Supporting the Digital Safety of At-Risk Users: Lessons Learned from 9+ Years of Research and Training

TARA MATTHEWS and ELIE BURSZTEIN, Google, Mountain View, California, USA
PATRICK GAGE KELLEY, Google, New York, New York, USA
LEA KISSNER, Google, Mountain View, California, USA
ANDREAS KRAMM, Google, Munich, Germany
ANDREW OPLINGER, Google, San Francisco, California, USA
ANDREAS SCHOU, MANYA SLEEPER, and STEPHAN SOMOGYI, Google, Mountain View, California, USA
DALILA SZOSTAK, Google, New York, New York, USA
KURT THOMAS, Google, Mountain View, California, USA
ANNA TURNER, Google, New York, New York, USA
JILL PALZKILL WOELFER and LAWRENCE L. YOU, Google, Mountain View, California, USA
IZZIE ZAHORIAN, Google, Seattle, Washington, USA
SUNNY CONSOLVO, Google, Mountain View, California, USA

Creating information technologies intended for broad use that allow everyone to participate safely online—which we refer to as *inclusive digital safety*—requires understanding and addressing the digital-safety needs of a diverse range of users who face elevated risk of technology-facilitated attacks or disproportionate harm from such attacks—i.e., *at-risk users*. This article draws from more than 9 years of our work at Google to understand and support the digital safety of at-risk users—including survivors of intimate partner abuse, people involved with political campaigns, content creators, youth, and more—in technology intended for broad use. Among our learnings is that designing for inclusive digital safety across widely varied user needs and dynamic contexts is a *wicked problem* with no "correct" solution. Given this, we describe frameworks and design principles we have developed to help make at-risk research findings practically applicable to technologies intended for broad use and lessons we have learned about communicating them to practitioners.

CCS Concepts: • **Security and privacy → Human and societal aspects of security and privacy**; • **Human-centered computing → Human computer interaction (HCI)**;

Lea Kissner, Jill Palzkill Woelfer, Lawrence L. You, and Izzie Zahorian completed their work on the At-Risk Research Program while they were at Google.

Authors' Contact Information: Tara Matthews (corresponding author), Google, Mountain View, California, USA; e-mail: taramatthews@google.com; Elie Bursztein, Google, Mountain View, California, USA; Patrick Gage Kelley, Google, New York, New York, USA; Lea Kissner, LinkedIn, Mountain View, California, USA; Andreas Kramm, Google, Munich, Germany; e-mail: akramm@google.com; Andrew Oplinger, Google, San Francisco, California, USA; Andreas Schou, Google, Mountain View, California, USA; Manya Sleeper, Google, Mountain View, California, USA; e-mail: manya@google.com; Stephan Somogyi, Google, Mountain View, California, USA; e-mail: somogyi@google.com; Dalila Szostak, Google, New York, New York, USA; Kurt Thomas, Google, Mountain View, California, USA; Anna Turner, Google, New York, New York, USA; Jill Palzkill Woelfer, Independent, Seattle, Washington, USA; e-mail: jillwoelfer@gmail.com; Lawrence L. You, Independent, Palo Alto, California, USA; Izzie Zahorian, United States Digital Service, Washington, District of Columbia, USA; e-mail: ezahorian@gmail.com; Sunny Consolvo, Google, Mountain View, California, USA; e-mail: sconsolvo@google.com.

ACM 1557-7325/2025/6-ART22
https://doi.org/10.1145/3716382

Additional Key Words and Phrases: At-risk users, digital safety, privacy, security, abuse, at-risk framework, user states framework, design principles, high-risk users, vulnerable users, technology-facilitated abuse

ACM Reference format:
Tara Matthews, Elie Bursztein, Patrick Gage Kelley, Lea Kissner, Andreas Kramm, Andrew Oplinger, Andreas Schou, Manya Sleeper, Stephan Somogyi, Dalila Szostak, Kurt Thomas, Anna Turner, Jill Palzkill Woelfer, Lawrence L. You, Izzie Zahorian, and Sunny Consolvo. 2025. Supporting the Digital Safety of At-Risk Users: Lessons Learned from 9+ Years of Research and Training. *ACM Trans. Comput.-Hum. Interact.* 32, 3, Article 22 (June 2025), 39 pages.
https://doi.org/10.1145/3716382

1 Introduction

Everyone should be able to participate online safely. But achieving this in practice is incredibly complex for those who create information technologies intended for broad use—whether through research, design, development, policy, or other domains. Users face a host of digital-safety[1] threats—such as privacy invasions, security attacks, abuse, and more [95]—that can lead to emotional, financial, relational, or even physical harm [88]. Such threats and corresponding needs are diverse and depend on an individual user's contextual circumstances [101]. One-size-fits-all digital-safety solutions can create barriers to some users protecting themselves, potentially exposing them to harm [101]. Key to creating safer technology *for everyone*—which we refer to as supporting *inclusive digital safety*—is understanding and supporting the digital-safety risks, reactions, and needs of a diverse range of users.

While users with "general" risk levels are commonly studied, a growing body of research about at-risk users has exposed considerable diversity in digital-safety experiences. *At-risk users* face circumstances that elevate their chances of experiencing a technology-facilitated attack and/or disproportionate harm from such an attack [101]. At-risk users' varied experiences reveal ways that assumptions about technology use, often based on an optimistic view of "general users" and their digital-safety risks,[2] can break down. For example, assuming devices are used by a *single individual* who can control how it is used breaks down for several at-risk groups—such as survivors of **intimate partner abuse (IPA)** who described being coerced to give abusers access to their devices [60]; and low-income New Yorkers who described economic constraints keeping them in living situations with people they did not trust which exposed them to a variety of privacy and security threats via their mobile devices [30].

Further, understanding the more severe digital-safety cases can shed light on risks all users face and how to prioritize solutions, to the benefit of both users who may be at elevated risk *and* general users. For example, some users are at-risk because they have extremely limited time to cope with digital-safety threats, like people involved with political campaigns [24] or emergency room staff [93]. Generally, anyone may find themselves in periods of greater-than-ordinary time constraints

[1] *Digital safety* (noun) and *digital-safety* (adjective) are umbrella terms, inclusive of topics pertaining to security, privacy, abuse, and/or any other safety-related topic in which technology is involved.

[2] "General user" is a construct of the kind of user that makes up the perceived majority of a technology's use, typically viewed as somewhere in the middle/average or even optimistically [104], for example, assuming that they have good access to trusted devices and reliable Internet, would benefit from increased technology use, and do not face extraordinary challenges. "General users" may not be real, instead reflecting how the holders of the concept view users—opening the concept up to biases, such as over-indexing on people from **Western, Educated, Industrial, Rich, Democratic (WEIRD)** societies [50]. The concept of general users is imprecise, but may persist as an attempt to design for the most users possible. We use "general users" as a stand-in for the assumptions described here, including that they experience lower levels of digital-safety risk when compared to at-risk users.

Table 1. Article Organization at a Glance

Section	Section's Focus
2	Inclusive digital safety: Problem and background —Inclusive digital safety is an important, but "wicked," problem —Context setting—Authors' background
3	Understanding inclusive digital safety through at-risk research —Our at-risk research —Related research in the field
4	Applying at-risk research: Frameworks and design principles —At-Risk Framework (10 factors known to increase a user's risk) —User States Framework (4 states of user emotions and needs surrounding digital-safety events) —Design principles (20 design building blocks for inclusive digital safety)
5	Communicating with practitioners about inclusive digital safety
6	Discussion: Charting progress and future work —Progress so far and known gaps —Recommendations and ideas about future work
A–D	Appendix A List of publications from our At-Risk Research Program B Examples of supporting inclusive digital safety in practice C Details on approaches to inclusive digital safety: user archetypes, stress cases, and design principles D Terminology: defining key terms used in this article

that limit or defer their ability to attend to digital-safety concerns. While all of these users may benefit from ways to quickly deploy digital-safety solutions for their situations, at-risk groups demonstrate critical needs. For these reasons, *designing for at-risk users can help improve digital safety for everyone.*

However, the complexities of designing for at-risk users' varied needs in practice—plus the dynamic nature of technology, risks, and people's contexts—help make inclusive digital safety a *wicked problem.* A wicked problem is a term introduced in social planning to describe complex societal problems for which issues, solutions, endpoints, and who is responsible are not agreed upon by experts or those involved [84]. The complexities of inclusive digital safety mean that there is often not a "correct" solution, so it is difficult or impossible to create best practices. However, there is still a need to translate what we, as a field, *do* know into guidance to help practitioners better support users. As Rittel and Webber articulate, professionals can orient toward solutions to wicked problems that are "better" or "good" [84].

In this article, we draw from more than 9 years of working within Google to understand and support the digital safety of at-risk users in technology intended for broad use. Since this is a wicked problem, we focus on how we have made progress through both research and practical guidance. Our primary contribution is a summary of lessons learned from implementing a research and education program that has included translating foundational at-risk research into practice to try to move toward inclusive digital safety in a large company, over nearly a decade (and counting).

In service of this contribution, we structure this article as shown in Table 1. In Section 2, we define inclusive digital safety as designing safer technology for everyone, including at-risk users. We discuss why this is a *wicked problem*, outlining key challenges in defining the problem (e.g., the context of technology, risks, and people are dynamic) and developing solutions (e.g., the lack of "correct" solutions that reconcile *all* user needs and constraints). We also share the authors' background to set context for the article. In Section 3, we highlight examples of insights relevant to inclusive digital safety from our own research and then synthesize related research from the broader community. In Section 4, we share frameworks and design principles we have developed

to help technology creators design safer technology informed by learnings from foundational at-risk research. In Section 5, we outline the types of actionable deliverables we have used to share our research, frameworks, and design principles with tech workers, and ways we have framed the content for different practitioner roles. Finally in Section 6, we discuss progress and gaps in research, arguing that the field is not ready to create "best" practices for inclusive digital safety. We recommend the field manage the wicked complexity of this problem with "good" practices and expert guidance toward safer solutions.

We believe that this article will be valuable to academics who would like their research to impact technology and policy intended for broad use; technology creators who wish to implement the principles herein; regulators who wish to develop policy in this complex space; and industry researchers who wish to implement similar programs to elevate inclusive digital safety. A cross-cutting theme in presenting the multiple parts of our long-running effort at Google is that it takes substantial investment—beyond publishing and presenting research—to meaningfully take even incremental steps to improve digital safety in broadly used technology, due to the nature of this wicked problem.

2 Inclusive Digital Safety: Problem and Background

Designing for inclusive digital safety means designing technology to be safer for *everyone*, which requires understanding a diverse range of at-risk and general user needs. We discuss why this is a *wicked problem*, outlining key challenges in defining the problem and developing solutions. We then set context by outlining the background we (the authors) bring to the guidance offered in this article. Note that we define terms used throughout this article in Appendix D.

2.1 Inclusive Digital Safety as an Important, but "Wicked," Problem

In their seminal work, Rittel and Webber [84] outline 10 characteristics of wicked problems, including that they have no definitive formulation, no "correct" solutions, no definitive test to assess how good solutions are, and more. Here, we articulate key aspects of what makes inclusive digital safety a wicked problem, aiming to align research and practitioner communities toward how to think about and discuss the problem responsibly and effectively. In particular, we emphasize that research findings, recommendations, and proposed practices should be qualified based on the wickedness of the problem, such as the need to account for a range of known and unknown contexts and needs beyond a single study.

Technology "solutions" can be used for good *and bad.* In their discussion of wicked problems, Rittel and Webber [84] pointed to optimism within their field (i.e., the social planning profession) as sustaining the belief that goals like planning for the betterment of society or the "perfectability" of their approaches were possible—goals they called into question. We draw attention to a related optimism in the technology field broadly, embedded in foundational approaches like design thinking [13], that technology will help solve people's problems, improve their experiences, and generally be deployed for good. However, technology can be (mis)used to cause harm, and understanding this sometimes requires a mindset shift for technology creators (a shift our frameworks and educational materials, described below, are intended to facilitate). Freed et al. [38] describe "UI-bound adversaries" that misuse functionality offered by a system to enact harm—for example, apps that allow parents to monitor and keep their children safer have been misused by abusers to stalk their intimate partners. Adversarial thinking and threat modeling that accounts for diverse digital-safety risks and needs can help teams explore how a feature intended for good might be (mis)used to cause harm [11, 38]. Even with threat modeling, tech misuse adds to the wicked complexity of inclusive digital safety.

Technology, digital-safety risks, and people's contexts are dynamic, so the problem defies definition. Technology is continually changing, as is the landscape for digital safety—this makes the problem hard or impossible to define and definitively solve, another attribute that makes inclusive digital safety a wicked problem [84]. Attackers constantly innovate ways to (mis)use technology to cause harm. Technology itself is an intervention that can impact the problem context over an extended time. *Who* is at-risk can change—a user's life circumstances can introduce risk factors at any time, temporarily or in an ongoing way. For example, a once happy relationship could become strained, making it less trustworthy or even abusive; or a user might post something online that goes viral, immediately growing their prominence and opening them up to harassment; or a person could lose their job and suddenly face serious economic hardship; and so on. In other words, *who is "at-risk" is dynamic and could potentially—over time—include anyone.* Such users may not realize that their digital-safety risk has changed and thus may not enact important protective measures until they realize they are under attack. This dynamic landscape suggests that the many, constantly changing issues surrounding inclusive digital safety are not completely knowable.

Tech solutions require judgments reconciling important and diverse needs and constraints. A major feature of wicked problems is that they do not have "correct" solutions (rather, solutions are "better or worse"), and there is no finite set of potential solutions, requiring practitioners to use their judgment in reconciling needs and constraints [84]. Rittel and Webber further discuss how context within society—composed of many sub-groups with differing values and needs—make it hard to weigh potential solutions to wicked problems [84]. Likewise, technology that serves billions of users worldwide must support an enormous diversity of users along multiple dimensions, including digital safety, but also utility, ease of use, cultural sensitivity, regulatory compliance, and more. Team and time resources are finite. These many needs and constraints must be reconciled. For example:

— *Different groups have different needs.* The needs of various user groups may differ or even sometimes conflict, and there are seemingly endless intersectional considerations, furthering the challenge of developing inclusive digital-safety solutions. As an example, modern browsers like Chrome offer a variety of ways users with different digital-safety needs can manage their browsing history [43], which supports a range of at-risk groups. Some survivors of IPA and women in South Asia may need to carefully curate their browsing history if they might be monitored by someone who could retaliate if they knew or even suspected that the browsing history had been deleted [60, 86]. The ability to delete individual browsing history entries could help them avoid an attacker or family member detecting changes. In contrast, an activist might need to quickly wipe their entire history if detained. These users could benefit from the ability to clear all of their browsing data at once. These examples hint at the complexity of designing safer technology for numerous groups and sub-groups of users, who experience a variety of risks that may intersect in unique ways. Here, focusing on the needs of only one group would not address the needs of the other group. Further, there are competing tradeoffs: adding options to an interface increases its complexity, putting the cognitive burden (of selecting among options) on all users. **User Experience (UX)** professionals are often taught that simpler designs, with fewer options, are better since they are usually easier to use, less overwhelming, and direct users toward commonly desired choices [42, 53]. Technology creators must reconcile these kinds of tradeoffs to make decisions about inclusive digital safety.

— *Stopping abuse on a platform vs. allowing it to continue.* Even commonly promoted and sometimes effective solutions can require reconsideration given the needs of certain at-risk groups. For example, one potential solution—which has been effective in many circumstances—is to disallow or otherwise stop the abuse on a platform (e.g., removing users from a platform,

censoring content). However, prior work describes some cases when stopping abuse on a platform may lead to unintended harm—such as if the target would benefit from collecting evidence of the abuse for legal proceedings, or if they want to keep track of the attacker's emotional state to assess the potential for escalation, or if shutting off the abuse on a platform might provoke the attacker to retaliate or otherwise escalate [35]. Also, ways of stopping abuse (like reporting or removing content) might be abused by attackers. In many cases, technology creators need to evaluate these tradeoffs to determine when stopping abuse might not be effective for users, when it is, and when/how to give users effective choices.

