"I chose to fight, be brave, and to deal with it": Threat Experiences and Security Practices of Pakistani Content Creators

Lea Gröber[1,5] Waleed Arshad[2] Shanza [2] Angelica Goetzen[3]
Elissa M. Redmiles[4] Maryam Mustafa[2] Katharina Krombholz[1]

[1]*CISPA Helmholtz Center for Information Security*
[2]*Lahore University of Management Sciences*
[3]*Max Planck Institute for Software Systems*
[4]*Georgetown University* [5]*Saarland University*

Abstract

Content creators are exposed to elevated risks compared to the general Internet user. This study explores the threat landscape that creators in Pakistan are exposed to, how they protect themselves, and which support structures they rely on. We conducted a semi-structured interview study with 23 creators from diverse backgrounds who create content on various topics. Our data suggests that online threats frequently spill over into the offline world, especially for gender minorities. Creating content on sensitive topics like politics, religion, and human rights is associated with elevated risks. We find that defensive mechanisms and external support structures are non-existent, lacking, or inadequately adjusted to the sociocultural context of Pakistan.

Disclaimer: This paper contains quotes describing harmful experiences relating to sexual and physical assault, eating disorders, and extreme threats of violence.

1 Introduction

Online content creators express themselves, reach broad audiences, raise awareness, or build careers [22,54] using services such as TikTok, Instagram, and YouTube. They cater to large audiences, share a sizable volume of private data and are consequently exposed to elevated risks [87]. Prior research explored the threats to US-based content creators [79,87] and the ways in which sensitive topics – such as sex work – affect the security and privacy of creators [38,62]. While it is estimated that approximately an equal number of men and women engage in content creation work [50], a survey by Thomas et al. [87] found that U.S. content creators who identify as women are more likely to experience sexual harassment, excessive negative reviews, stalking, spreading of rumors, and surveillance than those who identify as men. Furthermore, outside of content creation, a growing body of work establishes how gender impacts the digital security and privacy experiences [25,76–78,96]. In the global south [6,77,78,91], women are structurally disadvantaged and face unique threats, often tied to their socioeconomic status and literacy [6,70,78].

In this paper, we present a study with 23 semi-structured interviews investigating the intersectional marginalization for content creators across genders in Pakistan. Pakistan is a particularly interesting country in this regard: It ranks second to last in terms of gender parity [35]. Due to cultural and religious factors, the access to the regular labor markets is limited for women [19] as is their presence in public spaces in comparison to men [68]. To navigate these constraints many women choose to migrate to online spaces for forming social connections and for work and business opportunities [97]. Prior work reveals the vital importance of online communities in countries like Pakistan for women to discuss taboo narratives, find work, explore identities and form connections [11,97].

In contrast to general social media users, content creators are especially vulnerable and constitute a special case in two ways: (1) They cater to a large audience, and depending on the type of content they produce can (be perceived to) threaten social norms and structures.[1] (2) They *have to* stay online and maintain public profiles and personas to keep their business/activism going and thus cannot rely on affordances available to other users who have private profiles and are not dependent on making money as content creators.

Anecdotal evidence reports instances of women content creators in Pakistan facing severe harassment [21], group assault [31], or getting killed [55,82]. The prominent case of Qandeel Baloch demonstrates how content creation can be a powerful way for women from low socioeconomic backgrounds in Pakistan to emancipate themselves, but who also face lethal consequences when family, friends or extended relatives start to feel threatened or dishonoured by their presence in online spaces, particularly when their private data (identity) is leaked [55]. Additionally, content creators - both men and women - who engage with sensitive topics such as religion, sexuality or politics face harassment [21,45], may be forced to leave the country temporarily [3], and have even been killed for their work [47]. These stories showcase the elevated risk to which Pakistani content creators are exposed and under-

[1] Blasphemy laws in Pakistan carry severe penalties up to death [59].

score the critical need for research into how to best protect these marginalized content creators.

Our qualitative study investigates the security context of digital content creation in Pakistan including the threat landscape, how creators navigate concrete negative experiences, defensive mechanisms, and what support structures they rely on. Accordingly, we defined the following research questions:

RQ1: What does the threat landscape for content creators in Pakistan look like?

RQ2: What defensive mechanisms do Pakistani content creators implement to stay safe and secure? Which resources do they rely on?

RQ3: What gaps are present in existing security and privacy measures? Which interventions would be needed that are specific to the Pakistani social media ecosystem?

