] Adriana Alvarado Garcia, Alyson L Young, and Lynn Dombrowski. On making data actionable: How activists use imperfect data to foster social change for human rights violations in Mexico. *PACM HCI*, 1(CSCW), 2017.

[11] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 2016.

[12] Nazanin Andalibi, Ashley Lacombe-Duncan, Lee Roosevelt, Kylie Wojciechowski, and Cameron Giniel. LGBTQ Persons' Use of Online Spaces to Navigate Conception, Pregnancy, and Pregnancy Loss: An Intersectional Approach. *ACM Transactions on Computer-Human Interaction*, 29(1), 2022.

[13] Alissa N. Antle. The ethics of doing research with vulnerable populations. *Interactions*, 24(6), 2017.

[14] Mariam Asad. Prefigurative Design as a Method for Research Justice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019.

[15] Madeline Balaam, Rob Comber, Rachel E. Clarke, Charles Windlin, Anna Ståhl, Kristina Höök, and Geraldine Fitzpatrick. Emotion Work in Experience-Centered Design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[16] Lemi Baruh, Ekin Secinti, and Zeynep Cemalcilar. Online Privacy Concerns and Privacy Management: A Meta-Analytical Review: Privacy Concerns Meta-Analysis. *Journal of Communication*, 67, 2017.

[17] Catherine Barwulor, Allison McDonald, Eszter Hargittai, and Elissa M. Redmiles. "Disadvantaged in the American-dominated Internet": Sex, Work, and Technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[18] Rosanna Bellini, Simon Forrest, Nicole Westmarland, Dan Jackson, and Jan David Smeddinck. Choice-Point: Fostering Awareness and Choice with Perpetrators in Domestic Violence Interventions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[19] Rosanna Bellini, Simon Forrest, Nicole Westmarland, and Jan David Smeddinck. Mechanisms of Moral Responsibility: Rethinking Technologies for Domestic Violence Prevention Work. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[20] Rosanna Bellini, Patrick Olivier, and Rob Comber. "That Really Pushes My Buttons": Designing Bullying and Harassment Training for the Workplace. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[21] Rosanna Bellini, Emily Tseng, Nora McDonald, Rachel Greenstadt, Damon McCoy, Thomas Ristenpart, and Nicola Dell. "So-called privacy breeds evil": Narrative Justifications for Intimate Partner Surveillance in Online Forums. In *Proceedings of the ACM on Human-Computer Interaction*, volume CSCW3, 2021.

[22] Rosanna Bellini, Alexander Wilson, and Jan David Smeddinck. Fragments of the past: Curating peer support with perpetrators of domestic violence. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[23] Rasika Bhalerao, Vaughn Hamilton, Allison McDonald, Elissa M. Redmiles, and Angelika Strohmayer. Ethical Practices for Security Research with At-Risk Populations. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2022.

[24] Rasika Bhalerao, Nora McDonald, Hanna Barakat, Vaughn Hamilton, Damon McCoy, and Elissa Redmiles. Ethics and Efficacy of Unsolicited Anti-Trafficking SMS Outreach. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[25] Alex Bowyer, Kyle Montague, Stuart Wheater, Ruth McGovern, Raghu Lingam, and Madeline Balaam. Understanding the Family Perspective on the Storage, Sharing and Handling of Family Civic Data. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[26] Maia J. Boyd, Jamar L. Sullivan Jr., Marshini Chetty, and Blase Ur. Understanding the Security and Privacy Advice

Given to Black Lives Matter Protesters. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[27] Rebecca Campbell. "It's the Way That You Do It": Developing an Ethical Framework for Community Psychology Research and Action. *American Journal of Community Psychology*, 58(3-4), 2016.

[28] Stevie Chancellor, Michael L. Birnbaum, Eric D. Caine, Vincent M. B. Silenzio, and Munmun De Choudhury. A Taxonomy of Ethical Tensions in Inferring Mental Health States from Social Media. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 2019.

[29] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. The spyware used in intimate partner violence. In *Proc. IEEE S&P*, 2018.

