# Safer Algorithmically-Mediated Offline Introductions: Harms and Protective Behaviors

VERONICA A. RIVERA, Stanford University, USA
DARICIA WILKINSON, Microsoft Research, USA
AURELIA AUGUSTA, Max Planck Institute for Software Systems, DE
SOPHIE LI, Max Planck Institute for Software Systems, DE
ELISSA M. REDMILES*, Georgetown University, USA
ANGELIKA STROHMAYER*, Northumbria University, UK

People are increasingly introduced to each other *offline* thanks to *online* platforms that make algorithmically-mediated introductions between their users. Such platforms include dating apps (e.g., Tinder) and in-person gig work websites (e.g., TaskRabbit, Care.com). Protecting the users of these online-offline systems requires answering calls from prior work to consider 'post-digital' orientations of safety: shifting from traditional technological security thinking to consider algorithm-driven consequences that emerge throughout online and offline contexts rather than solely acknowledging online threats. To support post-digital safety in platforms that make algorithmically-mediated offline introductions (AMOIs), we apply a mixed-methods approach to identify the core harms that AMOI users risk facing, the protective safety behaviors they employ, and the prevalence of those behaviors. First, we systematically review existing work ($n$ = 93), synthesizing the harms that threaten AMOIs and the protective behaviors people employ to combat these harms. Second, we validate prior work and fill gaps left by primarily qualitative inquiry through a survey of respondents' definitions of safety in AMOI and the prevalence and implementation of their protective behaviors. We focus on two exemplar populations who engage in AMOIs: online daters ($n$ = 476) and in-person gig workers ($n$ = 451). We draw on our systematization and prevalence data to identify several directions for designers and researchers to reimagine defensive tools to support safety in AMOIs.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: safety, security, online dating, gig work, algorithmically-mediated interactions

**ACM Reference Format:**
Veronica A. Rivera, Daricia Wilkinson, Aurelia Augusta, Sophie Li, Elissa M. Redmiles, and Angelika Strohmayer. 2018. Safer Algorithmically-Mediated Offline Introductions: Harms and Protective Behaviors. In . ACM, New York, NY, USA, 41 pages. https://doi.org/XXXXXXX.XXXXXXX

## 1 INTRODUCTION

A subset of social computing systems involve *algorithmically-mediated offline introductions* (AMOIs) in which people are algorithmically matched for offline meetups, crossing the digital-physical divide. For example, people find romantic partners through online dating apps, household service providers (e.g., to repair household items or care for children) through online gig-work platforms, and housing through online marketplaces [204]. While the interaction initiates online

*Both authors advised this work equally


Manuscript submitted to ACM

via an app or online platform, the goal of the matched individuals is to gravitate towards in-person interactions. While social computing platforms that enable AMOIs have brought considerable benefits, like increasing social connections and amplifying opportunities for labor, they also facilitate various harms that transcend digital boundaries and into the physical world. In this work, we take a mixed-methods approach to study safety in AMOIs, focusing on two representative interactions: online dating and in-person gig work.

The harms enabled by social computing systems, especially those that facilitate AMOIs, have a significant negative impact on users. A growing body of research examines the impact of online hate and harassment [52, 186, 187], mis/disinformation [184], stalking [67], intimate partner violence [37, 199, 200, 227], censorship [58] and privacy violations by automated systems [226] on users' mental health and physical safety [36, 206]. In AMOIs, threats to users' safety cross between digital and physical realms: a client met through a gig work platform can engage in physical assault while online harassment on a dating platform can result in trauma experienced both digitally and physically [44]. Such threats impact people's overall well-being, future interactions with digital systems, and their trust in the digital platform that mediated the harm [44, 152, 178].

Because threats to people's safety in AMOIs cannot be easily classified as strictly-digital or strictly-physical, mitigating them requires a highly interdisciplinary understanding of both the harms and current strategies people use to stay safe. Prior work investigating these questions has been conducted in disparate areas including computer security, HCI/CSCW, psychology, sociology, and criminology. Systematization, or synthesis of existing work, is powerful in providing such an understanding, allowing the research community to approach measurements and technical solution-building with theoretical grounding and shared language, especially in areas where definitions of key concepts are still under discussion [42, 163, 186, 212, 222]. This work aims to contribute such a systematization to guide CSCW researchers and practitioners in better safety by design [13] for AMOI systems, including highlighting opportunities to embed useful protective interventions in online platforms that mediate offline interactions.

Prior work has systematized different aspects of digital harm. These works aim to help us make sense of the severity of harmful online content [163], the varied definitions of privacy harms [42], the types of online hate and harassment attacks [186], and the factors that increase users' risk of digital attack [212]. Our work contributes to the broader conversation in which these frameworks are situated by considering high-risk *interactions* that cross the digital-physical divide and the *behaviors* users engage in to mitigate the harms arising from these interactions.

To do so, we systematically review 93 prior works on AMOI to contribute a synthesized taxonomy of the harms people face (Section 4) and the protective behaviors in which they engage (Section 5) throughout the process of an algorithmically-mediated offline introduction including prior-to-meeting, during meeting, and post-meeting. Our analysis identifies several protective behaviors: self-disclosure, obfuscation, screening, vetting, environmental precautions, covering, emergency alerts, surveillance and documentation, blocking, and reporting – to defend against four harms: physical, emotional, financial, autonomy.

To validate our taxonomy and fill gaps in the existing body of chiefly qualitative work, we collect empirical data from online daters ($n$ = 476) and in-person gig workers ($n$ = 451) to measure the salience of the harms identified in prior work, how prevalent adoption of these behaviors is in online daters' and gig workers' safety workflows and the mechanisms by which they carry out the behaviors. For example, our model highlights not just why users self-disclose, but also where they do so (e.g., online messaging, in app profile, during offline meeting).

Drawing on our results, we present directions for future research and design opportunities to help guide researchers and practitioners at the intersection of CSCW and security and privacy (S&P) in supporting the safety of users who engage in AMOIs (Section 6). We discuss the tensions and tradeoffs in introducing novel technologies to mitigate

tech-facilitated harm and how to better align safety interventions with users' existing strategies. We also illustrate how users' protective behaviors in AMOIs relate to security behaviors used to combat phishing and attacks on organizations and how these similarities can be leveraged to design more proactive, rather than reactive, safety defense mechanisms within social computing systems that support AMOIs.

## 2 TERMINOLOGIES AND SCOPING

Here we provide context on key definitions we leverage in our work: safety and algorithmically-mediated offline introductions.

### 2.1 Definitions of Safety: From Critical to Context-Driven Perspectives

In this study, we ground our definition of safety to reflect the multidimensional nature of people's well-being while acknowledging the complexities of what it means to be safe online [178, 219]. Decades of interdisciplinary research across privacy, security, criminology, social sciences, and legal studies, have engaged in methodologies to understand factors that aid or hinder people's sense of safety. In psychology, Maslow's Hierarchy of Needs describes safety as psychological needs that differ for each individual based on their current stage of life and which can only be addressed after basic survival needs (e.g., access to food, water, shelter) are met [119, 144]. Advocates for public policy have leaned on rights-affirming frameworks to inform conceptualizations of safety to extend understandings for protection within digital spaces [86]. Along this line, the United Nations (UN) has championed expanded definitions of safety to acknowledge intersecting realities among technical and relational aspects of online safety. The intersecting nature of the dimensions of safety makes it difficult to study one aspect without missing important details about how the nature of harms affects people overall. Experts in human rights and gender-based violence from the UN have conducted studies around online harassment in multiple contexts and note that: *"This abuse is often overlapping in its forms, may involve offline threats and attacks, and can lead women and girls to limit their participation and sometimes withdraw completely from online platforms"* [136]. Thus, in scholarship and in practice, thoughts on online safety continue to evolve from static definitions towards more holistic orientations.

This evolution could be observed within computing disciples as well. Coles-Kemp and colleagues argue that the *post-digital* enmeshing of our digital and non-digital worlds and the use of technology in a multitude of daily contexts have also embedded security issues into day to day life [44]. Thus, they motivate a need for security research to be more reflexive and participatory by considering safety beyond solely technical contexts including digital threats perpetuated by bad actors, such as phishing, scams, and social engineering attacks. For example, we must understanding what makes people feel more secure rather than pushing "one size fits all" narratives of risk. Strohmayer et. al. "suggest a paradigm shift is needed; from a focus on security to safety in pervasive computing, which is necessary to meaningfully and proactively protect people who use technologies in our complex world" [178]. They draw on feminist security and justice-oriented safety literatures to present the topic of 'safety' as a challenge and vision for the future of security research in our post-digital world. Meanwhile, social computing scholarship on safety has focused on targeted interpersonal harm like bullying and hate speech or content-based harm from viewing offensive or undesirable content on social media [163]. More recently social computing studies have drawn connections between the behavior that threatens online safety to better characterize the resulting harms [95, 139, 163]. In our work we align with this approach, understanding and conceptualizing safety and harm within the context of the interaction and space in which it occurs.

### 2.2 AMOI Scoping

Algorithmically-mediated offline introductions (AMOIs) are a subset of a broader class of digitally-mediated interactions that have been previously considered by prior work. These digitally-mediated interactions include various interpersonal interactions via online systems, such as social networking sites and online forums. Usually, digitally-mediated interactions on these systems are carried out online; however, sometimes users may decide to meet offline. For example, a journalist may interact with a potential source on Twitter and later decide to talk with them offline; similarly, a content creator or social media influencer may decide to host a meet and greet with online followers in an offline location. While these examples illustrate digitally-mediated *offline* interactions, they are not examples of *algorithmically-mediated offline introductions* because the platform on which the interaction initiates does not actively match individuals for offline meetings. Furthermore, while individuals who meet on social networking sites and online forums may decide to meet offline, the platform is not inherently designed to serve this purpose, unlike platforms that support AMOIs.

In this paper we focus on two groups of individuals who engage in AMOIs, online daters and gig workers who perform in-person domestic jobs (e.g., cleaning, organizing, handiwork, carework) where they interact with clients inside their homes and/or other intimate settings. We chose to focus on these two groups because the interactions they engage in are clearly AMOIs: both groups seek matches on online platforms and apps with an intent to meet their matches offline. Following this definition, we excluded Airbnb hosts from consideration in this paper. Some hosts are engaged in AMOIs (e.g., those who deliver keys to renters in-person and those who cohabitate with renters in their homes). However, not all do (e.g., those who use contactless key delivery and rent out private apartment units). We also excluded rideshare drivers and food couriers. While these user groups also engage in AMOIs, there is a short window between when they are matched and when they must accept a job. We wanted to prioritize selecting groups that have the ability to engage in prior to meeting protective behaviors.

## 3 METHODS

To build a taxonomy of harms and protective behaviors in AMOIs, we systematically collect and analyze literature in this domain, as summarized in Section 3.1. To validate, expand, and take first steps toward quantifying key concepts in this taxonomy (e.g., use of particular behaviors) we deploy a quantitative survey informed by the literature. We use similar approaches to those of [186, 212] to collect and systematize the literature, and use the results of our survey to supplement gaps in prior work. Our methodology is summarized in Section 3.2. We conclude with a summary of our work's limitations in Section 3.3.

As our contribution centers on providing a taxonomy of post-digital safety in AMOIs, we take a somewhat non-traditional approach to structuring our paper. We present a structured taxonomy of the harms experienced and protective behaviors used in AMOIs, leveraging our analysis of both related work and our survey results to justify and quantify the components of this taxonomy.

### 3.1 Literature Review

We conducted an electronic search of academic literature to identify post-digital harms and protective behaviors. Due to the interdisciplinary nature of this work, our search was conducted in databases spanning computer and social sciences such as Google Scholar, ACM Digital Library, ScienceDirect, Springer Link, and IEEE Xplore Digital Library. We considered articles available in English but adopted no restrictions on publication dates or venues. Relevant keywords included strings that were appended by the relevant interaction across the digital-physical boundary and "safety,"

"harm," or "scams" (e.g. "online dating safety" or "gig work scams"). We found additional literature by reviewing the related work section of each paper in our dataset. We specifically sought papers that discussed the safety concerns, safety definitions, and/or protective behaviors of people engaged in AMOIs such as gig workers, sex workers, ctivists, those dating online, and those in abusive intimate relationships.

**Data Abstraction.** For all papers, we reviewed the titles, abstracts, and concluding arguments for relevance. Ultimately, we reviewed 93 papers and abstracted data related to (1) harms and (2) protective behaviors. In identifying harms, we examined *mechanisms of harm* (e.g., How are the harms caused? Who or what creates these harms?). We likewise sought to understand what *protective behaviors* are used to protect against these harms. This included identifying (a) the phase in which the behavior is used (e.g., What behaviors are used before, during, or after an offline interaction?), (b) the harm mitigated by the behavior; and (c) the protective mechanism (e.g., How is this behavior executed? What online or offline tools and resources are required?)

We performed affinity diagramming [89] to understand the relationship between harms, protective behaviors, and the phases in which they occur. We identified four types of harm (Section 4), one mechanism through which harm occurs (Section 4), and ten protective behaviors by which users try to mitigate harm (Section 5). These findings informed our survey questions, as further described in Section 3.2. Through this iterative process, we also uncovered key differences in different groups' experiences with safety and gaps in the body of work we reviewed.

### 3.2 Survey

The prior work in this area is highly partitioned: the majority focuses on the experiences of a single population (e.g., online daters) and/or a single aspect of safety (e.g., data privacy). As the vast majority of prior work uses qualitative methods to examine safety in AMOIs, quantification of the threats and behaviors encountered in AMOIs is limited. Further, detailed information on the implementation of behaviors is scattered, most often buried in participant quotes, if at all. To offer a larger-scale validation of the harms and protective behaviors detailed in prior work and fill in gaps of knowledge on these threats, behaviors, and their implementations across multiple populations we surveyed two representative AMOI populations that cover two different classes of interactions (romantic and labor) common to AMOI: online daters ($n$ = 476) and in-person gig workers ($n$ = 451). Our survey methods were approved by our institutional ethics review board.

**Survey Questionnaire.** To validate our systematization of the four harms and one mechanism of harm we identified in the literature review and to understand which harms were most salient to people engaged in AMOIs, we asked respondents to explain what safety means to them in the context of the interactions in which they engage. Additionally, we asked questions to assess the role of safety in respondents' decisions to engage in AMOIs and the prevalence of their unsafe experiences.

To understand the prevalence of the ten protective safety behaviors we identified in the literature review, we asked respondents several questions regarding whether they engage in those behaviors and how they implement them. Our questions and answer choices were informed by the results of our literature review when possible; in cases where behaviors and/or their implementation might not have been fully explored by prior work, we developed logical answer choices. We aimed to obtain a comprehensive understanding of how different behaviors are used in different contexts. Therefore, both our participant populations answered the same questions, except for minor wording changes and answer choice options to reflect differences in context. We also included one attention check question in each survey, following best practice in survey methodology [151]. We discarded responses from those who did not answer the attention check question correctly. The exact wording of our survey questions are in Appendix A.3.

**Data Collection.** We recruited our sample of online daters using Prolific, a crowdworking platform ($n$ = 372), and Lucid, a marketplace for survey panels ($n$ = 104). We recruited our sample of gig workers ($n$ = 451) only from Prolific. Our surveys ran for 4 months, from August to December 2021. We recruited respondents who met the following criteria: (1) were located in the U.S., (2) had used a dating or gig app within the past two years, respectively, and (3) had met in-person with someone they met on a dating or gig app. Because our survey respondents are online daters or gig workers in the U.S., we aimed to recruit samples with demographics roughly representing the U.S. using the 2020 Census [28]. For complete demographic information, see Appendix A.5.

For participants on Prolific, we first ran a short screening survey to identify participants who met these criteria; respondents were paid $0.15 for a 1 minute survey ($9/hour). Qualified respondents were sent our main survey and compensated $2.85 ($10.05/hour). For respondents recruited by Lucid, we are not privy to Lucid's compensation structure; we paid Lucid $5.50 per survey completed.

When analyzing our results, we noticed there were two questions we did not ask to both groups but should have: one of these questions was left out entirely from the gig work survey, another was only asked to a subset of respondents in both groups but should have been asked to everyone. There was also one question where we did not include the same answer choices to both daters and gig workers. To strengthen our analysis, we decided to re-field these questions in September 2022 to the Prolific respondents from our prior survey. We re-fielded a total of three questions to both groups (included in the survey questionnaire in Appendix A.3). To incentivize our original respondents to complete the survey, we paid at a slightly higher rate than for the original survey: $1 for a 3 minute survey ($20/hr). We received responses from 140 daters (38% of original sample) and 217 gig workers (48% of original sample) over a two week period. The year gap between fielding these two surveys is both a benefit – repeated measures surveys often use a 1 year gap period to reduce the likelihood of participants recalling prior answers to related questions in a previous round of the survey [111] (and we intentionally did not remind participants of their prior participation) – and a detriment: participants' behavior may have changed between the original survey and the follow up period. We reran one question from the original panel in our re-fielding to both groups; the response to this question – which measured the prevalence of vetting behavior – did not change significantly (Daters: $p = 0.873$, $X^2 = 0.025$; Gig workers: $p = 0.948$, $X^2 = 0.004$) between the original data collection and the re-fielding period: Among our original sample 85.3% of daters and 79.8% of gig workers reported vetting; in our re-fielded sample 84.3% of daters and 79.3% of gig workers reported vetting. Regardless, for full transparency and clarity, we use a dagger ($^\dagger$) when reporting our results to indicate re-fielded data.

**Analysis.** We used a mixed methods approach to analyze our data. First, we used deductive thematic analysis [22] to analyze respondents' responses to the open-ended question, *"What does safety mean to you in the context of [online dating/in-person gig work?]"*. We used the harms identified in our literature review as the initial codebook. Using the codebook, one researcher independently coded all responses from the dating survey, and a different researcher independently coded all responses from the gig work survey. The two researchers then reviewed a random sample of 100 responses in the survey data they did not code and evaluated inter-rater reliability, achieving an average Cohen's Kappa of 0.656 (substantial) across codes.

We then conducted descriptive analyses with statistical comparisons to analyze respondents' use of protective safety behaviors. We measured the proportion of people in each sample who reported engaging in a particular behavior and the tools they used to do so. The online dating and gig work surveys used the same questions. However, for some questions, we phrased answer choices differently (or presented slightly different answer choices), to be more applicable to each context. In Tables 2—4 we group corresponding answer choices to facilitate comparisons between samples. In Appendix A.4 we explain the groupings, including the full text for all survey questions and responses. We compare

proportions between the two groups using $\chi^2$ tests, where $p \leq 0.05$ indicates a significant difference. We applied Bonferroni-Holm correction to the resulting p-values to reduce the Type I error rate.

