

# Repurposing AI in Dating Apps to Augment Women’s Strategies for Assessing Risk of Harm

Isha Datey  
Oakland University  
Rochester, Michigan, USA  
ishadatey@oakland.edu

Hanan Aljasim  
Oakland University  
Rochester, Michigan, USA  
hkaljasi@oakland.edu

Douglas Zytko  
Oakland University  
Rochester, Michigan, USA  
zytko@oakland.edu

## ABSTRACT

In this paper we present emerging findings from an interview study with women in North America about how AI could be designed to prevent online dating-facilitated violence by augmenting their strategies for risk awareness. The study is motivated by gender disparities in harm through online dating, and the relative absence of dating app designs that prioritize women. Findings show that women are receptive to the notion of AI that augments their existing strategies for assessing risk of harm with meeting a particular user face-to-face. They outline various physical and non-physical harms that they attempt to reduce uncertainty about, and a range of data points as indicators of risk that could serve as the basis for a risk awareness AI model. Due to subjectivity in what is considered an indicator of risk, current findings suggest that risk awareness AI, if implemented, should allow women to train their own models.

## CCS CONCEPTS

• **Human-centered computing** → **Participatory design**; • **Social and professional topics** → **Women**.

## KEYWORDS

Social matching, online dating, dating apps, women, feminist HCI, harm, sexual violence, AI, risk

## ACM Reference Format:

Isha Datey, Hanan Aljasim, and Douglas Zytko. 2022. Repurposing AI in Dating Apps to Augment Women’s Strategies for Assessing Risk of Harm. In *Companion Computer Supported Cooperative Work and Social Computing (CSCW’22 Companion)*, November 8–22, 2022, Virtual Event, Taiwan. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3500868.3559472>

## 1 INTRODUCTION

Over 700 million women have been victims of sexual or physical violence in their lifetime [15]. Violence against women is a global issue and one that has caught the attention of HCI researchers seeking to understand the role of computer-mediated communication in gender-based violence [6, 23] and assess technological solutions [20, 24].

A prominent context for computer-mediated harm against women is online dating. The types of harms facilitated by dating apps pervade online and offline modalities, including sexual harassment online [21] and physical sexual harm during face-to-face meetings [25]. Women are disproportionately affected by these harms [2, 19]. Concerningly, research has shown an increase in online dating-facilitated sexual violence over time [1], which we can expect to continue given the broadening use and design of dating apps for social goals beyond dating [9].

Solutions to online dating-facilitated violence against women have not been particularly inventive, and they tend to be reactive rather than preventative as exemplified by user blocking and reporting features, as well as panic buttons [22] that have questionable utility [10].

We draw attention to artificial intelligence (AI) as a valuable design material in light of its use for automating detection of harm in other social computing contexts [11, 17] and the ever-improving capacity to involve stakeholders in design of artificially intelligent applications [12, 16]. While AI has been a core element of dating apps since their inception, it is typically used in recommendation of geographically nearby users for near-instantaneous face-to-face encounters. Advances in AI for safety-driven purposes are relatively rare [18]. Through the lens of feminist HCI [3] we critique the current AI implementations in online dating for implying a universally desirable user experience; one in which increasing discoverability to nearby strangers and increasing the pace towards face-to-face meetings are sought by all users. Pursuant to the feminist HCI qualities of participation and advocacy, it is essential to increase participation of women in design of safety-conscious AI given the disproportionate gendered impact of online dating harm.

Through a series of preliminary focus groups [4] and informal discussions with women, a frequently suggested AI use case for safety was *risk awareness*, or more specifically: an AI model that augments women’s existing strategies for reducing uncertainty about the risk of harm associated with meeting a particular online stranger in the physical world. Uncertainty reduction about potential meeting partners has been a goal of women in online dating for some time [7], however dating app designs to support uncertainty reduction for harm mitigation reasons are underdeveloped.

Towards ensuring that risk awareness AI would meet the needs and expectations of women, this paper presents emerging findings from an interview study with 20 cisgender and transgender women in North America about their current practices for reducing uncertainty about risk of harm in dating apps and how those strategies could be translated into an AI model that works in partnership with women to increase awareness of risk. The study’s research questions include: **RQ1**. *What are women’s existing strategies for*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CSCW’22 Companion, November 8–22, 2022, Virtual Event, Taiwan

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9190-0/22/11.

