Safe Digital Intimacy: A Research Agenda

Vaughn Hamilton*[1], Gabriel Kaptchuk*[2], Allison McDonald*[2], and Elissa M. Redmiles*[1]

[1]Max Planck Institute for Software Systems, {vhamilto,eredmiles}@mpi-sws.org
[2]Boston University, {kaptchuk,amcdon}@bu.edu

1 Introduction

People engage in sexual intimacy online in a variety of ways. For example, they may send intimate images to each other (sext) or they may work in or consume content produced by the sex industry. Engaging in digital intimacy is common: an estimated 1 in every 200 people has been a sex worker[1] in their lifetime and more than 80% of adults[2] have sexted.

Many risks plague online sexual intimacy: intimate content can be stolen and republished [7], online platforms are hostile to sex workers [3, 5, 6], and ensuring one's safety when interacting with someone sexually online can be challenging. Yet, safety while engaging in digital intimacy is necessary for authentic online engagement.

Those engaged in digital sexual intimacy—whether for work (sex workers) or recreation (sexters)—share many of the same risks and behaviors; however, the risks of digital intimacy are typically more pronounced, in volume and severity, for sex workers. We therefore focus on digital sex work as a lens through which to understand digital intimacy threats and elucidate potential research directions to mitigate them. By centering the most marginalized users when building solutions, we can increase safety for all.

In this piece, we provide background on digital sex work and present *threat models* characterizing the risks sex workers face. To develop these threat models, we draw on the rich academic and community-led work investigating sex workers' safety (e.g., [13, 3, 2, 12, 4, 5]).[3] We then provide 11 directions for cryptographic and systems-security research that could improve the security and privacy of sex workers. We conclude by discussing how these threat models and research directions may benefit other groups including those engaged in recreational digital intimacy.

We contextualize this piece with four points. First, this article addresses only digital intimacy initiated between consenting adults. Second, the article centers the voices of sex workers and those engaged in other forms of digital intimacy by aggregating empirical research that has directly studied these populations. Additionally, we engaged a sex worker consultant to develop and inform this paper at every level. Any research or technical development inspired by this article should similarly center the population served. Third, the legislative environment in which sex work and digital intimacy sit is complex. Sex work exists under a variety of legislative models ranging from legalized, regulated, decriminalized (regulated under regular labor law), to criminalized—for either the worker, the client, or both [14]. Recreational digital intimacy is largely unregulated, although there is growing legislative consideration of its consequences, e.g., sharing intimate

[*] Following the norm in math and theoretical computer science, authors are listed alphabetically.

[1] Sex work is broadly the exchange of sexual services for money [14], and includes work such as escorting, stripping, webcamming (performing online private or public shows ranging from dancing to pornographic performances [10]), sexting, and content production and sale, which includes both digital goods—e.g., images and videos—and physical goods such as used clothing.

[2] The most recent rigorous estimate of sexting behavior was produced in 2015: https://www.apa.org/news/press/releases/2015/08/common-sexting.

[3] The literature we draw upon is broader than we can cite; for further reading, we compiled a living reading list here: https://github.com/VaughnHamilton/SW_Research

imagery without the subject's consent.[4] Regardless of the legislative contexts, digital intimacy of any form is stigmatized, especially for women and when engaged in as labor; thus, victims are often framed as responsible for their own victimization [11]. Finally, the stigma of engaging in digital intimacy is highly intersectional, magnifying existing marginalization, inequity and potential harms such that the digital risks faced by many sex workers and others who are engaged in digital intimacy are not just related to their activities but to their very identity [11, 14]. Work in this space should be thoughtful about how additional identities such as being a person of color, LGBTQ+, a migrant, and/or disabled may make solutions more or less safe.

2 Background: Sex Work and Technology

Our goal in this section is to equip readers unfamiliar with sex work—or readers who have only encountered highly stylized and often stereotypical depictions of sex work in media—with the necessary background to understand our proposed research agenda.[5] While we do not attempt a comprehensive description of sex work, we do highlight the way sex workers leverage technology. In providing this overview, we make two choices: (1) Throughout this work we omit discussion of digital identity related technologies, such as technologies and techniques used to ensure consistent identity across platforms (e.g., age verification). These approaches are well-covered in prior work and have observed little meaningful progress. (2) We intentionally do not name existing platforms used by sex workers in an effort not to attract undue attention, particularly from law enforcement. Instead, we include mock-ups of the described platforms in Figure 1 to help visualize these platforms.

2.1 Sex Workers' Technology Use

Sex workers use technology to advertise, communicate with clients, vet clients, offer digital services [9], find peer support and learn harm-reducing information, e.g., about health and digital security [2, 12].

Client Acquisition: Advertising. Digital advertising is an important tool for sex workers to attract new clients (see Figure 1). Advertisements often contain photos and may list or allude to services, prices and client contact procedures. Sex workers post advertisements to both mainstream and sex-work specific platforms (Figure 1a) and sometimes maintain personal websites (Figure 1b).