For one survey question (*Screening Heuristic* in Table 2) some answer groups were relevant to just one sample. For example, "job pay rate" is only relevant in gig work; "availability of social media info" is only applicable in online dating platforms where users can link social media info to their app profile. Naturally, for these answers, we did not perform a comparison between groups.

### 3.3 Limitations

We carefully implemented safeguards in our research, but acknowledge its limitations. Our literature review may have missed work related to AMOIs that did not surface from our search terms. Therefore, there may be experiences and definitions of safety that are not represented in our results, affecting the validity of our taxonomy. We encourage future work examining additional relevant post-digital safety experiences and conceptions in AMOIs. The survey portion of our work faces limitations inherent to many survey studies, such as social desirability bias, under-reporting, and recall bias. To reduce the former we carefully worded questions to avoid suggesting there are right or wrong answers, instead asking respondents to answer the questions based on their personal experiences. To mitigate potential under-reporting and recall bias, we limited our survey to those who had engaged in AMOIs within the last two years, and frequently asked respondents to recall specific situations they may have encountered in the past. We also presented the questions strategically, making sure to ask all questions about a particular behavior together. While we purposefully recruited respondents to match the demographics of the US, our work might not capture the full range of age groups, cultural backgrounds, and types of interactions involved in AMOIs. Future work might consider expanding our work through interviews and co-design sessions with people who engage in AMOIs other than online dating and gig work.

## 4 TAXONOMY OF HARMS

For nearly all of those we surveyed (96.6% D (daters); 97.6% G (gig workers)), safety affects their decision to meet someone offline. Furthermore, 59.4% of daters and 51.9% of gig workers report having had an experience that made them feel unsafe while meeting someone offline. Thus, it is critical to consider the harms that make them feel unsafe and how they are manifested.

In this section, we make two contributions toward understanding harms in AMOIs. First, we synthesize findings from prior work to identify four AMOI harms: harm to physical, emotional or financial safety or harm to autonomy. Building on the tradition of threat modeling to understand security risks in software and technical systems [167, 175] and socio-technical systems [62, 66, 128, 169, 178], we describe not only the harms themselves but how these harms may be perpetrated by several actors: *platforms* that enable AMOIs; *Meets*, those who an individual intends to meet offline; and *scammers* and *aggressors* who pose as Meets to intentionally cause harm. We provide quotes from our respondents' definitions of safety to further describe each harm. Second, we measure the salience of these harms in online daters' and gig workers' definitions of safety in an effort to characterize people's priorities within the AMOI safety design space. For each harm, we report its salience in the appropriate subsection, and summarize the salience of all harms in Figure 1.

The harms we identify align closely with two existing taxonomies from other contexts: data privacy violations and interactions with offensive online content. In their taxonomy of privacy harms, Citron and Solove identified seven types of harms resulting from privacy violations, including physical, economic, psychological, and autonomy [42]. In line with Citron and Solove's conception of data privacy as a violation through which such threats occur, prior work and our participants call out data privacy violations as one mechanism through which these harms can occur. In their taxonomy

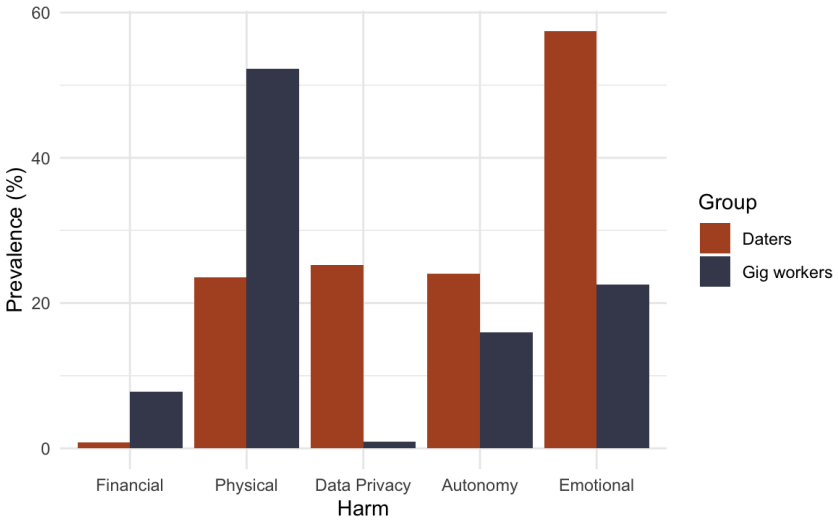


Fig. 1. Prevalence of the concerns our survey respondents reported in their definitions of safety

of harmful online content, Scheuerman et. al., similarly identified four types of harms resulting from interaction with offensive or undesirable content on social media: physical, emotional, relational, and financial [163]. Therefore, rather than redefining these categories, we reference these existing definitions and show how they manifest in AMOIs through prior work and quotes from our respondents' definitions of safety.

### 4.1 Physical harm

Physical harm has been defined as that which results in bodily injury [42, 163]. In AMOIs, physical harm includes assault and/or abuse (sexual and otherwise) [39, 70, 72, 114, 124, 154, 228], injury and/or death [2, 11, 25, 170], spread of disease [40, 48], and other forms of violence.

Physical harm may be (1) premeditated, (2) opportunistic, or (3) situational [179]. Premeditated physical harm arises when the Meet purposefully seeks to assault and uses digital platforms to find targets. Opportunistic physical harm occurs when the Meet becomes belligerent or violent, without initially intending to cause harm. This may occur if they become angry or upset, and compounded by other factors such as excessive alcohol consumption. An online dater well-described their concerns with these two forms of physical harm: *"For me, [safety] means not getting in a toxic (psychologically) relationship or a relationship where the woman can become physically dangerous in unconventional ways (ex. stalking, stealth attacking with a knife/gun). For others, it would be preventing physical/sexual violence, alcoholic coercion into sex, or abusive relationships" –D184.*

In the case of gig work, the nature of the work may also pose situational physical harm risks: for example, exposure to harsh cleaning chemicals [195, 198] or transporting packages on bikes [195]. Gig workers commonly described health and injury concerns arising from the spread of COVID-19. *"Safety means adhering to basic social distancing, sanitation, and mask protocols in order to minimize the spread of covid-19 for me and my client" –W199.* Physical harm was the most prevalent safety concern reported by gig workers (52.3%).

### 4.2 Financial harm

Financial harm in AMOIs is similar to both Citron and Solove's definition of economic harm [42] and Scheuerman et. al.'s definition of financial harm [163]. In particular, financial harm refers to monetary or material loss. Those engaged in AMOIs face financial harm from (1) scams that take place on the platform and (2) physical robbery. Scams on platforms include those tailored to the user's specific use case, such as scammers posing as a potential date or client [43, 134, 176, 195, 216]. As one respondent notes, *"A lot of online dating apps have a lot of scams and bots that will initiate a conversation with you and eventually try and get money. Or blackmail. Safety from these people [is] a must" –D225.* Platform scams may also exploit the vulnerabilities that lead the user to be engaging in AMOI such as financial precarity or curiosity and loneliness [150, 192, 216]. For example, prior work studying the risks of online dating notes: "a Tinder user wrote in their bio: 'Send me $5, see what happens' [and] manag[ed] to raise a significant amount of money" [176].

People may also be the targets of digitally-mediated physical robbery from actors they meet through the platform and others they interact with in the physical environment [11, 92, 164, 225]. Actors may leverage knowledge of how platforms work, or of how people use the platforms, to create harm. For example, actors may congregate in areas where they know gig platforms route their drivers [56]. One respondent in our study said, *"[Safety means] that I do not get harmed or robbed while being out and doing gig work" –W376.*

However, financial harm was not prominent in our respondents' safety definitions. Only 0.84% of daters and 7.76% of gig workers mentioned financial harm in their safety definition. This contrasts with the significant attention to financial scams in prior research on both groups, especially gig workers.

### 4.3 Autonomy harm

Citron and Solove define autonomy harms as those that restrict, undermine, inhibit, or unduly influence people's choices [42]. This includes coercion, manipulation, and limiting information such that an individual is not able to make choices freely.

In AMOIs, autonomy-related harm may stem from either the platform or Meet. Such harm can occur in either digital spaces (platforms and algorithms that control what users can and cannot do) or physical spaces (an individual may exert control over another by limiting access to their digital devices and accounts), or across an enmeshing of both. For example, platforms may limit individuals' autonomy through algorithmic management [9, 40] or deplatforming [1, 16, 18, 19, 24]. Algorithms may prioritize profiles based on features such as photo quality or the amount of personal information shared [194]. This pressures users to share more information than they would prefer, eliciting concerns around data privacy [113], and the increased amount of information a potential Meet will have access to [69]. On labor platforms, such management may also depend on clients' numeric ratings of workers; as a result workers feel pressured to maintain good reviews at the expense of safety [105, 225]. Workers often hesitate to stand up to or report a belligerent client for fear of receiving a low rating and reduced access to work opportunities [9, 112].

Individuals may limit another person's autonomy by controlling access to their devices, a common threat in intimate partner violence [37, 186]. In these situations an abuser may limit a target's access to resources (e.g., banking information), or try to prevent them from being able to document their experiences to report later [120, 186]. While this form of control usually occurs in digital space, the consequences can be experienced in the physical world [186].

Finally, autonomy may be harmed by lack of knowledge about the physical space of interaction. One commonly reported concern among our survey respondents was wanting to be familiar with locations where they meet an

individual so they can be self-sufficient in reacting to unsafe situations. For instance, several gig workers said that they want to know where the nearest emergency room is located so they can get help if they are injured on the job: *"Safety means I know where my working location is and where I can access emergency services close by" –W244.* Some online daters in our sample described wanting to feel that they can safely exit a space: *"It's important to always have an escape plan and ensure you don't get stuck" –D112.*

Interestingly, there is little to no prior work on daters' experiences with autonomy harms. We identified this harm (Section 4.3) based on literature studying various forms of gig work. Yet nearly a quarter (23.90%) of daters in our sample include autonomy harms when defining safety, compared to 16.0% of gig workers.

### 4.4 Emotional harm

Emotional harms refer to various negative mental responses, such as anxiety, fear, and worry. People who engage in AMOIs may suffer emotional harm as a result of hate and harassment, manipulation and deceit, and fears over the prospect of experiencing harm. In most cases these harms are caused by Meets. However, platforms may also cause emotional harm by pushing people towards behaviors and interactions that are harmful to their mental health or relational goals [26, 40, 198, 225], or by creating feelings of isolation, exploitation, or competition [75, 83, 112, 149, 164, 198, 225].

Emotional harm can occur even before a person engages with a Meet offline, via disparaging and disrespectful messages [117]. This form of online hate and harassment has offline effects; it may cause hesitation, distrust, and fear in future offline interactions even with a different person [186]. Similar harm occurs offline via insulting comments on physical appearance, or expressions of entitlement related to social class and gender [5, 126, 225]. One gig worker expressed their desire to engage with Meets who will be kind and supportive: *"[Safety means] Making sure that I, my client, anyone else around and the work space is clear, supportive, kind, loving, professional, and can work together to find a common goal" –W186.*

Meets may distort the truth via strategic manipulation and deceit, which can devastate the person believing the false reality [35]. Interacting with a Meet who has distorted the truth can lead a person to have unfavorable feelings towards future Meets [140]. For example, it may lead to self-other asymmetry – a bias where one believes that others are more likely to engage in deceptive behaviors than they would [165]. Among respondents, both gig workers and daters underlined the importance of avoiding deceitful Meets: *"Safety means making sure the person you are meeting is who they say they are, and can be trusted... If anything feels off, it's a sign to move on" –D343.* Similarly, just the prospect of experiencing harm can cause emotional harm [137, 217]. One gig worker explained, *"I'm scared of being set up and it being someone who just wants to hurt or assault me and not a job" –W196.* Emotional harm was the most prevalent safety concern reported by online daters (57.4%).

### 4.5 Data-privacy violations

Citron and Solove's harms taxonomy details the harms arising from *privacy violations* like misuse and accidental sharing of users' data [42]. People who engage in AMOIs are at risk of experiencing privacy violations: platforms that support AMOIs require the collection and distribution of personal details about users to effectively generate matches for romantic relationships, labor, and more. Thus, platforms may require users to share social media profiles, government identification, device permissions for location tracking, and other personal information as part of their profiles or the sign-up process [124, 161, 194].

| | Behavior | Citations | Mitigated Harms F | P | E | DP | A |
|---|---|---|---|---|---|---|---|
| Prior | Self-Disclosure | [34, 53, 61, 69, 129, 134, 194, 213–215] | ● | ● | ● | ● | ● |
| | Obfuscation | [1, 7, 16, 18, 19, 24, 43, 83, 91, 93, 101, 102, 129, 134, 148, 157, 161, 196, 197, 205, 208, 213–215, 230] | | ● | | ● | ● |
| | Screening | [4, 9, 48, 73, 76, 80, 91, 92, 124, 126, 131, 134, 145, 146, 195, 209, 229] | ● | ● | ● | ● | ● |
| | Vetting | [4, 7, 16, 43, 69, 80, 84, 91, 106, 112, 125, 132, 134, 141, 145, 161, 179, 180, 193, 195, 202, 216, 223, 224] | ● | ● | ● | | |
| During | Enviro. Precautions | [3–5, 7, 14, 60, 91, 92, 103, 110, 124, 130, 133, 134, 164, 185, 190, 221] | ● | ● | | | ● |
| | Covering | [30, 32, 48, 65, 91, 123, 124, 156, 164, 173, 182, 195] | | ● | | | |
| | Emergency Alerts | [33, 74, 97, 98, 115, 130, 195, 220] | | ● | | | |
| | Documentation | [7, 14, 82, 112, 124, 141, 145, 161, 162, 179, 180, 195, 209] | | ● | | | ● |
| Post | Blocking & Reporting | [7, 14, 38, 73, 82, 84, 112, 122, 124, 126, 141, 161, 162, 164, 166, 172, 177, 179, 180, 191, 193, 202, 207] | ● | ● | ● | ● | ● |

Table 1. Taxonomy of harms in AMOI—(F)inancial, (P)hysical, (E)motional, (D)ata (P)rivacy, and (A)utonomy—and the protective behaviors used to mitigate them. Citations are provided for each behavior.

Prior work and our empirical data show that people engaged in AMOIs are concerned about the misuse and abuse of personal information and the resulting harms [43, 91, 225]. Our survey respondents call out threats to data privacy as a mechanism through which harm can occur, and in some cases reference only threat to data privacy – encompassing the spread of resulting harms. Expressing this well, an online dater in our study defined safety as *"having enough information to learn about a person, but not enough to be able to locate and potentially interfere with someone's life unless they choose to specifically share that. Any personally identifying information that someone provides to a dating service should be very secure from intrusion" –D188.*

In our survey, few gig workers defined safety as related to data privacy (0.89%) compared to daters (26.2%). The lack of data privacy concerns among in-person gig-workers warrants further study, especially in light of prior work that has found significant data privacy concerns among crowdworkers: gig workers who perform primarily online work [158, 159, 161].

## 5 A TAXONOMY OF PROTECTIVE BEHAVIORS

In this section, we present ten protective behaviors, identified from prior work, that people who engage in AMOIs use to protect themselves from the harms described in Section 4. We systematize these behaviors by the phase of the interaction during which they are implemented: (1) prior to meeting, (2) during the meeting, and (3) post-meeting. *Prior to meeting* protective behaviors encompass all strategies that occur before meeting offline, including deciding whether to continue an interaction with a potential Meet and deciding to meet offline (Table 2). Methods for ensuring safety *during the meeting* – while meeting offline– can be set up beforehand, such as texting a friend about the meeting location or intended duration. They can also be triggered during the meeting, such as by using an emergency button on an alarm app or wearable device (Table 3). *Post-meeting* safety strategies occur after an offline meeting has concluded; these include blocking and reporting (Table 4). For each behavior, we also identify from the literature which harms from Section 4 users aim to protect against. Table 1 and Figure 2 summarize the results of our systematization. Table 1 focuses on the relationship between behaviors and the harms in Section 4. Figure 2 focuses on the temporal structure of how behaviors occur. Figure 2 is further used to discuss future work in Section 6.3.

We use our survey data to extend the literature by measuring the prevalence of respondents' adoption of the ten behaviors (Figure 3) and filling in gaps of knowledge regarding how they are implemented. Tables 2, 3, and 4 summarize these prevalence proportions for behaviors prior to meeting, during meeting, and post meeting, respectively.

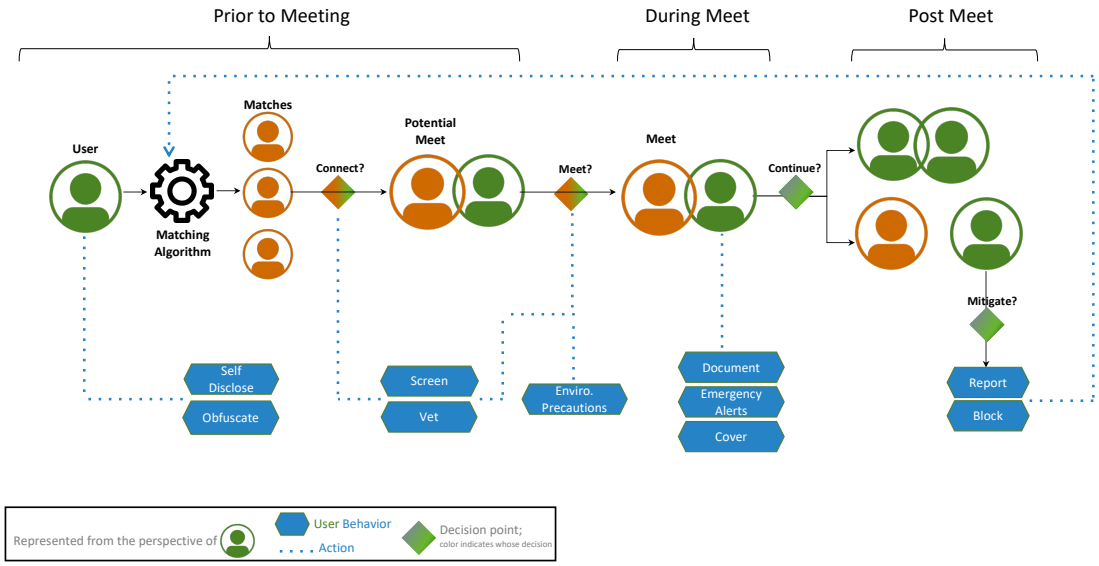


Fig. 2. Process flow diagram showing—in finer detail—when the behaviors of Table 1 occur during an AMOI. Note that the temporal structure of behaviors is complex: multiple behaviors can happen simultaneously, and some behaviors happen at multiple times.