— *Direct participation vs. limiting retraumatization in UX research.* The act of evaluating technology also involves tradeoffs. UX professionals are often taught to directly include representative users as participants in research related to design efforts [13, 53]. However, at-risk users are often survivors of trauma, and it is also important to limit retraumatizing them as part of research and design efforts [11, 16]. Trauma-informed approaches [16] can help UX professionals plan their research and design sensitively. It can also help for UX professionals to weigh the tradeoffs of engaging at-risk participants versus alternative approaches [11]. Even so, UX professionals should not rely on direct engagement with at-risk participants in exploring tactical questions pertaining to inclusive digital safety.

Wicked does not mean impossible; supporting users despite incomplete knowledge. Rittel and Webber state that "[e]very wicked problem is essentially unique," and directly applying prior solutions in new contexts could be problematic [84]. However, we can begin compiling "good" practices to benefit from what we and others have learned. This is a reason why it is valuable for us and others in the field to both understand the wickedness of the problem and share what we have learned toward designing for inclusive digital safety. In the following sections, we describe research insights and then operationalize our experiential knowledge in frameworks, high-level design principles, and examples in products, providing initial guidance to technology creators, even though what is "perfect" or "best" may not yet be achievable.

2.2 Authors' Backgrounds and Experience

The authors of this article all currently, or previously, worked at Google. Since 2015, we have been conducting research and creating guidance, both at Google and with external collaborators, to help improve digital safety for at-risk users.

When we began this effort in 2015, many of this article's authors were consulting with product teams about digital-safety-related UX. We were increasingly fielding questions about how a particular technology or feature could affect users who might not be considered "average" (e.g., political dissidents and others whose accounts were being targeted by nation-states). The vast majority of digital-safety research available for us to reference in the academic literature at that time was focused on "general" users, with some notable exceptions (e.g., [2, 7, 62]). We saw a need to build upon this work to better understand the digital-safety experiences of users whose needs might not be adequately captured by research focused on "general" users.

To address this need, we founded the *At-Risk Research Program* within Google.[3] This program aims to better understand the digital-safety concerns, needs, and experiences of at-risk users to help inform the design of technology and policy to improve their digital safety, and by extension, the digital safety of *all* users. Our team includes the authors of this article (while they worked at

[3]Google offers a wide range of products intended for broad use, including e-mail, messaging, productivity tools, operating systems, mapping, video, photo sharing, payments, and more.

Google), a cross-functional digital-safety-focused expert group in roles spanning research; UX; product management; privacy analysis and engineering; and privacy, safety, and security leadership.

We have conducted research involving numerous at-risk groups [24, 36, 37, 47, 60, 85–87, 92, 96] and engaged internal and external expert communities to raise awareness of our work and infuse it with diverse perspectives. As of January 2025, we have 18 peer-reviewed published or to appear on digital-safety research involving at-risk users, 13 of which were co-authored with academics and/or non-profit staff from many different organizations (see Appendix A, Table A1, for a list of publications from our At-Risk Research Program[4]). Outside of publication venues, we have presented our research in more than 20 external talks to raise awareness (e.g., [18–22]). We have also engaged with diverse experts across Google, including those specializing in particular products, security engineering, privacy engineering, anonymization, research, the law, policy, public relations, and ethics.

To make this research more useful within Google, members of our team have also worked to translate foundational research into actionable insights that employees of Google ("Googlers") can use to tactically improve the digital safety of technology intended for broad use. For example, we have consulted directly with numerous teams on specific product features, policies, and research studies. As another example, we developed a class to help Googlers learn about at-risk users and how they can support at-risk users' digital safety in their work, which has been taught over 80 times, with more than 1,800 employees enrolled worldwide since March 2018.

In this article, we draw upon these experiences to share what we have learned about supporting the digital safety of at-risk users, and we propose ways the research community might impact technology intended for broad use with their foundational research involving at-risk users.

3 Understanding Inclusive Digital Safety through At-Risk Research

Here, we highlight insights pertaining to inclusive digital safety drawn from examples of our research involving at-risk groups. Then, we cover related work on at-risk users and inclusive digital safety from the broader research community.

3.1 Our Research Involving At-Risk Users: Insights about Inclusive Digital Safety

We have conducted research with many at-risk groups, revealing insights about inclusive digital safety—that is, ways an optimistic view of "general" users do not adequately account for at-risk users—summarized in Table 2 and described in this section. The example studies highlighted here primarily used semi-structured interviews; methodology details, research artifacts, and ethical considerations are provided in the papers cited in Appendix A (Table A1).

Survivors of IPA [60]. In 2015, collaborating with **Community Overcoming Relationship Abuse (CORA)**, we conducted research involving survivors of IPA to understand their digital-safety motivations, practices, and challenges [60]. We found that survivors' digital-safety needs were different across three phases of abuse—when (a) under the *physical control* of their attacker; (b) taking steps to *escape* their attacker; and (c) in a *life apart* from their attacker, with ongoing privacy and security work to maintain safety and personal autonomy. In the physical control phase, many survivors could not safely prevent their attacker from physically accessing their devices or accounts. This insight has many implications for the role that "personal" devices play in a user's digital safety, such as: device locks cannot be depended on to keep all untrustworthy

[4]The program we describe in this article does not encompass all of Google's research involving at-risk users. Other efforts exist to support synergistic but different organizational goals, such as research from Google Research with people who are transgender and non-binary [103], from Google Trust and Safety on image-based sexual abuse [97], and from Google Jigsaw on online hate speech [58].

Table 2. Example Insights about Inclusive Digital Safety Derived from Several At-Risk Groups We Have Studied, and the Optimistic Viewpoints of General Users They Help Expand

At-Risk Group Studied	Optimistic View of General Users	Inclusive Digital-Safety Insights
Survivors of IPA [60]	Devices are used by an individual who completely controls their use	Some people cannot safely stop others from accessing their devices or accounts
	People in one's household are largely trusted	People in one's household may not be trustworthy
People living in transitional homeless shelters [92]	People have reliable access to technology (e.g., devices, Wi-Fi)	Some people have higher risk of losing their phone or service, and being locked out of accounts
	People will regularly have times of lower stress in their life (e.g., to maintain technology settings)	Some people live with chronic stress that makes it harder to use technology
Women in South Asia [85, 86]	People can safely share photos of themselves online	A photo a person shares of themselves can be misused for harm
	Most people will value privacy features and settings	Some people prioritize openness and may view "privacy" as conflicting with their values
People involved with political campaigns [24]	Organizations will have the time and expertise to set up robust security precautions	Campaigns are transient, have tight timelines and budget, and often limited or no IT staff
Content creators [87]	People will enact digital-safety protections before they share content online	Creators face a range of threats, but many only adopt protections after experiencing a serious digital-safety event

others out of a user's accounts; using a device as the key to accessing private information and account settings can leave some users vulnerable to attack; and more. It also emphasizes the home environment as potentially involving people who have malicious intent toward the user, which helps technology creators consider abuse scenarios for an array of computing devices and accounts used in the home. We also found that if the survivor escaped, the attacker could leverage intimate knowledge in remote attacks, to gain access to accounts or harm the survivor in other ways. In their life apart, survivors took great care to protect information that an attacker could use to find and control them—like their contact information, coarse location, or even their employer. These findings revealed how survivors in different circumstances—under physical control vs. living apart from their attacker—had different digital-safety needs. For example, implementing **two-factor authentication (2FA)** on their accounts may not keep their abuser from accessing their accounts while a survivor was under the attacker's physical control, but may be a top recommended security action for those living apart from their attacker.

People living in transitional homeless shelters [92]. Also in 2015 with help from a non-profit organization, we explored the digital-safety experiences of people living in transitional homeless shelters, finding four types of challenging circumstances associated with financial insecurity that impacted their digital safety—(a) limited financial resources, (b) limited access to reliable devices and the Internet, (c) the need to cope with untrusted relationships, and (d) chronic stress [92]. For example, this group risked losing access to their phone or phone service if they could not pay their monthly bill or their device was stolen while in an insecure housing situation, which could lead to being locked out of their accounts if they had enabled 2FA tied to their device or phone service. Providing back-up 2FA options (like Google's 2-Step Verification backup codes), easy-but-secure account recovery, and ways to educate and raise awareness of such options can

help these at-risk users, as well as general users. As another example, chronic stress experienced by this group highlighted the importance of ease of use, such as directed flows and easy reading levels.

Women in South Asia [85, 86]. Beginning in 2017, in collaboration with researchers at several universities around the world, we conducted a qualitative study of (a) the online abuse experiences and coping practices [85] and (b) perceptions and behaviors related to personal privacy on shared phones [86], with 199 people who identified as women in India, Pakistan, and Bangladesh. The results showed widespread experiences with online abuse among these South Asian women, and perspectives toward privacy that differed culturally with Western users—for example, valuing openness and viewing privacy as against their values. These results led to recommendations for more culturally inclusive identity models and abuse reporting policies that account for different sensitivities toward data. For example, content that would likely be seen as harmless in many Western contexts—like a photo of a fully clothed woman—could be used to harass or extort women in South Asia. Our results also suggested offering features to support what we might consider to be privacy practices in ways that are not presented as "privacy" (helping to inform Safe Folder [98]).

People involved with political campaigns [23, 24]. Digital safety for people involved in politics may be influenced by very different risk factors than the other groups we had studied up to 2019 [60, 85, 86, 92]. They are targeted by many types of attackers, including sophisticated nation-states [33], hacktivists, and cybercriminals, and are often prominently known. We sought to understand digital security experiences on political campaigns. Thus, we conducted research with people from across the political spectrum who were still or had recently been involved with U.S. political campaigns. We found that the unique threats, constraints, and work culture of this group differed from general users and other at-risk groups we had studied, and left them vulnerable to technology-facilitated attacks. In particular, a campaign's transience, tight timelines and budget, and amorphous boundaries (e.g., with consultants, vendors, volunteers) led to sensitive data being kept in a plethora of personal and work accounts, with ad hoc and limited adoption of strong passwords, 2FA, encryption, and access controls. Participants were concerned about security threats from well-funded, sophisticated attackers, especially nation-states. These findings highlighted how important it is for strong security protections—especially strong 2FA, passwords, and multi-tenant accounts—to be easy-to-use across provider accounts by professionals with limited technology expertise, time, and money. While such security improvements could help many professional users, particularly those with limited or even no IT support, the need for political campaigns was stark—an example of a "stress case" that we discuss later. See Appendix B for an example of how some of these needs have been addressed by Google Workspace [44].

Content creators [87, 96]. Content creators share their content on platforms such as Instagram, TikTok, Twitch, and YouTube. Their prominence among an online audience, combined with the social norms of accessibility and authenticity to that audience, contributes to the unique risks they face. We explored their digital-safety experiences in two studies [87, 96], finding that attacks were nearly universal among the creators studied, and they experienced a host of threats—toxic content, stalking, doxing, impersonation, account security attacks, and more—but most only adopted protections *after* experiencing a serious digital-safety event. Also, toxic content was something they dealt with regularly and often at-scale, using moderation tools as a key frontline defense. These findings strongly indicated a need for education to raise awareness among new creators of the threats they may face and the protections they should employ (leading to product impact discussed in Appendix B on YouTube Creator Safety Center [106]). This was especially important since many creators started with small audiences and shared personal information while they perceived their risk to be relatively low, but were then unable to "take back" that information as

their prominence grew. This finding helps demonstrate how any user can become at-risk as their circumstances change.

3.2 Related Work: Research on Inclusive Digital Safety More Broadly

Inclusion in technology design is a growing area of research in the broader academic community [17], including in privacy and security [83, 100]. Here, we outline related work in the **Human-Computer Interaction (HCI)**, Security, and Privacy literature on at-risk groups, syntheses across at-risk groups, and frameworks to support inclusion in technology research and design.

3.2.1 Research Focusing on Specific At-Risk Groups. Since we founded the At-Risk Research Program in 2015, the number of papers published in this space has dramatically increased (see the review in [11], which included 45 papers from 2022, 34 from 2021, and fewer in earlier years). The bulk of these papers seek to understand the digital-safety experiences of specific at-risk groups. For example, studies have been conducted to understand the digital-safety experiences of survivors of IPA [15, 38, 39, 48, 60], journalists [29, 63–65], people with disabilities [2–4, 49, 59, 81], refugees [91], children [40, 41, 56, 66, 67, 70, 108], sex workers [10, 94], and more. Our review of at-risk research cited 31 at-risk groups that had been the focus of recent studies [101]. These studies collectively have helped us examine a diverse set of risks, practices, and needs, providing a foundation for making sense of the complexities of inclusive digital safety. They may also directly inform specialized technologies (rather than the technology intended for broad use we have mentioned throughout this article) aimed to support specific at-risk groups, for example, [1, 29, 99].

Studies of groups facing intersectional risks can allow for a rich examination of how multiple intersecting identities, and risk factors can uniquely impact an at-risk user's digital-safety experiences. In recent years, researchers have advocated for attention to intersectionality in HCI and computing, for example, [31, 34, 55, 82, 89, 105]. Intersectionality [26] considers how multiple marginalized identities or circumstances can combine to create unique modes of discrimination. This concept is important for inclusive digital safety because it highlights the importance of capturing complex, realistic representations of users' identities and relating them to their surrounding context [89]. People have unique experiences influenced by their intersecting identities, which may be impacted by many axes of privilege, domination, and oppression (e.g., racism, educationalism, ageism, colorism, pro-natalism, and more) [31]. Attentiveness to intersectionality can also help researchers examine why some individuals may be more susceptible to harm via technology-facilitated abuse [61]. Relatedly, we have found that when a user's risk factors intersect, they yield new digital-safety risks or amplify existing risks [101]. Research with at-risk users that attends to intersectionality could help extend our understanding of how users might experience risks. Collectively, this literature emphasizes the continued need for foundational research focused on specific at-risk groups, especially those with intersecting risk factors.

3.2.2 Supporting At-Risk Groups with Broadly Used Technology. Individual companies have shared some ways they support specific at-risk groups in broadly used technology. Microsoft offers Microsoft 365 for Campaigns [71] to help political campaigns and national party committees more easily set up and afford the advanced security capabilities of Microsoft 365 Business. Facebook [73], Twitter (now X) [72], and Uber [74] worked with the National Network to End Domestic Violence to publish guides specifically for survivors of IPA on how to safely use their products. In general, these offerings aim to help at-risk groups by raising awareness or improving the ease-of-use and accessibility of safety features that support all users. This follows a principle that echoes a key message in this article—supporting the digital safety of at-risk users helps to make technology safer for everyone.

Such decisions that impact safety in broadly used technology must be carefully considered with a wide range of constraints and stakeholders in mind. While it is relatively common for research papers reporting single-group studies to make design recommendations to support those users, they often do not report an assessment of the impact on other at-risk groups or general users. In general, there is a gap between the knowledge derived from research about a specific group and improving inclusive digital safety in broadly used technology.

3.2.3 Syntheses and Frameworks across At-Risk Groups. While it is important for inclusive digital safety to understand a wide variety of users' digital-safety experiences, it can be difficult (and even impractical) to apply the findings from such a large number of studies in practice. To help, a smaller subset of emerging literature synthesizes existing knowledge to provide insight, frameworks, and taxonomies pertaining to digital safety *across* at-risk groups. These resources can help make research usable by practitioners, adding a basic structure to the ideas found across (potentially many) studies, and better aligning with the wicked problem of supporting inclusive digital safety.