[30] Christine Chen, Nicola Dell, and Franziska Roesner. Computer security and privacy in the interactions between victim service providers and human trafficking survivors. In *Proceedings of the 28th USENIX Conference on Security Symposium*, SEC'19, 2019.

[31] Janet X. Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. Trauma-Informed Computing: Towards Safer Technology Experiences for All. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[32] Taejoong Chung, Jinyoung Han, Daejin Choi, Ted Taekyoung Kwon, Jong-Youn Rha, and Hyunchul Kim. Privacy Leakage in Event-based Social Networks: A Meetup Case Study. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 2017.

[33] Lizzie Coles-Kemp, Nick Robinson, and Claude P. R. Heath. Protecting The Vulnerable: Dimensions of Assisted Digital Access. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[34] Sunny Consolvo, Patrick Gage Kelley, Tara Matthews, Kurt Thomas, Lee Dunn, and Elie Bursztein. "Why wouldn't someone think of democracy as a target?": Security practices & challenges of people involved with u.s. political campaigns. In *Proceedings of the 29th USENIX Security Symposium*, 2021.

[35] Raymundo Cornejo, Robin Brewer, Caroline Edasis, and Anne Marie Piper. Vulnerability, Sharing, and Privacy: Analyzing Art Therapy for Older Adults with Dementia. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, 2016.

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

[37] Dana Cuomo and Natalie Dolci. The TECC Clinic: An innovative resource for mitigating technology-enabled coercive control. *Women's Studies International Forum*, 92, 2022.

[38] Alaa Daffalla, Lucy Simko, Tadayoshi Kohno, and Alexandru G. Bardas. Defensive Technology Use by Political Activists During the Sudanese Revolution. In *2021 IEEE Symposium on Security and Privacy (SP)*, 2021.

[39] Michael Ann DeVito, Ashley Marie Walker, and Julia R. Fernandez. Values (Mis)alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[40] Periwinkle Doerfler, Andrea Forte, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, and Damon McCoy. "I'm a Professor, which isn't usually a dangerous job": Internet-facilitated Harassment and Its Impact on Researchers. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 2021.

[41] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. Social Justice-Oriented Interaction Design: Outlining Key Design Strategies and Commitments. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, 2016.

[42] Marilyn J. Field, Richard E. Behrman, and Institute of Medicine (US) Committee on Clinical Research Involving Children. *Glossary, Acronyms, and Laws and Regulations*. National Academies Press (US), London, UK, 2004.

[43] Diana Freed, Sam Havron, Emily Tseng, Andrea Gallardo, Rahul Chatterjee, Thomas Ristenpart, and Nicola Dell. "Is my phone hacked?": Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019.

[44] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[45] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW), 2017.

[46] Kelsey R Fulton, Samantha Katcher, Kevin Song, Marshini Chetty, Michelle L Mazurek, Chloé Messdaghi, and Daniel Votipka. Vulnerability discovery for all: Experiences of marginalization in vulnerability discovery. In *32nd USENIX Security Symposium (USENIX Security)*, 2023.

[47] Cally Gatehouse, Matthew Wood, Jo Briggs, James Pickles, and Shaun Lawson. Troubling Vulnerability: Designing with LGBT Young People's Ambivalence Towards Hate Crime Reporting. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[48] Christine Geeng, Mike Harris, Elissa Redmiles, and Franziska Roesner. "Like Lesbians Walking the Perimeter": Experiences of U.S. LGBTQ+ Folks With Online Security, Safety, and Privacy Advice. In *Proceedings of the 29th USENIX Security Symposium*, 2023.

[49] Christine Geeng, Jevan Hutson, and Franziska Roesner. Usable Sexurity: Studying people's Concerns and Strategies When Sexting. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, 2020.

[50] Matt Goerzen, Elizabeth Anne Watkins, and Gabrielle Lim. Entanglements and exploits: Sociotechnical security as an analytic framework. In *9th USENIX Workshop on Free and Open Communications on the Internet (FOCI 19)*, 2019.