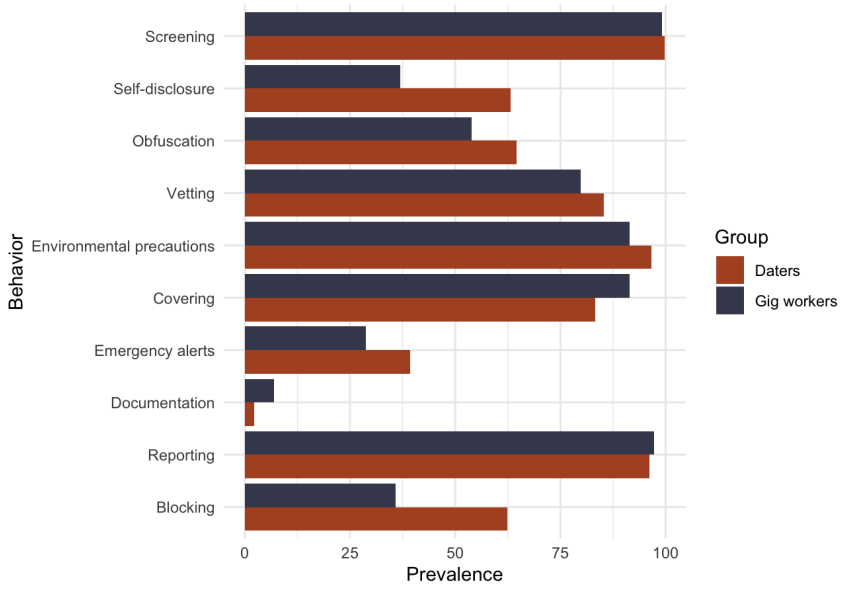


Fig. 3. Prevalence of safety behaviors our respondents engage in to mitigate the harms described in Section 4

.

### 5.1 Protective Behaviors Prior to Meeting

Prior to meeting in-person, people use a variety of protective behaviors in an effort to ensure they are safe when ultimately engaging in offline interactions. The goal of prior to meeting protective behaviors is to gauge the safety of a future offline interaction to prevent harm upfront. These behaviors include *self-disclosure* (intentionally revealing information

about themselves in hopes that the other person will use it to screen them); *obfuscation* (hiding or misrepresenting information about themselves); *screening* (using a variety of decision heuristics to evaluate the information presented by a digital platform about the person they are considering meeting); and *vetting* (seeking out additional information about a potential Meet to verify information obtained directly from the platform).

*5.1.1 Self-Disclosure.* Self-disclosure of personal information both serves to facilitate matching on AMOI platforms and as a protective safety behavior to protect against all harms in Section 4. The algorithms that curate matches in AMOIs rely on user-provided information to facilitate matches. The information people self-disclose determines who the matching platform or marketplace suggests they connect with, and ultimately who they interact with offline. Therefore, people purposefully choose to disclose aspects of their identity for the purpose of obtaining the most compatible matches and to make themselves more attractive to a potential Meet [134, 194]. For example, online daters might self-disclose information that would make them more attractive to a match (e.g., hobbies, physical attributes, etc.); gig workers might self-disclose information that would increase their likelihood of getting hired, which could relate to the type(s) of job they seek (e.g., self-disclosing the number of children they have on a carework site).

Self-disclosure as a protective safety behavior typically occurs in the initial stages of an interaction, when individuals are still interacting with the potential Meet strictly online; people may self-disclose information they believe may put them at risk of harm if it were discovered offline (e.g., LGBTQIA+ identity or race/ethnicity) [34, 53, 69]. They hope that by sharing this information, Meets will only agree to engage with them further if they are accepting of their identity [213]. However, self-disclosure can also be a continuously developing process; people might start to reveal more about themselves once a certain level of trust has been established [15, 153].

Individuals must balance self-disclosure with privacy and control over their personal information. For example, some people share their HIV status on social networking and dating apps [213–215]. However, they express privacy concerns around platforms having access to this sensitive information [213]. Additionally, self-disclosing can be emotionally taxing: individuals who share sensitive information may face stigma in their community [61, 129, 213]. We found no literature that describes self-disclosure behavior among gig workers, however our survey data suggests a notable proportion of gig workers who engage in AMOIs do so as detailed below.

*Self-Disclosure Prevalence.* 63.2% of daters and 36.9%$^{\dagger}$ of gig workers ($p < 0.001$) said they self-disclose information prior to meeting someone offline for safety reasons. People use three methods to self disclose (Table 2): in their app profile, in an online or text conversation with a potential Meet, or during the first introduction meeting offline. The most commonly reported method by which online daters self-disclose is within private online communication (e.g., online messaging and texting) (40.4%). Among gig workers, the most commonly reported method by which they self-disclose is in an online profile within the app (22.6%$^{\dagger}$). Overall, fewer gig workers report self-disclosing than online daters. This may be because most gig platforms offer workers little control over their profiles. The 36.9% of workers who report self-disclosing might be using one of the few platforms with customizable profiles (e.g., some carework and handiwork platforms). The differences between where participants self-disclose might be explained by the different digital communication tools used by each group to interact with Meets. Gig workers typically communicate through the app where they are hired, either because the platform requires it, or to mitigate financial harm from scams [94]. Online daters may be more likely to share phone numbers and social media information with a Meet, thus facilitating their ability to self-disclose via online messaging and texting.

| Question | Response | Daters (%) | Gig Workers (%) | p-value | Significance |
|---|---|---|---|---|---|
| Where Self-Disclose | In online profile within app | 35.1 | 22.6† | 0.084 | |
| | In text or online messaging | 40.4 | 15.2† | < 0.001 | *** |
| | During an offline meeting | 23.9 | 12.9† | 0.080 | |
| Where Obfuscate | In app (omitting) | 42.7 | 28.8 | 0.001 | ** |
| | In app (misrepresenting) | 17.4 | 12.0 | 1.000 | |
| | From text or online messaging conversation (ommitting) | 34.5 | 31.0 | 1.000 | |
| | From text or online messaging conversation (misrepresenting) | 20.2 | 14.9 | 1.000 | |
| | During offline meeting (omitting) | 23.4 | 22.0 | 1.000 | |
| | During offline meeting (misrepresenting) | 11.6 | 13.5 | 1.000 | |
| Screening Heuristic | Specific phrases in profile | 25.5 | | | |
| | Meet's location or hometown | 57.1 | | | |
| | SES of job location | | 39.2 | | |
| | Job pay rate | | 71.2 | | |
| | Similarity of job to others done in past | | 46.1 | | |
| | Availability of profile photo | 95.2 | 77.8 | < 0.001 | *** |
| | App profile is not blank | 83.2 | 63.0 | < 0.001 | *** |
| | Enough info in profile for online search | 34.3 | 52.1 | < 0.001 | *** |
| | Availability of social media info | 35.6 | | | |
| How Vet | Online: Search engine | 72.1 | 88.0 | < 0.001 | *** |
| | Online: Social media lookup | 86.7 | 80.2 | 1.000 | |
| | Online: Background check/court records search | 21.2 | 23.1 | 1.000 | |
| | Socially: Online whisper networks | 10.9 | 21.2 | 0.009 | ** |
| | Socially: Offline support networks | 40.2 | 38.4 | 1.000 | |
| | Directly: Ask for PII directly | 20.7 | 13.4 | 0.598 | |
| What Vet | Personal info | 72.0† | 51.7† | 0.052 | |
| | Additional media | 95.8† | 65.1† | < 0.001 | *** |
| | Location | 46.6† | 55.8† | 1.000 | |
| | Reputation | 28.0† | 75.0† | < 0.001 | *** |
| | Personality | 51.7† | 40.7† | 1.000 | |

Table 2. Protective behaviors used **prior to meeting**. Proportions shown are out of the total number of people who answered the question. Blanks indicate that proportions for those responses were not measured because those responses were not relevant for those groups. Proportions across the two groups are compared with $X^2$ tests corrected with the Holm–Bonferroni method. Zero * indicates no significant difference, one * indicates $p < 0.05$, two ** indicates $p < 0.01$, three *** indicates $p < 0.001$. † indicates refielded data (see Section 3).

*5.1.2 Obfuscation.* In contrast to self-disclosure, some people may wish to obfuscate personal information by either omitting [134] or misrepresenting [148, 197] parts of their identity. Related literature on privacy-preserving strategies among social media users suggests that obfuscating is used to protect against physical, autonomy, and data privacy harms [101, 205].

**Omission.** When using omission as an obfuscation tactic, people omit information they believe may cause an unsafe encounter with a potential Meet, such as religion, job, and sexual preferences [43]. Specific strategies used to omit information include removing online profiles [83, 102] and censoring, or hiding, images and personal information that they do not want accessed by a potential malicious actor. For example, people sometimes use emojis and specific in-group language in text communication or profiles [7, 129, 213–215] to hide information from potentially malicious actors while sharing it with people who they do want to access it.

**Misrepresentation.** When using misrepresentation as an obfuscation strategy, people present inaccurate information about themselves to Meets. Some may simply provide a different name, or lie about how they look and their qualifications for a job, sometimes referred to as privacy lies [157, 161]. Others use more involved impression management strategies such as maintaining differing online personas across multiple platform profiles with different social media information, photos, and phone numbers or devices [230].

Obfuscating comes at a cost: potential deplatforming if obfuscation violates platform rules. Deplatforming leads to loss of income in the case of digitally-mediated labor [1, 16, 18, 19, 24]. Therefore, individuals must weigh the financial costs they may incur from obfuscating information against the safety protections it could provide.

*Obfuscation Prevalence.* Rates of obfuscation among our samples are similar to their respective rates of self disclosure. 64.6% of online daters and 53.9% of gig workers ($p = 0.069$) obfuscate by either omitting (60.0% D; 50.3% G) or misrepresenting (34.0% D; 28.4% G) parts of their identity. Table 2 shows how respondents obfuscate, which we summarize below.

The most common place where daters omit information is in their profile within the app (42.7%). Significantly fewer gig workers report doing this (28.8%), perhaps because most gig platforms offer workers little control over their profiles, and opaque matching algorithms play a much larger role in determining which clients workers are visible to [93, 208]. Because of this, gig workers may be less able to decide what information they want to show or hide to others. Instead, gig workers are more likely to omit information in an online or text conversation with a potential Meet where they do have greater control over the sharing of their information (31%).

On the other hand, relatively few daters or gig workers misrepresent themselves in their profiles (17.4% D; 12.0% G). Instead, among both groups, the most common place where individuals present inaccurate information is in an online text or conversation (20.2% D; 14.9% G).

*5.1.3 Screening.* While deciding whether they want to meet with the person presented in a platform profile, people use certain heuristics, which we term decision heuristics, to evaluate, or screen, the potential Meet's profile [4, 92, 126, 134, 145, 146, 195, 229]. Screening is done using information the potential Meet has decided to self-disclose via the platform and that the platform has chosen to display; when screening, individuals are not actively seeking out information about a potential Meet but instead are evaluating the information presented to them via a series of heuristics. Screening is used to avoid entering into situations that can cause any of the harms described in Section 4. For example, a person might want to avoid someone who seems likely to cause emotional harm through racial or religious discrimination. Others may look for "red flags" that indicate potential for harassing, discriminatory or other unwanted behavior that could cause emotional or physical harm [4, 73, 145]. Prior work identifies three primary heuristics used to screen Meets: indicators of ethical alignment, informal lessons, and information availability. We describe each of those below.

**Ethical Alignment.** The heuristics most commonly reported in prior work are those used to screen for ethical alignment. Specifically, prior work suggests that in digital environments, emotional safety in particular is maintained by an iterative process of boundary regulation where lines "are drawn in relation to a shared set of affective and relational knowledge" [146]. Interactions are more likely to be perceived as emotionally beneficial when a connection is made by a party who respects and maintains established boundaries. Thus, people may look for details that indicate ethical alignment between themselves and a potential meet such as the presence of reassuring information and actions (e.g., use of the gender-inclusive term "cis-man" in a profile) [57, 142, 213]; the absence of concerning information or actions (e.g., no weapons or concerning objects in profile photos) [145, 146]; or perceived reciprocity and openness of the Meet regarding the sharing of personal information [209].

**Informal Lessons**. Additional prior work identifies that people draw on their previous experiences with similar AMOIs, or on the experiences of those they know, to determine whether an offline engagement with a Meet will be safe. This is similar to how non-expert computer users leverage family and friends' stories of experiences with security incidents to make security decisions [147]. This type of screening heuristic is context dependent. For example, gig workers often decide whether to engage with a Meet offline if the job pays similarly to other jobs they or other workers

have done in the past; a pay rate that is too much higher may indicate the job is a scam [131]. Similarly, in online dating, a potential Meet that lives overseas might indicate a catfishing attempt [49].

**Information Availability.** A small body of prior work reports that people use the information provided in a profile as a heuristic for safety and trustworthiness. People assess both the content provided in a profile, as well as the quantity of information provided. Prior work finds that in the context of identifying deepfake profiles on social media, users search for text inconsistencies in profile information such as repetitions, grammar mistakes, and contradictions [127, 192]. Other work finds that people look for the availability of specific information, such as a profile photo [48], as a heuristic. Not having this information decreases people's trust in the potential Meet's profile and can raise phishing concerns [168]. When platforms do not provide enough information about a potential Meet for people to screen them (as is often the case on gig work platforms [225]), individuals engage in vetting to seek out additional information (Section 5.1.4). Thus, prior work finds that people also use as a screening heuristic the presence of *enough* information on an individual's profile to vet them [145].

*Screening Prevalence.* Screening is a nearly universal behavior: nearly all respondents in both groups surveyed report screening Meets (99.8% D; 99.1% G; $p = 1.00$) using one or more of the three heuristics. Table 2 details the precise implementation of these heuristics, which we summarize below.

Ethical alignment factors were frequently reported as screening heuristics in prior work on online dating and our empirical data confirms this. For example, 25.5% of online daters report screening potential Meets based on specific phrases or words included in their profile. Because gig platforms are typically limited in the amount of information they share with workers about potential clients, we did not ask questions regarding ethical alignment factors to gig workers. Instead, gig workers prevalently report using informal learning as a screening heuristic. For example, 71.2% of gig workers report using the job's pay rate as a screening heuristic; 46.1% report using the similarity of the job to others they have done in the past.

Both gig workers and online daters use similar information availability heuristics. 83.2% of daters and 63.0% of gig workers look for a non-blank profile. Many people in both groups look specifically for the presence of a potential Meet's profile photo in the app, although daters do this significantly more than gig workers (95.2% D; 77.8% G). On the other hand, a significantly greater percentage of gig workers than online daters use whether there is enough information in the potential Meet's profile for them to do their own online search as a screening heuristic (34.3% D; 52.1% G). As mentioned above, online dating platforms typically provide more information about potential Meets to users of their app over gig work platforms, potentially reducing the relevance of this heuristic for those online dating.

*5.1.4 Vetting.* Individuals often want information about a potential Meet beyond that presented by the platform to support their screening process. Platforms may not provide enough information about a potential Meet for people to screen them passively. Therefore, people take action to seek out the information they need to trust a potential Meet sufficiently to proceed with the interaction. We call this practice of actively and externally seeking additional information beyond that available on the platform vetting.

Vetting is used to try to prevent financial, physical, and emotional harm [43, 84]. People use three types of strategies to vet: they seek information online, beyond the platform(s) they use for AMOIs; they draw on their social capital, consulting their social support networks (e.g., friends and family) and online communities sometimes termed *whisper networks*; and/or directly ask the potential Meet for information.

**Online information seeking.** When desired information (e.g., someone's full name, whether they have a criminal record) is not available for screening directly on the AMOI platform, people seek information online. Even when

screening information is available, people are aware that others may engage in obfuscation and thus may want to corroborate the information a platform presents about a potential meet to evaluate the validity of the information a Meet has presented on the platform and look for inconsistencies that could signal potential harm [4, 91, 134, 145, 224]. The same reasons that lead people to vetting pose challenges in carrying out the behavior. In particular, enough information to successfully search online is not always available; AMOI platforms may not require users to provide their full legal name on their profile or prevent individuals from displaying fake information. Therefore, people may lack enough information from the AMOI platform to conduct a useful search.

**Social Information Seeking.** Some people vet socially, particularly when seeking relational and reputational information that does not appear as the results of an internet, social media, or court records search, such as the Meet's reputation within similar communities (e.g., among other workers or daters).

Online, some people will consult *whisper networks* for information about a person. Whisper networks are informal online feedback networks that people use to communicate bad experiences [141]. They reached social prominence in the wake of #MeToo when many informal networks were created to communicate bad experiences with men, both after dates and more generally [84, 141, 202]. These networks may exist in the form of "bad date" or "bad client" lists. They lists are commonly built and maintained by sex workers as an alternative to the criminal justice system [179], but have also been prominent in the broader gig worker and online dating communities as well; in the latter two communities, bad date lists typically exist in spaces like online forums, where people share feedback and report negative experiences about those they have interacted with [7, 16, 161, 179, 180, 193]. Within online whisper networks, people may search for details about a potential Meet or directly ask others in the community whether they have had any negative experiences with the person [161, 195, 216].

Beyond whisper networks, people may also vet via close contacts, like friends and family members. For example, prior work has found that online daters may ask their friends for help screening a potential Meet's profile [117] or may ask friends engaged in similar activities (dating, gig work) whether they have experience with a potential Meet.

**Direct Inquiry.** Finally, a third strategy for finding information about a potential Meet is directly asking the person for the desired information, and explaining what the information will be used for [43, 134]. However, this behavior can be seen as intrusive and violating the privacy of the potential Meet [43].

*5.1.5 Vetting Prevalence.* More than three quarters of our survey respondents in both groups engage in vetting (85.3% D; 79.8% G; $p = 1.00$). Table 2 describes how respondents vet and what information they seek while vetting. We describe each of these in detail below.

Online information seeking behaviors include searching the internet (72.1% D; 88% G), social networking sites (86.7% D; 80.2% G), and background check services and court records (21.2% D; 23.1% G) to obtain information about a potential Meet.