We find that the online and offline threat landscape is tightly connected, with online threats frequently manifesting in offline harm. The degree of both online and offline threats are severe and in the socio-political context of Pakistan can become lethal, especially if content revolves around religion, sexuality, or politics. Online threats are accounted for with a mixture of *technical* and *behavioral defenses*. While defenses exist for prominent threat categories like *toxic content*, participants are dissatisfied with options to deplatform attackers. However, certain threat categories such as *impersonation* lack any defensive mechanisms, while exposing victims to severe threats. Our findings inform the further development of inclusive safety tooling for social media platforms tailored to different populations and risks that are specific to those populations. Moreover, our findings contribute to the improvement of information sources for at-risk populations.

2 Background and Related Work

2.1 Sociocultural Background of Pakistan

Pakistan is a particularly interesting country to study the challenges of publicly exposed figures such as content creators and frontiers of security and privacy measures given its complex patriarchal and religious landscape. In the recent 2022 gender gap report by the World Economic Forum Pakistan places second to last (Afghanistan is last) in terms of gender parity [35]. In Pakistan, offline and online privacy behaviors are heavily influenced by religious and conservative values and patriarchal norms. Notions of community, family honor, and piety as well as Islamic values heavily influence legal, political, and social norms [37]. Pakistan is a culturally and linguistically diverse country with significant region-specific differences within the country. It is also deeply class-based where wealth is not equally distributed, and the reality of high-income citizens is very different from those of low-income citizens [4]. There exists an educational divide between rural and urban citizens, and the majority of women have no formal education [1]. Mobile phones are equally available to both rural and urban households, however, rural households are three times less likely to have access to a computer or internet [1]. Islam is the state religion with approximately 95-98% of the population identifying as Muslim. The Islamic principles of *Purdah* and gender segregation often bleed into digital spaces as well. *Purdah* is broadly defined as the segregation of genders, and involves both modesty of the heart and the eye [40]. The practice of Islam in Pakistan values modesty, the segregation of genders and the covering of women's bodies [40]. These values also impact women's access to formal labor markets, public spaces, and digital spaces [56]. Prior work highlights the ways in which women in Pakistan navigate constraints on mobility, social networks, access to labor markets, and social support by leveraging digital spaces [69, 97]. Online spaces are a particularly critical pathway for women to access vital resources and gain financial independence. However, prior work also highlights the ways in which these spaces fail women in Pakistan [70].

Since 2013, the Pakistani national identity cards have a third gender category [2], and since 2018, Pakistan's Transgender Persons (Protection of Rights) Act theoretically strengthened the rights of transgender people. Although the transgender community – locally referred to as the *Khwaja Sira* – are officially recognized, they continue to face severe discrimination, are often excluded from the conventional job market, and are seen as beggars in public [71]. Despite legal recognition and a Supreme Court ruling allowing transgender people to be identified as a third gender, systemic social support is still lacking. Prominent representatives of the *Khwaja Sira* community argue that the Western LGBTQ+ acronym does not adequately capture their unique experiences, and they emphasize the separation of gender identity and sexual orientation [48]. In Pakistan, there are two main social attitudes towards the *Khwaja Sira*: conservative groups marginalize them for not conforming to a binary gender, while the liberal population often understands trans rights but through Western ideals [9].

2.2 Security and Privacy of Content Creators

In 2022, Thomas et al. [87] conducted a survey with 135 U.S. content creators from different platforms (esp. YouTube, Instagram, and TikTok) with a focus on quantifying the extent of negative experiences, reactions to attacks, and perceived gaps in protective solutions provided by platforms. They found that content creators in the U.S. are structurally exposed to risk, with one in three being the target of attacks and 95% experiencing hate or harassment at least once. In response, many content creators chose to ignore attacks, while others engaged in self-censorship, or leaving platforms altogether. A recent interview study by Samermit et al. [79] explored threats relevant to U.S. creators, and their protective practices. They found diverse threat models apply for creators, and that defenses are often adopted only after negative expe-