Scheuerman et al. [88] classified harm from online content into four types—physical, emotional, relational, and financial harm—and provided a framework for evaluating the severity of these harms using eight dimensions—perspectives, intent, agency, experience, scale, urgency, vulnerability, and sphere. At-risk users can be identified by their elevated chances of experiencing these harms—at all and potentially more severely—when compared to general users. The framework by Scheuerman et al. provides a way to consistently assess the kinds of harms that are involved and their severity when examining circumstances involving at-risk users. It is complementary to the frameworks we have developed (described later in this article) and one we have used to discuss harms. Levy and Schneier [57] focused on a class of privacy problems they refer to as *intimate threats*. Although Levy and Schneier [57] describe intimate threats as applying to anyone, these threats often apply to at-risk users—as we found when developing our At-Risk Framework (Section 4.2), which includes three relationship risk factors (relationship with the attacker, reliance on a third party, and access to other at-risk users) [101]. In our own prior work (in collaboration with several academics), we have synthesized the literature to define a taxonomy of hate and harassment attacks [95] and recommend strategies for safer research involving at-risk users [11]. Also while working with academic collaborators, we have studied how people with different backgrounds (e.g., people who identify as LGBTQ+, young adults) reacted to toxic content [54].

In this article, we expand this literature by discussing two additional frameworks which have been pivotal to our efforts within Google regarding inclusive digital safety—(1) The At-Risk Framework (Section 4.3) simplifies how to consider many digital-safety risks to at-risk users (previously published in [101]), and (2) The User States Framework (Section 4.3) facilitates empathizing with and supporting users' needs—with an emphasis on considering the emotions they are likely to be experiencing—regarding digital-safety events.

3.2.4 Approaches for Inclusion in HCI and Frameworks in Other Fields. Several orientations and approaches have been proposed for the creation of more equitable and inclusive technology. Aligned with this work, Wang [100] called for *inclusive security and privacy*, highlighting the need for guidance and community building among researchers and practitioners working in this space. Renaud and Coles-Kemp [83] built on the idea of *inclusive security*, analyzing how barriers to accessibility can contribute to security vulnerabilities, especially for people with a variety of disabilities. Both works acknowledge the complexity of designing for a wide variety of security needs and call for future work.

Also related, trauma-informed computing [16] raises awareness for widespread experiences of trauma among users and proposes principles for how to approach design to mitigate tech-facilitated harm to these users. Social justice-oriented design [28] proposes a set of sensitivities

and commitments technology creators can adopt to promote more equitable technology, especially when they impact complex political issues. Bardzell [9] discusses a set of feminist interaction design qualities, which align with central tenets of our situated work. For example, "pluralism" involves "encouraging a constructive engagement with diversity, and embracing the margins both to be more inclusive and to benefit from the marginal as resources for design solutions" [9]. Likewise, we have adopted an emphasis on understanding the diverse digital-safety needs of at-risk users and have found in practice that supporting them can mitigate key threats to at-risk users while benefiting all users. Aligning with other feminist design qualities, our work has also incorporated the participation of a wide range of at-risk groups, experts, and related stakeholders; emphasized transparency and clarity in our design principles; sought to balance diverse needs while advocating for the digital safety of at-risk users; and incorporated human emotion into our frameworks and resources.

The four states in our User States Framework (Section 4.3) are related to frameworks in disaster (or emergency) management [5], a well-established field with decades-old literature describing response to emergencies. The purpose of disaster management is to "[protect] communities by coordinating and integrating all activities necessary to build, sustain, and improve the capability to mitigate against, prepare for, respond to, and recover from threatened or actual natural disasters, acts of terrorism, or other man-made disasters" [32]. Relevant disasters include an earthquake, hurricane, flood, wildfire, pandemic, aircraft crash, bridge collapse, terrorist attack, or other emergency situations that impact society. A commonly referenced disaster management framework by Alexander [5] defines the following phases of a disaster: *mitigation* ("all actions designed to reduce the impact of future disasters"), *preparation* ("actions taken to reduce the impact of disasters when they are forecast or imminent"), *response* ("emergency actions taken during both the impact of a disaster and the short-term aftermath"), and *recovery* ("the process of repairing damage, restoring services, and reconstructing facilities after disaster has struck"). An expectation in the field is that stakeholders would follow an emergency plan, starting at the beginning with mitigations, then through the remaining phases as needed. The audience for disaster management is broad, including governments (at local, regional, and national levels), international bodies, private sector organizations and facilities, public infrastructure facilities, heritage sites, and more [6]. Other fields—like crisis communication [25] and cybersecurity crisis management [27]—have adapted and applied this disaster management framework and accompanying ideas within their domains. Although informed by this prior work, our User States Framework (Section 4.3) differs, in part, because it was developed for a different subject (at-risk users), purpose (improving digital safety for at-risk users), audience (tech workers and researchers), domain (HCI, Security, and Privacy), and is not intended to be linear.

4 Applying At-Risk Research: Frameworks and Design Principles

Research that focuses on understanding the digital-safety needs of different at-risk groups, and how those needs relate to the needs of general users, can provide a powerful foundation toward designing for inclusive digital safety. In this section, we share two frameworks and a set of design principles we developed to help tech workers apply such research and approach this wicked problem in practice, beginning with how we created them.

4.1 Method

4.1.1 At-Risk Framework. Drawing on an emerging body of peer-reviewed research involving at-risk groups, we wanted to explore the similarities and differences pertaining to the groups' digital-safety experiences. Working with researchers from the University of Maryland, we systematically identified and reviewed 95 papers from the HCI, security, and privacy research communities that

were focused on the digital-safety experiences of different at-risk groups. We inductively developed a codebook to analyze what factors were known to contribute to a user's digital-safety risk, the practices they employed to protect themselves, and the barriers that got in their way of enacting effective protections. We then coded the 95 papers in our dataset and analyzed the data for themes to create the At-Risk Framework.[5] Our paper [101] provides a more detailed description of the method we employed. We later applied findings from this foundational research effort to help practitioners create safer technology by simplifying how they might consider the many known risks users experience, at scale, in their work.

4.1.2 User States Framework and Design Principles for Inclusive Safety. Following our early studies with at-risk groups, we wanted to create actionable resources to help practitioners empathize with at-risk users and apply what research had shown about the digital-safety experiences and needs of at-risk users in product design. We drew on principles from design thinking [13], which can help designers cope with wicked problems [14]—this includes practicing empathy for users, paying close attention to "extreme" (in our case, at-risk) users, involving multiple disciplines from the start, accounting for business constraints, communicating with tech workers about the ideas, and iterating [13]. With this approach, our focus was not on academic rigor, which is a limitation. We instead emphasize that we wanted to make meaningful, positive progress on a wicked problem, do not consider this framework or design principles to represent "best" practices, and continue to iterate as we learn more. The value of the resulting framework and design principles has been demonstrated through their use in practice over many years to support at-risk users' digital safety (described more in Section 5 and through examples in Appendix B).

Data: Sources of knowledge about users and practical needs. To identify cross-group themes related to digital safety, we started with the results from our early studies of several at-risk groups—survivors of IPA [60], people living in transitional homeless shelters [92], homeless youth [104], content creators, and NGO staff supporting media activists. We also drew from known threats (e.g., hate and harassment in online comments) that members of our team or the experts we consulted with had worked to understand and design features around. Finally, we reviewed UX flows (i.e., the path through the user interfaces for a task) that users could deploy to cope with digital-safety threats (e.g., changing privacy settings, going through account recovery). We have iterated on the framework and principles over time, drawing on findings from all of the studies listed in Appendix A (Table A1).

Process: Developing, applying, and iterating on the framework and design principles. Core to our practical approach was (1) collaboration among experts and stakeholders across disciplines and (2) iterative improvement over time. The User States Framework and Design Principles for Inclusive Safety were developed by a team of experts (see Section 2.2), and updated with feedback from an extended group of stakeholders—including UX Researchers, Designers, and Writers; Product Managers; Software Engineers; Privacy Engineers; Lawyers; and more. Collectively, this group knew not only about user needs but also what product teams need and constraints such teams must consider.

We began using the earliest of the sources described above in 2015 with six researchers and designers, from across the company, discussing and developing a description of at-risk users, why it is important to design with them in mind, examples of their experiences and challenges, and initial design considerations. Structured by these ideas, two of these team members met regularly to examine and discuss common themes in the sources described above. These common themes were

[5]While systematically created, our dataset was not exhaustive of all relevant papers, may not report all challenges each at-risk group actually experiences, and (reflecting the sampled literature) skewed heavily toward Western—and predominantly U.S.—populations.

organized to represent users' experiences and needs surrounding digital-safety events, focusing on what would be important to UX Designers (i.e., our initial target audience). Other team members were regularly consulted during this time to ensure agreement with this analysis of common themes. Feedback was also incorporated from multiple tech workers who were experts in UX, digital safety, and/or at-risk users (beyond the core team), to improve the relevance of the analysis to tech workers. From these discussions and the documentation of common themes, we developed the *user states* (detailed below) of Prevention, Monitoring, Active Event (originally Crisis [52]), and Recovery.

In discussions, we "tested" the User States Framework against knowledge of specific groups from research results—for example, we walked through how a content creator might experience a digital-safety threat and then evaluated the framework definitions and structure to refine it. Once we had a solid framework that reasonably represented the at-risk groups from our research, we focused on developing relevant design principles and improving the effectiveness of how the user states were presented to product-focused teams. We first published the user states, design principles, and examples from products that demonstrated each design principle within Google in 2017.

The framework and design principles were then vetted and refined iteratively over time, based on new research (both ours and from the literature), use in practice, and use in the research process. For example, in a recent study we conducted with researchers from the University of Washington, we used the User States Framework in a qualitative analysis of image-based sexual abuse help-seeking requests on Reddit, identifying themes related to when and in what emotional state posters sought help [102]; we then updated the framework with what we learned during that analysis. We have also stress tested the framework and design principles over time in digital-safety consultations with product and policy teams. For example, see Appendix B for how specific products—who worked with us about how at-risk research could inform features in their products—exemplify certain design principles.

4.2 At-Risk Framework: What Puts Users at Risk?

The At-Risk Framework outlines 10 factors that are known to heighten risk for users (across groups) [101].[6] These risk factors are temporary or ongoing circumstances that increase a user's chances of being targeted by, or experiencing disproportionate harm from, a tech-facilitated attack. The At-Risk Framework should help scale how technology creators consider and mitigate users' digital-safety risks in technology intended for broad use; we posit that it is much easier for technology creators to consider 10 common risk factors than it is to think about the risks experienced by many at-risk groups. The 10 factors resulting from our analysis that can elevate a user's risk are as follows:

- *Legal or Political Factors.* Heightened risk of government, political, or legal scrutiny or regulation, potentially including sophisticated attacks from nation-states. *For example, activists, journalists, political campaign workers, political dissidents, refugees.*
- *Marginalization.* Pervasive negative treatment or exclusion due to an individual's identity attributes or life experiences, from a broad potential set of attackers. *For example, LGBTQ+ people, marginalized racial groups, people with disabilities.*
- *Social Norms.* Informal rules that govern behavior in society. Some norms restrict options for an at-risk user and differ from the expectations of technology creators—especially mismatches

[6]These 10 risk factors are not exhaustive and may be expanded in future research. The At-Risk Framework also includes practices at-risk users are known to take in an attempt to protect themselves and the barriers that get in their way of enacting effective safety protections (see Warford et al. [101] for more on those elements of the framework).

regarding an individual's agency, device ownership, or privacy expectations. *For example, women in repressive regions, children and teens, refugees.*

— *Relationship with the Attacker.* A personal relationship with an untrustworthy person may put the user at risk of attacks that take advantage of personal knowledge of the user, physical access to the user or their devices, or relational power dynamics. *For example, survivors of IPA, people with low socio-economic status, survivors of trafficking.*

— *Reliance on a Third Party.* By providing needed or digital-safety-focused help or care, a third party can (sometimes inadvertently) contribute to risk from the third party, through the third party, or an attacker impersonating the third party. *For example, children, older adults, people with disabilities.*

— *Access to Other At-Risk Users.* Having access to an at-risk user can make the individual with that access become at-risk themselves, as it increases the risk of stepping-stone attacks aimed to access or harm the other at-risk user. *For example, teachers, journalists, NGO staff, family members of political candidates.*

— *Prominence.* Standing out in a population, because they are well known publicly or have noticeable attributes (e.g., outspokenness, attractiveness), can expose an individual to a variety of attacks and amplify other risks. *For example, activists, political campaign workers, content creators, journalists.*

— *Resource or Time Constraints.* Atypical constraints related to either access to resources (e.g., money, devices, connectivity, accessible technology) or time to cope with threats can elevate risk and/or decrease one's ability to respond to risk. *For example, people with low socio-economic status, political campaign workers, emergency room staff.*

— *Underserved Accessibility Needs.* An accessibility need (e.g., due to a disability, neurodiversity, language barrier, or developmental maturity) unmet by technology can contribute to anxiety of and susceptibility to attacks. *For example, children and teens, people with disabilities, undocumented immigrants.*

— *Access to a Sensitive Resource.* Having access to a privileged resource (e.g., financial accounts, proprietary knowledge), often as part of a profession, increases the risk of attacks aimed to co-opt access. *For example, journalists, emergency room staff, political campaign workers, real estate agents, people with high socio-economic status.*

These risk factors have important implications for how users experience risk and the support they need mitigating those risks in technology, which we build upon in the following sections.

4.3 Users States Framework: How Do Users Experience Digital-Safety Risks and Events?

The User States Framework is intended to help technology creators design safer products and policies by accounting for users' emotional reactions and needs due to (the risk of) digital-safety events. We have been using and improving this framework for nearly 10 years to train others, analyze data, and inform technology creation within Google.[7] The framework's four states—*Prevention*, *Monitoring*, *Active Event*, and *Recovery*—outline the user's expected needs and emotional state of mind before, during, and after digital-safety events (e.g., account hijacking, doxing, harassment, surveillance, content leakage, and more). Figure 1 depicts the framework; each of the four states is described below.

Although we developed the User States Framework based largely on findings from at-risk research, we believe that it applies to *all users* since anyone can experience (or be concerned about) a digital-safety event. Prior work has shown that digital-safety events are stressful (or even panic-inducing) for general users and spur activities to resolve the event (similar to our Active

[7]The User States Framework has been used as an analysis tool by us and others at Google.

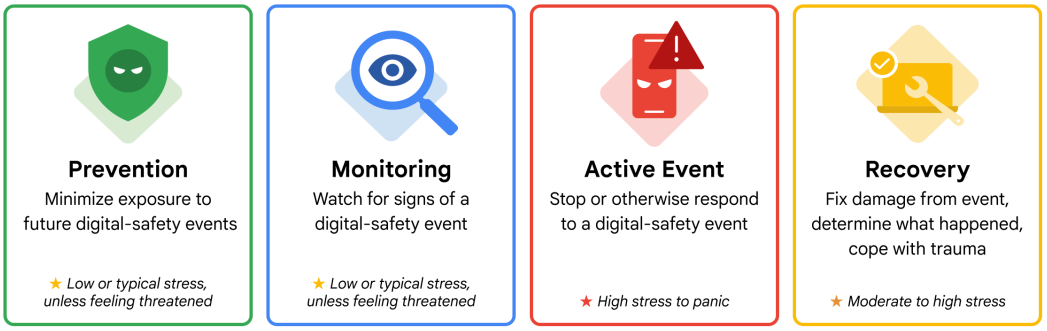

Fig. 1. User States Framework. The four user states describe at-risk users' (and likely all users') expected needs and emotions when responding to digital-safety risks and/or events. Note that the order of these states is not fixed or sequential (see Section 4.3.5), and not all users who experience or are concerned about a digital-safety event will go through all, or even most, states.

Event state) [8]. Likewise, the design principles we associate with the user states were developed to help all users, with at-risk users likely to benefit most given their elevated risk of harm.