Social information seeking behaviors include searching within online whisper networks and talking to social support networks (typically offline). Significantly more gig workers than daters rely on online whisper networks when vetting (10.9% D; 21.2% G). This may be explained by gig workers' common practice of collective algorithmic sensemaking, or self-organizing to collectively make sense of the platform they work on and its underlying algorithm [225]. Similar proportions of online daters and gig workers report seeking vetting support from social support networks like friends and family (40.2% D; 38.4% G).

Daters are more likely to ask a potential Meet for additional information directly than gig workers (20.7% D; 13.4% G). Yet, we find that the majority of respondents *rarely* or *never* tell those they vet that they have done so (65.0% D; 82.5% G; $p < 0.01$).

When searching for additional information about a potential Meet, online daters most commonly try to find additional media (e.g., videos, photos) about the other person (95.8%). Gig workers most commonly search for the potential Meet's reputation (75%), such as their criminal history and reputation among other workers in online whisper networks. Other types of information sought are personal information (e.g., the potential Meet's full name or phone number) (72.0% D; 51.7% G) and information about the Meet's location (e.g., home and work location) (46.6% D; 55.8% G). People may also seek information they could use to infer personality traits, like political affiliations (51.7% D; 40.7% G).

When people were unable to vet a potential Meet, they chose not to continue with the introduction in around one-quarter of interactions. 25.6% of daters and 22% of gig workers would *rarely* or *never* engage with an individual offline who they were unable to vet.

## 5.2 During-Meet Protective Behaviors

When people interact with Meets offline, they engage in various digital and non-digital behaviors to protect their safety. These behaviors include environmental precautions (using techniques to try to protect their autonomy), covering (sharing details of a meeting with trusted friend or family member), emergency alerts (using technology to send a distress signal or call for help), and surveillance/documentation (using technology to record interactions).

*5.2.1 Environmental Precautions.* When interacting with a Meet offline for the first time, people take a variety of precautions to make the interaction and environment safer [134, 190]. These precautions are typically not digitally-mediated. Instead, they include behaviors that people believe will help them stay in control of the meeting, thus avoiding physical, financial, and autonomy-related harm. Environmental precautions include bringing protective items to the meeting [7, 14, 124, 130], engaging in advanced planning (e.g., having a planned escape route, getting their own ride to the meeting location) [103], selectively choosing the meeting time and location [3, 7, 92, 134, 185], not going to the meeting alone [91, 164], and avoiding certain personal behaviors that may increase their risk of harm (e.g., drinking alcohol) [133].

When taking precautions to protect their autonomy, some people will carry items they feel give them greater control, such as lethal (e.g., firearms) and non-lethal (e.g., pepper spray and pocket knives) weapons [7, 14], even in situations where such items are prohibited or illegal [130]. The items people believe will protect them are dependent on the type of interaction they will engage in with a Meet. For example, sex workers and online daters may carry contraceptives, such as condoms, to protect themselves from health-related physical harm like STDs [124]. Others might try to protect their autonomy by engaging in advanced planning, such as using online mapping tools to examine the meeting location and identify a pre-determined escape route to quickly leave in an emergency or driving themselves to the meeting location.

When meeting someone for the first time, people try to meet them in locations that feel safe, such as in crowded, well-lit areas; they often meet in a public space, such as a coffee shop or busy park [3, 134], and avoid late meetings [92] or those in locations they deem unsafe, such as those in socioeconomically-disadvantaged areas [185]. Sometimes people will bring a trusted individual to the meeting [91, 164]. Some may also avoid having in-home meetings [5]. Some of these strategies may be more appropriate in some contexts than others. For example, some online daters note that meeting in public places is "useless" for those who are seeking hook-ups or casual sex from dating apps [4]. Gig workers

| Question | Response | Daters (%) | Gig Workers (%) | p-value | Significance |
|---|---|---|---|---|---|
| Environmental precautions: How | Bringing protective items to meeting | 27.7 | 30.4 | 1.000 | |
| | Advanced planning | 59.2 | 54.8 | < 0.001 | *** |
| | Selectively choosing time and location | 93.4 | 82.9 | < 0.001 | *** |
| | Not going alone | 29.4 | 39.0 | 0.159 | |
| | Personal behavior changes | 29.0 | 30.6 | 1.000 | |
| Covering: How | Sharing location details with guardian | 74.5 | 84.5 | 0.015 | * |
| | Sharing Meet details with guardian | 37.1 | 48.6 | 0.036 | * |
| | Sharing expected time back with guardian | 48.7 | 63.9 | < 0.001 | *** |
| | Live-sharing location with guardian | 30.0 | 37.9 | 0.811 | |
| | Asking guardian to check in | 35.0 | 30.4 | 1.000 | |
| Covering: What plans | Guardian will come get me | 64.4 | 56.5 | 1.000 | |
| | Guardian will contact police | 44.9 | 62.5 | 0.026 | * |
| | Guardian will contact a safety organization | 1.0 | 0.5 | 1.000 | |
| Emergency Alerts: What | A distress text or silent call/alarm | 12.3 | 6.4 | 0.205 | |
| | GPS coordinates | 14.2 | 17.7 | 1.000 | |
| | Audio/video recording | 5.5 | 4.7 | 1.000 | |
| | Details about the Meet | 14.6 | 10.0 | 1.000 | |
| | Audible Alarm | 4.0 | 2.0 | 1.000 | |
| | Fake call (from someone or an app) | 27.0 | 12.9 | < 0.001 | *** |

Table 3. Protective behaviors used **during the meeting**. Proportions shown are out of the total number of people who answered the question. These proportions are compared with $X^2$ tests corrected with the Holm–Bonferroni method. Zero * indicates no significant difference, one * indicates p < 0.05, two ** indicates p < 0.01, three *** indicates p < 0.001.

and sex workers often do not have a choice to meet in a public location, since their jobs require private interactions, often in people's homes or private spaces [124].

People sometimes avoid certain personal behaviors such as wearing makeup, drinking alcohol, and wearing headphones [110]. These are behaviors that individuals may typically engage in outside the context of AMOIs. But in this context, they believe engaging in the behavior may put them at greater risk of harm. For example, daters may choose to not drink alcohol to ensure they stay in control of the meeting. Gig workers are typically not in a context where drinking is part of their interaction with a Meet. Instead, they may carefully calibrate their gendered presentation, including avoiding wearing makeup and/or dress in a particular way as they may perceive these behaviors as protective. These behaviors may be influenced by victim-blaming culture [60, 221].

*5.2.2 Environmental Precautions Prevalence.* Nearly all respondents among both daters and gig workers report engaging in environmental precautions (96.6% D; 91.4% G; $p < 0.075$). Table 3 shows the methods by which respondents engage in environmental precautions. The most commonly reported environmental precaution among both groups is selectively choosing the meeting time and/or location (93.4% D; 82.9% G). The second most commonly reported environmental precaution among both groups is engaging in advanced planning (59.2% D; 54.8% G).

*5.2.3 Covering.* While interacting with Meets offline, people use a protective strategy termed "covering" to protect against physical harm [32, 124]. People cover by (1) having another person present (or close by) during a meeting [9, 80, 91, 123, 124, 164], or (2) sharing details about the Meet and meeting location with others [48, 65, 124]. In most cases people share this information with a trusted, close contact who can serve as a proactive bystander, or "guardian" [6]. Others may rely on individuals in the vicinity where the interaction takes place to serve as guardians. For example, in indoor settings, receptionists (and security cameras which they may be monitoring) add to the safety of sex workers. In the UK it was found that sex workers may ask for help from them with dangerous clients, or they may be points of contact and support after unsafe incidents [156]. Furthermore, in Canada, "third parties" involved in sex work such as venue owners and managers were found to often be current or former sex workers themselves, and ultimately were

important individuals in the safety ecosystem by providing client screening, additional security, and sexual health resources [123].

People often share information with guardians, such as where they are going (including live location sharing [7, 161]), personal details about the Meet, what time they are expected to be back, and instructions for the guardian check in periodically [124, 182, 195]. Sometimes people may also ask their guardian to contact the police if they suspect something went wrong. However, in gig work literature, particularly work that relates to criminalized and stigmatized communities, it is often noted that people may be hesitant to contact police in emergency situations. This is especially true when they believe the police may blame them or criminalize their work [30, 124, 164, 173].

*5.2.4 Covering Prevalence.* 83.3% of online daters and 91.4% of gig workers surveyed engage in covering ($p = 0.026$). The details people share with trusted contacts (Table 3) include where they are going (74.5% D; 84.5% G) or live-sharing their location (30.0% D; 37.9% G); personal details about the Meet (37.1% D; 48.6% G); or what time they expect to be back (48.7% D; 63.9% G). 35.0% of online daters and 30.4% of gig workers give their guardian instructions to check in on them during an offline meeting.

39.3% of daters and 28.8% of gig workers ($p = 1.00$) said they make emergency plans with their trusted contacts in case the guardian suspects something went wrong during the meeting. The most prevalent emergency plan made by daters is for their guardian to meet them at the meeting location (64.4%); among gig workers the most prevalent emergency plan is for the guardian to contact the police (62.5%). The difference between the percentage of gig workers and online daters who instruct their guardian to contact the police on their behalf is significant and adds nuance to prior work suggesting people may be hesitant to instruct guardians to contact law enforcement.

*5.2.5 Emergency Alerts.* When people feel their physical safety is threatened during an offline interaction they may send emergency alerts to trusted contacts or law enforcement. There are two common types of emergency alerts used in unsafe situations, both in AMOIs and more broadly: (1) sending information to a guardian [6, 195], and (2) playing an audible alarm [74, 97, 130, 220]. Unlike covering, where individuals send information to a guardian as a *proactive* measure, sending information to a guardian via an emergency alert is a *reactive* measure; it is done once harm has occurred (or appears imminent) to enable the guardian to intervene in the situation. For example, sending information can support a future investigation. Similarly, playing an alarm can distract the perpetrator or attract help.

Several apps and wearable technologies exist to enable people to send emergency alerts during unsafe situations. For example, some apps give the user a fake call [115], contain "panic buttons" that call local law enforcement when triggered [98], and allow users to quickly share their GPS location with a list of pre-determined contacts [33]. However, prior work has not measured how prominently these safety apps feature in people's protective model.

*5.2.6 Emergency Alerts Prevalence.* 39.3% of online daters and 28.8% of gig workers report sending emergency alerts to trusted contacts or law enforcement ($p = 0.066$). Among our survey respondents, people most often use emergency alerts to share information with trusted contacts such as personal details about the Meet (14.6% D; 10.0% G) or GPS coordinates/other location details (14.2% D; 17.7% G) (see Table 3). People also use emergency alerts to send a silent distress alert (12.3% D; 6.4%) or to trigger a fake phone call that can provide an excuse to leave the situation (27.0% D; 12.9% G). Only rarely do people start recording in an emergency situation (5.5% D; 4.7% G) or play an audible alarm (4.0% D; 2.0% G).

To extend the prior work on safety apps and wearable technologies, we asked respondents about their familiarity with eight emergency alert technologies (Appendix Table 6), selected through an extensive search and our collective

knowledge of the area. We find that only a minority of survey respondents in our sample know about and use such apps. 28.6% of daters and 37.3% of gig workers have heard of at least one of the technologies. Of those, only 12.7% of daters and 15.5% of gig workers reported feeling that they knew enough about at least one technology to explain what it does, and only 4.57% of daters and 2.66% of gig workers personally use at least one emergency alert technology.

*5.2.7 Documentation.* People sometimes record elements of their offline interactions with a Meet to protect against physical harm and gain more autonomy and control over the interaction [180]. For example, people may proactively (vs. emergency-alert recording which begins once someone starts to feel unsafe during a meeting) record conversations and the Meet's behaviors [7] visibly, or without the Meet's knowledge [14, 195]. People who visibly record hope the presence of a recording device might deter a Meet's harmful behavior [7, 195]. For example, some gig workers working for rideshare apps commonly use dashboard cameras to video record passengers during the drive [7]. People may choose to record an interaction without the Meet's knowledge to obtain evidence for later reporting [7, 161, 195], especially since many platforms fail to provide adequate documentation and reporting mechanisms [112].

*5.2.8 Documentation Prevalence.* Interestingly, despite recording interactions being commonly reported strategies in prior work [7, 112, 161, 209], we found that only 2.3% of daters and 6.9% of gig workers reported using a recording device ($p < 0.097$). Perhaps one of the reasons people may hesitate to record interactions is the high level of legal risk involved, especially if recording without the Meet's consent. In many locations, recording interactions without consent is illegal. This may lead to deplatforming and legal charges [161].

### 5.3 Post-Meet Protective Behaviors

After an unsafe offline interaction with a Meet, people may discuss this experience with others or warn others about their experiences with the individual, report their experiences to platforms, hotlines, and law enforcement, and/or take steps to prevent the Meet from contacting them again [82, 124, 141, 162, 179]. It is important to note that these methods cannot rectify the harms an individual has experienced; rather they work toward preventing similar experiences from happening to the same or a different person in the future.

*5.3.1 Reporting and Blocking.* Prior work finds that people report harmful or unpleasant experiences with a Meet to the platform that facilitated the interaction [7, 112] or safety organizations such as NGOs that provide helplines for sexual assault [180]. They also share negative experiences with many of the same online and offline whisper networks from whom they seek advice and information when vetting [7, 14, 124, 126, 164, 179, 180]. While prior work does find that people will instruct guardians to contact the police if they suspect something went wrong during an offline meeting (see Section 5.2.3), we did not find prior work suggesting people report negative experiences to law enforcement *themselves*. It is important to note that reporting to law enforcement, and similarly platforms, requires that individuals have trust in those groups. Prior work does discuss why some people who engage in AMOIs, such as those from marginalized groups, may not trust these institutions to support them [124, 161, 164]. For example, they may have experienced previous violence from these institutions and individuals, expect discriminatory and stigmatizing treatment, or believe these institutions will not support them at all [112, 122, 172, 179, 207].

People have different goals when reporting to different institutions: they report to platforms to try to seek retribution for the offending Meet [7, 112]; they report to whisper and social support networks to help others avoid similar harm and to share safety information not provided by the platform [14, 82, 84, 141, 162, 202]. Sometimes people also share their negative experiences with online and offline communities to take care of themselves following tech-related harm;

| Question | Response | Daters (%) | Gig Workers (%) | p-value | Significance |
|---|---|---|---|---|---|
| Reporting: To Whom | Platform | 29.2 | 49.3 | < 0.001 | *** |
| | Police | 9.7 | 10.3 | 1.000 | |
| | Safety organization(s) | 6.2 | 8.5 | 1.000 | |
| | Social support network | 88.4 | 80.9 | 0.131 | |
| | Online whisper network | 12.1† | 40.6† | 1.000 | |
| How Block | On app | 53.4 | 27.3 | < 0.001 | *** |
| | On social media | 44.5 | 19.1 | < 0.001 | *** |
| | On phone/messaging app | 42.6 | 25.5 | < 0.001 | *** |

Table 4. Protective behaviors used **post meeting**. Proportions shown are out of the total number of people who answered the question. These proportions are compared with $X^2$ tests corrected with the Holm–Bonferroni method. Zero * indicates no significant difference, one * indicates p < 0.05, two ** indicates p < 0.01, three *** indicates p < 0.001. † indicates refielded data (see Section 3).

they find that sharing provides a cathartic emotional release and supports their emotional well-being as they navigate post-harm consequences [38, 191, 193].

Finally, prior work finds that in addition to reporting, people may also block the offending party from contacting them again via the platform and/or personal devices [73, 166, 177].

*5.3.2 Reporting and Blocking Prevalence.* Overall, among the respondents in our survey, 96.1% of daters and 97.7%[1] of gig workers have reported a Meet ($p = 1.00$). People are most likely to describe these experiences to social support networks (88.4% D; 80.9% G). Some also report their negative experiences to online whisper networks [7, 14, 124, 126, 164, 179, 180] (12.1% D; 40.6% G). The percentage of respondents who report to online whisper networks is relatively low compared to the percentage who seek vetting support from these same networks (40.2% D; 38.4% G; see Section 5.1.4). One explanation may be that while these groups can be very helpful in making decisions about whether or not an offline meeting will be safe, there can exist toxicity within them, especially when someone reports a negative experience, like falling prey to a scam [216]. This may make people hesitant to report their negative experiences. More than a quarter of daters and nearly half of gig workers report harmful or unpleasant experiences to platforms [7, 112] (29.2% D; 49.3% G). Few report to safety organizations [180] (6.3% D; 8.5% G) or law enforcement (9.7% D; 10.3% G). The difference between the proportion of our sample who instruct trusted individuals to contact police on their behalf if something goes wrong (44.9% D; 62.5% G; see Section 5.2.3) vs. the proportion who report to law enforcement after an incident has occurred is stark. It may be that people mistrust police and other law enforcement agents, but feel reassured when someone else can contact those authorities on their behalf. Or perhaps in a situation of immediate danger during a meeting, people see no other recourse vs. after the harm has already occurred.

Finally, 62.4% of online daters and 35.9% of gig workers block the offending Meet from contacting them again ($p < 0.001$). Table 4 contains the details of how blocking is implemented. Both groups most commonly block Meets on the app itself (53.4% D 27.3% G), although the percentage of online daters who report doing this is about twice the percentage of gig workers. This difference is significant and may be explained by the fact that blocking features are not built into all gig platforms.

[1]These proportions include data from our original (reporting to platform, police, safety organizations, social support network) and refielded (reporting to online whisper networks) samples. Since not all original survey respondents answered the online whisper networks question when it was asked in the refielded survey, this proportion may be an under-count.

## 6 DISCUSSION

In this work we draw on a systematic literature review of $n = 93$ papers to systematize the four harms (and one vehicle of harm) of AMOIs: harms to autonomy, to physical, emotional, and financial safety, and data privacy violations as a mechanism through which these harms can occur. In a survey of nearly 500 online daters and 500 in-person gig workers, we find some harms more salient to these groups' definitions of safety than others. Daters' definitions of safety focus on emotional, physical, and autonomy harm as well as data privacy as a mechanism through which such harms can occur. Gig workers focus on physical, emotional and autonomy harms. We also systematize from the literature, the 10 protective behaviors in which people engage to protect themselves from these harms. We find that these behaviors are widely adopted: with all behaviors, except documentation, used by at least 25% of each population we surveyed. The most popular five behaviors – screening, vetting, environmental precautions, covering and reporting, including to personal communities – adopted by at least 75% of those we surveyed.