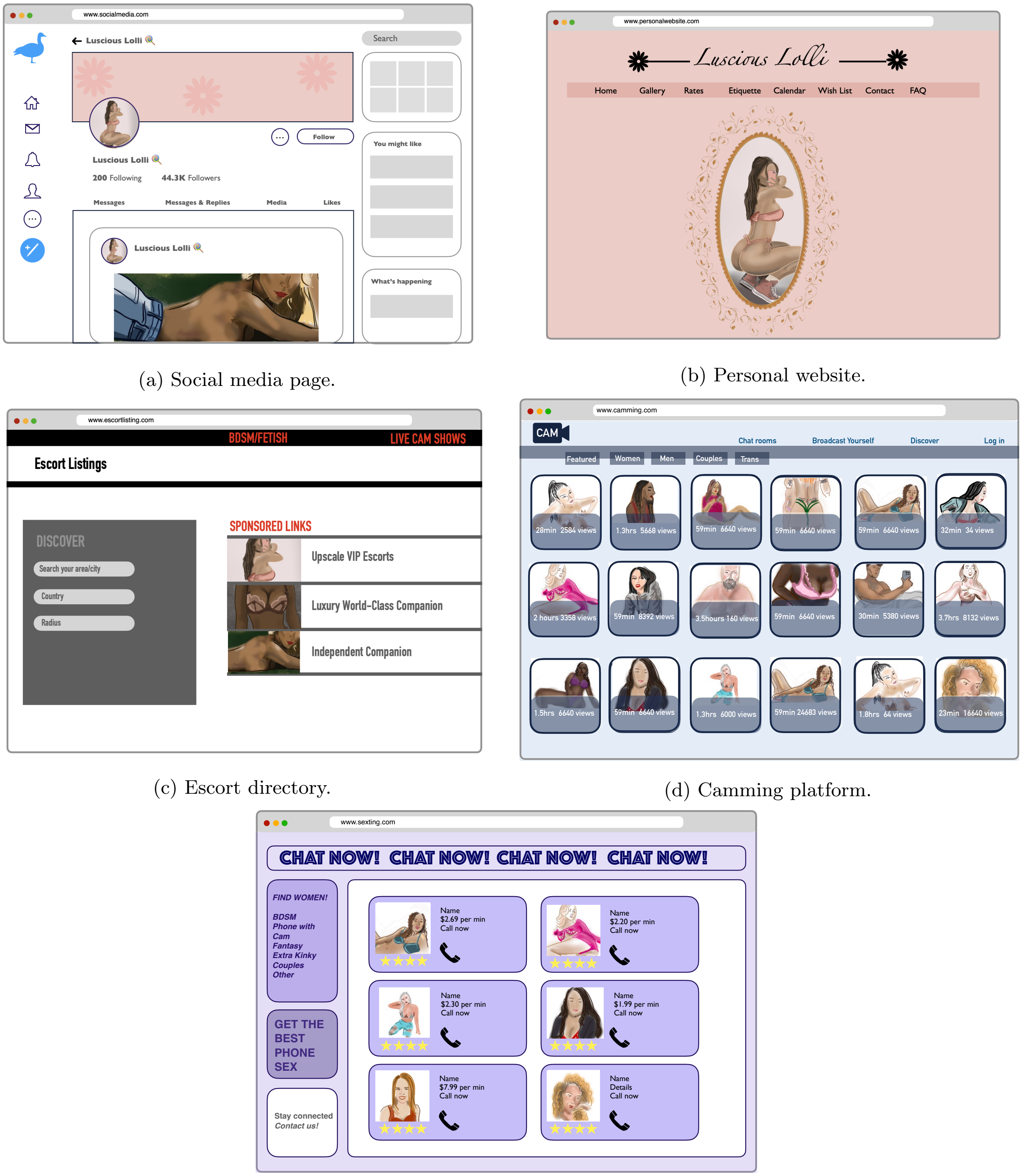
Sex workers working for an employer (such as a brothel, strip club, or escort service) may leverage the organization itself to advertise (Figure 1c). Similarly, some sex workers work on digital platforms like camming platforms or phone sex services, that may promote the sex worker's content or profile to prospective clients through advertisements on porn websites or listings directly on the camming site (Figure 1d and Figure 1e). Workers may have the option to pay to promote their listing (e.g., per click or for a particular rank in the listings).

Client Acquisition: Vetting. Before interacting with clients in person, sex workers will generally attempt to ascertain the risk associated with meeting a new potential client. Vetting strategies vary but may include requiring state-issued identification, a workplace profile, or references from colleagues, either provided directly or through a vetting platform [13, 12].

Client Services. When providing services in person, sex workers leverage technology to maintain physical safety. Many use a practice called "covering," where they share their whereabouts and booking information with a colleague or trusted contact—either with location sharing through their phone or describing their

[4] See `https://cybercivilrights.org/nonconsensual-pornagraphy-laws/` and `https://revengepornhelpline.org.uk/information-and-advice/about-intimate-image-abuse/intimate-image-abuse-laws/`

[5] We note that a deeper understanding of sex work is necessary for actively pursuing some of the proposed research; interested readers can find additional resources at `https://github.com/VaughnHamilton/SW_Research`



(a) Social media page.