[51] Moritz Gruber, Christian Höfig, Maximilian Golla, Tobias Urban, and Matteo Große-Kampmann. "We may share the number of diaper changes": A Privacy and Security Analysis of Mobile Child Care Applications. *Proceedings on Privacy Enhancing Technologies*, 2022(3), 2022.

[52] Tamy Guberek, Allison McDonald, Sylvia Simioni, Abraham H. Mhaidli, Kentaro Toyama, and Florian Schaub. Keeping a Low Profile? Technology, Risk and Privacy among Undocumented Immigrants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[53] Diana L. Gustafson and Fern Brunger. Ethics, "Vulnerability," and Feminist Participatory Action Research With a Disability Community. *Qualitative Health Research*, 24(7), 2014.

[54] Vaughn Hamilton, Hanna Barakat, and Elissa M. Redmiles. Risk, Resilience and Reward: Impacts of Shifting to Digital Sex Work, 2022.

[55] Heidi Hartikainen, Afsaneh Razi, and Pamela Wisniewski. Safe Sexting: The Advice and Support Adolescents Receive from Peers regarding Online Sexual Risks. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[56] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. Clinical computer security for victims of intimate partner violence. In *Proc. USENIX Security*, 2019.

[57] Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang. Cooperative Privacy and Security: Learning from People with Visual Impairments and Their Allies. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security*, 2019.

[58] Jeremy Heyer, Zachary Schmitt, Lynn Dombrowski, and Svetlana Yarosh. Opportunities for Enhancing Access and Efficacy of Peer Sponsorship in Substance Use Disorder Recovery. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[59] Geng Hong, Zhemin Yang, Sen Yang, Xiaojing Liaoy, Xiaolin Du, Min Yang, and Haixin Duan. Analyzing Ground-Truth Data of Mobile Gambling Scams. In *2022 IEEE Symposium on Security and Privacy (SP)*, 2022.

[60] Dominik Hornung, Claudia Müller, Irina Shklovski, Timo Jakobi, and Volker Wulf. Navigating Relationships and Boundaries: Concerns around ICT-uptake for Elderly People. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 2017.

[61] James R. Hébert, William A. Satariano, Daniela B. Friedman, Cheryl A. Armstead, Allen Greiner, Tisha M. Felder, Thomas A. Coggins, Sora Tanjasiri, and Kathryn L. Braun. Fulfilling Ethical Responsibility: Moving Beyond the Minimal Standards of Protecting Human Subjects from Research Harm. *Progress in community health partnerships : research, education, and action*, 9(0), 2015.

[62] IEEE. IEEE Symposium on Security and Privacy 2023. https://sp2023.ieee-security.org/cfpapers.html.

[63] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S Ackerman, and Eric Gilbert. Yes: Affirmative consent as a theoretical framework for understanding and imagining social platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[64] Jane Im, Sarita Schoenebeck, Marilyn Iriarte, Gabriel Grill, Daricia Wilkinson, Amna Batool, Rahaf Alharbi, Audrey Funwie, Tergel Gankhuu, Eric Gilbert, and Mustafa Naseem. Women's Perspectives on Harm and Justice after Online Harassment. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[65] Janne J. Jensen and Mikael B. Skov. A review of research methods in children's technology design. In *Proceedings of the 2005 conference on Interaction design and children*, IDC '05, 2005.

[66] Rikke Bjerg Jensen, Lizzie Coles-Kemp, and Reem Talhouk. When the Civic Turn turns Digital: Designing Safe and Secure Refugee Resettlement. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[67] Crescent Jicol, Julia Feltham, Jinha Yoon, Michael J Proulx, Eamonn O'Neill, and Christof Lutteroth. Designing and Assessing a Virtual Reality Simulation to Build Resilience to Street Harassment. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[68] Erin Kenneally and David Dittrich. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *SSRN Electronic Journal*, 1, 1, 2012.