4.3.1 Prevention. Users in the *Prevention* state aim to minimize their future exposure to digital-safety events. Users tend to engage in prevention when their stress levels are at a steady state, typical of times they are not actively coping with or recovering from a digital-safety event, unless they are feeling threatened.[8] Such users may think about prevention after receiving a recommendation to take action from a news article, a friend, or a training from their employer, or they may think about prevention after realizing that their circumstances may be changing in a way that puts them at higher risk (e.g., starting a new job or ending a relationship). Some users may engage in prevention after they or someone they know experiences an event, with increased awareness of the risks and motivation to avert further harm.

During Prevention, users aim to perform actions they hope will lessen their chance of experiencing a tech-facilitated attack. For example, a user might set up or strengthen 2FA on their accounts, accept or reject permissions requested by an app, modify digital-safety-related defaults while creating an account, change account privacy or security settings, set up a password manager to help them use strong and unique passwords across their accounts, delete content they have shared online, and more. Tech workers can support users in Prevention by making these tasks easy to find and complete, giving users options to suit diverse digital-safety concerns, and more covered in Section 4.4.

4.3.2 Monitoring. In the *Monitoring* state, users watch for signs of potential digital-safety events. Users in this state will likely feel low to typical stress levels, unless they are in the Monitoring state because they feel threatened. Monitoring can overlap with Active Event or Recovery—after a user experiences an event, a typical response is to vigilantly monitor for more to occur. For example, after explicit images portraying them were non-consensually shared, survivors commonly looked for the images to be shared in new contexts or for additional images to be shared [102]. When Monitoring overlaps with higher stress user states like Active Event or Recovery, it also becomes higher stress.

[8]There are exceptions where user stress will be higher than expected during Prevention, for example, if they are experiencing overlapping digital-safety events or are engaging in prevention to counter a (perceived) threat. Also, users may experience a "new normal" of higher stress after experiencing a digital-safety event(s). In addition, some at-risk groups experience baseline stress levels that are always relatively high due to challenging circumstances [60, 92].

Users in the Monitoring state can be supported by technology that lets them easily keep track of their online presence (e.g., via their accounts, devices, mentions on social media) and readily identify signs of an attack (e.g., alerts of changed security or privacy settings, account lockout, toxic content, image abuse, surveillance). Monitoring actions a user may take include checking their settings to ensure that only trusted devices have access to their account, viewing their account activity, checking privacy and security settings, scanning for tracking devices, reviewing email notifications for privacy and security setting changes or new sign-ins, turning on alerts, auditing previously shared social media posts and profiles, or looking for content posted by others about them online.

4.3.3 Active Event. In the *Active Event* state, users recognize that they are experiencing a tech-facilitated attack (or attacks)—such as cyberstalking, online impersonation, account hijacking, harassment, and more [95]—and often want to stop it.[9] Users in this state are probably experiencing much higher than their normal amount of stress, even to the point of panic. This stress can make it hard (or impossible) for users to think rationally or take the actions needed to alleviate the situation. Some users start the Active Event state confused and worried—suspecting something is wrong and trying to understand it—and their stress may increase as they learn more. Users in this state can be supported by helping them understand what is happening, providing practical guidance to stop the attack, making designs especially easy to use for users who are experiencing high levels of stress or panic (including helping them prioritize what to do and avoiding overwhelming them), and being considerate with terminology and visuals to avoid exacerbating stress. Tasks users might perform in this state include reaching out to people they trust (online or in person), searching for help on what to do, changing their account password, restricting access to their content, trying to get content about them that has been shared by a third-party taken down, and blocking or muting an online harasser.

While we expect most users in the Active Event state to prioritize stopping the attack, some at-risk users' circumstances may make it difficult or problematic to do so. In these cases, a user may want to know about the attack and/or monitor its progress, but not actually take action to stop it. In some cases, stopping an attack could lead to worse outcomes, as discussed in Section 2.1. A user may also be unable to stop an event, because they do not know how (e.g., they are being surveilled from an unknown device or mechanism) or there is no mechanism to do so (e.g., personal information or images leaked on the Internet or sent through messaging may be impossible to "take back" [87, 102]).

4.3.4 Recovery. In the *Recovery* state, users have stopped the event, the event ended through other means, or they have determined that they are unable to (or should not) stop the event and are ready to learn about what they *can* do. Recovering users often want to determine what went wrong, fix the damage caused, and cope with ongoing trauma. Recovery tasks may include restoring an account; creating a new identity; reporting abuse; contacting others or communities for support; learning about how to prevent this or other types of events in the future; and more.

People in Recovery may have calmed somewhat from the fear or panic of the Active Event state, but are likely still experiencing higher than normal stress and ongoing trauma. Users in the Recovery state can be supported by making relevant user interfaces clear and actionable, encouraging prevention tasks to lessen the risk of future events, and facilitating monitoring tasks if the user is worried about the potential for additional attacks, and more covered in Section 4.4. For

[9]Note that users may be subject to an active attack they are not yet aware of. For the purposes of the User States Framework, a user is not in the Active Event state until they are aware (or suspect) they are being attacked, since this framework focuses on what the user knows, feels, and needs.

technology creators, it can be beneficial to adopt trauma-informed design approaches [16] to help account for the emotional needs of users in all states, and especially when the user is experiencing or recovering from an event.

4.3.5 Diverse Paths through States. Users may take diverse paths through these states due to their contexts (e.g., threats, resource constraints, past experiences, and more). The order in which a user experiences these states is not fixed or sequential, and some users will not experience all or even most of the states. For example, a user might start with Prevention and never experience an Active Event. Or they might start with an Active Event, move to Recovery, and then decide they need to set up Prevention. This latter path was common for content creators [87], many of whom did not enact preventative measures until *after* experiencing a serious digital-safety event. Alternatively, a user might start in the Active Event state, get stuck or abandon the affected technology, and not go through any of the other states.

A user may also experience attacks that are simultaneous, in quick succession, or overlapping, making the boundaries between states hard to distinguish. For example, content creators described exposure to toxic content as a frequent and sustained attack they coped with cumulatively over time [87]. Recovering from and preventing such sustained attacks often did not involve stopping the attackers entirely, but instead slowing them (e.g., with moderation tools or blocking features) and developing a "thick skin" to emotionally cope. Such overlapping attacks blurred the states, with moderation tools being used to both stop active attacks and prevent future ones.

The timelines for each state can also vary. At-risk users are often simultaneously dealing with competing priorities, such as physical safety, figuring out how to pay for necessities like food and shelter, working with law enforcement, and so on. For example, people living in homeless shelters described that it could take them years to recover from attacks—like identity theft or online scams—due to a lack of time and monetary resources [92]. As another example, a user who experiences a severe event or multiple events (such as survivors of IPA [16, 39, 60]) may need more time to stop the event(s) and recover, take time to set up a more extensive set of prevention mechanisms, and/or be primed to put more effort into ongoing monitoring. As yet another example, survivors of recorded sexual assault sometimes described ongoing emotional trauma and extended timelines before reaching out for help to recover [102].

4.4 Design Principles for Inclusive Digital Safety

Frameworks like those described above can be helpful for understanding the structure of key ideas from research, and *design principles* can help tech workers get started applying these ideas with more concrete building blocks for design. Operationalizing the risk factors and user states presented above, we share design principles—outlined in Table 3 and expanded in Appendix C.3—intended for technology creators to use to help improve digital safety and maintain usability for all users, while providing substantial benefit to at-risk users. We have refined these principles over the years as we have used them in classes and consultations within Google. Following our discussion of inclusive digital safety as a wicked problem (Section 2.1), note that these principles do not address all safety concerns, are not intended to be prescriptive, and require considerable thought to translate into safe,[10] effective, and workable solutions.

These design principles help address select digital-safety *challenges* documented in prior at-risk research, focusing on those with implications for technology and oriented toward making progress on a wicked problem. Examples of these challenges are shown in Table 3. For instance, the challenge, *cannot reasonably stop others from accessing their devices or accounts*, is especially relevant

[10]For example, when designing solutions such as controls or warnings, it is important to consider that an attacker may have compromised a user's accounts or devices.

for users with the risk factors *social norms* and *relationship with the attacker*, among others (Table 3). Some women in South Asia faced both of these, with family members (such as a mother-in-law, husband, or brother) who tended to monitor their devices and accounts [86]. This monitoring may be intended, for example, to protect the woman or observe her behavior online. This challenge has design implications—e.g., *provide private spaces on shared devices*, which applies in the Prevention state (Table 3). A product feature that exemplifies designing for this challenge is Safe Folder [98], which provides users with a private place to store files on a shared device.

Referring to Table 3, the design principles are organized by the four *user states*—Prevention, Monitoring, Active Event, and Recovery—because they align with users' needs and emotions in each state. For example, *provide safety education* is a design principle for the Prevention state, suggesting that technology creators consider how to help users (or their allies) learn about protections they can set up while they are likely in a relatively calm state of mind. They are less likely, for example, to spend time with educational content when experiencing high stress from an Active Event, especially if the content does not clearly and directly advise on their particular circumstance in a simple way.

Also listed for each principle are the *challenges* and *risk factors* for which they should be especially beneficial. As an example of a design principle, *safety education* is expected to be helpful for all users, but especially for at-risk users who face the challenge of *sophisticated and/or motivated attackers*—like those with the risk factors *legal or political*, *prominence*, *relationship with the attacker*, or *access to sensitive resources*. Another challenge for these at-risk users is that they may *not have the knowledge needed* to counter the sophisticated, repeated, and diverse attacks associated with their risk factors. This design principle is exemplified by the YouTube Creator Safety Center [106], which offers safety education for content creators who experience the *prominence* risk factor (see Appendix B for more on this).

As another design principle example, *design for strong digital safety by default* involves considering how default settings for a technology could affect user security, privacy, and digital safety (for both general and at-risk users). Product defaults are important for anyone since users may not modify them. But they are expected to be especially beneficial for at-risk users with risk factors *legal or political*, *prominence*, *relationship with the attacker*, or *access to sensitive resources*, because they *may not have the knowledge needed to counter sophisticated or motivated attackers*. For example, Security Templates in Google Workspace [44] provide default settings of existing Workspace security features that are easily deployable by campaign workers who have the *legal or political* risk factor and need robust security protections (see Appendix B for more on this).

As a final example, the *make data sharing transparent* principle aims to make it hard for an attacker to access a user's information or activity without the user's awareness by providing enhanced transparency across multiple channels, such as via indicators when data are being shared, e-mail reminders (or other notifications) at random intervals, or notifications on multiple surfaces. Such transparency should be beneficial to all users, especially those who are more likely targeted with surveillance tactics—like those with the risk factors *relationship with the attacker*, *reliance on a third party*, and *social norms*. For example, the Location Sharing function in Google Maps offers enhanced transparency through regular e-mails informing users when location sharing is "on," sent at random times to hopefully increase the chance that a user would see the e-mail without it being intercepted by an attacker who has access to the user's account (see Appendix B for more on this).

See Appendix C.3 for a definition of each design principle in Table 3. Technology creators should consider these principles *along with* other constraints that may apply to broadly used technology, including (but not limited to) legal and regulatory requirements, how particular solutions function, usability, utility, and more. For example, a technology's utility is likely to be a top priority for many users, whereas safety features are sometimes better designed to be available, but unobtrusive, until

Table 3. Design Principles for Inclusive Digital Safety, Organized by User State

User State	Design Principle	Expected to help all users, but especially users with these challenges and risk factors Challenges	 Risk Factors
Prevention	Provide digital-safety education	Knowledge of risks and digital-safety options is limited Targeted by sophisticated or motivated attackers	–Legal or political –Prominence – Relationship with the attacker –Access to sensitive resources
	Make digital-safety options easy to find	Setting up digital-safety options competes with other urgent priorities Knowledge of risks and digital-safety options is limited	–Relationship with the attacker –Legal or political –Prominence –Resource constrained
	Design for strong digital safety by default	Knowledge of risks and digital-safety options is limited Targeted by sophisticated or motivated attackers	–Legal or political –Prominence –Relationship with the attacker –Access to sensitive resources
	Provide multiple digital-safety options, especially ways to delete	Safety concerns are contextual Cannot reasonably stop others from accessing their devices or accounts	–Relationship with the attacker –Reliance on third party –Social norms –Legal or political
	Provide private spaces on shared devices	Cannot reasonably stop others from accessing their devices or accounts	–Relationship with the attacker –Social norms –Resource constrained
	Reduce the size of security updates	Unable to avoid using an older device	–Resource constrained –Legal or political (in some geos)
	Consider supporting older devices and OSs	Unable to avoid using an older device	–Resource constrained –Legal or political (in some geos)
	Let users trade functionality for digital safety	Targeted by sophisticated or motivated attackers	–Legal or political –Prominence –Relationship with the attacker
Monitoring	Show the reach of info the user shares online	Copes with toxicity or privacy concerns when participating online	–Marginalization –Prominence –Relationship with the attacker
	Make data sharing transparent	Targeted with surveillance tactics	–Relationship with the attacker –Reliance on third party –Social norms
	Show digital-safety controls in relevant places	Knowledge of risks and digital-safety options is limited	–All
	Warn about potential digital-safety concerns	Knowledge of risks and digital-safety options is limited	–All
Active Event	Help users understand what is happening and show help info in context	Users are panicked	–All
	Provide a clear recommended action to stop the event, if possible	Users are panicked	–All
	Tailor help info to the user's device and OS	Unable to avoid using an older device	–Resource constrained –Legal or political (in some geos)
Active Event and Recovery	Design for users experiencing high to extreme stress	Users are panicked or traumatized	–All
Recovery	Provide redundancy in digital-safety-critical flows, especially account recovery	Does not have consistent access to technology (e.g., a personal device, phone number, or Wi-Fi)	–Relationship with the attacker –Resource constrained –Legal or political
	Use trauma-informed approaches	Users are traumatized	–Relationship with the attacker
	Set clear expectations when users take a digital-safety action	Users are traumatized	–All
	Provide recommendations for the future	Setting up digital-safety options competes with other urgent priorities Knowledge of risks and digital-safety options is limited	–All

These design principles are aimed to elevate digital safety for all users, but should be especially helpful for at-risk users. The design principles are initial building blocks for technology creators, but do not address all digital-safety issues, are non-prescriptive, and require careful thought to translate into safe, effective, workable solutions.

there is a need. Overall, these design principles can be used as an initial set of building blocks for technology creators that, at a high level, account for some of the diverse safety needs across general users and various at-risk groups.

5 Communicating with Practitioners about Inclusive Digital Safety

In this section, we share lessons learned from how we help tech workers use the frameworks and design principles described in Section 4, as well as build empathy for at-risk users (from research, such as the examples in Section 3). This includes how to communicate effectively to practitioner audiences via actionable and concrete educational programs and with framing for different roles (e.g., engineers, product management), and the need to spend extra time and involve experts when conducting work that involves people from at-risk groups. We derived these lessons from our efforts over the past 9+ years of training and consulting with tech workers on products and policies.

5.1 Actionable and Concrete Deliverables Practitioners Can Use

The ideas presented in research papers can have more impact on products and policies if they are translated into deliverables and services that are readily usable by practitioners. To communicate research on inclusive digital safety to practitioners, we have translated the ideas into a live class, self-serve educational materials, and consulting services where we work directly with teams.

Our *live class* (mentioned in Section 2.2) introduces tech workers to who at-risk users are, covers the User States Framework, and presents select challenges and design principles from Section 4. We communicate these ideas by adding concrete examples—including *user archetypes* and *example solutions*. The user archetypes are hypothetical examples of at-risk users, based on research, meant to help attendees empathize with these users, their emotional states, and needs—see Appendix C.1 for examples of five user archetypes from our class. The example solutions describe specific examples of products that demonstrate the design principles in Table 3—see Appendix B for three example product solution descriptions. Feedback from class attendees confirms that tech workers find these elements of the class valuable, because they are motivating, add concreteness, and make the class content relevant to their work.[11]

We also offer *self-serve educational materials* that Googlers can review on their own, which provide guidance relevant to understanding, designing inclusively for, and conducting research involving at-risk users. And since we started the At-Risk Research Program in 2015, we have offered *consulting services* for Googlers aimed to elevate inclusive digital safety in their particular projects, whether in research, product, or policy contexts. Any Googler can reach out for help about supporting at-risk users and inclusive digital safety. Our consulting team includes experts in the digital safety of at-risk users plus other areas: UX, privacy engineering, product management, and various products; in some cases, we have connected Googlers directly with NGOs for additional expert input. A network of relevant experts and consulting teams across the company know about our consulting services and refer teams to us. We have also promoted our services on e-mail groups. A range of product and policy teams across Google have sought advice from our consulting team. As we have discussed, inclusive digital safety is a wicked problem and best practices do not yet exist, so this consulting effort still requires time from experts to develop each one-off solution, and even then, the "right" solution is not always clear.