(b) Personal website.

(c) Escort directory.

(d) Camming platform.

(e) Phone sex website.

Figure 1: Stylized visualizations of sex work platforms.

booking details via a messaging app—and check in after the booking. Some apps and wearables offer similar functionality but are typically aimed at recreational online dating.[6]

Technology is also involved in creating and distributing intimate digital media (e.g., cameras, media editing applications, file hosting services, sex-work advertising platforms). For synchronous services, creation of content happens simultaneously with distribution, often mediated by a sex work platform that either offers its own video or phone calling infrastructure or links worker profiles to other video streaming platforms. Asynchronous services usually involve a sex worker creating video, audio and image content (and potentially physical goods such as used clothing) and then later listing that content for sale on an adult or mainstream platform (e.g., Figure 1a). Distribution of this content might be managed through the platform or manually by the sex worker (e.g., sharing password protected files hosted on cloud storage accounts).

Client Communication and Client Maintenance. Sex workers use many traditional communication platforms (e.g., email, SMS, and direct messaging) to communicate with clients. Due to the marginalized nature of sex work, some sex workers opt to use adult-specific platforms to communicate with clients or use technologies that provide stronger privacy protections, like end-to-end encrypted email services and messaging apps. Sex workers may choose privacy preserving systems because mainstream services might forbid "illicit or immoral" communication, regardless of whether that communication is work-related or personal. Sex workers may also use technology to communicate with clients while providing services, for example using chatbots, messaging scheduling, or even hiring human assistants or moderators to reply to messages on their behalf.

Payments. Sex workers use multiple methods for payment including cash, gift cards (digital or physical), payment or banking apps, checks or direct deposits (e.g., from a sex work platform), through platforms through which they offer digital services, and to a limited extent direct payment processing. Many major payment platforms forbid payments associated with sex work, regardless of whether that work is conducted legally. As a result, many transactions for in-person work are conducted in cash. Sex workers may also encourage clients to buy them gifts, accept gold, cryptocurrencies, or access to the client's own credit cards, bypassing interacting with payment processors or platform intermediaries altogether. Workers' choice of payment method is influenced by client preferences, workers' need for privacy, and what payment apps or platforms the worker can access.

Community. Sex workers also use technologies such as groups in messaging apps and mainstream and adult forum platforms to communicate with each other. Their primarily text-based communications range from water-cooler workplace conversation, to peer support for serious safety issues, to organizing for advocacy and justice [2].

2.2 Case Studies

The following stylized case studies of sex worker experiences serve to recall the human factors that drive and dictate the use of the technologies described above and illustrate the potential for harm.

Case Study 1: Digital-Only Sex Work

> K spent the last few summers as a food server. They were surprised to find food service wasn't quite what they expected. They got tipped better when they flirted with customers and dressed in ways that aligned with gender- and beauty-norms. Delivering food correctly seemed less important than personality. K spent a lot of their free time on <social media platform> and saw people talking about selling nude photos through <social media platform>. K describes themselves as pretty confident and comfortable with their sexuality. They felt selling these photos was not much different from their food service work and might pay better, so they decided to try it. They created a new account and successfully started selling nudes directly via DM. Eventually, they started seeing posts that others had started accounts on <sex work platform> through which

[6] See, for example, Flare: `https://getflare.com/`.

people could more easily buy their content and even buy a subscription. K decided to start an account on this platform.

K is comfortable with what they do and has made quite a lot of money doing it. Because they worried they could not control who would discover their work or see their content, K let their family know what they were going to do before starting. They also tried to keep people in their personal networks from finding out by proactively blocking contacts from seeing their sex work account. Eventually contacts did find their content and harassed them for it. They ignored the haters, but wish there was some way they could protect themselves more thoroughly.

Case Study 2: Risks of Being Outed

M has provided BDSM services to clients in-person for many years. They are a permanent resident of the country in which they work and legally registered to run their BDSM business. They immigrated from a country where being queer is criminalized. If anyone from their home country were to learn about their business, it might create a risk of violence for their family still living there and prevent M from returning. Thus, they take great care—using multiple different phones, wiping any relevant information from their phone before crossing borders, being vigilant about their social media posts and likes—to ensure nothing connects their personal identity to their work. M wishes they could be "out" (public) about their sex work within the country in which they live so they could feel more visible in their community and could do activism to improve working conditions for other sex workers, but the risk is simply too high.

Case Study 3: Whatever it takes

R is a single parent and has custody of her two toddlers every other week. She needs a source of income that allows her to spend that whole week with her children. In the week when she doesn't do childcare, she has experimented with different kinds of sex work to try to find the one with the most flexibility and highest income. She crossed state borders to strip outside her community, worrying that if someone who knew her found out they might tell her ex-husband and she could face losing custody. She sometimes accepts high value jobs from a discreet escort agency that doesn't require her to share pictures of her face. For these, she flies interstate or abroad. When these jobs were slow, she opened a content sale account, sharing heavily-edited lewds and nudes of herself, but she blocks her whole home state and the state where her ex lives. She promotes her content account lightly on social media but doesn't share pictures of her face, and carefully removes metadata from all her photos.