[69] Vera Khovanskaya, Phoebe Sengers, and Lynn Dombrowski. Bottom-Up Organizing with Tools from On High: Understanding the Data Practices of Labor Organizers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[70] Hyeok Kim, Youjin Hwang, Jieun Lee, Youngjin Kwon, Yujin Park, and Joonhwan Lee. Personalization Trade-offs in Designing a Dialogue-based Information System for Support-Seeking of Sexual Violence Survivors. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[71] Yong Ming Kow, Yubo Kou, Bryan Semaan, and Waikuen Cheng. Mediating the Undercurrents: Using Social Media to Sustain a Social Movement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 2016.

[72] Sandjar Kozubaev, Fernando Rochaix, Carl DiSalvo, and Christopher A. Le Dantec. Spaces and Traces: Implications of Smart Technology in Public Housing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[73] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How Effective is anti-phishing Training for Children? In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security*, 2017.

[74] Celine Latulipe, Ronnie Dsouza, and Murray Cumbers. Unofficial Proxies: How Close Others Help Older Adults with Banking. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[75] Stevens Le Blond, Alejandro Cuevas, Juan Ramón Troncoso-Pastoriza, Philipp Jovanovic, Bryan Ford, and Jean-Pierre Hubaux. On Enforcing the Digital Immunity of a Large Humanitarian Organization. In *2018 IEEE Symposium on Security and Privacy (SP)*, 2018.

[76] Debora de Castro Leal, Angelika Strohmayer, and Max Krüger. On Activism and Academia: Reflecting Together and Sharing Experiences Among Critical Friends. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[77] Ada Lerner, Helen Yuxun He, Anna Kawakami, Silvia Catherine Zeamer, and Roberto Hoyle. Privacy and Activism in the Transgender Community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[78] Zi Li and Bonnie Nardi. "There Should Be More Than One Voice in A Healthy Society": Infrastructural Violence and Totalitarian Computing in China. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 2021.

[79] Niklas Luhmann, Rhodes Barrell, Nico Stehr, and Gollhard Bechmann. *Risk: A Sociological Theory*. Routledge, 2017.

[80] William R. Marczak and Vern Paxson. Social Engineering Attacks on Government Opponents: Target Perspectives. In *Proceedings on Privacy Enhancing Technologies*, 2017.

[81] Hiroaki Masaki, Kengo Shibata, Shui Hoshino, Takahiro Ishihama, Nagayuki Saito, and Koji Yatani. Exploring Nudge Designs to Help Adolescent SNS Users Avoid Pri-

vacy and Safety Threats. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[82] Tara Matthews, Kathleen O'Leary, Anna Turner, Manya Sleeper, Jill Palzkill Woelfer, Martin Shelton, Cori Manthorne, Elizabeth F. Churchill, and Sunny Consolvo. Stories from Survivors: Privacy & Security Practices when Coping with Intimate Partner Abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 2017.

[83] Allison McDonald, Catherine Barwulor, Michelle L. Mazurek, F. Schaub, and Elissa M. Redmiles. "It's stressful having all these phones": Investigating Sex Workers' Safety Goals, Risks, and Practices Online. In *Proceedings of the 30th USENIX Security Symposium*, 2021.

[84] Nora McDonald, Alison Larsen, Allison Battisti, Galina Madjaroff, Aaron Massey, and Helena Mentis. Realizing Choice: Online Safeguards for Couples Adapting to Cognitive Challenges. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, 2020.

[85] Susan E. McGregor, Polina Charters, Tobin Holliday, and Franziska Roesner. Investigating the Computer Security Practices and Needs of Journalists. In *24th USENIX Security Symposium (USENIX Security)*, 2015.

[86] Susan E. McGregor, Elizabeth Anne Watkins, Mahdi Nasrullah Al-Ameen, Kelly Caine, and Franziska Roesner. When the Weakest Link is Strong: Secure Collaboration in the Case of the Panama Papers. In *26th USENIX Security Symposium (USENIX Security)*, 2017.

[87] Susan E. McGregor, Elizabeth Anne Watkins, and Kelly Caine. Would You Slack That? The Impact of Security and Privacy on Cooperative Newsroom Work. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 2017.