<https://doi.org/10.1145/3500868.3559472>

*reducing uncertainty around risk of harm during online dating? RQ2. How could these strategies serve as a foundation or blueprint for an eventual risk awareness AI model?*

Emerging findings build on prior research about how online daters' impression formation practices [5] intersect with safety, such as for evaluating deception [7] and unsafe sexual situations [8, 14]. Our study provides evidence that users are receptive to AI augmenting their existing strategies for risk awareness. Findings indicate that a singular risk awareness AI model may be impractical, however, in light of subjectivity and variability in the information that women perceive to be valuable indicators of risk. We thus encourage a pluralist [3] approach to risk awareness AI that allows women to tailor a personal model based on their own conceptualizations of risk.

## 2 METHOD

We have conducted interviews with 20 women to discuss their existing strategies for reducing uncertainty about risk of harm in online dating. During these interviews we also worked with participants to translate their strategies into mental models of inputs, decision rules, and outputs that could serve as the basis for an eventual risk awareness AI model. A total of 20 woman-identifying individuals have participated in the study, who were recruited through word-of-mouth campaigns amongst the social circles of the woman-identifying members of the research team and online advertisements through social media and email lists associated with our university. Ages have ranged from 18 to 41. Participants identified as black (6), white (6), Asian (4), mixed race (2), Native American (1), and middle eastern (1). All participants identified as women, with two voluntarily disclosing during interviews that they were transgender. Participants resided in 11 different states around the United States, and one in Canada. Interviews have ranged from 1 hour and 51 minutes to 2 hours and 19 minutes. All participants had experience with online dating through a range of applications (some of which were not necessarily intended for dating), including: Tinder (9), Bumble (4), Instagram (4), OkCupid (3), Hinge (3), Meetup (2), Facebook and Facebook Dating (2), Snapchat (1), Patio App (1), Coffee Meets Bagel (1), Muzmatch (1), Match (1), Christian Mingle (1), eharmony (1), and Zoosk (1).

Interviews began by having participants broadly reflect on their experiences with using online dating, followed by encouragement to expand on a particular experience that they found risky or that culminated in harm. This was used as a springboard for the participant to articulate what qualifies as harm, and the strategies that they practice—if any—for assessing such risks. While a participant unpacked their conceptualizations of risk and strategies for risk awareness the research assistants visually documented corresponding information into what we call a “human intelligence model” (in contrast to an artificial intelligence model). This carries inspiration from Lee and colleagues' approaches to participatory AI model building [12, 13]. We recurrently asked the participant for feedback on the model. Afterwards we introduced the concept of AI and how it is typically incorporated into dating apps, followed by reflection on if and how the participant's risk awareness strategies could be augmented by AI. All interviews were audio recorded and transcribed. The transcripts are in the process of being subjected

to open coding in which we apply line-by-line coding in Dedoose and organize quotes via Miro to determine emergent themes.

## 3 FINDINGS

All participants have been receptive to the idea of AI augmenting their strategies for reducing uncertainty about risk of harm with meeting an online dater face-to-face. Our ongoing analysis has shown that women attempt to reduce uncertainty about a range of harms that fall into two categories: physical harm and non-physical harm. The indicators of risk that they assess in dating apps, which would serve as factors in an eventual AI model to augment their risk awareness, pertain to information about the potential meeting partner but also the location of a potential meeting. Some of these predictors are objective in nature, such as crime rate of the surrounding area, while others are more subjective and informed by women's past experiences with harm such as political views of the meeting partner.

### 3.1 Reducing Uncertainty About Physical Harm

Participants primarily conceptualized physical harm as sexual, such as unwanted bodily touch or sexual assault. Those that identify as transgender were also concerned with physical violence due to their gender identity.

Information about an eventual meeting location was a strong focus for many women in order to reduce uncertainty about risk of physical harm associated with, or amplified by, the location. They typically had to resort to memory or third party data collection (e.g., Google Maps, restaurant reviews, opinions of friends) to perform this risk assessment. Common examples included crime rate of the surrounding area as well as presence of alcohol because some associated that with potentially reckless behavior of bystanders and reduced inhibitions of their meeting partner. Some women also reduced uncertainty about physical risk by their perceived capacity to eject from a harmful situation through easily accessible exits and the walkability of a meeting location. As P11 described: *“I can leave when I want to, it's not like I'm going to be locked in there. It's just the ability to freely come and go, I guess is one of the things that can make a person feel safe.”* Some women attempted to assess likelihood of families being present at the meeting location, particularly those with children, because they considered these people to be safe bystanders to turn to should they need emergency assistance. According to P2, they could also deter the meeting partner from trying to inflict harm: *“The most dead giveaway when it comes to safety is regardless of how ill-intentioned a person might be it's very difficult to expect that anyone is going to do anything that stupid, crazy or awful in front of children.”* Relatedly, a few women mentioned cellphone reception as impactful to their assessment of risk because it is their only mode of contacting friends or family if a meeting partner attempts to inflict harm. P2: *“I'm a city girl, I feel safer in the city, because my brain sort of already assesses those risks. You know, I have my phone, I know where to go. [...] if I was in a remote area, I would feel very uncomfortable. For me, if there's no cell reception, that's very remote.”*