3 Mechanisms of Harm

Prior work [12] details seven components of safety for sex workers:

- **Boundaries**: ability to enforce boundaries about what type of services are provided, or to whom a worker will not provide services (e.g., a blackmailing landlord).
- **Financial**: ability to not have content stolen, to not lose access to mainstream platforms on which they promote their business, to purchase services without stigma.
- **Physical**: not being assaulted by law enforcement or clients.
- **Privacy**: ability to decide who knows they are a sex worker or their sexual identity. Privacy also includes the ability to control what information—like location or legal name—a client knows about them. Lack of privacy can lead to reduced autonomy (e.g., being banned from entering certain countries, including country of origin) or physical and emotional safety (e.g., stalking).

– **Legality**: ability to work without fear of legal repercussions.

– **Respect**: protection from discrimination and stigma.

– **Access to Community**: ability to build community with other sex workers.

Technology can both cause and facilitate harm. Attackers may use technology to create safety violations through these *mechanisms of harm*:

– **Deplatforming**: having an account or content removed or suppressed (e.g., shadowbanning [6]) from a digital platform. This violates *financial* safety if a worker loses the advertising base they have built and cannot conduct business or access funds and prevents *access to community* if a worker is not able to connect with others.

– **Payment Inoperability**: inability to receive payment from clients or pay for necessary goods and services because payment formats are incompatible.[7] This violates *financial* safety by preventing workers from being able to pay for essentials such as food and housing, and may violate *privacy* by forcing a sex worker to reveal personal information in order to receive payment (e.g., by bank transfer).

– **Outing & Context Collapse**: having one's intimate content, identity as a sex worker (or related information such as sexual identity), or personal information (e.g., address, legal name) exposed without consent. Such *privacy* violations can threaten *physical* safety in cases of stalking or being outed to family.

– **Content Theft**: having content (e.g., images, videos, or ad copy) stolen or republished without consent,[8] which can violate *boundaries* and in turn threaten *physical* safety if a client is led to expect a service the worker does not offer from a fraudulent ad. Content theft also affects *financial* safety by directly stealing revenue from the worker, and can violate *privacy* if images are republished in a place that is likely to out the worker.

3.1 Harm Surfaces

Prior work (e.g., [13, 3, 2, 12, 4, 5]) reveals a number of ways in which the technologies sex workers use can be a source of harm. Here, we summarize the categories of *technological interfaces* involved in sex work that require implicit trust relationships (illustrated in Table 1).

Devices. Sex workers have digital interactions with a wide number of people, some of whom are interested in stalking, doxxing, or obtaining information from them. As a result, their devices (e.g., cell phones, computers, and/or cameras) may become infected with malware, spyware, or ransomware. Device compromise may out a worker by exposing personal information or leaking intimate media.

In addition to device compromise, those working in-person may be harmed by their clients' devices, which could be used to secretly record their engagement and distribute the resulting content without the worker's knowledge or consent. Further, a sex worker may be forced to surrender devices to law enforcement for inspection (e.g., when crossing international borders), which may out them as a sex worker, putting them in legal and physical danger.

Finally, devices may pose usability challenges. For example, to avoid potential cross over of identifying information (IP address, MAC address, Bluetooth identity, etc.) between sex work and non-sex work accounts and profiles some sex workers use multiple devices [12], a burdensome practice that is difficult to implement perfectly.

[7] Workers paid through gift cards cannot use them to e.g., pay for rent or medical care.

[8] A person's initial act of sharing intimate content, either as work or recreation, does not condone later unfettered sharing on the part of the initial recipient. Similarly, engaging in intimate activity in one context (i.e., on a sex work platform) does not automatically indicate consent to have that information made public.

Table 1: A summary of the technologies used by sex workers and the purposes for which they are used, as well as the mechanisms through which they may cause harm and what harms may be caused.