[88] Bridget Christine McHugh, Pamela J. Wisniewski, Mary Beth Rosson, Heng Xu, and John M. Carroll. Most Teens Bounce Back: Using Diary Methods to Examine How Quickly Teens Recover from Episodic Online Risk Exposure. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 2017.

[89] Hamid Mehmood, Tallal Ahmad, Lubna Razaq, Shrirang Mare, Maryem Zafar Usmani, Richard Anderson, and Agha Ali Raza. Towards Digitization of Collaborative Savings Among Low-Income Groups. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019.

[90] Helena M. Mentis, Galina Madjaroff, and Aaron K. Massey. Upside and Downside Risk in Online Security for Older Adults with Mild Cognitive Impairment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[91] Aparna Moitra, Syed Ishtiaque Ahmed, and Priyank Chandra. Parsing the 'Me' in #MeToo: Sexual Harassment, Social Media, and Justice Infrastructures. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[92] Wendy Moncur. The emotional wellbeing of researchers: considerations for practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, 2013.

[93] Philip Moons, Eva Goossens, and David R. Thompson. Rapid reviews: the pros and cons of an accelerated review process. *European Journal of Cardiovascular Nursing*, 20(5), 2021.

[94] Cosmin Munteanu, Heather Molyneaux, Wendy Moncur, Mario Romero, Susan O'Donnell, and John Vines. Situational Ethics: Re-thinking Approaches to Formal Ethics Requirements for Human-Computer Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, 2015.

[95] Collins W. Munyendo, Yasemin Acar, and Adam J. Aviv. "Desperate Times Call for Desperate Measures": User Concerns with Mobile Loan Apps in Kenya. In *2022 IEEE Symposium on Security and Privacy (SP)*, 2022.

[96] Savanthi Murthy, Karthik S. Bhat, Sauvik Das, and Neha Kumar. Individually Vulnerable, Collectively Safe: The Security and Privacy Practices of Households with Older Adults. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[97] Tyler Musgrave, Alia Cummings, and Sarita Schoenebeck. Experiences of Harm, Healing, and Joy among Black Women and Femmes on Social Media. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[98] National Academy of Sciences, National Academy of Engineering (US) and Institute of Medicine (US) Committee on Science, Engineering, and Public Policy. *On Being a Scientist: A Guide to Responsible Conduct in Research: Third Edition*. National Academies Press (US), 2009.

[99] James Nicholson, Ben Morrison, Matt Dixon, Jack Holt, Lynne Coventry, and Jill McGlasson. Training and Embedding Cybersecurity Guardians in Older Communities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[100] Fayika Farhat Nova, Michael Ann DeVito, Pratyasha Saha, Kazi Shohanur Rashid, Shashwata Roy Turzo, Sadia Afrin, and Shion Guha. "Facebook Promotes More Harassment": Social Media Ecosystem, Skill and Marginalized Hijra Identity in Bangladesh. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 2021.

[101] Borke Obada-Obieh, Lucrezia Spagnolo, and Konstantin Beznosov. Towards Understanding Privacy and Trust in Online Reporting of Sexual Assault. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, 2020.

[102] National Institute of General Medical Sciences. Medical scientist training program, 2022. https://nigms.nih.gov/training/instpredoc/Pages/PredocOverview-MSTP.aspx,.

[103] Kentrell Owens, Anita Alem, and Franziska Roesner. Electronic Monitoring Smartphone Apps: An Analysis of Risks from Technical, Human-Centered, and Legal Perspectives. In *31st USENIX Security Symposium (USENIX Security)*, 2022.

[104] Kentrell Owens, Camille Cobb, and Lorrie Cranor. "You Gotta Watch What You Say": Surveillance of Communication with Incarcerated People. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[105] Xinru Page, Andrew Capener, Spring Cullen, Tao Wang, Monica Garfield, and Pamela J. Wisniewski. Perceiving Affordances Differently: The Unintended Consequences When Young Autistic Adults Engage with Social Media. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[106] Jessica Pater, Casey Fiesler, and Michael Zimmer. No Humans Here: Ethical Speculation on Public Data, Unintended Consequences, and the Limits of Institutional Review. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 2022.