Women in the study reported meeting online daters for a variety of reasons, beyond just dating, which leads them to meeting users of different gender identities. They consider a user's gender as crucial

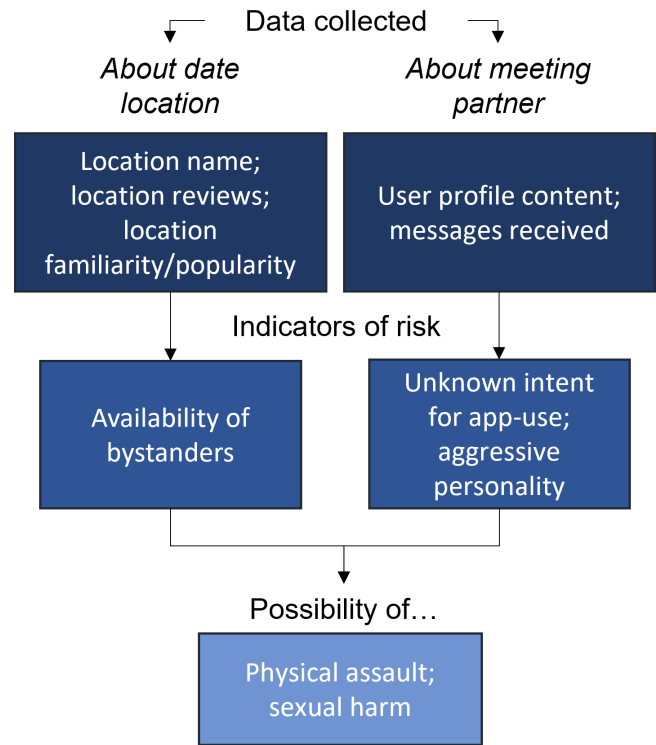
to assessing risk, with cisgender men universally being considered higher risk than users of any other gender identity. In P12’s words: *“In my experience, misogyny is alive and well and flourishing. And there are a lot of men out there [...] just being really hateful towards women for not meeting like this standard, I guess.”*

As implied by P12’s quote, indicators of risk were often informed by past dating and social experiences. Another example is political views, particularly when a potential meeting partner exhibits opposing political views through messaging in the dating app or their profile. Some women considered political views as reliable indicators of behavior towards marginalized groups and stances on issues germane to dating such as abortion rights and LBGTQ rights. P7 described an aversion to meeting users with conservative political views for these reasons: *“If there’s something that really I would disagree with, politically or socio politically, so they are sort of like human values, just being a respectful human being, a kind human.”* Education level was also mentioned by a few participants as a subjective indicator of physical risk. These participants tended to draw perceived connections between education level and other risk factors, such as likelihood of drug use or engagement in crime. As P4 indicated: *“When someone is more educated, I think they’re more trustworthy, because they, from a crime standpoint aren’t [...] going to be leaning on drugs for money, or probably interacting really closely with people that have a crime-based lifestyle.”*

Several women honed on perceived intentions for dating app-use as one of the most important uncertainty reduction factors for physical risk. They paid particular attention to profile content and utilized messaging interaction to assess whether a potential meeting partner was primarily using the dating app for casual sexual encounters. This was considered a strong indicator of risk by women who were not using dating apps for sexual reasons because an incompatibility in reasons to meet face-to-face could result in a forced sexual encounter. Data that women considered indicators of casual sexual intentions went beyond blatant statements of interest in sex and included more subtle signals such as the perceived effort to construct one’s profile. As P5 described: *“If you can’t be bothered to put in a little bit of effort [to properly complete a profile] in order to like, meet your future boyfriend, girlfriend? Like, what’s your real, um, like, what’s your real end goal for being on this site?”* Women who did acknowledge sex as a motivation for app-use still considered their assessment of a meeting partner’s intentions to be important to risk awareness. They were particularly concerned about users who did not clarify their intentions because it prevented them from reducing uncertainty about the user’s behavior during a potential face-to-face meeting.