Though AMOI-related harms may occur in offline realms, the behaviors used to protect against these harms are primarily digital or digitally-mediated by nature. Of the behaviors we synthesize, all but one (environmental precautions) are predominantly technologically-mediated and four are practiced digitally before any offline interaction occurs. AMOIs are an example of what Coles-Kemp and colleagues define as a *post-digital* enmeshing of our digital and non-digital worlds [44]. In a post-digital world, threats to security and privacy are no longer bound by space or contained in separated "online" and "offline" locales [178]. Instead, such threats – and defenses against them – cross between the digital and physical realms to affect people's overall sense of safety [152, 178]. Trauma we experience digitally also impacts our physical bodies, likewise, violence we experience physically – e.g., jobs that are digitally-mediated – will impact our relation to the digital platform that mediated this harm.

This concept of *post-digital safety* builds on decades of research across security as well as criminology, social sciences, and legal studies that attempt to both complicate and clarify what we mean by 'safety': from Maslow's hierarchy of needs, to human rights frameworks, to security frameworks for online content, at-risk groups, and in-game interactions [99, 119, 124, 163, 186, 212]. Broadly, safety is understood as a basic human need, which requires us to live in environments that are free of violence, threats, harms, and other intolerable risks which may be self-directed, interpersonal, or collective [108]. Building on this literature, post-digital safety encompasses the existence of safety threats that manifest in interactions intended to be exclusively online (e.g., stalking, doxxing) as well as those that reside in the contexts away-from-keyboard [155] during digitally-mediated offline interaction (DMOI) for e.g., dating, relationships, or labor.

In this section, we: (1) identify overlaps between safety in AMOIs and other previously studied contexts, (2) discuss how power dynamics impact the types of protective behaviors users are able to engage with, and (3) build on those power implications to suggest recommendations for shifting the burden of protection in AMOIs from people to platforms.

### 6.1 Contextualizing the Safety of Algorithmically-Mediated Offline Interactions

Respondents most commonly defined safety in AMOIs as centering on physical, emotional, and autonomy-related harms. The relationship between physical and emotional harms and safety is not new: social science researchers have explored the significance of "safe spaces" in physical settings [47, 108], and defined safety as including aspects of human well-being such as economic development, social justice, and environmental protection [10, 171]. An emerging body of work on digital safety has shown that emotional and physical safety is part of digital safety on social media [20, 81, 152, 163, 186], in online communities [55, 78, 162], and when doing work with communities that are marginalized [124, 187, 203]. However,

our survey results surface differences between the concerns most commonly expressed in respondents' definitions of safety and what prior work has focused on. These nuances further emphasize the importance of understanding users safety preferences in their context rather than imposing researchers' definitions [45].

Only a small subset of our respondents mentioned financial harm and data-privacy violations as core to their definition of what it means to be safe, especially among gig workers (see Section 4): fewer than 1% of gig workers surveyed defined safety as related to data privacy violations. These results may be explained by the physical nature of the interactions respondents engage in; even if workers who interact with clients in-person are concerned about data privacy violations, they may be more worried about the downstream physical impacts of such a violation than the violation itself.

The harms we include in our framework of safety in AMOIs overlap with those of Scheuerman et. al.'s framework of the severity of harmful online content [163] and Citron and Solove's framework of privacy harms [42]. In particular, our analysis of existing work on harms in AMOIs finds that prior work has considered harms to autonomy and to physical, emotional, and financial safety, as well as data privacy violations as a mechanism through which these harms can occur. These four harms align with both Scheuerman et. al.'s and Citron and Solove's frameworks; our conceptualization of data privacy violations also aligns with the latter. However, unlike these two frameworks, in our work we additionally consider respondents' definitions of safety and how the salience of the harms they describe align with the harms prior work has considered. In doing so, we were able to surface two interesting insights: (1) that very few gig workers defined safety as related to data privacy violations relative to the volume of prior work on crowdworker privacy, and (2) that close to a quarter of online daters described autonomy-related harms when defining safety, despite little to no prior work considering these in the online dating literature.

We observe contextual nuance in how respondents define autonomy harm, in particular, in contrast to prior work. Our respondents' characterization of autonomy harm varies slightly from that of Citron and Solove; while Citron and Solove define autonomy harm as an impairment on a person's ability to freely make informed choices about their data, the autonomy-related harm our respondents described relates to control over their physical bodies in offline spaces (e.g., the ability to physically leave an unsafe situation). Prior work on autonomy-related harm in AMOIs has focused on digital autonomy harms, such as the ways in which platforms use opaque algorithms to exert control over users [8]. Our work expands the meaning of autonomy-related harm by revealing how people's priorities within AMOIs differ from the digital-first focus of existing work.

The behaviors in our taxonomy also overlap with some of the protective practices described in Warford et. al.'s framework of at-risk users [212]. Warford et. al.'s framework describes two distancing strategies – "censoring online sharing" and "reducing one's digital footprint" – which overlap with the omission approach to obfuscation that we describe. Further, Warford et al. identify five social strategies at-risk users engage in to overcome digital safety threats, including: "preemptive disclosure for control," which we term self-disclosure, and "Vetting identities to avoid potential attackers," which we also term vetting. These overlaps suggest that even if a person is not otherwise "at-risk," engaging in an interaction they perceive as high-risk may lead them to engage in similar strategies to those at-risk users use in other interactions.

### 6.2 Power in AMOIs

The harms users experience in AMOIs reflect the lack of power they have over their experiences on the platform. Algorithmic systems, like AMOI platforms, are a black box [29], whose inner workings are largely inscrutable to users [41]. In the context of AMOIs, users may control some of the information they provide to the platform. However,

they have little visibility into how that information is used. Companies may disclose data sharing practices in privacy policies but extensive studies have shown that people rarely read these documents and even when they do it is difficult to understand the extent of data sharing practices [135]. Users often have to exercise good faith with actors who repeatedly sell sensitive, but profitable, data points to data brokers with limited transparency around which government entities or malicious actors might access them [12]. In addition to increased concerns about limited institutional privacy, the methods used to generate matches do not allow people to indicate what values (e.g., safety, compatibility, etc.) they want to prioritize in those matches. This puts them at risk of blindly walking into a situation where they may experience harm. This risk is compounded by two factors. First, a platform is financially incentivized to generate matches, even if those are not in a user's best interests. Second, users have limited agency to exercise informed refusal [17] over the use of their data for primarily corporate gain.

Platforms influence which protective behaviors are available to users through the design of the system. For instance, rating mechanisms make gig workers hesitant to address bias and harassment from clients for fear of the financial repercussions associated with a negative review [112]. Also, some protective behaviors (e.g., vetting, self-disclosure, and blocking) depend on the existence of specific platform features (e.g., certain user profile fields and a blocking mechanism). If these features are not available, using said protective behaviors becomes impossible, or at least harder [72, 85, 116, 183].

Even when protective behaviors are supported by a platform, the lack of transparency and accountability reduces user power. For instance, some platforms offer reporting mechanisms for flagging harm and reporting this to the platform. However, users often have little visibility into how these reports are handled and few actionable options once a report is submitted. It is common for users to file a report and never hear back from the platform [112]. This has implications for users' ability to seek justice and for regulators' ability to understand harms. Greater transparency into reporting systems could show how harms evolve or how new harms emerge as platforms add new features. Because of the platform's opacity, users often "report," or share, their experiences with other groups instead, such as closed, private whisper networks [96, 141]. This kind of disclosure has its own limitations. For example, it hides said harms from those who are not in these networks, such as those with limited social capital [116]. Moreover, the lack of cross-platform accountability impedes harm reduction; perpetrators can continuously re-enter online spaces (e.g., by remaking accounts).

Social status also influences protective behaviors. There are many examples of this. First, greater social capital may further enhance one's ability to use protective behaviors such as vetting and post-harm reporting (e.g., via whisper networks). Second, access to trusted contacts may facilitate covering and emergency alerts. Third, financial resources may determine whether someone can pay for background check services. Finally, a user's location may determine their ability to take environmental precautions (e.g., meeting in a crowded place may be harder in rural areas than in a large city). Furthermore, all protective behaviors we study (Sec. 5) are performed by users; thus, they shift the responsibility of protection from platforms to users, minimizing platform responsibility [64]. Prior research finds that this type of invisible labor falls unequally on those from marginalized groups because are more likely to experience privacy violations and associated harms [116]. A similar argument could be made about safety. People from marginalized populations are more likely to experience safety harms from AMOIs [84, 143, 182]. Thus, they might need to go to greater lengths to protect themselves.

Perhaps the ultimate protective strategy is not using AMOI platforms at all, but some users do not have the power to do this. Non-use of social media and other internet services has been considered to mitigate privacy harms [87, 138]. Such research has found that non-use has significant limitations. Many services have become so ubiquitous that disconnecting from them substantially disrupts daily life [87]. Oftentimes, non-users are cut off from valuable resources that a large digital network could provide [138]. Given the cost of disengagement, power is implicit in the question of

who can successfully disengage. Despite shortcomings, technology can sometimes be empowering to individuals from marginalized populations. For instance, gig platforms give individuals who are not able to work traditional office jobs (e.g., those with disabilities and childcare duties) greater control over when and where they work [77, 92, 160]. Also, online dating expands a user's pool of potential partners, which can benefit those with limited social capital [65] and from stigmatized communities [146]. This makes it all the more important to support the use of AMOI platforms in ways that better consider the impact of power and users' experiences and values. In the next subsection, we present recommendations for shifting the burden of protection in AMOIs from people to platforms.

### 6.3 Building Safer Algorithmically-Mediated Offline Introductions

Because AMOIs are digitally mediated and thus exist within a post-digital framing, there are multiple potential directions through which computing research and development may support safety. While this work focuses on technological approaches to supporting safety that can be taken by platforms, we note that such technological changes exist within a broader legal and societal ecosystem. While people have attempted to hold platforms responsible for the harms they create by matching users together [51, 71], such attempts have largely been unsuccessful and the legislative landscape defining platforms' responsibility for the harms their users experience as a result of the matches the platform creates is murky [63, 68, 107]. Legal shifts in how responsibility is placed are often preceded by societal shifts in related attitudes such as victim blaming and acknowledgement of the seriousness of harms that have a digital component [54, 64]. To create safer AMOIs, technological change alone will not be enough: changes in society, law, *and* technology are required [210].

**Leverage technical security techniques to mitigate harm in AMOIs.** Existing safety technologies, such as safety apps, aim to address AMOI-related harm, but have no actual mechanisms for prevention. Instead they function as alerts to use during or after the harm has occurred. Meanwhile, people do engage in more proactive safety behaviors, as described Section 5.1. However, these are rather ad-hoc; people appropriate specific technologies (e.g., the Meet's online profile within the matching app, social media sites, search engines, online forums) for safety themselves. Beyond this safety work that falls onto users, there are very few options for preventive safety tools.

Computing, in particular techniques from computer security, can help to limit people's exposure to offline harm and reduce the individual responsibility for safety placed on users [116]. For example, researchers and platforms should consider the potential to use existing defensive security techniques, such as authentication [79], training to detect potentially malicious interactions (e.g., phishing) [88], and automated detection of malicious accounts/users [211]. Future work in CSCW may consider how to build support for collective proactive safety defenses. For example, how may we create collective or social authentication protocols, similar to those described in [222], that leverage the community's experience with an individual to determine what features, services, or individuals they have access to on the platform? Similarly, how can we crowdsource the detection of abusers based on people's reports of negative experiences with Meets in bad client/aggressor lists? In some security contexts, collective detection of harm has been found to be effective, indicating potential. In particular, prior work finds that using employees as a collective phishing detection mechanism in large organizations is effective; it leads to fast detection of new phishing campaigns with an acceptable operational load on the organization and the employees [104].

Threat modeling is another defensive security technique that may be appropriate. Threat models are used to identify and communicate information about the threats to a system [167, 175]. An early step in threat modeling is often to design process flow [189] and data flow [46] diagrams that decompose system components to show their interactions.

A process flow diagram may illustrate the inputs and outputs to a system, the paths along a decision tree, the users involved, and/or the time elapsed [189]. A data flow diagram shows which components touch which data [46]. Both diagram types can help reveal vulnerabilities, potential attack vectors, and bad actors. This helps security professionals design defenses.

In the context of AMOIs, threat modeling may support the development of approaches for harm prevention, for example, by enabling analysis of how weaknesses in platform design, actor motivations, and resource availability interact to cause harm. Following conventions in computer security, our process flow diagram (Fig. 2) illustrates the inputs and outputs to an AMOI interaction (user behavior), decision tree paths (decision points and actions), users involved (users, matches, and matching algorithm), and time elapsed (stages of interaction). Future work may build on our diagram to develop threat models that describe how bad actors may leverage specific platform features, or compromise protective behaviors to cause harm.

Security techniques can also be used to support people's existing safety behaviors. For example, there may be certain information that people want to self-disclose to potential Meets, but do not want to share with the platform. We might consider how to develop privacy-preserving ways for users to share potentially sensitive information only with Meets. Perhaps differential privacy approaches may be helpful here; platforms can collect aggregated data for particular information fields in the app over all users, rather than for each user separately [50].

Finally, to implement technical approaches to support safety in AMOIs, it will be important to understand how users are already protecting themselves and the context in which harms occur. This is where future researchers can leverage our survey findings on the salience of harms in users' definitions of safety and the implementation mechanisms of their protective behaviors. For example, developing an automated risk detection system for AMOIs may require understanding what people look for (e.g., people's screening heuristics in Section 5.1.3). Future work can extend our results by examining the factors that increase people's vulnerability to safety harms in AMOIs. This may be useful in targeting support towards those populations, similar to how understanding phishing risk factors in security has led to more targeted detection and mitigation approaches [23].

**Trading off privacy and safety.** Platforms that enable AMOIs are notoriously bad at protecting users' safety [112]. This leads individuals to feel they must go outside the platform to protect themselves. For example, our participants reported vetting potential Meets by looking them up using search tools and social media, and consulting whisper networks. While this may help users find more information about a potential Meet to make the interaction safer, it can also violate the privacy of the Meet. This echos results reported in [43], where researchers found that safety concerns might lead online daters to behave in ways that violate their own or others' privacy.

The results of our survey suggest that even though individuals often look up others using the information Meets have shared on the platform, they are wary of sharing their own information and hide personal information in the pre-meet stage. While participants reported doing this for safety reasons, this behavior makes it difficult for Meets to protect their safety by vetting them in return. These behaviors create an interesting design paradox that needs to be considered when thinking about what safeguards platforms should implement to support safety in AMOIs.

One approach we encourage future research to consider is consensual access to vetting information. This could alleviate the need to "stalk" Meets outside the platform prior to meeting, but raises concerns around information sharing and abuse. These should be examined through a trust and abusability lens using the toolkit described in [181]. For instance, future work would need to consider the types of data that are appropriate and necessary to collect from users for a consensual vetting system. If such a system is centralized, platforms will serve as consensual vetting brokers

(as they already do to some degree); this requires that users trust the company running the platform, and believe the company cares about protecting their safety. Therefore, future work should also explore the degree to which users trust platforms to support their safety, and how their level of trust influences the information they are willing to share.

Consensual vetting systems will be vulnerable to abuse. Users and/or malicious actors may take advantage of their access to others' information to cause a variety of harms such as coercive control and manipulation. While it will never be possible to entirely design out this harm [178], engineers and designers will need to examine the abusability of the system at various stages of its implementation and whether there are risks of greater harm than those that may be mitigated. Learning from community-owned and community-run harm reduction tools such as sex workers' bad client and aggressor lists [179] and similar fora [14, 124] may be useful here – the work necessary to build and maintain these trusted, cared for, and community-maintained systems is built on notions of restorative justice rather than the traditionally punitive systems we build into security and other digital safety tools.

**Reconsider the design of existing safety tech.** There have been significant efforts made to build safety apps to protect people from offline harms, such as physical assault [121]. These include tracking apps to know if someone is in danger [21, 118, 218] and panic buttons/alarms that alert emergency authorities of an unsafe situation when triggered by a user [98]. Our research, however, shows that few people are aware of these apps, and even fewer actually use them.

Research suggests there are limitations to what these technologies can do and how useful users find them to be in a moment of crisis [98, 121]. For example, many safety apps are reactive rather than proactive – they can only be used once an unsafe event has occurred rather than preventing one from happening. Users have expressed that this support occurs too late to be useful [98, 112]. Likewise, users often find tracking devices too inaccurate to use effectively [21]. Finally, safety apps that do rely on tracking may be seen by users as harmful surveillance technologies, rather than helpful safety apps [109, 172]. Surveillance is not the same as safety, and may cause harm, especially to individuals from marginalized groups who face stigma due to their race, gender, and sexuality [172]. Our work corroborates these findings and offers additional insight into why people may hesitate to use existing safety apps to protect their safety. We use those insights to make design recommendations.

Respondents frequently rely on collective strategies, such as consulting whisper networks and relying on trusted contacts to intervene when an offline meeting becomes unsafe. Yet, design of existing safety technologies is highly individualistic and relies on centralization (e.g., emergency alerts that contact law enforcement). Respondents' collective strategies do not require the use of specialized technology beyond a device for communication and accessing online whisper networks. Further, while some people instruct trusted individuals to contact police on their behalf if a meeting becomes unsafe, very few choose to report unsafe experiences to police themselves. Thus, future research and development on safety technologies should carefully consider how to enable collective action and decentralized protection. Our process flow diagram (Fig. 2) may inform this work, supporting research that explores the attitudinal and behavioral antecedents that influence protective behaviors and decision-making.

Finally, researchers and platforms should involve users (especially from communities most at risk of experiencing harm) in co-creating safety technologies. Lack of platform and safety app support for safety is not just an interface problem. It also relates to institutional and power dynamics. Users must be empowered to advocate for protection against their safety concerns. Future work may draw on Stein et al.'s research on worker-designed data institutions to support AMOI platform users in designing safer infrastructure [174], rather than standalone safety technologies. Other work may draw on Keyes et al.'s anarchist HCI framework to help victims of harm construct meaning from that experience [100]. For example, this enables users to define what harm is, especially as platforms with new capabilities

emerge, and allows them to inform how platforms and standalone tools address those harms. These participatory approaches would serve to counter the power that platforms have over users' safety.

**Cautious Design.** It is clear that AMOIs constitute a complex ecosystem with various actors (e.g., the platform, users, and bystanders in the offline environment where interactions occur). Each of these actors has different goals and priorities, which may factor into their perceptions of and experiences with safety. We urge researchers to consider what is ethical when researching solutions to safety and security issues. Researchers should be sure to examine their own privilege and take active steps to mitigate their own biases in deciding whose safety and well-being to protect and reflecting on the consequences of those decisions. We encourage researchers to consider a feminist orientation to safety, following the guidelines described in [178], and to follow recent guidelines on trauma-informed computing [38].