	Business Technologies	Safety Strategies & Technologies	Mechanisms of Harm	Harms
Client Acquisition	Personal Website Mainstream Platforms Social Media Sex Work Platforms Devices Media	Vetting via: Professional Credentials (e.g., work ID) References (e.g., from other sex workers) Sex Work Platforms or Bad Client Lists Paywalls Age Gates Using Multiple Accounts & Devices Self-Censorship	Deplatforming Content Theft Outing & Context Collapse	Financial Privacy Physical Boundaries
Client Services	Payment Platforms Mainstream Platforms File Sharing Services Streaming Services Sex Work Platforms Devices Media	Covering via. . . Alarm Apps Messaging Apps & SMS Security Cameras On-Platform Blocking Preventing Unauthorized Sharing via... Google Alerts Incorporating Client Name into Media Watermarking Paywall Age Gate	Deplatforming Payment Inoperability Content Theft Outing & Context Collapse	Financial Privacy Physical Boundaries
Client Communication	Mainstream Platforms Email SMS & Messaging Apps On-Platform Direct Message Social Media Sex Work Platforms On-Platform Direct Message Live chat (e.g., during a virtual show) Chatbots Devices Media	Using Multiple Accounts & Devices Encryption (Email & Messaging Apps) Limit Available Communication Channels Blocking Human or Automated Moderators	Deplatforming Content Theft	Financial Physical Boundaries
Payments	Cash Gifts/Gift Cards (Wishlists) Cryptocurrencies Payment Platforms (e.g., Venmo, Cashapp)	Using Multiple Accounts & Devices Self-Censorship	Deplatforming Payment Inoperability Outing & Context Collapse	Financial Privacy
Community	Mainstream Platforms Messaging Apps Social Media Sex Work Platforms Forums	Using Multiple Accounts & Devices Self-Censorship	Deplatforming	Community

Media. Media (images, videos, audio recordings, personal webpages, advertising copy) can be a source of harm when they are non-consensually produced or shared. Further, contextual information (e.g., image backgrounds) and metadata may be used to out a worker.

Mainstream Platforms. Many mainstream platforms have a history of selectively enforcing policies to the detriment of sex workers [3, 9]. This includes abruptly deplatforming sex workers or failing to support sex workers targeted by online harassment and abuse. Because platforms rarely provide clear guidance on what content will lead to deplatforming, sex workers must gamble on what content they can share on mainstream platforms to maximize their followings without risking deplatforming.

Mainstream social media platforms also pose privacy risks for sex workers. A worker may be outed by a platform recommending they connect with clients on their personal social media, or advertising their work accounts to family. While there are strategies for preventing this (e.g., proactive blocking, as K did in Case

Study 1), they are not 100% effective. Context collapse is difficult to predict and prevent because there is very little transparency around the data aggregation and prediction tools used by these platforms.

Sex Work Platforms. Sex work platforms may fail to implement protections to prevent content theft or to allow workers to effectively screen and remove harassing clients. Additionally, platforms often require government ID and/or legal name and address from workers for age verification and payment. Breaches of platform information could out a worker by leaking these details, particularly linked together with a worker's work persona and/or content, which are easily identifiable to their clients and may also out them to others in their life.[9]

Scraping. The threat of scraping, a consequence of aggregating sex-work-related content, is significant for sex workers. There is evidence that sex-work-related content is regularly compiled in massive databases. For example, anti-trafficking organizations, corporate entities, or law enforcement may scrape workers' personal websites and other content to create databases of worker information for use in criminal prosecution, mass-text campaigns, deplatforming,[10] and border control [4].

Payment Platforms. Payment apps may become inoperable, freeze payments, or deplatform sex workers. This is true regardless of the legality of the services paid for and is not required by U.S. law.[11] While it is not known how precisely payment processors flag payments, hypotheses include keywords in the notes provided with payments, patterns of payments, and networks of payment connections. In the U.S., specifically, banks may also freeze or delay sex workers' payments or may close their accounts (often retaining the money stored in them) altogether.

4 Concrete Opportunities for Technical Research

We present several opportunities for research to support safer digital intimacy.

4.1 Deplatforming

As explained above, deplatforming is a significant risk to sex workers, which suggests multiple research directions, including:

(1) **Filter Analysis.** Payment platforms and other mainstream platforms deploy automated scanning technology that flags accounts it determines are associated with sex workers.[12] Unfortunately, the uses of filtering technologies are not well understood. An intriguing line of research would be generating tools that test the behavior of a filter and continue to adapt to the filter as it changes. In the hands of a sex worker, this tool could increase the probability that their transactions or profiles would not be flagged. Platforms themselves could even consider offering these pre-screening tools, as content creators and platforms may be aligned in their goals: content-creators seek not to lose their accounts over disallowed content and platforms seek to avoid having such content on the platform [9]. However, for more circumvention-oriented tools (e.g., to help users avoid algorithmic profiling), creating and maintaining such a tool once platforms become aware of its existence will be challenging.

[9] An example of a data breach of sex worker information was Pornwikileaks, which no longer exists `https://www.cnet.com/culture/pornwikileaks-reveals-identities-of-porn-stars/`

[10] For example, see this patent: `https://patents.google.com/patent/US10019653B2`.

[11] See analyses of existing legislation and the impacts of payment deplatforming such as `https://www.nswp.org/sites/default/files/fosta_briefing_note_2018.pdf`, `https://www.aclu.org/news/lgbtq-rights/how-mastercards-new-policy-violates-sex-workers-rights`, `https://oneill.law.georgetown.edu/unpacking-the-dangers-of-mastercards-push-to-exclude-sex-workers-from-safer-sex-trade-spaces/`.

[12] For example, see this patent filed by AirBnB, which seeks to identify sex workers and those with disabilities: `https://patents.google.com/patent/US9070088B1`.