riences. While these works offer insight into experiences of content creators in the US, a full understanding of content creator safety and security requires an understanding of the experiences of non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic [44]) creators. Centering the most marginalized users allows us to ensure that we fully capture the risks faced by content creators in designing interventions; as McDonald et al. [63] explain, "often, the privacy risks of vulnerable populations are not fully considered in the design of systems because those risks and potential harms are not fully understood (nor necessarily prioritized) by those responsible for research and design." For instance, recent work has begun to explore the unique safety and security experiences of digital sex workers. These content creators exist at the intersection between digital content creation and sex work; they utilize social media to create monetized adult content as well as connect with other creators and engage in community support [34, 38, 39, 89]. In addition to negative experiences faced by content creators at large [87], digital sex workers are also subject to social stigma and strict content moderation policies [17], producing additional negative outcomes like being de-platformed from social media sites [16, 38]. In general, algorithmic fairness is an issue for marginalized creators [26, 33, 49]. For instance, creators with disabilities face challenges with demonetization [49] and transfeminine TikTok creators need to navigate visibility traps [33].

2.3 Research beyond WEIRD populations

Prior works exploring the privacy and security concerns of non-WEIRD populations reveal the deeply gendered ways in which privacy is understood, enacted and preserved [15, 70, 77, 78]. In these studies religion, literacy and gender play a pivotal role in privacy perceptions and behaviours. In the Pakistani context, studies reveal the ways in which women negotiate the creation of gendered women-only spaces online as a way of protecting their modesty from unfamiliar men, as prescribed by Islamic teachings [70, 97]. A recent study on young Pakistanis' perceptions on security and cyber crime found that men and women prioritize threats differently, and form networks to cope with lacking platform affordances [15]. Studies exploring Arab Gulf citizens' privacy behaviours reveal Arabic social media users frequently establish private online accounts to uphold personal or familial honor, in line with social norms [5]. Prior work also reveals low-income, low-literate women in Pakistan, India, and Bangladesh associate privacy with *western values* [78] or with shame [70], often believing privacy is only for people who have done something that they want to hide. These studies highlight the mechanisms women in Pakistan, India and Bangladesh employ to protect themselves including but not limited to avoidance, deletions, limited lists, private profiles, multiple accounts and the use of private modes [70, 77, 78]. In contrast to these populations our work focuses on a particularly vulnerable group, content creators, who due to the nature of their online engagement are unable to use many of the provided platform affordances like private modes, limited lists, private profiles or avoidance.

2.4 Research on Vulnerable Populations

We place our work in the broader context of populations exposed to elevated risks. In both the HCI and security communities researchers have identified the need to study vulnerable populations to understand the structural threat landscape these people are exposed to, and to investigate possible shortcomings in protective means [63, 94]. In a meta study, Warford et al [94] identified *contextual risk factors* that can be used to categorize at-risk populations. *Societal factors* describe how users are at-risk due to their cultural and social embeddedness and public roles they take on. Examples of people facing elevated risk due to taking legal or political action include, but are not limited to, people involved with political campaigns in the US [28], activists [30, 51, 57, 85], and journalists [64–66]. Moreover, research focused on vulnerable marginalised groups such as LGBTQ+ people [20, 53, 80], sex workers [18, 84], and populations that are at-risk due to low socioeconomic status (e.g., non-Western woman [32, 77, 78, 92], people in developing regions [7, 67, 74, 91], and people in developed regions [27, 72, 73, 93]). Apart from that, Warford et al. [94] identified the *relationships* of people (e.g., intimate partner abuse [24, 36, 41, 60]), and the *personal circumstances* (e.g., underserved accessibility needs [8, 46, 58]) as potential risk factors. In our study we find that content creators who use their voice to make political statements, or otherwise comment on societal norms such as religion or sexuality are exposed to elevated risk in the context of Pakistan.

3 Methodology

We conducted a qualitative semi-structured interview study ($n = 23$) to explore content creators' security and privacy perceptions, needs, and practices in the socio-cultural context of Pakistan. The following sections detail our recruitment, interview procedure, pilot tests, and analysis.

3.1 Recruitment and Participants

Our target population are content creators matching the following criteria: (1) living and working in Pakistan, (2) generating some form of income from their content (e.g., monetary or through compensation with physical goods). We recruited Pakistani content creators across platforms (YouTube, Instagram, TikTok, Facebook, Twitter, Snapchat, Spotify) and targeted a diverse sample across several dimensions (income level, literacy, gender, follower count, content type) to ensure we captured a range of relevant experiences. For recruitment, we manually compiled short lists of creators of different genders based on profile data[2] ensuring broad coverage of the above

[2] In Pakistan, few people openly disclose their gender online. We validated our assumption with participants' self-reported gender data.