Further, while our work focuses on individuals' experiences with safety in offline interactions broadly, we note that women, non-binary folks, and people of color are at especially pronounced risk of harm in AMOIs [84, 143, 182]. People of color and women experience physical harm such as sexual assault, rape, and murder at a higher rate [11, 84]. Risk of emotional harm related to hate and harassment may be especially pronounced among marginalized groups [7, 186, 212]. These groups often have to engage in further emotionally taxing "safety work" [73, 85, 183] to try to protect their safety, and manage post-harm trauma. We encourage future work to both consider how individuals' identities (e.g., gender, race, education level) influence protective behaviors and to engage and amplify marginalized voices in the creation of safer AMOIs. When we design for those who are most at risk for harm, we are creating safer spaces not only for them, but for everyone.

Finally, perhaps the most important questions to consider when designing for post-digital safety in AMOIs are whether people want to use technology at all. We encourage future work to critically consider when it is appropriate to introduce technology and when it is most appropriate to abstain and make space for other forms of action. In some cases, new regulations may be more appropriate to counter the power imbalance between platforms and users, given misaligned incentives between them and the weakness of current safety regulations [209]. For instance, to increase transparency into platforms' reporting processes, regulatory bodies may dictate that platforms must maintain clear reporting mechanisms, and/or make anonymous datasets of reports and their outcomes available to all users.

In other cases, empowering other safety experts may be the best course of action. Homewood [90] lays out a plan about the opportunities that *inaction* as a design decision brings to a research space, and Strohmayer et al. [179] argue that vital human interaction that aims to reduce harms in AMOIs may be digitally mediated, but should not be replaced with novel digital technologies. Especially when considering safety in our post-digital world, 'inaction' on the novel technology-development front becomes an important point of reflection. This 'inaction' can create space of action for the improvement of existing technologies and, importantly, non-digital services: creating space to empower safety experts in anti-violence and post-violence support services and/or reallocate funds away from technology and into violence-reduction work. For instance, third-party organizations focused on anti-violence work could be empowered to maintain reporting systems or other tools to support safety in AMOIs, similar to Australia's independent regulator for online safety, eSafety [59]. Such an organization could be involved in helping shed light on new forms of harm emerging in users' reports and provide guidelines for researchers, platforms, and regulatory bodies to better protect users' safety in AMOIs.

## 7 CONCLUSION

This work aims to formalize a protective model for algorithmically-mediated offline interactions (AMOIs): those in which a platform's algorithm matches a pair of people for a purposeful offline interaction such as for dating or household labor. By systematizing 93 prior works, we synthesize the harms people risk facing in these interactions and the steps they take to protect themselves.

Offline harms emerging from AMOIs are a technology problem: platforms and apps play a significant role in matching strangers and encouraging their interactions with one another, with limited safeguards to protect them. Addressing them requires understanding the nuances between harms in AMOIs and the ways in which we currently conceive of digital safety and harm. Previously, security researchers have focused on studying and addressing harms abusers cause using technology, including both traditional security considerations such as financial harms from scams and fraud [31, 188, 201], as well as emotional or physical harm via online hate and harassment [52, 186, 187] and technology-facilitated intimate partner violence [37, 199, 200, 227]. Security defenses against these harms may in some cases center on avoiding post-digital spillover of threats (e.g., an online abuser causing physical/sexual violence to a target) by reducing the likelihood for offline interaction. However, in AMOIs such offline interaction is the goal of using systems, and thus preventing such interaction is not an appropriate mitigation strategy. In this work we offer suggestions for supporting safety in AMOIs and shifting the burden of protection from users to platforms.

However, this work is but a step in the journey towards post-digital security and safety by design. As new technologies and technology-mediated spaces emerge, we will need to continuously refine what digital safety means. In doing so it is imperative to involve users in co-constructing the meaning of safety in those environments. Our empirical work is a step in that direction. We urge future work to actively involve users in co-creating safer post-digital spaces, taking a broad definition of the scope of safety. Such work may draw on participatory design methods [27] to create safety-mitigating technologies, strategies, and policy. It may also involve carefully examining the abusability of a system following the guidelines presented in [181] and building trauma-aware systems following the guidelines in [38].

## REFERENCES


[1] Kendra Albert, Emily Armbruster, Elizabeth Brundige, Elizabeth Denning, Kimberly Kim, Lorelei Lee, Lindsey Ruff, Korica Simon, and Yueyu Yang. 2020. FOSTA in legal context. *Columbia Human Rights Law Review* 52 (2020), 1084.

[2] Martin R Albrecht, Jorge Blasco, Rikke Bjerg Jensen, and Lenka Mareková. 2021. Collective Information Security in Large-Scale Urban Protests: The Case of Hong Kong. In *30th USENIX Security Symposium (USENIX Security 21)*. 3363–3380.

[3] Kath Albury and Paul Byron. 2016. Safe on My Phone? Same-Sex Attracted Young People's Negotiations of Intimacy, Visibility, and Risk on Digital Hook-Up Apps. *Social Media + Society* 2, 4 (Oct. 2016), 2056305116672887. https://doi.org/10.1177/2056305116672887

[4] Kath Albury, Paul Byron, Anthony McCosker, Tinonee Pym, Jarrod Walshe, Kane Race, Doreen Salon, Tim Wark, Jessica Botfield, Daniel Reeders, and Christopher Dietzel. 2019. *Safety, risk and wellbeing on dating apps: Final report.* Report. Swinburne University of Technology. https://apo.org.au/node/268156

[5] Kath Albury, Anthony McCosker, Tinonee Pym, and Paul Byron. 2020. Dating apps as public health 'problems': Cautionary tales and vernacular pedagogies in news media. *Health Sociology Review* 29, 3 (Sept. 2020), 232–248. https://doi.org/10.1080/14461242.2020.1777885

[6] Hanan Khalid Aljasim and Douglas Zytko. 2023. Foregrounding Women's Safety in Mobile Social Matching and Dating Apps: A Participatory Design Study. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (2023), 1–25.

[7] Mashael Yousef Almoqbel and Donghee Yvette Wohn. 2019. Individual and Collaborative Behaviors of Rideshare Drivers in Protecting their Safety. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 217:1–217:21. https://doi.org/10.1145/3359319

[8] Micah Altman, Alexandra Wood, and Effy Vayena. 2018. A harm-reduction framework for algorithmic fairness. *IEEE Security & Privacy* 16, 3 (2018), 34–45.

[9] Ira Anjali Anwar, Joyojeet Pal, and Julie Hui. 2021. Watched, but Moving: Platformization of Beauty Work and Its Gendered Mechanisms of Control. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 250:1–250:20. https://doi.org/10.1145/3432949

[10] Kofi Annan. 2001. Definitions of human security. https://www.gdrc.org/.

[11] Avi Asher-Schapiro. 2022. U.S. gig worker murders expose apps' safety gaps, says labor group. https://news.trust.org/item/20220406101735-ixk5v/.

[12] Noah Ashman. 2020. Outed by Advertisements: How LGBTQ Internet Users Present a Case for Federal Data Privacy Legislation. *Or. L. Rev.* 99 (2020), 523.

[13] Australia eSafety Commissioner. 2021. Principles and background. https://www.esafety.gov.au/industry/safety-by-design/principles-and-background. (Accessed on 07/17/2023).

[14] Hanna Barakat and Elissa M Redmiles. 2022. Community under surveillance: Impacts of marginalization on an online labor forum. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 16. 12–21.

[15] John A Bargh, Katelyn YA McKenna, and Grainne M Fitzsimons. 2002. Can you see the real me? Activation and expression of the "true self" on the Internet. *Journal of social issues* 58, 1 (2002), 33–48.

[16] 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*. 1–16.

[17] Ruha Benjamin. 2016. Informed refusal: Toward a justice-based bioethics. *Science, Technology, & Human Values* 41, 6 (2016), 967–990.

[18] Danielle Blunt and Ariel Wolf. 2020. Erased: The impact of FOSTA-SESTA and the removal of Backpage on sex workers. *Anti-trafficking Review* 14 (2020), 117–121.

[19] Danielle Blunt, Ariel Wolf, Emily Coombes, and Shanelle Mullin. 2020. Posting into the void: Studying the impact of shadowbanning on sex workers and activists. https://hackinghustling.org/posting-into-the-void-content-moderation/.

[20] Jessica E Bodford, Cameron J Bunker, and Virginia SY Kwan. 2021. Does perceived social networking site security arise from actual and perceived physical safety? *Computers in Human Behavior* 121 (2021), 106779.

[21] Julie Boesen, Jennifer A Rode, and Clara Mancini. 2010. The domestic panopticon: Location tracking in families. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. 65–74.

[22] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[23] Cristian Bravo-Lillo, Lorrie Faith Cranor, Julie Downs, and Saranga Komanduri. 2010. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy* 9, 2 (2010), 18–26.

[24] Carolyn Bronstein. 2021. Deplatforming sexual speech in the age of FOSTA/SESTA. *Porn Studies* 8, 4 (2021), 367–380.

[25] Kellen Browning. 2022. At least 50 people have been killed doing gig driving since 2017, report says. https://www.nytimes.com/2022/04/06/business/uber-lyft-driver-deaths.html.

[26] Jed R Brubaker, Mike Ananny, and Kate Crawford. 2016. Departing glances: A sociotechnical account of 'leaving' Grindr. *New Media & Society* 18, 3 (2016), 373–390.

[27] Caroline Bull, Hanan Aljasim, and Douglas Zytko. 2021. Designing Opportunistic Social Matching Systems for Women's Safety During Face-to-Face Social Encounters. In *CSCW '21: Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA). 23–26. https://doi.org/10.1145/3462204.3481751

[28] United States Census Bureau. 2020. 2020 Census Demographic Profile. https://www.census.gov/data/tables/2023/dec/2020-census-demographic-profile.html#data.

[29] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society* 3, 1 (2016), 2053951715622512.

[30] Yvette Butler. 2020. Aligned: Sex Workers' Lessons for the Gig Economy. *Michigan Journal of Race & Law* 26 (2020), 337.

[31] Mark Button, Carol McNaughton Nicholls, Jane Kerr, and Rachael Owen. 2014. Online frauds: Learning from victims why they fall for these scams. *Australian & New Zealand Journal of Criminology* 47, 3 (2014), 391–408.

[32] Lindsey D Cameron, Bobbi Thomason, and Vanessa M Conzon. 2021. Risky business: Gig workers and the navigation of ideal worker expectations during the COVID-19 pandemic. *Journal of Applied Psychology* 106, 12 (2021), 1821.

[33] Lauren F Cardoso, Susan B Sorenson, Olivia Webb, and Sara Landers. 2019. Recent and emerging technologies: Implications for women's safety. *Technology in Society* 58 (2019), 101108.

[34] Bronwyn Carlson. 2020. Love and hate at the cultural interface: Indigenous Australians and dating apps. *Journal of Sociology* 56, 2 (2020), 133–150.

[35] Avner Caspi and Paul Gorsky. 2006. Online deception: Prevalence, motivation, and emotion. *CyberPsychology & Behavior* 9, 1 (2006), 54–59.

[36] Stevie Chancellor, Jessica Annette Pater, Trustin Clear, Eric Gilbert, and Munmun De Choudhury. 2016. # thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*. 1201–1213.

[37] 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 *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 441–458.

[38] 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 2022 CHI conference on human factors in computing systems*. 1–20.

[39] Edmond Pui Hang Choi, Janet Yuen Ha Wong, and Daniel Yee Tak Fong. 2018. An emerging risk factor of sexual abuse: The use of smartphone dating applications. *Sexual Abuse* 30, 4 (2018), 343–366.

[40] Nicola Christie and Heather Ward. 2019. The health and safety risks for people who drive for work in the gig economy. *Journal of Transport & Health* 13 (2019), 115–127.

[41] Angèle Christin. 2020. The ethnographer and the algorithm: beyond the black box. *Theory and Society* 49, 5-6 (2020), 897–918.

[42] Danielle Keats Citron and Daniel J Solove. 2022. Privacy harms. *BUL Rev.* 102 (2022), 793.

[43] Camille Cobb and Tadayoshi Kohno. 2017. How Public Is My Private Life? Privacy in Online Dating. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1231–1240. https://doi.org/10.1145/3038912.3052592

[44] Lizzie Coles-Kemp, Rikke Bjerg Jensen, and Claude P. R. Heath. 2020. *Too Much Information: Questioning Security in a Post-Digital Society*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376214

[45] Jessica Colnago, Lorrie Faith Cranor, and Alessandro Acquisti. 2023. Is there a reverse privacy paradox? an exploratory analysis of gaps between privacy perspectives and privacy-seeking behaviors. *Proceedings on Privacy Enhancing Technologies* 1 (2023), 455–476.

[46] Larry Conklin. 2024. Threat Modeling Process. https://owasp.org/www-community/Threat_Modeling_Process.

[47] Karen Corteen. 2002. Lesbian safety talk: Problematizing definitions and experiences of violence, sexuality and space. *Sexualities* 5, 3 (2002), 259–280.

[48] Danielle Couch and Pranee Liamputtong. 2007. Online dating and mating: Perceptions of risk and health among online users. *Health, Risk & Society* 9, 3 (2007), 275–294.

[49] Danielle Couch, Pranee Liamputtong, and Marian Pitts. 2012. What are the real and perceived risks and dangers of online dating? Perspectives from online daters: Health risks in the media. *Health, Risk & Society* 14, 7-8 (2012), 697–714.

[50] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. 2021. " I need a better description": An Investigation Into User Expectations For Differential Privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 3037–3052.

[51] Maria Dinzeo. 2018. Ninth Circuit Takes New Look at Duty to Warn in Match Gone Bad. https://www.courthousenews.com/ninth-circuit-takes-new-look-at-duty-to-warn-in-match-gone-bad/. (Accessed on 07/16/2023).

[52] Periwinkle Doerfler, Andrea Forte, Emiliano De Cristofaro, Gianluca Stringhini, Jeremy Blackburn, and Damon McCoy. 2021. "I'm a Professor, which isn't usually a dangerous job": Internet-facilitated Harassment and Its Impact on Researchers. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–32.

[53] Stefanie Duguay. 2016. "He has a way gayer Facebook than I do": Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society* 18, 6 (2016), 891–907.

[54] Stefanie Duguay, Jean Burgess, and Nicolas Suzor. 2020. Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence* 26, 2 (2020), 237–252.

[55] Brianna Dym and Casey Fiesler. 2020. Social norm vulnerability and its consequences for privacy and safety in an online community. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–24.

[56] Josh Dzieza. 2021. Revolt of the Delivery Workers. https://www.curbed.com/article/nyc-delivery-workers.html.

[57] Nicole B Ellison, Jeffrey T Hancock, and Catalina L Toma. 2012. Profile as promise: A framework for conceptualizing veracity in online dating self-presentations. *New media & society* 14, 1 (2012), 45–62.

[58] Sheena Erete, Yolanda A Rankin, and Jakita O Thomas. 2021. I can't breathe: Reflections from Black women in CSCW and HCI. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–23.

[59] Australia eSafety Commissioner. 2024. eSafety Commissioner. https://www.esafety.gov.au/.

[60] Kimberly Fairchild. 2015. "But look at what she was wearing!": Victim blaming and street harassment. In *Gender, Sex, and Politics*. Routledge, 22–32.

[61] Julia R. Fernandez and Jeremy Birnholtz. 2019. "I Don't Want Them to Not Know": Investigating Decisions to Disclose Transgender Identity on Dating Platforms. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 226:1–226:21. https://doi.org/10.1145/3359328

[62] Lauren S Ferro, Andrea Marrella, and Tiziana Catarci. 2021. A Human Factor Approach to Threat Modeling. In *International Conference on Human-Computer Interaction*. Springer, 139–157.

[63] Nataliia Filatova-Bilous. 2021. Once again platform liability: On the edge of the" Uber" and" Airbnb" cases. *Internet Policy Review* 10, 2 (2021), 1–27.

[64] Eric Filice, Kavishka D Abeywickrama, Diana C Parry, and Corey W Johnson. 2022. Sexual violence and abuse in online dating: A scoping review. *Aggression and violent behavior* 67 (2022), 101781.

[65] Eli J Finkel, Paul W Eastwick, Benjamin R Karney, Harry T Reis, and Susan Sprecher. 2012. Online dating: A critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest* 13, 1 (2012), 3–66.

[66] 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*. 1–13.

[67] Andrea Gallardo, Hanseul Kim, Tianying Li, Lujo Bauer, and Lorrie Cranor. 2022. Detecting iPhone Security Compromise in Simulated Stalking Scenarios: Strategies and Obstacles. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. 291–312.

[68] Kira M Geary. 2020. Section 230 of the Communications Decency Act, Product Liability, and a Proposal for Preventing Dating-App Harassment. *Penn St. L. Rev.* 125 (2020), 501.

[69] Jennifer L Gibbs, Nicole B Ellison, and Chih-Hui Lai. 2011. First comes love, then comes Google: An investigation of uncertainty reduction strategies and self-disclosure in online dating. *Communication Research* 38, 1 (2011), 70–100.

[70] Louisa Gilbert, Aaron L Sarvet, Melanie Wall, Kate Walsh, Leigh Reardon, Patrick Wilson, John Santelli, Shamus Khan, Martie Thompson, Jennifer S Hirsch, et al. 2019. Situational contexts and risk factors associated with incapacitated and nonincapacitated sexual assaults among college women. *Journal of Women's Health* 28, 2 (2019), 185–193.

[71] Eden Gillespie and Tamsin Rose. 2022. Call for dating apps to require criminal checks as Australian government plans summit on safety | Australian police and policing. https://www.theguardian.com/australia-news/2022/dec/24/call-for-dating-apps-to-require-criminal-checks-as-australian-government-plans-summit-on-safety. (Accessed on 07/16/2023).

[72] Rosalie Gillett. 2018. Intimate intrusions online: Studying the normalisation of abuse in dating apps. In *Women's Studies International Forum*, Vol. 69. Elsevier, 212–219.

[73] Rosalie Gillett. 2021. "This is not a nice safe space": Investigating women's safety work on Tinder. *Feminist Media Studies* 23, 1 (2021), 1–17.