[11] Attendees could respond to a survey after taking the class, and 1,034 people have responded in total. Of respondents, 94.4% agreed or strongly agreed that "Overall, this experience was worth my time," indicating positive reactions to what they learned. Respondents to the class survey could also optionally write feedback about what they found most valuable about the class. The 105 people who wrote a response most commonly noted the value in (a) developing an awareness of at-risk users, (b) the user archetypes as examples of at-risk users, (c) the User States Framework (Prevention, Monitoring, Active Event, Recovery), and (d) the technology design examples that illustrated solutions to user challenges.

5.2 Communicating to Different Audiences

Support for at-risk users and inclusive digital safety is a wicked problem that requires many sectors and expert perspectives. To effectively engage these audiences, research knowledge and recommendations should be tailored to the various practitioners involved, when possible. For example, our peer-reviewed research papers are not targeted at software engineers, product managers, or other product or policy roles; rather, we have had to translate those research findings into materials that resonate with practitioner audiences. Here, we share examples of specific framings that have often been effective for two key audiences: software engineers and decision makers (like product managers and team leads). While this is not exhaustive of all relevant roles (e.g., we do not cover legal and regulatory requirements, policy makers, public relations), it provides an introduction to communicating to practitioners in this complex space.

Stress cases, not edge cases. To communicate with software engineers, we have learned to reframe talk of at-risk users' needs being "edge cases"—as may be a common instinct on tech teams, but suggests niche concerns—and toward the concept of "stress cases" to increase product resilience considering the risk factors, attacks, and challenges diverse users face. *Stress tests* are commonly known to software engineers for testing the robustness and reliability of a system under heavy loads. In the context of at-risk users, *stress cases* challenge how digital safety is handled and highlight potential abuse issues, helping teams make safer product and policy designs for everyone. Importantly, focusing only on general users could miss key needs of at-risk users, while ensuring designs can handle stress cases involving at-risk users helps users with a broad range of risk. We have found this framing to be very effective for tech workers, especially those in technical roles. See Appendix C.2 for details on how we use stress cases in technology creation.

Reframe the user base. To better support at-risk users, it helps to reframe broad user bases for decision makers and product managers by explaining why at-risk groups' needs are pervasive and that designing for-risk user needs early-on can improve actual and perceived digital safety for all users, dispelling the misconception that it would result in niche designs. To this end, we commonly give examples of "general" users who benefit from digital-safety features highlighted by at-risk user needs: for example, supporting prominent content creators as they manage toxic comments can help anyone whose content is unexpectedly seen by more people than intended. Not only can any user become at-risk at some point, but digital-safety features can help users feel safer and better protect themselves. Additionally, we emphasize that at-risk groups are *not* niche groups—they make up a substantial portion of a broad user base. It can help to present numbers that represent the size of different at-risk groups where possible—such as the WHO estimates that 1 in 3 women worldwide who have been in a relationship have experienced physical or sexual violence by an intimate partner [78]; the U.S. government estimates that 37% of U.S. adults would not be able to easily access $400 [76]; the World Bank estimates that in 2022, 3.6 billion people worldwide lived on less than $6.85 per day [46]; the WHO estimates that 1 in 6 people worldwide (i.e., 1.3 billion people) experience significant disability [79]; the United Nations reports that, in 2021, 1 in 10 people worldwide were over the age of 65, and this population is expected to grow to 1 in 6 people by 2050 [75]; and many more at-risk groups have been documented [101]. These numbers help dispel worries that at-risk user digital-safety challenges are niche concerns, making a strong case that it is important for teams to understand and actively support their needs.

5.3 Interventions That Support At-Risk Users Require Time and Expertise

Research, policy, and technology designs that can impact at-risk users tend to require more time and expertise to plan and execute safely [11]. We emphasize these points when communicating

with practitioners, as a way of setting expectations for how tech workers should approach the problem, as follows:

Foremost, we suggest that tech workers *work with trusted experts* when designing interventions (e.g., research, technology, or policy) that are likely to impact at-risk users' digital safety [11]. Experts can help assess an intervention's risks to at-risk groups, determine appropriate digital-safety mitigations, co-design interventions, facilitate interactions with at-risk groups to evaluate interventions, help communicate to and from at-risk groups, and more. Different expertise can provide invaluable support: experts on the at-risk group can advocate for the group's needs; experts in security, privacy, and safety can provide invaluable technical advice; and experts in areas such as accessibility, mental health, legal, ethics, and more can elevate digital safety in topics specific to the intervention [11].

To cope with the many factors making inclusive digital safety a wicked problem, creating new interventions requires *extra time and expertise*—from accounting for different digital-safety needs across user groups, to ensuring strong usability and function, to predicting potential misuse of features, and more (see Section 2.1). We have found that it can help with the complex decisions technology teams must make to ground risk assessments, in part, on risk factors and user states that apply across at-risk groups, and to emphasize that accounting for these should help everyone. Our frameworks presented in Section 4 provide a simpler way to think through these issues than examining dozens of at-risk groups. We encourage interested teams to use these frameworks first, as well as reviewing the literature, before considering their own direct research with at-risk groups [11].

Advocacy is an important part of the time spent to support at-risk groups' digital-safety needs in broadly used technology, policy, and research. Prior work relatedly documents the importance of "security champions" or advocates in organizations to encourage the adoption of good security practices [51, 80]. We have put considerable effort into communicating with tech workers and academic researchers about the value of supporting at-risk users and raising awareness for the time and care needed to do so. Overall in our experience, tech workers are incredibly receptive to learning about and better supporting at-risk users, and this advocacy work is about framing the issues to motivate safe action, providing actionable guidance, and raising awareness to grow a network of advocates.

6 Discussion: Charting Progress and Future Work

Given our many years of work on the At-Risk Research Program, we have learned that improving inclusive digital safety—across at-risk groups and all users—is a *wicked problem* [84], and while there is much work left to do, we can examine progress toward understanding and addressing the problem.

Gaps in foundational research on at-risk groups and inclusive digital safety. Despite growing literature on at-risk user digital safety, gaps exist in our foundational knowledge. The literature is biased toward Western contexts with gaps in understanding digital-safety needs in non-Western geographies and cultures [101]. Also, there is limited research on how multiple intersecting risk factors change users' digital-safety experiences and needs. Certain at-risk groups have received relatively little (if any) attention in the digital-safety literature, including a range of marginalized groups based on multiple axes of privilege and oppression [69], and people with disabilities other than visual impairments [77]. While there is still much more to learn—including about groups that are already represented in the at-risk research literature—we have collectively developed enough knowledge to organize what is known in this space into an initial set of frameworks and design principles for inclusive digital safety. We call for future research expanding these kinds of cross-group syntheses with new knowledge of at-risk groups, their risks, needs, and ideally what

makes solutions more effective. Going forward, we advocate that new recommendations made based on single-group research should be considered in light of other groups and risk factors, *if* the authors aim to impact technology intended for broad use.

Progress on solutions for inclusive digital safety is formative; the field is not ready to create best practices. Despite substantial progress in foundational research, more work is needed to determine technology and policy solutions that more effectively support digital safety for all users. We have not found a one-size-fits-all solution or reusable checklist that will resolve digital-safety risk for at-risk users, and the literature on wicked problems suggests that this is not a feasible goal [84]. Instead, designing for inclusive digital safety in broadly used technology is wickedly complicated, since context matters, technology is dynamic and can be misused, risk can be intersectional, needs across user groups are diverse and may contradict, and more as outlined in Section 2.1. This suggests that, even with extensive effort and consultations with experts in the creation of technology, it is likely that digital-safety risk will still exist for at-risk and general users. Also given this wickedness, we argue that the field is not yet prepared to develop best practices for designing inclusively safe technology and policy that adequately accounts for *all* users, including those who we consider to be at-risk. Formulating inclusive digital safety as a wicked problem will help the research community make more nuanced, appropriate, and ultimately safer statements about related problems, goals, study findings, and recommendations.

Recommendations: Manage wicked complexity with good practices and expert guidance. Many of the problems surrounding inclusive digital safety are at a societal level. Technology alone cannot "fix" this wicked problem, but we can help manage the complexity with good practices and expert guidance. Thus, our research and training efforts described above have focused on producing a range of "good practices"—research, frameworks, and design principles—for considering and mitigating risks that arise from users' circumstances, which can serve as a starting point for designing "better" research, technology, and policy. These good practices still need careful consideration and may require expert consultation, to apply on a case-by-case basis in practice. We have found that HCI and technical expertise offer only a subset of the considerations needed to deploy a product intended for broad use. Product constraints from legal, public relations, policy and regulatory requirements, ethics, and more are essential to consider. For researchers to have impact, they must consider this complexity—otherwise their recommendations run the risk of not being realistic, useful, or even safe in technology intended for broad use.

Future research questions. Inclusive digital safety is a wicked problem that extends beyond the purview of a single product, policy, or research effort. This problem needs the tech industry, academia, regulators, and other advocates to help further our knowledge toward safer solutions. With this in mind, we propose the following research questions for the community to explore:

- —In what ways might new research about at-risk groups expand our understanding of user risk factors, reactions to threats, and barriers to protections [101]?
- —How does intersectional risk—especially from yet un- or under-studied combinations of risk factors—impact user digital safety and needs?
- —What can we learn about supporting user digital safety from research with under-studied at-risk groups—such as marginalized groups [69] and people with disabilities [77]?
- —What prioritized advice would help various at-risk groups effectively react to or set up digital-safety measures for Prevention, Monitoring, coping with Active Events, or Recovery?
- —What technology design principles can support digital safety for a range of at-risk groups, accounting for the highly contextual nature of digital-safety threats and needs (perhaps building on our proposals in this article)?

—What approaches can help build inclusive digital safety into technology creation from start to finish?
—How can users seeking help for digital-safety concerns or events receive effective support? What type of help (e.g., technical, social, legal) is needed? Who should provide it? How should it be provided?
—How might generative AI impact digital safety for at-risk groups? For example, how might it add risk or assist with solutions?

7 Conclusion

In this article, we have described research insights, frameworks, design principles, and lessons learned from more than 9 years of working within Google to elevate at-risk users' digital-safety experiences and needs. In service of this contribution, we emphasized the importance of understanding at-risk users' needs to improving inclusive digital safety and described it as a *wicked problem*; presented two frameworks and a set of design principles to help technology creators apply findings from foundational research to the design and development of safer technology; shared lessons learned about communicating findings from foundational at-risk research with practitioners; and charted a path forward toward improving inclusive digital safety in technology intended for broad use. We believe that this article will be valuable to academic researchers who want their research to impact technology intended for broad use, and technology creators who want to elevate the digital safety of at-risk users in their offerings.

Acknowledgments

We would like to extend a very big thank you to the many people who have supported or provided feedback on this work over the years, including Amanda Walker, C. J. Adams, Garen Checkley, Elizabeth Churchill, Lee Dunn, Keith Enright, Jen Gennai, Bryant Gipson, Beth Goldberg, Yasmin Green, Amelia Hassoun, Shane Huntley, Justin Kosslyn, Ashley Ladner, Stefan Linssen, Sarah Meiklejohn, Matt Moore, Katie O'Leary, Ryan O'Toole, Jenni Kim Park, Dan Russell, Nithya Sambasivan, Patrawat Samermit, Alex Sayde, Martin Shelton, Jeff Shih, Nina Taft, Dave Vorhaus, Mia Vu, Vanessia Wu, our academic and NGO partners, the participants from our research, other internal and external experts who provided feedback on or otherwise supported our efforts, colleagues from the product and policy teams we have worked with, the reviewers and publishers of our academic papers, the hosts and audiences of our talks, and so many more, including you for your interest in supporting at-risk users' digital safety.

References

[1] Syed Ishtiaque Ahmed, Steven J. Jackson, Nova Ahmed, Hasan Shahid Ferdous, Md. Rashidujjaman Rifat, A. S. M. Rizvi, Shamir Ahmed, and Rifat Sabbir Mansur. 2014. Protibadi: A platform for fighting sexual harassment in urban Bangladesh. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, 2695–2704. DOI: https://doi.org/10.1145/2556288.2557376

[2] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. 2015. 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)*. ACM, New York, NY, 3523–3532. DOI: https://doi.org/10.1145/2702123.2702334

[3] Tousif Ahmed, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. 2016. Addressing physical safety, security, and privacy for people with visual impairments. In *Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS '16)*. USENIX Association, Berkeley, CA, 341–354. Retrieved from https://www.usenix.org/conference/soups2016/technical-sessions/presentation/ahmed

[4] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. 2020. "I am uncomfortable sharing what I can't see": Privacy concerns of the visually impaired with camera based assistive applications. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security '20)*. USENIX Association, Berkeley, CA, 1929–1948. Retrieved from https://www.usenix.org/conference/usenixsecurity20/presentation/akter