[107] Justin Petelka, Lucy Van Kleunen, Liam Albright, Elizabeth Murnane, Stephen Voida, and Jaime Snyder. Being (In)Visible: Privacy, Transparency, and Disclosure in the Self-Management of Bipolar Disorder. In *Proceedings of*

the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, 2020.

[108] Lara Schibelsky Godoy Piccolo, Pinelopi Troullinou, and Harith Alani. Chatbots to Support Children in Coping with Online Threats: Socio-technical Requirements. In *Designing Interactive Systems Conference 2021*, DIS '21, 2021.

[109] Office for Human Research Protections (OHRP). The Belmont Report, 2010.

[110] Mohammad Rashidujjaman Rifat, Mahiratul Jannat, Mahdi Nasrullah Al-Ameen, S M Taiabul Haque, Muhammad Ashad Kabir, and Syed Ishtiaque Ahmed. Purdah, amanah, and gheebat: Understanding privacy in bangladeshi "pious" muslim communities. In *ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '21, 2021.

[111] Sabirat Rubya and Svetlana Yarosh. Interpretations of Online Anonymity in Alcoholics Anonymous and Narcotics Anonymous. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 2017.

[112] Nithya Sambasivan, Amna Batool, Nova Ahmed, Tara Matthews, Kurt Thomas, Laura Sanely Gaytán-Lugo, David Nemer, Elie Bursztein, Elizabeth Churchill, and Sunny Consolvo. "They Don't Leave Us Alone Anywhere We Go": Gender and Digital Abuse in South Asia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[113] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchill. "Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, 2018.

[114] Pedro Sanches, Vasiliki Tsaknaki, Asreen Rostami, and Barry Brown. Under Surveillance: Technology Practices of those Monitored by the State. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 2020.

[115] Shruti Sannon and Dan Cosley. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[116] Shruti Sannon and Dan Cosley. Toward a More Inclusive Gig Economy: Risks and Opportunities for Workers with Disabilities. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[117] Shruti Sannon and Andrea Forte. Privacy research with marginalized groups: What we know, what's needed, and what's next. In *Proceedings of the ACM on Human-Computer Interaction*, CSCW2, 2022.

[118] Morgan Klaus Scheuerman, Stacy M. Branham, and Foad Hamidi. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 2018.

[119] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R Brubaker. A framework of severity for harmful content online. *Proceedings of the ACM on Human-Computer Interaction*, CSCW, 2021.

[120] Zachary Schmitt and Svetlana Yarosh. Participatory Design of Technologies to Support Recovery from Substance Use Disorders. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 2018.

[121] Irina Shklovski and Volker Wulf. The Use of Private Mobile Phones at War: Accounts From the Donbas Conflict. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, 2018.

[122] Emrys Shoemaker, Gudrun Svava Kristinsdottir, Tanuj Ahuja, Dina Baslan, Bryan Pon, Paul Currion, Pius Gumisizira, and Nicola Dell. Identity at the margins: examining refugee experiences with digital identity systems in Lebanon, Jordan, and Uganda. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, COMPASS '19, 2019.

[123] Lucy Simko, Ada Lerner, Samia Ibtasam, Franziska Roesner, and Tadayoshi Kohno. Computer Security and Privacy for Refugees in the United States. In *2018 IEEE Symposium on Security and Privacy (SP)*, 2018.

[124] Manya Sleeper, Tara Matthews, Kathleen O'Leary, Anna Turner, Jill Palzkill Woelfer, Martin Shelton, Andrew Oplinger, Andreas Schou, and Sunny Consolvo. Tough times at transitional homeless shelters: Considering the impact of financial insecurity on digital security and privacy. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI'19, 2019.

[125] Julia Slupska, Scarlet Dawson Dawson Duckworth, Linda Ma, and Gina Neff. Participatory threat modelling: Exploring paths to reconfigure cybersecurity. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[126] Katta Spiel, Eva Hornecker, Rua Mae Williams, and Judith Good. ADHD and Technology Research – Investigated by Neurodivergent Readers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 2022.