demographics. Recruitment followed an exploratory, iterative approach. We used the results of the ongoing qualitative analysis to expand the shortlists with candidates that could bring new perspectives to the study. For example, we found that content creators frequently engage in self-censorship to protect themselves. To complement our perspective, we then specifically targeted people who post content on sensitive topics that other content creators avoid, such as gender and politics, but also people who only post seemingly uncritical content, such as food or pets. Moreover, after each interview, we asked the participant if they knew someone whom we should interview for this study. We reached out through content creators' designated contact channels and asked them to fill out a three-minute screening survey to ensure they fit our criteria. The screening survey collected personal information (like age, location information, gender, disabilities, education, and income), and platform information (where they create content, compensation for content creation, description of content topic, posting habit, and follower count). The positive response rates per shortlist were: 21.4% *woman*, 30% *man*, 100% *Khwaja Sira*.

3.2 Interviews

Two interviewers conducted 23 one-hour semi-structured interviews in Pakistan from December 2022 to May 2023. This resulted in a total of 1261 minutes, with a median of 56 minutes. We compensated each participant with PKR 5000 (USD 19.02), resulting in a median hourly wage of PKR 5258 or USD 20.37. To comply with socio-cultural norms, the interviewers were a man and a woman Pakistani co-author who respectively spoke to the men and women participants. Khwaja Sira chose their interviewer. This practice was intended to create a room where participants could feel safe and talk openly. We conducted interviews in English and Urdu, depending on the participants' preferences. Interviews took place in person or online via Zoom.

We developed the interview guideline with our exploratory research questions in mind. We also took the results of previous studies into account [87], building on them while exploring the topic broadly through semi-structured interviews. Figure 2 in Appendix B provides an overview of the interview flow and contents. We describe the interview sections below and provide the interview guideline in the supplementary material. Each interview started with a fixed introduction that established the purpose and procedure of the interview and allowed the participant to ask organizational questions. Afterward, we obtained the participant's consent to record the interview and proceeded with the interview guideline. The interview was structured as follows:

Verification Section: We revisit questions from the screening survey about the scope of content creation. This serves both as opening questions to get participants talking and to re-verify that creators fall within our target population and are, indeed, the creators we reached out to. No participants were screened out.

Defining Their Work: This section establishes our participants' context and content creation habits. In the beginning, participants define their job title (e.g., *content creator* vs. *influencer*) by which we address them throughout the interview. Further, they describe posting habits and the type of content they create. We provide an overview of participants' background information in Section 4.1.

Negative Experiences: We ask participants to describe their personal definition of safety and privacy in relation to content creation. Then, we ask about their biggest digital and physical security and safety concerns. To make the conversation more tangible, we focus on concrete negative experiences that participants encountered in connection to their content creation work. Moreover, we ask who they identified as potential adversaries, and how they protected themselves in these scenarios. Based on this, we categorize the negative experience of content creators in Pakistan in Section 4.2.

Defense and Mitigation Strategies: We explore how participants aim to prevent these threats from occurring via *behavioral* (e.g., adjustment of content types, or leaving platforms) and *technical* defenses (e.g., platform features, or authentication schemes). Then, we explore any additional support structures participants relied on to cope with negative experiences. Finally, we explore (gaps in) participants' socio-technical approaches to mitigating harm from the negative experiences they have endured (see Section 4.3).

3.3 Pilot Testing

To pilot test the interview guideline, one researcher took on the persona of a Pakistani content creator and we conducted two pilot tests, one with each interviewer. We tested the flow of the interview guideline in terms of the overall structure and order of questions. In response, we restructured minor parts of the interview guideline to remove redundancies. As the study progressed, we further developed the interview guideline by incorporating participant feedback. Changes included minor rewording to make questions more open-ended and changes to the question order. No substantial changes were made to the interview guideline.

3.4 Analysis

We transcribed the interview recordings manually and using a GDPR-compliant transcription service (Amberscript). Urdu parts were first transcribed and then translated into English by the researchers of our team. Four researchers were involved in the coding process: the two co-authors from Pakistan who conducted the interviews and two co-authors with German and U.S. backgrounds, respectively. Three of them are computer scientists; one is a social scientist. We conducted a thematic analysis where data collection (interviewing) and analysis was