[5] David E. Alexander. 2002. *Principles of Emergency Planning and Management*. Oxford University Press.
[6] David E. Alexander. 2016. *How to Write an Emergency Plan*. Dunedin Academic Press.
[7] Deena Alghamdi, Ivan Flechais, and Marina Jirotka. 2015. Security practices for households bank customers in the kingdom of Saudi Arabia. In *Proceedings of the 11th USENIX Conference on Usable Privacy and Security (SOUPS '15)*. USENIX Association, Berkeley, CA, 297–308.
[8] Julio Angulo and Martin Ortlieb. 2015. "WTH.!?!" Experiences, reactions, and expectations related to online privacy panic situations. In *Proceedings of the 11th Symposium on Usable Privacy and Security (SOUPS '15)*. USENIX Association, Berkeley, CA, 19–38. Retrieved from https://www.usenix.org/conference/soups2015/proceedings/presentation/angulo
[9] Shaowen Bardzell. 2010. Feminist HCI: Taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, 1301–1310. DOI: https://doi.org/10.1145/1753326.1753521
[10] Catherine Barwulor, Allison McDonald, Eszter Hargittai, and Elissa M. Redmiles. 2021. "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)*. ACM, New York, NY, Article 563, 1–16. DOI: https://doi.org/10.1145/3411764.3445378
[11] Rosanna Bellini, Emily Tseng, Noel Warford, Alaa Daffalla, Tara Matthews, Sunny Consolvo, Jill Palzkill Woelfer, Patrick Gage Kelley, Michelle L. Mazurek, Dana Cuomo, et al. 2024. SoK: Safer digital-safety research involving at-risk users. In *Proceedings of the 2024 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, Los Alamitos, CA, 74–74. DOI: https://doi.org/10.1109/SP54263.2024.00071
[12] Marion Brown Brittany O'Shea, Rebecca Feicht, and Matthew Numer. 2024. Rethinking sexual violence labels: Exploring the impact of 'victim' and 'survivor' discourse. *European Journal of Psychotraumatology* 15, 1 (2024), 2296329. DOI: https://doi.org/10.1080/20008066.2023.2296329
[13] Tim Brown. 2008. Design thinking. *Harvard Business Review* 86, 6 (2008), 84.
[14] Richard Buchanan. 1992. Wicked problems in design thinking. *Design Issues* 8, 2 (1992), 5–21. Retrieved from http://www.jstor.org/stable/1511637
[15] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. 2018. The spyware used in intimate partner violence. In *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP)*, 441–458. DOI: https://doi.org/10.1109/SP.2018.00061
[16] Janet X. Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A. Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-informed computing: Towards safer technology experiences for all. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '22)*, 1–20. DOI: https://doi.org/10.1145/3491102.3517475
[17] Elizabeth F. Churchill and Mikael Wiberg. 2024. Design for all, design by all. *Interactions* 31, 5 (Aug. 2024), 5. DOI: https://doi.org/10.1145/3689344
[18] Sunny Consolvo. 2016. Stories from Survivors of Intimate Partner Abuse. HCIC '16 Keynote.
[19] Sunny Consolvo. 2017. Privacy and Security Practices of Individuals Coping with Intimate Partner Abuse. Talk at Enigma '17. Retrieved from https://www.usenix.org/conference/enigma2017/summit-program/presentation/consolvo
[20] Sunny Consolvo. 2021. Digital Security and U.S. Political Campaigns. Session: Securing Democracy, Talk at Enigma '21. Retrieved from https://www.usenix.org/conference/enigma2021/presentation/consolvo
[21] Sunny Consolvo. 2021. Stories from Survivors: Privacy and Security Practices When Coping with Intimate Partner Abuse. Talk at Safe@Home Hackathon.
[22] Sunny Consolvo. 2023. A Framework for Unifying At-Risk User Research. Talk at CNIL Privacy Research Day 2023. Retrieved from https://www.cnil.fr/en/privacy-research-day-discover-program-and-register-free-attend-event
[23] Sunny Consolvo, Patrick Gage Kelley, and Tara Matthews. 2020. Digital Security and U.S. Political Campaigns: Expert Roundtable. Google Civics Training Center. Retrieved May 21, 2024 from https://civicsresources.withgoogle.com/digital-security-expert-roundtable
[24] Sunny Consolvo, Patrick Gage Kelley, Tara Matthews, Kurt Thomas, Lee Dunn, and Elie Bursztein. 2021. "Why wouldn't someone think of democracy as a target?": Security practices and challenges of people involved with U.S. political campaigns. In *Proceedings of the 30th USENIX Security Symposium (USENIX Security '21)*. USENIX Association, Berkeley, CA, 1181–1198. Retrieved from https://www.usenix.org/conference/usenixsecurity21/presentation/consolvo
[25] W. Timothy Coombs and Sherry J. Holladay (Eds.). 2010. *The Handbook of Crisis Communication*. Wiley-Blackwell.
[26] Kimberlé Crenshaw. 1989. Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum* 1989, Article 8 (1989), 139–168. Retrieved from https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8
[27] Deloitte. 2016. Cyber Crisis Management: Readiness, Response, and Recovery. Retrieved January 29, 2025 from https://www2.deloitte.com/kz/en/pages/risk/articles/cyber-crisis-management.html

[28] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. 2016. Social justice-oriented interaction design: Outlining key design strategies and commitments. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, 656–671. DOI: https://doi.org/10.1145/2901790.2901861

[29] Kasra Edalatnejad, Wouter Lueks, Julien Pierre Martin, Soline Ledésert, Anne L'Hôte, Bruno Thomas, Laurent Girod, and Carmela Troncoso. 2020. DatashareNetwork: A decentralized privacy-preserving search engine for investigative journalists. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security '20)*. USENIX Association, Berkeley, CA, 1911–1927. Retrieved from https://www.usenix.org/conference/usenixsecurity20/presentation/edalatnejad

[30] Ame Elliott and Sara Brody. 2015. *Straight Talk: New Yorkers on Mobile Messaging and Implications for Privacy*. Technical Report. Simply Secure. Retrieved from https://simplysecure.org/resources/techreports/NYC15-MobMsg.pdf

[31] Sheena Erete, Aarti Israni, and Tawanna Dillahunt. 2018. An intersectional approach to designing in the margins. *Interactions* 25, 3 (2018), 66–69. DOI: https://doi.org/10.1145/3194349

[32] FEMA. 2024. Emergency Management: Definition, Vision, Mission, Principles. Retrieved May 17, 2024 from https://emilms.fema.gov/is_0908/media/63.pdf

[33] Belfer Center for Science and International Affairs. 2017. Cybersecurity Campaign Playbook. Retrieved from https://www.belfercenter.org/publication/cybersecurity-campaign-playbook

[34] Sarah Fox, Amanda Menking, Stephanie Steinhardt, Anna Lauren Hoffmann, and Shaowen Bardzell. 2017. Imagining intersectional futures: Feminist approaches in CSCW. In *Proceedings of the Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17 Companion)*. ACM, New York, NY, 387–393. DOI: https://doi.org/10.1145/3022198.3022665

[35] Cynthia Fraser, Erica Olsen, Kaofeng Lee, Cindy Southworth, and Sarah Tucker. 2010. The new age of stalking: Technological implications for stalking. *Juvenile and Family Court Journal* 61, 4 (2010), 39–55.

[36] Diana Freed, Natalie N. Bazarova, Sunny Consolvo, Eunice J. Han, Patrick Gage Kelley, Kurt Thomas, and Dan Cosley. 2023. Understanding digital-safety experiences of youth in the U.S. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, Article 191, 1–15. DOI: https://doi.org/10.1145/3544548.3581128

[37] Diana Freed, Sunny Consolvo, Dan Cosley, Patrick Gage Kelley, Ender Ricart, Kurt Thomas, and Natalie N. Bazarova. 2025. Help-seeking and coping strategies for technology-facilitated abuse experienced by youth. *Proceedings of the ACM on Human-Computer Interaction*. CSCW.

[38] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2018. "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)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3173574.3174241

[39] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2017. Digital technologies and intimate partner violence: A qualitative analysis with multiple stakeholders. *Proceedings of the 1st ACM on Human-Computer Interaction*. CSCW, Article 46, 1–22. DOI: https://doi.org/10.1145/3134681

[40] Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J. LaViola Jr, and Pamela J. Wisniewski. 2018. Safety vs. surveillance: What children have to say about mobile apps for parental control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3173574.3173698

[41] Arup Kumar Ghosh, Charles E. Hughes, and Pamela J. Wisniewski. 2020. Circle of trust: A new approach to mobile online safety for families. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3313831.3376747

[42] Noah J. Goldstein, Steve J. Martin, and Robert Cialdini. 2009. *Yes!: 50 Scientifically Proven Ways to Be Persuasive*. Free Press.

[43] Google. 2024. Check and Delete Your Chrome Browsing History. Retrieved April 5, 2024 from https://support.google.com/chrome/answer/95589

[44] Google. 2024. Google Workspace. Retrieved May 17, 2024 from https://workspace.google.com/

[45] Google. 2024. Share Your Real-Time Location With Others in Google Maps. Retrieved January 29, 2025 from https://support.google.com/maps/answer/15437054

[46] World Bank Group. 2024. Poverty and Inequality Update. Retrieved May 17, 2024 from https://thedocs.worldbank.org/en/doc/69d007a1a509633933b92b3804d0e504-0350012024/original/poverty-and-inequality-spring-update-6.pdf

[47] Amelia Hassoun, Ian Beacock, Sunny Consolvo, Beth Goldberg, Patrick Gage Kelley, and Daniel M. Russell. 2023. Practicing information sensibility: How gen Z engages with online information. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, Article 662, 1–17. DOI: https://doi.org/10.1145/3544548.3581328

[48] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2019. Clinical computer security for victims of intimate partner violence. In *Proceedings of the 28th USENIX Conference on Security Symposium (SEC'19)*. USENIX Association, Berkeley, CA, 105–122.

[49] Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang. 2019. Cooperative privacy and security: Learning from people with visual impairments and their allies. In *Proceedings of the 15th Symposium on Usable Privacy and Security (SOUPS '19)*. USENIX Association, Berkeley, CA, 1–20. Retrieved from https://www.usenix.org/conference/soups2019/presentation/hayes

[50] Joseph Henrich, Steven J. Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *The Behavioral and Brain Sciences* 33, 2-3 (2010), 61–135. DOI: https://doi.org/10.1017/S0140525X0999152X

[51] Martin Gilje Jaatun and Daniela Soares Cruzes. 2021. Care and feeding of your security champion. In *Proceedings of the 2021 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, 1–7. DOI: https://doi.org/10.1109/CyberSA52016.2021.9478254

[52] Google Jigsaw. 2016. Designing for an At-Risk World. Retrieved March 18, 2024 from https://medium.com/jigsaw/designing-for-an-at-risk-world-75081c6fa061

[53] Steve Krug. 2009. *Don't Make Me Think! A Common Sense Approach to Web Usability*. Pearson Education.

[54] Deepak Kumar, Patrick Gage Kelley, Sunny Consolvo, Joshua Mason, Elie Bursztein, Zakir Durumeric, Kurt Thomas, and Michael Bailey. 2021. Designing toxic content classification for a diversity of perspectives. In *Proceedings of the 17th USENIX Conference on Usable Privacy and Security (SOUPS '21)*. USENIX Association, Berkeley, CA, Article 16, 1–19.

[55] Neha Kumar and Naveena Karusala. 2019. Intersectional computing. *Interactions* 26, 2 (2019), 50–54. DOI: https://doi.org/10.1145/3305360

[56] Priya C. Kumar, Marshini Chetty, Tamara L. Clegg, and Jessica Vitak. 2019. Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3290605.3300537

[57] Karen Levy and Bruce Schneier. 2020. Privacy threats in intimate relationships. *Journal of Cybersecurity* 6, 1 (2020). DOI: https://doi.org/10.1093/cybsec/tyaa006

[58] Yonatan Lupu, Richard Sear, Nicolas Velásquez, Rhys Leahy, Nicholas Johnson Restrepo, Beth Goldberg, and Neil F. Johnson. 2023. Offline events and online hate. *PLoS ONE* 18, 1 (2023), e0278511. DOI: https://doi.org/10.1371/journal.pone.0278511

[59] Sonali Tukaram Marne, Mahdi Nasrullah Al-Ameen, and Matthew K. Wright. 2017. Learning system-assigned passwords: A preliminary study on the people with learning disabilities. In *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS '17)*. USENIX Association, Berkeley, CA. Retrieved from https://www.usenix.org/conference/soups2017/workshop-program/wips2017/marne

[60] Tara Matthews, Kathleen O'Leary, Anna Turner, Manya Sleeper, Jill Palzkill Woelfer, Martin Shelton, Cori Manthorne, Elizabeth F. Churchill, and Sunny Consolvo. 2017. Stories from survivors: Privacy and security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 2189–2201. DOI: https://doi.org/10.1145/3025453.3025875

[61] Nora McDonald, Karla Badillo-Urquiola, Morgan G. Ames, Nicola Dell, Elizabeth Keneski, Manya Sleeper, and Pamela J. Wisniewski. 2020. Privacy and power: Acknowledging the importance of privacy research and design for vulnerable populations. In *Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. ACM, New York, NY, 1–8. DOI: https://doi.org/10.1145/3334480.3375174

[62] Susan E. McGregor, Polina Charters, Tobin Holliday, and Franziska Roesner. 2015. Investigating the computer security practices and needs of journalists. In *Proceedings of the 24th USENIX Security Symposium (USENIX Security '15)*. USENIX Association, Berkeley, CA, 399–414. Retrieved from https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/mcgregor

[63] Susan E. McGregor, Franziska Roesner, and Kelly Caine. 2016. Individual versus organizational computer security and privacy concerns in journalism. In *Proceedings of Privacy Enhancing Technologies*, 418–435. DOI: https://doi.org/10.1515/popets-2016-0048

[64] Susan E. McGregor, Elizabeth Anne Watkins, Mahdi Nasrullah Al-Ameen, Kelly Caine, and Franziska Roesner. 2017. When the weakest link is strong: Secure collaboration in the case of the Panama papers. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security '17)*. USENIX Association, Berkeley, CA, 505–522. Retrieved from https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/mcgregor

[65] Susan E. McGregor, Elizabeth Anne Watkins, and Kelly Caine. 2017. Would you slack that? The impact of security and privacy on cooperative newsroom work. *Proceedings of the 1st ACM on Human-Computer Interactions*. CSCW, Article 75, 1–22. DOI: https://doi.org/10.1145/3134710

[66] Brenna McNally, Priya Kumar, Chelsea Hordatt, Matthew Louis Mauriello, Shalmali Naik, Leyla Norooz, Alazandra Shorter, Evan Golub, and Allison Druin. 2018. Co-designing mobile online safety applications with children. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, 1–9. DOI: https://doi.org/10.1145/3173574.3174097

[67] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 5197–5207. DOI: https://doi.org/10.1145/3025453.3025735

[68] Merriam-Webster. 2024. Best Practice. Retrieved October 2, 2024 from https://www.merriam-webster.com/dictionary/best%20practice

[69] Kathryn Pauly Morgan. 1996. Describing the emperor's new clothes: Three myths of educational (in-)equity. In *The Gender Question in Education: Theory, Pedagogy, and Politics* (1st ed.). Ann Diller, Barbara Houston, Kathryn Pauly Morgan, and Maryann Ayim (Eds.). Routledge, New York, NY. DOI: https://doi.org/10.4324/9780429496530

[70] Carol Moser, Tianying Chen, and Sarita Y. Schoenebeck. 2017. Parents' and children's preferences about parents sharing about children on social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 5221–5225. DOI: https://doi.org/10.1145/3025453.3025587

[71] Jan Neutze. 2019. Microsoft 365 for Campaigns now available. Microsoft Blog. Retrieved March 8, 2024 from https://blogs.microsoft.com/on-the-issues/2019/06/19/microsoft-365-for-campaigns-now-available/

[72] NNEDV. 2016. Safety and Privacy on Twitter: A Guide for Victims of Harassment and Abuse. Retrieved March 7, 2025 from https://www.techsafety.org/blog/2017/6/30/nnedv-resource-highlight-safety-on-social-media

[73] NNEDV and Facebook. 2015. Privacy and Safety on Facebook: A Guide for Survivors of Abuse. Retrieved May 17, 2024 from https://www.techsafety.org/resources-survivor/facebook

[74] NNEDV and Uber. 2023. Using Uber: Safety and Privacy Considerations for Survivors. Retrieved May 17, 2024 from https://www.techsafety.org/uber-safety-and-privacy

[75] United Nations Department of Economic and Social Affairs. 2023. World Social Report 2023: Leaving No One Behind in an Ageing World. Retrieved from https://digitallibrary.un.org/record/4000104

[76] Board of Governors of the Federal Reserve System. 2023. Economic Well-Being of U.S. Households in 2022. Retrieved from https://www.federalreserve.gov/publications/files/2022-report-economic-well-being-us-households-202305.pdf

[77] World Health Organization. 2002. ICF Beginner's Guide: Towards a Common Language for Functioning, Disability and Health. Retrieved May 17, 2024 from https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health

[78] World Health Organization. 2013. Global and Regional Estimates of Violence Against Women: Prevalence and Health Effects of Intimate Partner Violence and Non-partner Sexual Violence. Retrieved from http://www.who.int/iris/handle/10665/85239

[79] World Health Organization. 2023. Disability. Retrieved May 17, 2024 from https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[80] OWASP. 2024. OWASP Security Champions Guide. Retrieved Oct 21, 2024 from https://owasp.org/www-project-security-champions-guidebook/

[81] John R. Porter, Kiley Sobel, Sarah E. Fox, Cynthia L. Bennett, and Julie A. Kientz. 2017. Filtered out: Disability disclosure practices in online dating communities. *Proceedings of the 1st ACM on Human-Computer Interactions*. CSCW, Article 87, 1–13. DOI: https://doi.org/10.1145/3134722

[82] Yolanda A. Rankin and Jakita O. Thomas. 2019. Straighten up and fly right: Rethinking intersectionality in HCI research. *Interactions* 26, 6 (2019), 64–68. DOI: https://doi.org/10.1145/3363033

[83] Karen Renaud and Lizzie Coles-Kemp. 2022. Accessible and inclusive cyber security: A nuanced and complex challenge. *SN Computer Science* 3, 346 (2022), 82–87. DOI: https://doi.org/10.1007/s42979-022-01239-1

[84] Horst W. J. Rittel and Melvin M. Webber. 1973. Dilemmas in a general theory of planning. *Policy Sciences* 4, 2 (1973), 155–169.