[127] Denny L. Starks, Tawanna Dillahunt, and Oliver L. Haimson. Designing Technology to Support Safety for Transgender Women & Non-Binary People of Color. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, DIS '19 Companion, 2019.

[128] Enno Steinbrink, Lilian Reichert, Michelle Mende, and Christian Reuter. Digital Privacy Perceptions of Asylum Seekers in Germany: An Empirical Study about Smartphone Usage during the Flight. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 2021.

[129] Angelika Strohmayer, Jenn Clamen, and Mary Laing. Technologies for Social Justice: Lessons from Sex Workers on the Front Lines. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[130] Sharifa Sultana, Mitrasree Deb, Ananya Bhattacharjee, Shaid Hasan, S.M.Raihanul Alam, Trishna Chakraborty, Prianka Roy, Samira Fairuz Ahmed, Aparna Moitra, M Ashraful Amin, A.K.M. Najmul Islam, and Syed Ishtiaque Ahmed. Unmochon;: A Tool to Combat Online Sexual Harassment over Facebook Messenger. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[131] Sharifa Sultana, François Guimbretière, Phoebe Sengers, and Nicola Dell. Design within a patriarchal society: Opportunities and challenges in designing for rural women in bangladesh. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018.

[132] Sharifa Sultana, Sadia Tasnuva Pritha, Rahnuma Tasnim, Anik Das, Rokeya Akter, Shaid Hasan, S.M. Raihanul Alam, Muhammad Ashad Kabir, and Syed Ishtiaque Ahmed. 'ShishuShurokkha': A Transformative Justice Approach for Combating Child Sexual Abuse in Bangladesh. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[133] Kaiwen Sun, Yixin Zou, Jenny Radesky, Christopher Brooks, and Florian Schaub. Child Safety in the Smart Home: Parents' Perceptions, Needs, and Mitigation Strategies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 2021.

[134] Julia Słupska, Ruba Abu-Salma, Selina Cho, and Nayanatara Prakash. "They Look at Vulnerability and Use That to Abuse You": Participatory Threat Modelling with Migrant Domestic Workers. In *31st USENIX Security Symposium (USENIX Security)*, 2022.

[135] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, Damon McCoy, Sarah Meiklejohn, Thomas Ristenpart, and Gianluca Stringhini. SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. In *2021 IEEE Symposium on Security and Privacy (SP)*, 2021.

[136] Kurt Thomas, Patrick Gage Kelley, Sunny Consolvo, Patrawat Samermit, and Elie Bursztein. "It's common and a part of being a content creator": Understanding How Creators Experience and Cope with Hate and Harassment Online. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 2022.

[137] Alexandra To, Wenxia Sweeney, Jessica Hammer, and Geoff Kaufman. "They Just Don't Get It": Towards Social Technologies for Coping with Interpersonal Racism. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 2020.

[138] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. In *29th USENIX Security Symposium (USENIX Security)*, 2020.

[139] Emily Tseng, Diana Freed, Kristen Engel, Thomas Ristenpart, and Nicola Dell. A Digital Safety Dilemma: Analysis of Computer-Mediated Computer Security Interventions for Intimate Partner Violence During COVID-19. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[140] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. Care Infrastructures for Digital Security in Intimate Partner Violence. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[141] Paul C Van Oorschot. *Computer Security and the Internet: Tools and Jewels from Malware to Bitcoin*. Springer, 2021.

[142] Tracey Varker, David Forbes, Lisa Dell, Adele Weston, Tracy Merlin, Stephanie Hodson, and Meaghan O'Donnell. Rapid evidence assessment: increasing the transparency of an emerging methodology. *Journal of Evaluation in Clinical Practice*, 21(6), 2015.

[143] Aditya Vashistha, Abhinav Garg, Richard Anderson, and Agha Ali Raza. Threats, Abuses, Flirting, and Blackmail: Gender Inequity in Social Media Voice Forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 2019.