(2) **Facial Recognition Circumvention.** A growing number of facial recognition techniques are used to automatically identify and filter sex workers and other "undesirable" persons from using digital platforms. Further research is needed to develop usable adversarial machine learning techniques for image and video content that help users protect themselves from facial recognition algorithms.[13]

4.2 Payment Inoperability

Frozen or delayed payments endanger sex workers' livelihoods. Sex workers are already highly innovative in discovering payment methodologies that are not subject to tracking or delays. However, as clients increasingly move to traceable digital payment technologies, there is a need for further research. For example, we suggest:

(3) **Usable Privacy-Preserving Cryptocurrencies.** Non-traditional, distributed, digital payment platforms, such as cryptocurrencies, are in theory a promising tool for sex workers. Many cryptocurrencies offer either formal privacy properties or at least pseudonymity. Despite the potential of this technology—and the significant resources the security and privacy community has devoted to developing cryptocurrencies—cryptocurrencies are rarely used by sex workers. There are significant usability barriers for these privacy-preserving systems; sex workers must pay for goods and services with the payments they receive, which is challenging with cryptocurrencies. Moreover, many clients may be less technologically sophisticated or unwilling to use cryptocurrencies. While recent work has started to examine usability concerns, e.g., the usability of cryptocurrency wallets [15], significant work remains. If a client is unable to understand how to purchase and send cryptocurrency, it is impractical for sex workers to demand its use. Additionally, cryptocurrency valuations are highly volatile and exchanging cryptocurrencies for legal tender creates the risk of revealing the sex worker's identity. Finally, in cases where sex workers do chose to accept cryptocurrencies, they may be deplatformed if they use online wallets designed to make cryptocurrencies usable by non-experts.

4.3 Outing and Context Collapse

Due to the stigmatized nature of sex work, many sex workers want to maintain tight control over their identity as a sex worker. The increasingly digital nature of sex work makes this difficult. Many sex workers publicly share images of themselves on both mainstream and sex work platforms. While such images might be necessary as advertisements, they significantly increase the risk that the sex worker might be outed to their personal community. This issue suggests many different research directions:

(4) **Automated Blocking of Contacts.** One technique sex workers use to minimize their risk of being outed is to proactively block friends and family on social media. However, it is not clear what the *best* approach to proactive blocking would be. Should sex workers only block close contacts? What about the social media "friends" of those contacts? Creating an automated blocking tool, configurable according to a sex worker's personal needs, would reduce the effort and anxiety associated with this task. Moreover, measuring the privacy utility of blocking differing network distances of contacts could provide better transparency into the effectiveness of this approach.

(5) **Robust Image Modification.** Sex workers who hope to keep their identity secret often blur their faces and other identifying features to prevent contacts from recognizing them. Creating tools that automatically scrub identifiable features from content would reduce the chances that the sex worker accidentally leaks their identity. Additionally, it would be valuable to better understand the real privacy value of this approach. For example, it may be possible for a machine learning algorithm to predict the facial features (or even identity) of a sex worker, even when their content is blurred. Studying the most effective way to shield a sex worker's identity from these algorithms could be very valuable.

[13] Efforts have already begun in both directions. Patents for such malicious uses of facial recognition have been field, such as `https://patents.google.com/patent/US10019653B2` and research efforts have been funded to combat such uses, e.g., `https://www.nsf.gov/awardsearch/showAward?AWD_ID=2144988`.

Such efforts importantly seek to prevent *human* identification vs. *machine* identification. The latter is addressed in direction (2) and may be most effective when combined with efforts on combating human detection.

(6) **Privacy-Preserving Multi-Profile Support.** Past research has focused on access control to separate user profiles on a single device or across devices. However, there is a lack of deep analysis on whether such approaches are sufficient and usable in-the-wild to protect identity in cases where a single person wants to heavily utilize two completely separate, unconnected personas.

   (a) **Software approaches.** Existing software approaches to keeping identities or personas separate include browser profiles and Android profiles. However, prior work suggests that such profiles may be insufficient to fully protect user identity both from social exposure and algorithmic exposure, and that marginalized users such as sex workers do not trust such software-level strategies [8, 12]. Future research is necessary to consolidate such approaches and increase user trust.

   (b) **Hardware approaches.** Existing isolation models such as trusted execution environments (TEEs) may not fully address people's need to protect two or more profiles on a single device or across devices. Further, such isolation models have not been evaluated for usability [1]. Thus, future work is necessary to develop more nuanced and usable persona protections at the hardware level.

(7) **Verification of Privacy Settings.** Many techniques currently used by sex workers to prevent outing rely heavily on platforms' privacy protection mechanisms functioning correctly. However, the effect of many privacy preferences is completely invisible to users and impossible to verify in practice. As such, prior work suggests [12] that sex workers find it difficult to trust that activating privacy features actually achieves the desired result. This distrust could be overcome by designing privacy preferences that are *auditable*, either directly by a user or through third-party software. Sex workers, or trusted organizations within the sex worker community, could then verify that the protections provided by certain privacy configurations meet sex worker needs. Importantly, the auditing mechanism must simultaneously (1) be robust, in that the auditing methods must be significantly more convincing than just a UI change, (2) not allow auditors access to previously protected information, and (3) be sufficiently understandable that it engenders trust. We note that this idea can extend beyond privacy preferences; for example, it would be valuable to verify that metadata removal software and services actually accomplish their goal.