[85] Nithya Sambasivan, Amna Batool, Nova Ahmed, Tara Matthews, Kurt Thomas, Laura Sanely Gaytán-Lugo, David Nemer, Elie Bursztein, Elizabeth Churchill, and Sunny Consolvo. 2019. "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)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3290605.3300232

[86] Nithya Sambasivan, Garen Checkley, Amna Batool, Nova Ahmed, David Nemer, Laura Sanely Gaytán-Lugo, Tara Matthews, Sunny Consolvo, and Elizabeth Churchil. 2018. "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 14th USENIX Conference on Usable Privacy and Security (SOUPS '18)*. USENIX Association, Berkeley, CA, 127–142. Retrieved from https://dl.acm.org/doi/10.5555/3291228.3291240

[87] Patrawat Samermit, Anna Turner, Patrick Gage Kelley, Tara Matthews, Vanessia Wu, Sunny Consolvo, and Kurt Thomas. 2023. "Millions of people are watching you": Understanding the digital-safety needs and practices of creators. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security '23)*. USENIX Association, Berkeley, CA, 5629–5645. Retrieved from https://www.usenix.org/conference/usenixsecurity23/presentation/samermit

[88] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R. Brubaker. 2021. A framework of severity for harmful content online. *Proceedings of the 5th ACM on Human-Computer Interactions*. CSCW2, Article 368, 1–33. DOI: https://doi.org/10.1145/3479512

[89] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. 2017. Intersectional HCI: Engaging identity through gender, race, and class. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, 5412–5427. DOI: https://doi.org/10.1145/3025453.3025766

[90] Renee Shelby, Shalaleh Rismani, Kathryn Henne, AJung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, et al. 2023. Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. ACM, New York, NY, 723–741. DOI: https://doi.org/10.1145/3600211.3604673

[91] Lucy Simko, Ada Lerner, Samia Ibtasam, Franziska Roesner, and Tadayoshi Kohno. 2018. Computer security and privacy for refugees in the United States. In *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP)*, 409–423. DOI: https://doi.org/10.1109/SP.2018.00023

[92] Manya Sleeper, Tara Matthews, KathleenO'Leary, Anna Turner, Jill Palzkill Woelfer, Martin Shelton, Andrew Oplinger, Andreas Schou, and Sunny Consolvo. 2019. 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)*. ACM, New York, NY, 1–12. DOI: https://doi.org/10.1145/3290605.3300319

[93] Elizabeth Stobert, David Barrera, Valérie Homier, and Daniel Kollek. 2020. Understanding cybersecurity practices in emergency departments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, 1–8. DOI: https://doi.org/10.1145/3313831.3376881

[94] Angelika Strohmayer, Jenn Clamen, and Mary Laing. 2019. 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)*. ACM, New York, NY, 1–14. DOI: https://doi.org/10.1145/3290605.3300882

[95] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, et al. 2021. SoK: Hate, harassment, and the changing landscape of online abuse. In *Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 247–267. Retrieved from https://ieeexplore.ieee.org/document/9519435

[96] Kurt Thomas, Patrick Gage Kelley, Sunny Consolvo, Patrawat Samermit, and Elie Bursztein. 2022. "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)*, 1–15. DOI: https://doi.org/10.1145/3491102.3501879

[97] Rebecca Umbach, Nicola Henry, Gemma Faye Beard, and Colleen M. Berryessa. 2024. Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, Article 779, 1–20. DOI: https://doi.org/10.1145/3613904.3642382

[98] Joris van Mens and Pranay Bhatia. 2020. Making Privacy Personal with Files by Google. Retrieved January 29, 2025 from https://blog.google/around-the-globe/google-asia/making-privacy-personal-files-google/

[99] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. 2019. EarTouch: Facilitating smartphone use for visually impaired people in mobile and public scenarios. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3290605.3300254

[100] Yang Wang. 2018. Inclusive security and privacy. *IEEE Security and Privacy* 16, 4 (2018), 82–87. DOI: https://doi.org/10.1109/MSP.2018.3111237

[101] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L. Mazurek, Manya Sleeper, and Kurt Thomas. 2022. SoK: A framework for unifying at-risk user research. In *Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP)*, 2344–2360. DOI: https://doi.org/10.1109/SP46214.2022.9833643

[102] Miranda Wei, Sunny Consolvo, Patrick Gage Kelley, Tadayoshi Kohno, Tara Matthews, Sarah Meiklejohn, Franziska Roesner, Renee Shelby, Kurt Thomas, and Rebecca Umbach. 2024. Understanding help-seeking and help-giving on social media for image-based sexual abuse. In *Proceedings of the 33rd USENIX Security Symposium (USENIX Security '24)*. USENIX Association, Philadelphia, PA, 4391–4408. Retrieved from https://www.usenix.org/conference/usenixsecurity24/presentation/wei-miranda-understanding

[103] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L. Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring care: How trans and non-binary people meet health and well-being needs through technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, Article 489, 1–17. DOI: https://doi.org/10.1145/3544548.3581040

[104] Jill Palzkill Woelfer and David G. Hendry. 2010. 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 (CHI '10)*. ACM, New York, NY, 1291–1300. DOI: https://doi.org/10.1145/1753326.1753520

[105] Marisol Wong-Villacres, Arkadeep Kumar, Aditya Vishwanath, Naveena Karusala, Betsy DiSalvo, and Neha Kumar. 2018. Designing for intersections. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, 45–58. DOI: https://doi.org/10.1145/3196709.3196794

[106] YouTube. 2024. Creators Safety Center. Retrieved May 15, 2024 from https://www.youtube.com/creators/safety/

[107] YouTube. 2024. Learn about Comment Settings. Retrieved May 15, 2024 from https://support.google.com/youtube/answer/9483359

[108] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 2019. 'I make up a silly name': Understanding children's perception of privacy risks online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, 1–13. DOI: https://doi.org/10.1145/3290605.3300336

Appendices

A List of Papers from the At-Risk Research Program

Table A1. List of Peer-Reviewed Papers Published or to Appear as Part of Google's At-Risk Research Program (as of January 2025)

Title	Topic	Authors	Affiliations	Venue	Year
Supporting the Digital Safety of At-Risk Users: Lessons Learned from 9+ Years of Research and Training	Program synthesis	T. Matthews, E. Bursztein, P. G. Kelley, L. Kissner, A. Kramm, A. Oplinger, A. Schou, M. Sleeper, S. Somogyi, D. Szostak, K. Thomas, A. Turner, J. P. Woelfer, L. L. You, I. Zahorian, S. Consolvo	Google	TOCHI	2025
Help-Seeking and Coping Strategies for Technology-Facilitated Abuse Experienced by Youth	Youth	D. Freed, S. Consolvo, D. Cosley, P. G. Kelley, E. Ricart, K. Thomas, N. N. Bazarova	Brown U Cornell U Google NSF	CSCW	To appear
Understanding Help-Seeking and Help-Giving on Social Media for Image-Based Sexual Abuse	Image-based sexual abuse	M. Wei, S. Consolvo, P. G. Kelley, T. Kohno, T. Matthews, S. Meiklejohn, F. Roesner, R. Shelby, K. Thomas, R. Umbach	Google U of Washington	USENIX Security	2024
SoK: Safer Digital-Safety Research Involving At-Risk Users	Cross-group synthesis	R. Bellini, E. Tseng, N. Warford, A. Daffalla, T. Matthews, S. Consolvo, J. P. Woelfer, P. G. Kelley, M. L. Mazurek, D. Cuomo, N. Dell, T. Ristenpart	Cornell Tech Google JumpCloud Lafayette College U of Maryland	IEEE S&P (Oakland)	2024
"Millions of people are watching you": Understanding the Digital-Safety Needs and Practices of Creators	Creators	P. Samermit, A. Turner, P. G. Kelley, T. Matthews, V. Wu, S. Consolvo, K. Thomas	Google	USENIX Security	2023
Understanding Digital-Safety Experiences of Youth in the U.S.	Youth	D. Freed, N. N. Bazarova, S. Consolvo, E. J. Han, P. G. Kelley, K. Thomas, D. Cosley	Cornell U Google NSF	CHI	2023
Practicing Information Sensibility: How Gen Z Engages with Online Information	Youth	A. Hassoun, I. Beacock, S. Consolvo, B. Goldberg, P. G. Kelley, D. M. Russell	Gemic Google/Jigsaw	CHI	2023
SoK: A Framework for Unifying At-Risk User Research	Cross-group synthesis	N. Warford, T. Matthews, K. Yang, O. Akgul, S. Consolvo, P. G. Kelley, N. Malkin, M. L. Mazurek, M. Sleeper, K. Thomas	Google U of Maryland	IEEE S&P (Oakland)	2022
"It's common and a part of being a content creator": Understanding How Creators Experience and Cope with Hate and Harassment Online	Creators	K. Thomas, P. G. Kelley, S. Consolvo, P. Samermit, E. Bursztein	Google	CHI	2022

(continued)

Table A1. continued

Title	Topic	Authors	Affiliations	Venue	Year
"Why wouldn't someone think of democracy as a target?": Security Practices and Challenges of People Involved with U.S. Political Campaigns	Political campaigns	S. Consolvo, P. G. Kelley, T. Matthews, K. Thomas, L. Dunn, E. Bursztein	Google	USENIX Security	2021
Designing Toxic Content Classification for a Diversity of Perspectives	Cross-group synthesis, Hate and harassment	D. Kumar, P. G. Kelley, S. Consolvo, J. Mason, E. Bursztein, Z. Durumeric, K. Thomas, M. Bailey	Google Stanford U U of Illinois, U-C	SOUPS	2021
SoK: Hate, Harassment, and the Changing Landscape of Online Abuse	Cross-group synthesis, Hate and harassment	K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. G. Kelley, D. Kumar, D. McCoy, S. Meiklejohn, T. Ristenpart, G. Stringhini	Boston U Cornell Tech Figma, Inc. Google New York U Stanford U U College London U of Illinois, U-C	IEEE S&P (Oakland)	2021
Toward Gender-Equitable Privacy and Security in South Asia	Women in South Asia	N. Sambasivan, N. Ahmed, A. Batool, E. Bursztein, E. Churchill, L. S. Gaytán-Lugo, T. Matthews, D. Nemar, K. Thomas, S. Consolvo	Google Info Tech U North South U U de Colima U of Virginia	IEEE Security and Privacy Magazine	2019
"They don't leave us alone anywhere we go": Digital Abuse Challenges and Coping Practices Among South Asian Women	Women in South Asia	N. Sambasivan, A. Batool, N. Ahmed, T. Matthews, K. Thomas, L. S. Gaytán-Lugo, D. Nemer, E. Bursztein, E. Churchill, S. Consolvo	Google Info Tech U North South U U de Colima U of Kentucky	CHI	2019
Tough Times at Transitional Homeless Shelters: Considering the Impact of Financial Insecurity on Digital Security and Privacy	People who are financially insecure	M. Sleeper, T. Matthews, K. O'Leary, A. Turner, J. P. Woelfer, M. Shelton, A. Oplinger, A. Schou, S. Consolvo	Google	CHI	2019
"Privacy is not for me, it's for those rich women": Performative Privacy Practices on Mobile Phones by Women in South Asia	Women in South Asia	N. Sambasivan, G. Checkley, N. Ahmed, A. Batool, D. Nemer, L. S. Gaytán-Lugo, T. Matthews, S. Consolvo, E. Churchill	Google Info Tech U North South U U de Colima U of Kentucky	SOUPS	2018
Security and Privacy Experiences and Practices of Survivors of Intimate Partner Abuse	Survivors of Intimate Partner Abuse	T. Matthews, K. O'Leary, A. Turner, M. Sleeper, J. P. Woelfer, M. Shelton, C. Manthorne, E. F. Churchill, S. Consolvo	CORA Google	IEEE Security and Privacy Magazine	2017
Stories from Survivors: Privacy and Security Practices when Coping with Intimate Partner Abuse	Survivors of Intimate Partner Abuse	T. Matthews, K. O'Leary, A. Turner, M. Sleeper, J. P. Woelfer, M. Shelton, C. Manthorne, E. F. Churchill, S. Consolvo	CORA Google	CHI	2017

B Examples of Supporting Inclusive Digital Safety in Practice

We have leveraged the research-based insights, frameworks, and design principles described above to influence product and policy decisions within Google. One of the more direct ways we can identify impact is when we consult directly with teams to help them consider how to better support at-risk users' digital-safety needs. Here, we share some stories of how our foundational research has impacted products.

Location Sharing in Maps. Around the time we began studying at-risk users, the Location Sharing [45] team wanted to offer users the ability to share their location in Google Maps, while mitigating potential abuse of the feature. They wanted to consider IPA as a possible abuse scenario and used our research with survivors of IPA [60] to inform their design of two Location Sharing features. First, the default time interval for Location Sharing needed to balance digital safety and usefulness.

This led to a default of 1 hour, which was short enough to not open users up to long-term sharing, and long enough to be useful (for those wishing to share their location). Second, it was important that the feature be transparent with users that their location was being shared and with whom, to reduce the possibility that it could be kept "on" without the user's awareness. This led to randomly timed e-mails that informed users when location sharing was "on," to hopefully increase the chance that a user would see the e-mail without it being intercepted by an abuser. Members of the product team found it particularly important to know about the phases of abuse [60], especially that a user could go through multiple periods of time where their abuser could physically access (and tamper with) their device. The Location Sharing team innovated these ways to potentially mitigate abuse risk in their product, and some of their team members have become active champions, instructors, and consultants in the At-Risk Research Program.

Security Templates in Workspace. The findings from our research with people involved with U.S. political campaigns [24] helped inform the Security Templates feature in Google Workspace [44]. This product team was motivated to improve election-related digital safety and could see from our research that campaigns struggled to deploy existing security and digital-safety features given their aggressive timelines and limited technical expertise. Thus, Security Templates in Workspace streamlines the setup of existing Workspace security features with a template configuration that aligns with the digital-safety needs of campaigns workers, guided by our research and consultations with members of our team and external experts. For example, by default, 2FA is set to "on," sharing files publicly is disallowed, strong passwords are enforced, reusing passwords is disallowed, and more. These default settings are easily deployable by campaign workers who may not have professional IT experience since campaigns rarely have these skills on board. These security settings are also available to other types of professionals who may benefit from them (such as small businesses), demonstrating that supporting the digital safety of at-risk groups can help other users too.

Creator Safety Center and moderation features in YouTube. Our research with content creators [87, 96] indicated a need to raise their awareness of the digital-safety threats creators commonly face and the recommended protections to employ. This helped inform the development of the YouTube Creator Safety Center [106], launched in 2022, which provides education for creators on top digital-safety concerns and ways to mitigate them. Our research also highlighted the importance of moderation as a protection against toxic content, informing YouTube moderation features [107] to help creators cope with toxic comments at scale. For example, creators can increase the strictness threshold for the automatic detection of inappropriate comments that are held for review to scale comment moderation. In these ways, the digital-safety research with creators showed ways to improve digital safety for anyone who creates online content.

C Details on Approaches to Inclusive Digital Safety

Here, we expand on our approaches toward inclusive digital safety presented in Sections 4 and 5, by sharing example user archetypes, how stress cases can be applied, and definitions of the design principles.

C.1 User Archetypes

The User States Framework—Prevention, Monitoring, Active Event, and Recovery—is more usable by tech workers, in our experience, when they have concrete examples of who experiences these states. Thus, we often couple the user states with a set of at-risk user archetypes—which were developed based on research—that incorporate aspects of the user states (e.g., user emotions and needs) and a set of challenges (like those listed in Table 3). Five user archetype examples are shown in Table C2. These archetypes are non-exhaustive, and we continue to expand and evolve them.