[74] Nancy Glass, Amber Clough, James Case, Ginger Hanson, Jamie Barnes-Hoyt, Amy Waterbury, Jeanne Alhusen, Miriam Ehrensaft, Karen Trister Grace, and Nancy Perrin. 2015. A safety app to respond to dating violence for college women and their friends: The MyPlan study randomized controlled trial protocol. *BMC Public Health* 15, 1 (2015), 1–13.

[75] Paul Glavin, Alex Bierman, and Scott Schieman. 2021. Über-alienated: Powerless and alone in the gig economy. *Work and Occupations* 48, 4 (2021), 399–431.

[76] Mark Graham, Isis Hjorth, and Vili Lehdonvirta. 2017. Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research* 23, 2 (2017), 135–162.

[77] Mary L Gray and Siddharth Suri. 2019. *Ghost work: How to stop Silicon Valley from building a new global underclass*. Eamon Dolan Books.

[78] Ana M Giménez Gualdo, Simon C Hunter, Kevin Durkin, Pilar Arnaiz, and Javier J Maquilón. 2015. The emotional impact of cyberbullying: Differences in perceptions and experiences as a function of role. *Computers & Education* 82 (2015), 228–235.

[79] Cheng Guo, Brianne Campbell, Apu Kapadia, Michael K Reiter, and Kelly Caine. 2021. Effect of Mood, Location, Trust, and Presence of Others on {Video-Based} Social Authentication. In *30th USENIX Security Symposium (USENIX Security 21)*. 1–18.

[80] Shruti Gupta. 2020. Gendered Gigs: Understanding the Gig Economy in New Delhi from a Gendered Perspective. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development* (Guayaquil, Ecuador) *(ICTD2020)*. Article 7, 10 pages. https://doi.org/10.1145/3392561.3394635

[81] Oliver L Haimson, Justin Buss, Zu Weinger, Denny L Starks, Dykee Gorrell, and Briar Sweetbriar Baron. 2020. Trans Time: Safety, Privacy, and Content Warnings on a Transgender-Specific Social Media Site. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–27.

[82] Oliver L. Haimson, Dykee Gorrell, Denny L. Starks, and Zu Weinger. 2020. Designing Trans Technology: Defining Challenges and Envisioning Community-Centered Solutions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3313831.3376669

[83] Vaughn Hamilton, Hanna Barakat, and Elissa M Redmiles. 2022. Risk, Resilience and Reward: Impacts of Shifting to Digital Sex Work. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW (2022), 1–37.

[84] Kenneth R. Hanson. 2021. Becoming a (Gendered) Dating App User: An Analysis of How Heterosexual College Students Navigate Deception and Interactional Ambiguity on Dating Apps. *Sexuality & Culture* 25, 1 (Feb. 2021), 75–92. https://doi.org/10.1007/s12119-020-09758-w

[85] Bridget A Harris and Delanie Woodlock. 2019. Digital coercive control: Insights from two landmark domestic violence studies. *The British Journal of Criminology* 59, 3 (2019), 530–550.

[86] Steven Hick, Edward Halpin, and Eric Hoskins. 2016. *Human rights and the Internet*. Springer.

[87] Kashmir Hill. 2020. I Tried to Live Without the Tech Giants. It was Impossible. https://www.nytimes.com/2020/07/31/technology/blocking-the-tech-giants.html.

[88] Grant Ho, Asaf Cidon, Lior Gavish, Marco Schweighauser, Vern Paxson, Stefan Savage, Geoffrey M Voelker, and David Wagner. 2019. Detecting and characterizing lateral phishing at scale. In *28th USENIX Security Symposium (USENIX Security 19)*. 1273–1290.

[89] Karen Holtzblatt and Hugh Beyer. 1997. *Contextual Design: Defining Customer-centered Systems*. Elsevier.

[90] Sarah Homewood. 2019. Inaction as a Design Decision: Reflections on Not Designing Self-Tracking Tools for Menopause. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. 1–12. https://doi.org/10.1145/3290607.3310430

[91] Julie Hui, Kentaro Toyama, Joyojeet Pal, and Tawanna Dillahunt. 2018. Making a Living My Way: Necessity-Driven Entrepreneurship in Resource-Constrained Communities. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW, Article 71 (Nov 2018), 24 pages. https://doi.org/10.1145/3274340

[92] Abigail Hunt, Emma Samman, Sherry Tapfuma, Grace Mwaura, Rhoda Omenya, Kay Kim, Sara Stevano, and Aida Roumer. 2019. *Women in the gig economy: Paid work, care and flexibility in Kenya and South Africa*. Technical Report. ODI. https://odi.org/en/publications/women-in-the-gig-economy-paid-work-care-and-flexibility-in-kenya-and-south-africa/

[93] Mohammad Hossein Jarrahi and Will Sutherland. 2019. Algorithmic management and algorithmic competencies: Understanding and appropriating algorithms in gig work. In *International Conference on Information*. Springer, 578–589.

[94] Mohammad Hossein Jarrahi, Will Sutherland, Sarah Beth Nelson, and Steve Sawyer. 2020. Platformic management, boundary resources for gig work, and worker autonomy. *Computer supported cooperative work (CSCW)* 29, 1 (2020), 153–189.

[95] Jialun Aaron Jiang, Morgan Klaus Scheuerman, Casey Fiesler, and Jed R Brubaker. 2021. Understanding international perceptions of the severity of harmful content online. *PloS one* 16, 8 (2021), e0256762.

[96] Carrie Ann Johnson. 2023. The purpose of whisper networks: a new lens for studying informal communication channels in organizations. *Frontiers in Communication* 8 (2023), 1089335.

[97] Nicole Kalms. 2017. Digital technology and the safety of women and girls in urban space: Personal safety Apps or crowd-sourced activism tools? In *Architecture and Feminisms*. Routledge, 112–121.

[98] Naveena Karusala and Neha Kumar. 2017. Women's Safety in Public Spaces: Examining the Efficacy of Panic Buttons in New Delhi. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3340–3351. https://doi.org/10.1145/3025453.3025532

[99] Douglas T. Kenrick, Vladas Griskevicius, Steven L. Neuberg, and Mark Schaller. 2010. Renovating the Pyramid of Needs: Contemporary Extensions Built upon Ancient Foundations. *Perspectives on Psychological Science* 5, 3 (2010), 292–314. https://doi.org/10.1177/1745691610369469

[100] Os Keyes, Josephine Hoy, and Margaret Drouhard. 2019. Human-Computer Insurrection: Notes on an Anarchist HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300569

[101] Alfred Kobsa, Sameer Patil, and Bertolt Meyer. 2012. Privacy in instant messaging: An impression management model. *Behaviour & Information Technology* 31, 4 (2012), 355–370.

[102] Roman Kuhar and Mojca Pajnik. 2019. Negotiating professional identities: Male sex workers in Slovenia and the impact of online technologies. *Sexuality Research and Social Policy* 16, 2 (2019), 227–238.

[103] Isak Ladegaard, Alexandrea J Ravenelle, and Juliet Schor. 2022. 'God Is Protecting Me… And I Have Mace': Defensive Labour In Precarious Workplaces. *The British Journal of Criminology* 62, 3 (2022), 773–789.

[104] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in organizations: Findings from a large-scale and long-term study. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 842–859.

[105] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1603–1612.

[106] Rachel Lerman. 2021. Tinder's upcoming feature for singles? A criminal background check on potential dates. https://www.washingtonpost.com/technology/2021/03/16/tinder-match-background-check/.

[107] Karen Levy and Solon Barocas. 2017. Designing against discrimination in online markets. *Berkeley Technology Law Journal* 32, 3 (2017), 1183–1238.

[108] Ruth Lewis, Elizabeth Sharp, Jenni Remnant, and Rhiannon Redpath. 2015. 'Safe spaces': Experiences of feminist women-only space. *Sociological Research Online* 20, 4 (2015), 105–118.

[109] Calvin Liang, Jevan Hutson, and Os Keyes. 2020. Surveillance, stigma & sociotechnical design for HIV. *arXiv preprint arXiv:2006.04882* (2020).

[110] Jennifer Hickes Lundquist and Celeste Vaughan Curington. 2019. Love me Tinder, love me sweet. *Contexts* 18, 4 (2019), 22–27.

[111] Peter Lynn. 2009. *Methodology of Longitudinal Surveys*. Wiley Online Library.

[112] Ning F. Ma, Veronica A. Rivera, Zheng Yao, and Dongwook Yoon. 2022. "Brush It Off": How Women Workers Manage and Cope with Bias and Harassment in Gender-Agnostic Gig Platforms. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Article 397, 13 pages. https://doi.org/10.1145/3491102.3517524

[113] Zhuang Ma, Woon Kian Chong, and Linpei Song. 2022. How Arousing Benefits and Ethical Misgivings Affect AI-Based Dating App Adoption: The Roles of Perceived Autonomy and Perceived Risks. In *International Conference on Human-Computer Interaction*. Springer, 160–170.

[114] Michael Maffie and Allison Elias. 2019. Platform design as a managerial act: Analyzing sexual harassment in the gig economy. *LERA For Libraries* 23, 2 (2019), 18–23.

[115] Sridhar Mandapati, Sravya Pamidi, and Sriharitha Ambati. 2015. A mobile based women safety application (I Safe Apps). *IOSR Journal of Computer Engineering (IOSR-JCE)* 17, 1 (2015), 29–34.

[116] Alice E. Marwick. 2023. *The Private is Political: Networked Privacy and Social Media*. Yale University Press.

[117] Christina Masden and W Keith Edwards. 2015. Understanding the role of community in online dating. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 535–544.

[118] Sarah Maslen. 2021. (Dis) connected parenting: Other-tracking in the more-than-human sensorium. *The Senses and Society* 16, 1 (2021), 67–79.

[119] A. H. Maslow. 1943. A Theory of Human Motivation. *Psychological Review* 50, 4 (1943), 370–396. https://doi.org/10.1037/h0054346

[120] 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 & security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2189–2201.

[121] Lucy Maxwell, Alex Sanders, Jason Skues, and Lisa Wise. 2020. A content analysis of personal safety apps: Are they keeping us safe or making us more vulnerable? *Violence Against Women* 26, 2 (2020), 233–248.

[122] Bronwyn McBride, Kate Shannon, Brittany Bingham, Melissa Braschel, Steffanie Strathdee, and Shira M Goldenberg. 2020. Underreporting of violence to police among women sex workers in Canada: Amplified inequities for im/migrant and in-call workers prior to and following end-demand legislation. *Health and Human Rights* 22, 2 (2020), 257.

[123] Bronwyn McBride, Kate Shannon, Alka Murphy, Sherry Wu, Margaret Erickson, Shira M Goldenberg, and Andrea Krüsi. 2021. Harms of third party criminalisation under end-demand legislation: Undermining sex workers' safety and rights. *Culture, Health & Sexuality* 23, 9 (2021), 1165–1181.

[124] Allison McDonald, Catherine Barwulor, Michelle L Mazurek, Florian Schaub, and Elissa M Redmiles. 2021. "It's stressful having all these phones": Investigating Sex Workers' Safety Goals, Risks, and Practices Online. In *30th USENIX Security Symposium (USENIX Security 21)*. 375–392.

[125] Metropolitan Police. *n.d.*. Request information under Clare's Law: Make a Domestic Violence Disclosure Scheme (DVDS) application. https://www.met.police.uk/advice/advice-and-information/daa/domestic-abuse/alpha2/request-information-under-clares-law/. (Accessed on 08/31/2022).

[126] Ruth Milkman, Luke Elliott-Negri, Kathleen Griesbach, and Adam Reich. 2021. Gender, Class, and the Gig Economy: The Case of Platform-Based Food Delivery. *Critical Sociology* 47, 3 (May 2021), 357–372. https://doi.org/10.1177/0896920520949631 Publisher: SAGE Publications Ltd.

[127] Jaron Mink, Licheng Luo, Natã M Barbosa, Olivia Figueira, Yang Wang, and Gang Wang. 2022. DeepPhish: Understanding User Trust Towards Artificially Generated Profiles in Online Social Networks. In *Proc. of USENIX Security*.

[128] Eva Moog. 2021. *Design for Safety*. A Book Apart.

[129] Max Morris. 2021. The limits of labelling: Incidental sex work among gay, bisexual, and queer young men on social media. *Sexuality Research and Social Policy* 18, 4 (2021), 855–868.

[130] Tatenda Mpofu, Pitso Tsibolane, Richard Heeks, and Jean-Paul Van Belle. 2020. Risks and Risk-Mitigation Strategies of Gig Economy Workers in the Global South: The Case of Ride-Hailing in Cape Town. In *Information and Communication Technologies for Development (IFIP Advances in Information and Communication Technology)*, Julian M. Bass and P. J. Wall (Eds.). Springer International Publishing, Cham, 26–38. https://doi.org/10.1007/978-3-030-65828-1_3

[131] Katie Myhill, James Richards, and Kate Sang. 2021. Job quality, fair work and gig work: The lived experience of gig workers. *The International Journal of Human Resource Management* 32, 19 (2021), 4110–4135.

[132] National Sex Offender Database. *n.d.*. Keeping Children Safe from Sexual Offenders. https://www.meganslaw.com/. (Accessed on 08/31/2022).

[133] Rory T Newlands, Dominic M Denning, Kaiya S Massey, and Lorraine T Benuto. 2022. Safe dating in the digital era: Protective behavioral strategies in dating behaviors facilitated by dating applications. *Violence and Victims* 37, 2 (2022), 185–200.

[134] Borke Obada-Obieh, Sonia Chiasson, and Anil Somayaji. 2017. "Don't Break My Heart!": User Security Strategies for Online Dating. In *Workshop on Usable Security (USEC)*.

[135] Jonathan A Obar and Anne Oeldorf-Hirsch. 2020. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23, 1 (2020), 128–147.

[136] UN Office of the High Commissioner. 2017. UN experts urge States and companies to address online gender-based abuse but warn against censorship. https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=21317

[137] Yok-Fong Paat and Christine Markham. 2021. Digital crime, trauma, and abuse: Internet safety and cyber risks for adolescents and emerging adults in the 21st century. *Social Work in Mental Health* 19, 1 (2021), 18–40.

[138] Xinru Page, Pamela Wisniewski, Bart P Knijnenburg, and Moses Namara. 2018. Social media's have-nots: an era of social disenfranchisement. *Internet Research* 28, 5 (2018), 1253–1274.

[139] Jessica A Pater, Moon K Kim, Elizabeth D Mynatt, and Casey Fiesler. 2016. Characterizations of online harassment: Comparing policies across social media platforms. In *Proceedings of the 2016 ACM International Conference on Supporting Group Work*. 369–374.

[140] Kun Peng, Wan-Ying Lin, and Hexin Chen. 2022. Consequences of deceptive self-presentation in online dating. *Chinese Journal of Communication* 15, 4 (2022), 1–29.

[141] Maria Verena Peters. 2020. From the Whisper Network to #MeToo—Framing Gender, Gossip and Sexual Harassment. *European Journal of American Studies* 15, 4 (Dec. 2020). https://doi.org/10.4000/ejas.16587

[142] Anh Phan, Kathryn Seigfried-Spellar, and Kim-Kwang Raymond Choo. 2021. Threaten me softly: A review of potential dating app risks. *Computers in human behavior reports* 3 (2021), 100055.

[143] Kamarah Pooley and Hayley Boxall. 2020. Mobile dating applications and sexual and violent offending. *Trends and Issues in Crime and Criminal Justice* 612 (2020), 1–16.

[144] Bob Poston. 2009. Maslow's hierarchy of needs. *The surgical technologist* 41, 8 (2009), 347–353.

[145] Urszula Pruchniewska. 2020. "I Like That It's My Choice a Couple Different Times": Gender, Affordances, and User Experience on Bumble Dating. *International Journal of Communication* 14 (April 2020), 18. https://ijoc.org/index.php/ijoc/article/view/12657

[146] Tinonee Pym, Paul Byron, and Kath Albury. 2021. 'I still want to know they're not terrible people': Negotiating 'queer community' on dating apps. *International Journal of Cultural Studies* 24, 3 (May 2021), 398–413. https://doi.org/10.1177/1367877920959332

[147] Emilee Rader, Rick Wash, and Brandon Brooks. 2012. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. 1–17.

[148] Giulia Ranzini and Christoph Lutz. 2017. Love at first swipe? Explaining Tinder self-presentation and motives. *Mobile Media & Communication* 5, 1 (2017), 80–101.

[149] Noopur Raval and Paul Dourish. 2016. Standing Out from the Crowd: Emotional Labor, Body Labor, and Temporal Labor in Ridesharing. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. 97–107. https://doi.org/10.1145/2818048.2820026

[150] Alexandrea J Ravenelle, Erica Janko, and Ken Cai Kowalski. 2022. Good jobs, scam jobs: Detecting, normalizing, and internalizing online job scams during the COVID-19 pandemic. *New Media & Society* 24, 7 (2022), 1591–1610.

[151] Elissa M Redmiles, Yasemin Acar, Sascha Fahl, and Michelle L Mazurek. 2017. *A summary of survey methodology best practices for security and privacy researchers*. Technical Report. University of Maryland.

[152] Elissa M Redmiles, Jessica Bodford, and Lindsay Blackwell. 2019. "I just want to feel safe": A Diary Study of Safety Perceptions on Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 405–416.

[153] Larry D Rosen, Nancy A Cheever, Cheyenne Cummings, and Julie Felt. 2008. The impact of emotionality and self-disclosure on online dating versus traditional dating. *Computers in Human Behavior* 24, 5 (2008), 2124–2157.

[154] Janine Rowse, Caroline Bolt, and Sanjeev Gaya. 2020. Swipe right: The emergence of dating-app facilitated sexual assault. A descriptive retrospective audit of forensic examination caseload in an Australian metropolitan service. *Forensic Science, Medicine and Pathology* 16, 1 (2020), 71–77.

[155] Legacy Russell. 2020. Glitch feminism: A manifesto. *London: Verso* 11 (2020), 29–29.

[156] Teela Sanders and Rosie Campbell. 2007. Designing out vulnerability, building in respect: violence, safety and sex work policy. *The British journal of sociology* 58, 1 (2007), 1–19.

[157] Shruti Sannon, Natalya N. Bazarova, and Dan Cosley. 2018. Privacy Lies: Understanding How, When, and Why People Lie to Protect Their Privacy in Multiple Online Contexts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. 1–13. https://doi.org/10.1145/3173574.3173626

[158] Shruti Sannon and Dan Cosley. 2018. "It was a shady HIT": Navigating Work-Related Privacy Concerns on MTurk. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.