4.4 Content Theft

There is a real and pressing need for techniques that give sex workers greater control over their media. Note that this is not a traditional access control problem—the capacity to manipulate or share content is unavoidably shared when workers post their content publicly or share it with a client. Instead, tools must enable a sex worker to *detect* when their content is shared without permission or allow a sex work platform to proactively scan content to determine if it being posted with permission. Critically, these tools must be *usable*, which requires a low false positive rate and built-in protections against spamming.

While several digital rights management (DRM) solutions exist, there are notable differences between this setting and the archetypal DRM setting. Most DRM solutions are used by highly resourced organizations attempting to control a few high value pieces of content (e.g., blockbuster movies or TV shows). These organizations have the time and expertise to enforce those rights. The situation we consider is the opposite: there are a large number of sex workers creating a vast amount of content. Sex workers often lack the resources—in terms of time or money—to effectively assert their content rights; indeed, platforms, not fearing reprisal from marginalized sex workers, might actively make it difficult to enforce content rights. Moreover, the nature of the content produced is quite different: while it is difficult for an individual to assert that they own a well-known film, it may be difficult to ascertain who owns the content produced by sex workers just by looking at it. As such, applying traditional DRM solutions may not efficiently address the issue of content theft in digital sex work. Research directions include:

(8) **Certificate Infrastructure for Content.** Digital sex work could be supplemented with a certificate ecosystem for media management.

   (a) **As infrastructure.** Sex work platforms could require that all content uploads be accompanied by documentation affirming that the content creator intended for that content to appear on the platform, such as a certificate linked to the creator that is digitally signed using standard cryptographic techniques. While such an approach would be robust, significant research is required to understand how to manage digital identity in such a system. Indeed, a thief could steal a sex worker's content and generate a false certificate claiming ownership of the content, providing the thief has control of a legitimate digital identity. Research would also need to consider whether such an ecosystem can exist while still preserving privacy for sex workers who want plausible deniability in their work.

   (b) **For proof of ownership.** Even without a strong identity management system, creating some form of certificate-based content management system could make it easier for sex workers to issue DMCA take-down requests (as the ownership of content could be verified cryptographically).

(9) **Robust Content Matching.** Without a certificate ecosystem, there are still ways to identify media duplicated in unauthorized contexts. To build scanning systems that can help sex workers detect when their information has been shared (publicly) on the internet, we require a robust mechanism that can efficiently identify a sex worker's content. Existing robust content matching techniques to match "semantically equivalent" images have been developed in the context of detecting child sexual abuse material (e.g., Microsoft's PhotoDNA, Facebook's PDQ, or Apple's NeuralHash). The effectiveness of these techniques is questionable: the algorithms either require secrecy (PhotoDNA) or are more brittle than intended when made public (PDQ and NeuralHash). In all cases, the efficacy of these algorithms requires that the media for which the scanner is searching remains secret; if the fingerprint of the media were revealed, it would be easy to modify the media such that it no longer matched.

    It is not clear if existing approaches could be adapted to our setting, where different images could be semantically equivalent (e.g., different sex workers depicting the same activity) and fingerprints are not centrally curated. Alternately, cryptographic or steganographic watermarking to embed robust tags within the content could be used to identify its origin. While these techniques have been discussed in theory, to our knowledge they have not be meaningfully deployed.

(10) **Independent Content Scanning.** Reputation management firms[14] may help search for their clients' content across the internet and issue take-down requests. However, these can be expensive and may rely heavily on human labor rather than automation. One challenging research direction would be to explore how detection of known content can be automated, without the participation of other entities (e.g., clients, platforms). The primary challenges would be in scanning sufficient content to have reasonable coverage of the erotic content ecosystem, with the resources of an individual or small organization (e.g., sex work support groups). We note that this tool could be abused to compile information on a particular individual. Thus, a critical part of this research direction is defending against such abuses, for example using directions described in (3).

(11) **Anti-Theft Technology.** It is not possible to completely prevent content theft—viewers can always take screenshots, photograph or record using another device, and most anti-pirating mechanisms can be bypassed. However, it is worth investing in technologies that make theft onerous. There are multiple ways to hinder downloading static content in HTML, and many commercial streaming services use anti-recording technology. Researchers could investigate how to make these solutions usable for small content-creators and small-scale software developers.

[14] For example, `https://reputationmanagement.co`.

4.5 Relevance to Other Communities

The directions we summarize are relevant not only to protecting sex workers, but other marginalized groups as well. Efforts around preventing or recovering from content theft in particular will benefit people participating in both sex work and recreational digital intimacy. Technology that allows people to cryptographically prove ownership over images (8b) and systems for content matching (9) and scanning (10) may help those who suspect that their intimate images were made public.