### 3.2 Reducing Uncertainty About Non-Physical Harm

Participants spent ample time in interviews describing harms that are unrelated to physical violence. Some examples involved emotional harm stemming from hurtful comments during face-to-face dates, particularly around gender as mentioned by our transgender participants. P8, who identified as transgender, described this risk most succinctly: *“Risk isn’t only about physical safety, but it’s also about emotional safety in a way, not wanting to get one’s feelings hurt or one’s [self] confidence shattered.”* Gender played a role in



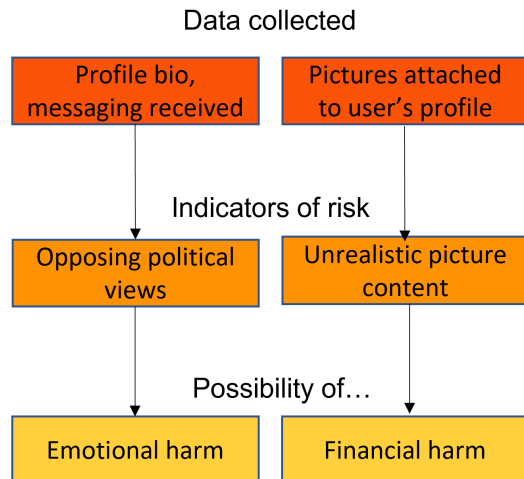
**Figure 1: A visual model for how a participant reduced uncertainty around risk of physical harm with meeting a given user. The box at the bottom includes harms they want to avoid, the middle boxes are indicators that reduce or elevate risk of such harms, and the boxes at the top are data that inform the indicators.**

reducing uncertainty around emotional harm, but in a different way than described for physical harm. P9, who is transgender, described how cisgender women sometimes consider *them* to be a threat because of their physical stature, which can lead to emotional harm through verbal comments about their appearance: *“...but as far as like usually with meeting women, especially cis women online, [...] it is more likely that it is you who will be perceived as the threat and that is what reads as a more immediately dangerous circumstance. It’s people making assumptions based on your appearance. [...] Instead of perceiving physical risk to yourself, you’re thinking like I could be perceived as threatening or I could be judged negatively.”*

Another non-physical harm mentioned by participants was reputational damage, particularly by vengeful meeting partners in retaliation for a rejected sexual or romantic advance. Some considered how social media could be used to disseminate hurtful and inaccurate comments that may jeopardize existing social and professional relationships. In P9’s words: *“One more aspect of risk could be reputation. People can say terrible things [on social media]. And if someone is upset, [...] maybe the person didn’t give them the validation that they were looking for [...] and they want to try to get back at them, then they could try to publicly slander them.”* Participants most often attempted to reduce uncertainty of this risk through

assessment of political views and “aggressive” personality as interpreted through messaging interaction. For example, P12 assessed personality through demands that a user makes over messaging, such as for physically revealing pictures: “I see the personality is like aggressive [when] some of the statements are demanding. They ask for certain things that I feel are maybe like inappropriate. Those are the kind of things I watch for.”

A commonly mentioned non-physical harm was financial scams, or the tricking of users into sending money to a perpetrator under a false premise. All women that mentioned this harm attempted to reduce uncertainty of financial scams through assessment of the potential for catfishing or fake profile content. Beyond obvious indicators like “picture verification” badges that are already included on profiles, participants mentioned detection of profile content that seemed “too good to be true.” P17 described such an experience: “One of the pictures that was used [in a profile] was of a man holding a \$25 million check that he had won from the lottery. And I thought, oh, no, this is like somebody posing as someone else so that they can have all these women who need money [...] send them their bank account information.”



**Figure 2: A visual model for how a participant reduced uncertainty around risk of non-physical harm with a given user. The boxes in the bottom layer include harms, the middle boxes are indicators of risk, and the boxes at the top are data that inform the indicators.**

#### 4 DISCUSSION AND FUTURE WORK

In this paper we presented early findings from an interview study with women to inform the design of AI that augments their existing strategies for reducing uncertainty about risk of harm in online dating. The variability and subjectivity in how women assess risk as discovered through the study would suggest that a singular risk awareness AI model is impractical, nor would one be encouraged

through a pluralist lens from feminist HCI [3]. Risk awareness AI should instead allow users to train their personal model to reflect their subjectively valued indicators of risk. Towards supporting women in producing personalized models for risk assessment our ongoing research involves participatory model building with women and data donation to allow them to demonstrate content that indicates risk.