[144] Krishna Venkatasubramanian, Jeanine L. M. Skorinko, Mariam Kobeissi, Brittany Lewis, Nicole Jutras, Pauline Bosma, John Mullaly, Brian Kelly, Deborah Lloyd, Mariah Freark, and Nancy A. Alterio. Exploring A Reporting Tool to Empower Individuals with Intellectual and Developmental Disabilities to Self-Report Abuse. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 2021.

[145] Jessica Vitak, Yuting Liao, Mega Subramaniam, and Priya Kumar. 'I Knew It Was Too Good to Be True": The Challenges Economically Disadvantaged Internet Users Face in Assessing Trustworthiness, Avoiding Scams, and Developing Self-Efficacy Online. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 2018.

[146] Jessica Vitak, Nicholas Proferes, Katie Shilton, and Zahra Ashktorab. Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards. *Journal of Empirical Research on Human Research Ethics*, 12(5), 2017.

[147] Ye Wang, Zhicong Lu, and Roger Wattenhofer. Gay Dating on Non-dating Platforms: The Case of Online Dating Activities of Gay Men on a Q&A Platform. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 2022.

[148] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L. Mazurek, Manya Sleeper, and Kurt Thomas. SoK: A framework for unifying at-risk user research. In *Proc. IEEE S&P*, 2022.

[149] Noel Warford, Collins W. Munyendo, A. Mediratta, Adam J. Aviv, and Michelle L. Mazurek. Strategies and Perceived Risks of Sending Sensitive Documents. In *30th USENIX Security Symposium (USENIX Security'21)*, 2021.

[150] Miranda Wei, Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. Anti-Privacy and Anti-Security Advice on TikTok: Case Studies of Technology-Enabled Surveillance and Control in Intimate Partner and Parent-Child Relationships. In *The Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021.

[151] Daricia Wilkinson and Bart Knijnenburg. Many Islands, Many Problems: An Empirical Examination of Online Safety Behaviors in the Caribbean. In *CHI Conference on Human Factors in Computing Systems*, 2022.

[152] Pamela Wisniewski, Heng Xu, Mary Beth Rosson, and John M. Carroll. Parents Just Don't Understand: Why Teens Don't Talk to Parents about Their Online Risk Experiences. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, 2017.

[153] Jill Palzkill Woelfer and David G. Hendry. Homeless young people's experiences with information systems: Life and work in a community technology center. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010.

[154] Jill Palzkill Woelfer, Amy Iverson, David G. Hendry, Batya Friedman, and Brian T. Gill. Improving the safety of homeless young people with mobile phones: Values, form and function. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 2011.

[155] Minhui Xue, Gabriel Magno, Evandro Cunha, Virgilio Almeida, and Keith W. Ross. The Right to be Forgotten in the Media: A Data-Driven Study. In *Proceedings of the Twelth USENIX Conference on Usable Privacy and Security*, 2016.

[156] Yixin Zou, Allison McDonald, Julia Narakornpichit, Nicola Dell, Thomas Ristenpart, Kevin Roundy, Florian Schaub, and Acar Tamersoy. The Role of Computer Security Customer Support in Helping Survivors of Intimate Partner Violence. In *30th USENIX Security Symposium (USENIX Security)*, 2021.

Appendix A.
Meta-Review

A.1. Summary

This paper presents a systematic analysis of 203 academic articles on digital safety research with at-risk users. By identifying 14 research risks and 36 safety practices, as well as consulting with 12 domain experts, the authors provide consolidated guidance for researchers in this field. The study highlights the need for more consistent reporting and suggests future areas of research on at-risk user research.

A.2. Scientific Contributions

- Provides a Valuable Step Forward in an Established Field
- Other: Systematization of Knowledge and Recommendations for Researchers

A.3. Reasons for Acceptance

1) The paper is a well-written and organized resource that provides valuable insights and guidance for researchers interested in studying at-risk populations. Reviews deem it useful and worth sharing with students and researchers in the field.
2) The engagement of domain experts in the research process adds credibility and provides practical strategies for at-risk research. The reviews highlight this aspect, noting that it enhances the overall contribution of the paper.