On the preventative side, people creating images for commercial or recreational use may benefit from anti-theft technology (11) and usable and effective image modification technology (5). For example, someone on a dating platform may more comfortably engage in intimate exchanges if those photos are less likely to be screenshot and/or they are less recognizable in the photos. We note that both commercial and recreational creators may have complex use cases beyond those we explore here, such as creation of intimate media by or with another person, which complicates image ownership.

Marginalized groups such as activists and racial and gender minorities also face deplatforming and biased, heavy-handed content moderation. Filter analysis (1) and facial recognition circumvention (2) could be used to identify biases, and their causes, in content moderation. Further, tools that prevent outing such as automated contact blocking (4) and multi-profile support (6) would be extremely valuable for other populations who frequently manage multiple, non-intersecting social spheres. For example, a transgender person who is in the process of coming out incrementally to friends and family needs to be able to control which self-presentation is visible to whom and in what context in order to stay safe. Finally, some directions, like the verification of privacy settings (7), stand to benefit all users.

5 Conclusion

As we demonstrate, there are many opportunities for systems security and cryptography research to support the safety of digital intimacy. As researchers embark on this work, we again offer several important considerations. First, as with all research on privacy, anonymity, and circumvention, some of this research may increase risks for some communities, potentially by aiding harmful behavior such as abuse. Researchers must pay attention to the potential impact of new tools more broadly and work to minimize the risk of misuse. Second, fundamental problems that sex workers face—stigma, deliberate discrimination, and criminalization—are rooted in misogyny and discrimination against sex workers. Technology alone will not change the prevalence of these risks, only impact their online manifestations. To fundamentally reduce risks for these communities, we must also advocate for substantive social change. Finally, and most importantly, centering affected communities directly in research conception and deployment is imperative to building systems that increase safety rather than harm. The people directly impacted by the products of the work will know best whether solutions reduce risk, increase risk, or merely rearrange it. There are a range of appropriate methods for centering marginalized users. Purely theoretical work (i.e., improving cryptographic primitives) should carefully leverage empirical work to appropriately ground theoretical use-cases, and applied work should engage directly with the people who may use—or be harmed by—the products of their work.

References

[1] Alessandro Armando, Gabriele Costa, Luca Verderame, and Alessio Merlo. Securing the "bring your own device" paradigm. *Computer*, 47(6):48–56, 2014.

[2] Hanna L Barakat and Elissa M Redmiles. Community under surveillance: Impacts of marginalization on an online labor forum. In *Proceedings of the International AAAI Conference on Web and Social Media*, 2021.

[3] Catherine Barwulor, Allison McDonald, Eszter Hargittai, and Elissa M Redmiles. "Disadvantaged in the American-dominated internet": Sex, work, and technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2021.

[4] Rasika Bhalerao, Nora McDonald, Hanna Barakat, Vaughn Hamilton, Damon McCoy, and Elissa M Redmiles. Ethics and efficacy of unsolicited anti-trafficking SMS outreach. *Proceedings of the ACM on Human-Computer Interaction*, (CSCW), Forthcoming.

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

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

[7] Asia Eaton, Holly Jacobs, and Yanet Ruvalcaba. *2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration.* Cyber Civil Rights Initiative, Inc., 2017.

[8] Hana Habib, Jessica Colnago, Vidya Gopalakrishnan, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, and Lorrie Faith Cranor. Away from prying eyes: Analyzing usage and understanding of private browsing. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 159–175, 2018.

[9] Vaughn Hamilton, Hanna Barakat, and Elissa M Redmiles. Risk, resilience and reward: Impacts of shifting to digital sex work. *Proceedings of the ACM on Human-Computer Interaction*, (CSCW), Forthcoming.

[10] Angela Jones. *Camming.* New York University Press, 2020.

[11] Sarah Esther Lageson, Suzy McElrath, and Krissinda Ellen Palmer. Gendered public support for criminalizing "revenge porn". *Feminist Criminology*, 14(5):560–583, 2019.

[12] Allison McDonald, Catherine Barwulor, Michelle L Mazurek, Florian Schaub, and Elissa M Redmiles. "It's stressful having all these phones": Investigating sex workers' safety goals, risks, and practices online. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 375–392. USENIX Association, 2021.

[13] Teela Sanders, Jane Scoular, Rosie Campbell, Jane Pitcher, and Stewart Cunningham. *Internet Sex Work: Beyond the Gaze.* Springer, 2018.

[14] Molly Smith and Juno Mac. *Revolting Prostitutes: The Fight for Sex Workers' Rights.* Verso Trade, London, 2018.

[15] Artemij Voskobojnikov, Oliver Wiese, Masoud Mehrabi Koushki, Volker Roth, and Konstantin Beznosov. The U in crypto stands for usable: An empirical study of user experience with mobile cryptocurrency wallets. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.