[159] Shruti Sannon and Dan Cosley. 2019. Privacy, Power, and Invisible Labor on Amazon Mechanical Turk. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. 1–12. https://doi.org/10.1145/3290605.3300512

[160] Shruti Sannon and Dan Cosley. 2022. Toward a More Inclusive Gig Economy: Risks and Opportunities for Workers with Disabilities. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.

[161] Shruti Sannon, Billie Sun, and Dan Cosley. 2022. Privacy, Surveillance, and Power in the Gig Economy. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Article 619, 15 pages. https://doi.org/10.1145/3491102.3502083

[162] Morgan Klaus Scheuerman, Stacy M. Branham, and Foad Hamidi. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 155:1–155:27. https://doi.org/10.1145/3274424

[163] Morgan Klaus Scheuerman, Jialun Aaron Jiang, Casey Fiesler, and Jed R Brubaker. 2021. A framework of severity for harmful content online. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–33.

[164] Bhavani Seetharaman, Joyojeet Pal, and Julie Hui. 2021. Delivery Work and the Experience of Social Isolation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 64:1–64:17. https://doi.org/10.1145/3449138

[165] Liesel L Sharabi and John P Caughlin. 2019. Deception in online dating: Significance and implications for the first offline date. *New Media & Society* 21, 1 (2019), 229–247.

[166] Liesel L Sharabi and Tiffany A Dykstra-DeVette. 2019. From first email to first date: Strategies for initiating relationships in online dating. *Journal of Social and Personal Relationships* 36, 11-12 (2019), 3389–3407.

[167] Adam Shostack. 2014. *Threat Modeling: Designing for Security.* John Wiley & Sons.

[168] Mariah Simmons and Joon Suk Lee. 2020. Catfishing: a look into online dating and impersonation. In *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis: 12th International Conference, SCSM 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*. Springer, 349–358.

[169] Julia Slupska and Leonie Maria Tanczer. 2021. Threat modeling intimate partner violence: Tech abuse as a cybersecurity challenge in the Internet of Things. In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*. Emerald Publishing Limited.

[170] Ashley Southall. 2018. He Used Tinder to Hunt the Women He Raped and Killed, Police Say. *The New York Times* (2018).

[171] Elizabeth Anne Stanko. 2003. *The Meanings of Violence.* Routledge London.

[172] Zahra Stardust, Rosalie Gillett, and Kath Albury. 2022. Surveillance does not equal safety: Police, data and consent on dating apps. *Crime, Media, Culture* (2022), 17416590221111827.

[173] Zahra Stardust, Carla Treloar, Elena Cama, and Jules Kim. 2021. 'I wouldn't call the cops if I was being bashed to death': Sex work, whore stigma and the criminal legal system. *International Journal for Crime, Justice and Social Democracy* 10, 3 (2021), 142–157.

[174] Jake M L Stein, Vidminas Vizgirda, Max Van Kleek, Reuben Binns, Jun Zhao, Rui Zhao, Naman Goel, George Chalhoub, Wael S Albayaydh, and Nigel Shadbolt. 2023. 'You Are You and the App. There's Nobody Else.': Building Worker-Designed Data Institutions within Platform Hegemony. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 281, 26 pages. https://doi.org/10.1145/3544548.3581114

[175] Rock Stevens, Daniel Votipka, Elissa M Redmiles, Colin Ahern, Patrick Sweeney, and Michelle L Mazurek. 2018. The battle for New York: A case study of applied digital threat modeling at the enterprise level. In *27th USENIX Security Symposium (USENIX Security 18)*. 621–637.

[176] Maria Stoicescu and Cosima Rughinis. 2021. Perils of digital intimacy. A classification framework for privacy, security, and safety risks on dating apps. In *2021 23rd International Conference on Control Systems and Computer Science (CSCS)*. 457–462. https://doi.org/10.1109/CSCS52396.2021.00081

[177] Elisabetta Stringhi. 2022. Addressing gendered affordances of the platform economy: The case of UpWork. *Internet Policy Review* 11, 1 (2022), 1–28.

[178] Angelika Strohmayer, Rosanna Bellini, and Julia Slupska. 2022. Safety as a Grand Challenge in Pervasive Computing: Using Feminist Epistemologies to Shift the Paradigm From Security to Safety. *IEEE Pervasive Computing* (2022).

[179] 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* (Glasgow, Scotland Uk) *(CHI '19)*. 1–14. https://doi.org/10.1145/3290605.3300882

[180] Angelika Strohmayer, Mary Laing, and Rob Comber. 2017. Technologies and Social Justice Outcomes in Sex Work Charities: Fighting Stigma, Saving Lives. In *2017 ACM SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3025453.3025615

[181] Angelika Strohmayer, Julia Slupska, Rosanna Bellini, Lynne Coventry, Tara Hairston, and Adam Dodge. 2021. *Trust and Abusability Toolkit: Centering Safety in Human-Data Interactions.* Northumbria University.

[182] Chiu-Ping Su and Tsung-Chiung Wu. 2020. The Dark Side of Solo Female Travel: Negative Encounters with Male Strangers. *Leisure Sciences* 42, 3-4 (July 2020), 375–392. https://doi.org/10.1080/01490400.2020.1712277

[183] Lisa Sugiura and April Smith. 2020. Victim blaming, responsibilization and resilience in online sexual abuse and harassment. In *Victimology*. Springer, 45–79.

[184] Jacob Thebault-Spieker, Stevie Chancellor, Michael Ann DeVito, Niloufar Salehi, Alex Leavitt, David Karger, and Katta Spiel. 2021. Do We Fix it or Burn it Down? Towards Practicable Critique at CSCW. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 234–237.

[185] Jacob Thebault-Spieker, Loren G. Terveen, and Brent Hecht. 2015. Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. 265–275. https://doi.org/10.1145/2675133.2675278

[186] 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 *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 247–267.

[187] 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 *CHI Conference on Human Factors in Computing Systems*. 1–15.

[188] Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. 2013. Trafficking Fraudulent Accounts: The Role of the Underground Market in Twitter Spam and Abuse. In *22nd USENIX Security Symposium (USENIX Security 13)*. 195–210.

[189] ThreatModeler. 2019. How to Process Flow Diagram for Threat Modeling. https://threatmodeler.com/how-to-process-flow-diagram-for-threat-modeling/.

[190] Juhu Thukral. 2005. Behind closed doors: An analysis of indoor sex work in New York City. *Siecus Report* 33, 2 (2005), 3.

[191] Julia Ticona. 2015. Strategies of control: Workers' use of ICTs to shape knowledge and service work. *Information, Communication & Society* 18, 5 (2015), 509–523.

[192] Julia Ticona. 2022. Red flags, sob stories, and scams: The contested meaning of governance on carework labor platforms. *New Media & Society* 24, 7 (2022), 1548–1566.

[193] Julia Ticona and Alexandra Mateescu. 2018. How Domestic Workers Wager Safety in the Platform Economy. *Fast Company* (2018). https://www.fastcompany.com/40541050/how-domestic-workers-wager-safety-in-the-platform-economy

[194] Julia Ticona and Alexandra Mateescu. 2018. Trusted strangers: Carework platforms' cultural entrepreneurship in the on-demand economy. *New Media & Society* 20, 11 (Nov. 2018), 4384–4404. https://doi.org/10.1177/1461444818773727

[195] Julia Ticona, Alexandra Mateescu, and Alex Rosenblat. 2018. *Beyond disruption: How tech shapes labor across domestic work and ridehailing*. Technical Report.

[196] Catalina L Toma and Jeffrey T Hancock. 2010. Looks and lies: The role of physical attractiveness in online dating self-presentation and deception. *Communication Research* 37, 3 (2010), 335–351.

[197] Catalina L Toma, Jeffrey T Hancock, and Nicole B Ellison. 2008. Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin* 34, 8 (2008), 1023–1036.

[198] Molly Tran and Rosemary K Sokas. 2017. The gig economy and contingent work: An occupational health assessment. *Journal of Occupational and Environmental Medicine* 59, 4 (2017), e63.

[199] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2020. The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums. In *29th USENIX Security Symposium (USENIX Security 20)*. 1893–1909.

[200] Emily Tseng, Diana Freed, Kristen Engel, Thomas Ristenpart, and Nicola Dell. 2021. A Digital Safety Dilemma: Analysis of Computer-Mediated Computer Security Interventions for Intimate Partner Violence During COVID-19. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Article 71, 17 pages. https://doi.org/10.1145/3411764.3445589

[201] Huahong Tu, Adam Doupé, Ziming Zhao, and Gail-Joon Ahn. 2019. Users really do answer telephone scams. In *28th USENIX Security Symposium (USENIX Security 19)*. 1327–1340.

[202] Deborah Tuerkheimer. 2019. Unofficial Reporting in the #MeToo Era Law in the Era of #MeToo. *University of Chicago Legal Forum* 2019 (2019), 273–298. https://heinonline.org/HOL/P?h=hein.journals/uchclf2019&i=277

[203] Zeynep Tufekci. 2017. Security in the Wild for Low-Profile Activists. https://www.usenix.org/conference/enigma2017/conference-program/presentation/tufekci.

[204] Chad Van De Wiele and Stephanie Tom Tong. 2014. Breaking boundaries: The uses & gratifications of Grindr. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 619–630.

[205] Jessica Vitak. 2015. Balancing privacy concerns and impression management strategies on Facebook. In *Symposium on Usable Privacy and Security (SOUPS)*. 22–24.

[206] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1231–1245.

[207] Stephen Walcott. 2020. *Victimisation and Fear of Crime in the Gig Economy.* Technical Report. The Police Foundation.

[208] Matthias Waldkirch, Eliane Bucher, Peter Kalum Schou, and Eduard Grünwald. 2021. Controlled by the algorithm, coached by the crowd–how HRM activities take shape on digital work platforms in the gig economy. *The International Journal of Human Resource Management* 32, 12 (2021), 2643–2682.

[209] Ari Ezra Waldman. 2019. Law, privacy, and online dating: "Revenge porn" in gay online communities. *Law & Social Inquiry* 44, 4 (2019), 987–1018.

[210] Ari Ezra Waldman. 2021. Navigating Privacy on Gay-Oriented Mobile Dating Applications. In *The Emerald international handbook of technology-facilitated violence and abuse.* Emerald Publishing Limited, 369–381.

[211] Gang Wang, Tianyi Wang, Haitao Zheng, and Ben Y Zhao. 2014. Man vs. machine: Practical adversarial detection of malicious crowdsourcing workers. In *23rd USENIX Security Symposium (USENIX Security 14).* 239–254.

[212] Noel Warford, Tara Matthews, Kaitlyn Yang, Omer Akgul, Sunny Consolvo, Patrick Gage Kelley, Nathan Malkin, Michelle L Mazurek, Manya Sleeper, and Kurt Thomas. 2021. SoK: A Framework for Unifying At-Risk User Research. *arXiv preprint arXiv:2112.07047* (2021).

[213] Mark Warner, Andreas Gutmann, M. Angela Sasse, and Ann Blandford. 2018. Privacy Unraveling Around Explicit HIV Status Disclosure Fields in the Online Geosocial Hookup App Grindr. *Proceedings of the ACM Conference on Human-Computer Interaction* 2, CSCW, Article 181 (Nov 2018), 22 pages. https://doi.org/10.1145/3274450

[214] Mark Warner, Agnieszka Kitkowska, Jo Gibbs, Juan F. Maestre, and Ann Blandford. 2020. Evaluating 'Prefer Not to Say' Around Sensitive Disclosures. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20).* 1–13. https://doi.org/10.1145/3313831.3376150

[215] Mark Warner, Juan F. Maestre, Jo Gibbs, Chia-Fang Chung, and Ann Blandford. 2019. Signal Appropriation of Explicit HIV Status Disclosure Fields in Sex-Social Apps Used by Gay and Bisexual Men. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19).* 1–15. https://doi.org/10.1145/3290605.3300922

[216] Elizabeth Anne Watkins. 2022. "Have you learned your lesson?" Communities of practice under algorithmic competition. *New Media & Society* 24, 7 (2022), 1567–1590.

[217] Monica T Whitty and Tom Buchanan. 2016. The online dating romance scam: The psychological impact on victims–both financial and non-financial. *Criminology & Criminal Justice* 16, 2 (2016), 176–194.

[218] Sarah Widmer and Anders Albrechtslund. 2021. The ambiguities of surveillance as care and control: Struggles in the domestication of location-tracking applications by Danish parents. *Nordicom Review* 42 (2021).

[219] Daricia Wilkinson and Bart Knijnenburg. 2022. Many Islands, Many Problems: An Empirical Examination of Online Safety Behaviors in the Caribbean. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–25.

[220] Mark A Wood, Stuart Ross, and Diana Johns. 2021. Primary crime prevention apps: A typology and scoping review. *Trauma, Violence, & Abuse* (2021), 1524838020985560.

[221] Jane E Workman and Robin L Orr. 1996. Clothing, sex of subject, and rape myth acceptance as factors affecting attributions about an incident of acquaintance rape. *Clothing and Textiles Research Journal* 14, 4 (1996), 276–284.

[222] Yuxi Wu, W Keith Edwards, and Sauvik Das. 2022. SoK: Social Cybersecurity. In *2022 IEEE Symposium on Security and Privacy (SP).* IEEE, 1863–1879.

[223] Shuzhe Yang and Andreas Albers. 2013. Overcoming information overload in online reputation management: A systematic literature review. In *ECIS 2013 - Proceedings of the 21st European Conference on Information Systems.*

[224] Stella Zaryan. 2017. Truth and Trust: How audiences are making sense of Fake News. (student paper).

[225] Angie Zhang, Alexander Boltz, Chun Wei Wang, and Min Kyung Lee. 2022. Algorithmic Management Reimagined For Workers and By Workers: Centering Worker Well-Being in Gig Work. In *CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22).* Article 14, 20 pages. https://doi.org/10.1145/3491102.3501866

[226] Tianqing Zhu, Dayong Ye, Wei Wang, Wanlei Zhou, and S Yu Philip. 2020. More than privacy: Applying differential privacy in key areas of artificial intelligence. *IEEE Transactions on Knowledge and Data Engineering* 34, 6 (2020), 2824–2843.

[227] Yixin Zou, Allison McDonald, Julia Narakornpichit, Nicola Dell, Thomas Ristenpart, Kevin Roundy, Florian Schaub, and Acar Tamersoy. 2021. The role of computer security customer support in helping survivors of intimate partner violence. In *30th USENIX Security Symposium (USENIX Security 21).* 429–446.

[228] Douglas Zytko and Hanan Aljasim. 2022. Designing AI for Online-to-Offline Safety Risks with Young Women: The Context of Social Matching. *arXiv preprint arXiv:2204.00688* (2022).

[229] Douglas Zytko, Nicholas Furlo, Bailey Carlin, and Matthew Archer. 2021. Computer-Mediated Consent to Sex: The Context of Tinder. In *Proceedings of the ACM on Human-Computer Interaction.* 189:1–189:26. https://doi.org/10.1145/3449288

[230] Douglas Zytko, Sukeshini A. Grandhi, and Quentin Jones. 2014. Impression Management Struggles in Online Dating. In *Proceedings of the 18th International Conference on Supporting Group Work* (Sanibel Island, Florida, USA) *(GROUP '14).* 53–62. https://doi.org/10.1145/2660398.2660410

# A APPENDIX

## A.1 Codebook

| Theme | Sub-Theme |
|---|---|
| Financial | Fraud |
| Physical | Physical |
| | Physical Violence |
| | Harassment |
| | Stalking |
| | Robbery |
| | Sexual harm |
| | Health |
| Emotion | Impersonation |
| | Fraud |
| | Emotion |
| | Authentication |
| | Mismatched Expectation |
| Environment | Environment |
| | Public |
| | During Day |
| | Autonomy |

Table 5. Codes resulting from qualitative analysis

## A.2 List of Safety Apps

| Technologies | Key features |
|---|---|
| Noonlight | share details about meeting location and time, create a safety network of friends & family to alert, silently call for help. |
| Kitestring | periodic check-ins, create a safety network of friends & family to alert |
| Circle of 6/Circulo | share details about meeting location, send fake phone call, create a safety network to ask for help |
| Flare | detects physical safety incident (e.g. fall), alerts emergency contacts |
| invisaWear | wearable that can text GPS location to emergency contacts upon triggered, alerts local emergency officials |
| Athena | wearable that can text GPS location to emergency contacts upon triggered |
| Birdie | personal safety alarm with flashing light |
| Sabre | share details about meting location and time with contacts, contact local emergency officials |

Table 6. List of emergency alert apps we included in our survey.

## A.3 Survey Questions

Linked is a PDF version of our survey instrument, including both original and re-fielded questions. Kindly note that this contains survey options for both user populations observed in the study (i.e. daters and gig workers).

## A.4 Survey Analysis

Linked in the file titled "Survey analysis groupings" is a description of how we grouped question responses across each sample to measure the proportions of the mechanisms/resources in Tables 2, 3, and 4.

## A.5 Respondent Demographics

| Gender | Daters | Workers | Census |
|---|---|---|---|
| Woman | 46.0% | 49.2% | 50.9% |
| Man | 49.8% | 48.1% | 49.1% |
| Agender | <1% | <1% | |
| Genderqueer | 1.47% | <1% | |
| Non-binary | 2.94% | 1.33% | |
| Other | <1% | <1% | |
| Prefer not to say | 0% | <1% | |

| Education | Daters | Workers | Census |
|---|---|---|---|
| Less than high school graduate | 1.26% | <1% | 9.8% |
| High school graduate | 16.2% | 11.8% | 27.8% |
| Some college, no degree | 26.3% | 28.2% | 17.5% |
| Associate's degree | 8.61% | 9.31% | 10.1% |
| Bachelor's degree | 35.3% | 38.1% | 22.1% |
| Advanced degree | 12.4% | 12.0% | 12.7% |

| Ethnicity | Daters | Workers | Census |
|---|---|---|---|
| American Indian or Alaska Native | 3.57% | 2.22% | 1.1% |
| Asian or Asian American | 8.19% | 12.4% | 6.0% |
| Black or African American | 9.66% | 15.1% | 12.4% |
| Hispanic or Latino | 10.7% | 12.6% | 18.7% |
| Native Hawaiian or Pacific Islander | <1% | <1% | 0.2% |
| White | 74.2% | 66.1% | 61.6% |
| Other | <1% | <1% | |
| Prefer not to say | <1% | <1% | |

Table 7. Participant demographics. We give gender, education, and ethnicity percentages for our gig worker participants, our online dater participants, and the US population (according to the 2020 Census).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009