Table C2. Five Example At-Risk User Archetypes

Nicole **Survivor of intimate partner abuse (largely based on findings from [60])** "Nicole" recently left an abusive partner. He wants to find her and convince her to return. She is afraid. This is not the first time Nicole has tried to leave him, but he has found her and brought her back. Nicole does not know how he found her after she escaped, but she suspects some kind of digital surveillance. When they lived together, he always checked her phone—anytime she got a message, he had to check to make sure that it was not another guy. He regularly reviewed her call logs, her e-mail, and her messaging accounts. He had many opportunities to tamper with her device and her account settings. He also knows a lot about her. Nicole contemplates how to protect herself against this attacker—a top priority for her is to keep him from finding her new location and contact information.	**Rahim** **Media activist in a repressive country (largely based on findings from internal studies)** "Rahim" is a media activist in a repressive country, who uses social media to speak out about political issues. Rahim's main concern is that the government will detain him or his family members. The police once confiscated, then returned his smartphone. He destroyed it, since he suspected it was compromised and his activities on it would be surveilled. Since then, he has not been able to afford more than a very old device that no longer accepts security updates. Rahim keeps all of his data in the cloud, just in case his phone or his laptop are confiscated. But he avoids services provided by local ISPs since he believes these are monitored by his government. To access some censored sites, Rahim uses a VPN on his laptop.
Jackson **Online celebrity (largely based on findings from [87, 96])** "Jackson" has been trying to establish himself as a Creator for the last few years. He has had a channel with a small, but loyal audience where he talks about his personal life, his daily struggles, and his successes. All of a sudden, one of his videos goes viral, and he quickly gains popularity and online celebrity. At first, Jackson is thrilled with the increased attention. But, then he notices undesired side effects that he had not anticipated. He had created many of his older videos without expecting a larger audience. His new fans go through his old videos and figure out his home address. Some show up at his house, and he has to call the police. Some of Jackson's older videos also included friends and family members, and some of Jackson's fans start to stalk and harass an ex-girlfriend who is no longer involved in his channel. Jackson also had an eating disorder that he had talked about on his videos in the past, and now he is receiving a lot of negative comments about it. This creates a lot of stress for him. He had not thought about large audiences seeing those videos when he originally posted them. While Jackson's channel is very popular, it is not producing the kind of revenue he would need to hire bodyguards or to get help from an expert with digital security or privacy. He is stressed about these issues, but also depends on the income from his content. He worries about how to balance sharing enough to keep true to his brand while also protecting himself, his friends, and his family.	**Sally** **Politician in the United States (largely based on findings from [24])** "Sally" is a politician in the United States. She is currently elected to a local office in her state, but she is considering a run for the U.S. House of Representatives in the future. She has been in politics for over a decade and has dealt with her share of complicated incidents. Her social media pages have had the occasional hate and harassment incidents over the years. Once, an angry man showed up at a campaign field office shouting and making threatening gestures, but the police came and no one was hurt, just shaken. She knows that sometimes her constituents do not like her policies, but she is also concerned that disgruntled citizens from other parts of the country—or maybe even bots—are attempting to create a false sense of hostility. Her staff help her manage her social media accounts, and also have access to some of her campaign and fundraising e-mail accounts. She does not have the resources for dedicated IT support staff or physical security. As she looks toward her potential run for federal office, she is worried about raising her profile at the national level, and the increased attention it will bring, as well as the potential to be targeted by advanced, savvy attackers from foreign governments. She feels like she has a pretty good grasp on technology, and gets PIN codes texted to her sometimes when she logs in, but worries it may not be enough.
George **Struggling to get by (largely based on findings from [92])** "George" is a married veteran in the United States. He temporarily lives in a homeless shelter with his wife and two young kids while looking for affordable housing. Although George and his wife work, their salaries are very low for the area. The family lives paycheck to paycheck. George and his wife are looking for better paying jobs. George's family shares a laptop and two phones. But those devices are old and often run out of memory. He has trouble freeing up enough space for security updates. George spends a lot of time looking for housing online, but he is not very tech savvy. He is concerned about online scams, having fallen prey to one before. He also worries that a relative, who once stole money from him, might get access to his accounts.	

These at-risk user archetypes are non-exhaustive, and we continue to expand and evolve them.

C.2 Stress Cases

Stress cases are a process tool for using knowledge of at-risk users from research and the frameworks in Section 4, to evaluate and improve designs in light of how an at-risk user could experience them.

This is akin to *stress testing* in software engineering (i.e., determining the robustness of software by testing beyond the limits of normal operation). Here, we build on this concept to include stress testing UX designs to account for digital-safety needs of at-risk users.

User states, challenges, risk factors (from Table 3), and user archetypes can be used to help tech workers think about how at-risk users may experience—and stress—their designs. For example, tech workers could go through user interface flows for tasks in the user states (Prevention, Monitoring, Active Event, Recovery) and imagine how relevant challenges (e.g., not being able to stop others from accessing their devices or accounts, being targeted by sophisticated or motivated attackers) could be experienced by the relevant user archetypes. They can ask questions like: In what ways did the archetype's circumstances stress the flow's design? How could the flow be improved, to be safer and more inclusive for at-risk users?

The process of evaluating stress cases is open ended, but can be more effective when it is considered early in the technology design process and revisited throughout. Teams can manage the complexity of evaluating stress cases by prioritizing the most relevant ones and those that are more likely to exacerbate user harm. In our experience, this process can be more productive for teams who engage with experts, such as those with expertise in security, privacy, abuse prevention, UX, and at-risk groups. Such experts can help teams focus on relevant stress cases for their technology and help ensure that important stress cases are not missed.

As an example stress case, consider a flow that helps users recover from monetary loss, like a financial scam. Then imagine the archetype, George, using this kind of flow to recover from an event in which a scammer took his money, engaging questions such as: How easy would it be for George, who likely feels elevated stress, to recover his money? Does the flow assume a financial safety net or certain standing with the financial institution? If so, is it possible to safely speed up the recovery process to avoid putting unnecessary strain on George (who may need the money imminently for food or rent)?

C.3 Design Principles

We have developed a set of design principles, presented in Section 4.4 and defined below, to help tech workers account for some key challenges across the four user states. In our training, we share a selection of these with tech workers as examples to help them get started thinking through how they can better support at-risk users and design for inclusive digital safety. Importantly, given that this is a wicked problem, these principles are not intended to be prescriptive or exhaustive; instead, technology creators should consider them along with other constraints that may apply to broadly used technology, such as legal and regulatory requirements, how particular solutions function, usability, and more.

C.3.1 Prevention.

— *Provide digital-safety education.* Consider how to educate at-risk users or allies (e.g., non-profit staff advocates) to help at-risk users protect themselves from risks.
— *Make digital-safety options easy to find.* Make options that users can deploy to manage their digital safety easy to find, even for users experiencing ongoing stress. This may involve introducing digital-safety options in the product's onboarding, in just-in-time flows (e.g., right before the feature is used), at relevant moments (e.g., when a user goes to a relevant location), and/or during regular checkups about settings.
— *Design for strong digital safety by default.* Consider how default settings affect user security, privacy, and safety (for users in low- to high-risk situations). Product defaults are important since users may not modify them. Think about ways to make default settings easy to understand and limited in scope (when applicable).

— *Provide multiple digital-safety options, especially ways to delete.* Consider providing multiple privacy, security, or other relevant digital-safety options for users to choose from, to support their diverse digital-safety needs. Providing multiple digital-safety options can help at-risk users find options that align with their digital-safety goals and help manage a variety of threats. When deletion is an option, consider supporting multiple ways to delete content, such as all at once, select items, items within a range or topic, and so on as appropriate.
— *Provide private spaces on shared devices.* Consider providing the ability to divide devices or accounts that may be shared into separate sections—for example, with multi-account login, family or household account structures, guest mode, or secure folders. Think about how these might be employed by users who cannot reasonably stop others from accessing their devices or accounts.
— *Reduce the size of security updates.* Consider ways to reduce the data and storage required to install critical security and privacy updates. Installing such updates is important preventive measures, but users with limited resources may be using older devices with extremely limited space.
— *Consider supporting older devices and OSs.* Consider ways to provide support for older devices and OSs—such as ongoing security (and other) updates, and help center content that is inclusive of older devices and OSs. Think about how this might support users who may not be able to avoid using an older device.
— *Let users trade functionality for digital safety.* Consider functionality tradeoffs some groups may want to make to counter serious digital-safety threats, such as targeting by a nation-state adversary or abusive intimate partner. For example, a user may want to turn off features that share information to the public, avoid location information being saved to their account, limit data sharing to applications installed on their device, and more.

C.3.2 Monitoring.

— *Show the reach of information the user shares online.* Consider providing reminders to help users assess the potential privacy impact of their actions in the moment, especially when they are about to share information. Most fundamentally, this includes what information will be shared and where it will be shared. To further support monitoring, it can be helpful if users know where to find the information and are able to make changes later.
— *Make data sharing transparent.* Consider ways to make it hard for an attacker to access a user's information or activity without the user's awareness by providing enhanced transparency across multiple channels—such as via indicators when data are being shared, e-mail reminders at random intervals, or notifications on multiple surfaces. Think about how this might support users who are more likely targeted with surveillance tactics.
— *Show digital-safety controls in relevant places.* Consider surfacing helpful security, privacy, and safety controls in context, making them easily accessible as the user monitors their accounts, devices, content, or presence online. For example, if a user receives an e-mail alerting them of account activity (such as a new sign in), the alert could provide a link that takes them directly to where they need to go to check the activity and take action to protect their account.
— *Warn about potential digital-safety concerns.* Consider sending users digital-safety warnings before they take an action (such as sharing information or visiting a website) if there is evidence of a digital-safety concern.

C.3.3 Active Event.

— *Help users understand what is happening and show help information in context.* Users experiencing an active event do not always understand what is happening. They may benefit

from information that helps them understand what event is occurring and what they can do in response. Consider surfacing such help information, to assist users through common digital-safety events, where they are likely to seek it. This may be through search results, via links in account-related alerts, next to information the user shared or posted online, or any place they are likely to be looking when they discover a digital-safety event.
— *Provide a clear recommended action to stop the event, if possible.* Help stressed users understand what to do next by promoting a simple recommended action (e.g., changing their password after experiencing an account compromise). Once they have completed that action, other relevant actions can be thoughtfully suggested (e.g., turning on 2FA).
— *Tailor help information to the user's device and OS.* Consider how to provide users with, and direct users to, digital-safety help information and options tailored to their device and operating system.
— *Design for users experiencing high to extreme stress.* Users coping with the acute stress of an active event will benefit from simple language, clear recommendations, and strong usability. This principle is also particularly relevant to Recovery (next).

C.3.4 Recovery.

— *Provide redundancy in digital-safety-critical flows, especially account recovery.* Consider how to build redundancy into digital-safety-critical flows. For example, when possible, provide digital-safety options that work independent of Internet access or a reliable phone number.
— *Use trauma-informed approaches.* Assess how a traumatized user may experience an interface, and consider how to reduce triggering content, support transparency in how the system is functioning, and provide control to users in how they experience the system and manage their digital safety [16]. This principle can also be relevant to the other user states, particularly Active Event.
— *Set clear expectations when users take a digital-safety action.* Let users know what to expect next, if appropriate, when they take a digital-safety-related action. For example, if a user reports an issue, let them know if they should expect follow-up communication.
— *Provide recommendations for the future.* Consider presenting options that encourage users to manage their security, privacy, and digital safety moving forward.

D Terminology

Here, we define key terms used throughout this article. Because terms used to label individuals or groups are fraught and often debated—since referring to a person or group with a single term can flatten important aspects of their experiences—we provide clear definitions to capture the nuances and breadth of user circumstances explored in this article and position this work among discussions in the digital-safety literature.

At-risk users: People who face circumstances that elevate their chances of experiencing a tech-facilitated attack and/or suffering disproportionate harm from such an attack, compared to a baseline of "general" users [101]. A user may be "at risk" for a moment in time, their entire lifetime, or somewhere in between.

Attack/Tech-facilitated attack: Any means used by an attacker to abuse or compromise someone's digital safety (intentionally or unintentionally) that involves the use of technology and can lead to harm. Example attacks include phishing, hacking, doxing, overloading, surveillance, online harassment, and more [95].

Attacker: Anyone who introduces tech-facilitated attacks or harms, regardless of the severity or intention (i.e., for the purposes of this article, we consider someone who causes harm—whether

or not it is intentional—to be an attacker). This could include a broad range of individuals or groups, such as a family member, friend, stranger, nation-state, and more.

Best practices: "A procedure that has been shown by research and experience to produce optimal results and that is established or proposed as a standard suitable for widespread adoption" [68].

Digital safety/Digital-safety: An umbrella term, inclusive of topics pertaining to security, privacy, abuse, and/or any other safety-related topic for which technology is involved. "Digital safety" is a noun and "digital-safety" is an adjective.

General users: A construct of the kind of user that makes up the perceived majority of a technology's use, typically viewed as somewhere in the middle/average or even optimistically [104], for example, assuming that they have good access to trusted devices and reliable Internet, would benefit from increased technology use, and do not face extraordinary challenges. "General users" may not be real, instead reflecting how the holders of the concept view users—opening the concept up to biases, such as over-indexing on people from WEIRD societies [50]. The concept of general users is imprecise, but may persist as an attempt to design for the most users possible. We use "general users" as a stand-in for the assumptions described here, including that they experience lower levels of digital-safety risk when compared to at-risk users.

Good practices: Practices that, when carefully applied on a case-by-case basis, can lead to better or improved solutions for a wicked problem. Good practices are not established or standardized best practices.

Harm: A negative consequence(s) resulting, at least in part, from a tech-facilitated attack (e.g., one framework classified online harm into physical, emotional, relational, and financial harm [88]; a taxonomy of sociotechnical harms of algorithmic systems discussed harms that were representational, allocative, quality of service, interpersonal, and societal [90]).

Inclusive digital safety: Digital safety that accounts for diverse risks (pertaining to security, privacy, abuse, or other tech-facilitated safety risks) and supports the digital-safety needs of a broad user-base, including at-risk *and* general users.

Risk: The potential for a person to experience danger or harm.

Risk factor: Circumstances that increase a person's risk of tech-facilitated attacks or disproportionate harm from such an attack, such as marginalization, prominence, having a relationship with their attacker, being reliant on a third party in some way, having access to sensitive resources, and more as defined by Warford et al. [101] and described in Section 4.2.

Survivor: An individual who has experienced tech-facilitated harm, especially in the context of interpersonal attacks, like IPA or trafficking.[12]

Tech workers: Anyone working in the technology industry, including those responsible for creating technology, defining technology policies, conducting research, providing infrastructure to support products, promoting products, or other roles.

Technology creation: Any activities related to the creation of technology, whether in academic, commercial, non-profit, or other settings. This could involve a spectrum of functions and stages related to creating technology, such as early inventions, policy-guiding technology requirements, prototypes, and deployed products or services.

Technology creators: Anyone who engages in technology creation (see above definition).

[12]We use survivor instead of victim, because victim has been critiqued as being associated with negative attributes—like weakness or powerlessness—that can be internalized [12] The term *survivor* has been shown to be associated with strength, agency, and other positive attributes. However, any identity label comes with limitations, as it associates a person's experiences of abuse or violence with their identity, and it can impact how they are viewed by others or how they view themselves. We choose to avoid the more negative term victim, while acknowledging the downsides of survivor.

Wicked problem: A term introduced in social planning to describe complex societal problems for which issues, solutions, endpoints, and who is responsible are not agreed upon by experts or those involved [84]. Rittel and Webber [84] defined 10 characteristics of wicked problems, including that they have no definitive formulation, no stopping rule, no "correct" solutions, no definitive test to assess how good solutions are, no enumerable set of solutions, and more.

Received 21 May 2024; revised 8 November 2024; accepted 2 January 2025