#### REFERENCES

- [1] National Crime Agency. 2016. Emerging new threat in online dating: Initial trends in internet dating-initiated serious sexual assaults.
- [2] Monica Anderson, Emily A. Vogels, and Erica Turner. 2020. Online Dating: The virtues and downsides of online dating. *Pew Research Center: Internet, Science amp; Tech* (Oct 2020). <https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/>
- [3] Shaowen Bardzell. 2010. Feminist HCI: Taking Stock and Outlining an Agenda for Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (CHI '10). Association for Computing Machinery, New York, NY, USA, 1301–1310. <https://doi.org/10.1145/1753326.1753521>
- [4] Caroline Bull, Hanan Aljasim, and Douglas Zytko. 2021. Designing Opportunistic Social Matching Systems for Women’s Safety During Face-to-Face Social Encounters. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) (CSCW '21). Association for Computing Machinery, New York, NY, USA, 23–26. <https://doi.org/10.1145/3462204.3481751>
- [5] Nicole Ellison, Rebecca Heino, and Jennifer Gibbs. 2006. Managing impressions online: Self-presentation processes in the online dating environment. *Journal of computer-mediated communication* 11, 2 (2006), 415–441.
- [6] Ryan J Gallagher, Elizabeth Stowell, Andrea G Parker, and Brooke Foucault Welles. 2019. Reclaiming stigmatized narratives: The networked disclosure landscape of# MeToo. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–30.
- [7] 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.
- [8] Mark J Handel and Irina Shklovski. 2012. Disclosure, ambiguity and risk reduction in real-time dating sites. In *Proceedings of the 17th ACM international conference on Supporting group work*. 175–178.
- [9] Joey Chiao-Yin Hsiao and Tawanna R Dillahunt. 2017. People-nearby applications: How newcomers move their relationships offline and develop social and cultural capital. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 26–40.
- [10] 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.
- [11] Seunghyun Kim, Afsaneh Razi, Gianluca Stringhini, Pamela J Wisniewski, and Munmun De Choudhury. 2021. A Human-Centered Systematic Literature Review of Cyberbullying Detection Algorithms. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–34.
- [12] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, et al. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.
- [13] Min Kyung Lee, Ishan Nigam, Angie Zhang, Joel Afriyie, Zhizhen Qin, and Sicun Gao. 2021. Participatory algorithmic management: Elicitation methods for worker well-being models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 715–726.
- [14] 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.
- [15] World Health Organization et al. 2021. Violence against women prevalence estimates, 2018: global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women. (2021).
- [16] Afsaneh Razi, Ashwaq AlSoubai, Seunghyun Kim, Nurun Naher, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. 2022. Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.
- [17] Afsaneh Razi, Seunghyun Kim, Ashwaq AlSoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J Wisniewski. 2021. A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. *Proceedings of the ACM on Human-Computer*

- Interaction* 5, CSCW2 (2021), 1–38.
- [18] Nicolas Rivero. 2021. Tinder is using AI to monitor DMS and Tame the creeps. *Quartz* (May 2021). <https://qz.com/2011998/tinder-is-using-ai-to-monitor-dms-and-tame-the-creeps/>
- [19] 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.
- [20] Muhammad Yasir Sarosh, Muhammad Abdullah Yousaf, Mair Muteeb Javed, and Suleman Shahid. 2016. Mehfoozaurat: Transforming smart phones into women safety devices against harassment. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*. 1–4.
- [21] Laura Thompson. 2018. "I can be your Tinder nightmare": Harassment and misogyny in the online sexual marketplace. *Feminism & Psychology* 28, 1 (2018), 69–89.
- [22] Tinder. 2020. Tinder introduces safety-focused updates. *Tinder Newsroom* (Jan 2020). <https://blog.gotinder.com/tinder-introduces-safety-updates/>
- [23] 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.
- [24] Chaeyoon Yoo and Paul Dourish. 2021. Anshimi: Women's Perceptions of Safety Data and the Efficacy of a Safety Application in Seoul. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.
- [25] Douglas Zytko, Nicholas Furlo, Bailey Carlin, and Matthew Archer. 2021. Computer-mediated consent to sex: the context of Tinder. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.