an iterative process. Our goal was to collect a set of threat patterns which are associated with in-depth data about individual experiences. Researchers used a combination of "top-down" qualitative content analysis [52, 61, 75, 81, 90] (informed by previous frameworks [86]) and "bottom-up" analysis inspired by "open coding" [23, 29, 83] allowing for emerging themes. Three researchers jointly established a first codebook taking into account a male and female interview. We followed an iterative approach where at least two researchers coded one interview independently. Then they discussed and resolved all disagreements and updated the codebook accordingly. As we coded, we wrote summaries and memos to collect and systematize potential themes. Once coding was complete, we jointly discussed themes by revisiting memos and grouping codes in the codebook. Thereby we identified five axial categories (negative experiences, concerns, attackers, defense practices, support). We reached stability in threat patterns after having interviewed 16 men and women. We continued interviewing but focused recruitment efforts on Khwaja Sira creators, to explore if (individuals of) this gender experienced additional threat patterns (because we found that gender is likely an influencing factor on negative experiences). However, we did not find additional threat patterns in the last six interviews (two Khwaja Sira and four men/women), although we learned about extreme instances of already known threats. Therefore, we closed data collection and concluded reaching thematic saturation [42, 43] with respect to threat patterns after 16 interviews. The final codebook contains 495 codes.

3.5 Limitations

Interview studies are limited by self-reported data, which may lead to under- or over-reporting. To address this, we designed the interview guideline to focus on specific experiences and provided prompts to aid memory recall. Participants sometimes felt uncomfortable reporting negative incidents, and we respected their decision. Therefore, we acknowledge that our findings may not fully capture extreme negative experiences. To mitigate social desirability bias, we reassured participants that we were interested in their experiences as content creators and would not judge their actions or responses to threats. Given the strict social gender norms in Pakistan (Section 2.1), we conducted interviews in a same-gender setting (selected gender for Khwaja Sira) to create a safe space for discussing sensitive topics. Our convenience sample does not necessarily represent the wider population of Pakistani content creators. While qualitative studies like ours do not strictly require representation, we made deliberate efforts to recruit diverse participants. We defined recruitment criteria based on previous research and anecdotal evidence to identify potential high-risk populations [2, 37], resulting in a diverse sample that includes Khwaja Sira creators. However, future work should quantitatively validate our findings. Moreover, our sample does not include participants who decided to withdraw from creating content, or who faced lethal consequences.

Ethical Considerations We obtained approval for this study by Saarland University's ethical review board. This research touches upon critical negative experiences, thus potentially bringing up past trauma. To obtain informed consent, we thoroughly explained the process of the study, including how we recorded, anonymized, and stored the collected data according to GDPR. We paid special attention to anonymizing participant quotes and present only aggregated demographical information to avoid potential harm due to de-anonymization.

4 Results

Findings are illustrated with participants quotes (Gx); x denotes the participant id within the gender group G (**M**an, **W**oman, **K**hawaja **S**ira).

4.1 Sample Descriptives

We recruited 12 women, 9 men, and 2 Khwaja Sira (see Section 2.1). Participants are young with 83% ages 18-24 and 17% ages 25-34; and educated with highest degree high school (48%) or University (48%). One person was still pursuing a high school degree. 56% of interviews contained answers in both English and Urdu, 26% were solely conducted in English, and 17% in Urdu alone. 48% of participants have 10k - 50k total followers across platforms, while the biggest creator has between 300k - 400k followers. Everyone earned some form of income from content creation, but only 36% were comfortable sharing annual income details. 48% of participants reported earning upwards of 100k PKR annually. Table 2 in the Appendix reports demographics; Table 3 presents a breakdown of the platforms participants create content on. All participants use Instagram, followed by TikTok (65%), and Youtube (61%). Participants post various content topics: from fashion and lifestyle to awareness of socio-political issues. Some participants posted multiple types of content which fall into a common broader category, e.g., fashion, lifestyle, etc. Most commonly participants create lifestyle content (39%), followed by fashion (35%), comedy (17%), and music, life stories, vlogs, and socio-political and human rights issues (13% each). It is important to note that socio-political issues include topics on gender, sexuality, and religious minorities, politics, and human rights include topics like women and trans rights. Two of our participants posted dance videos. The remaining categories include makeup, food, mentoring, photography, pets/animals, family content, and documentaries (4.3% each).

4.2 Threat Landscape and Negative Experiences

We structured the threat landscape of Pakistani creators following Thomas et al.'s taxonomy of online hate and harassment attacks [86] and expanded it towards offline threats. Threats were identified based on the axial categories *negative experiences*, *concerns*, and *attackers*. Figure 1 provides an

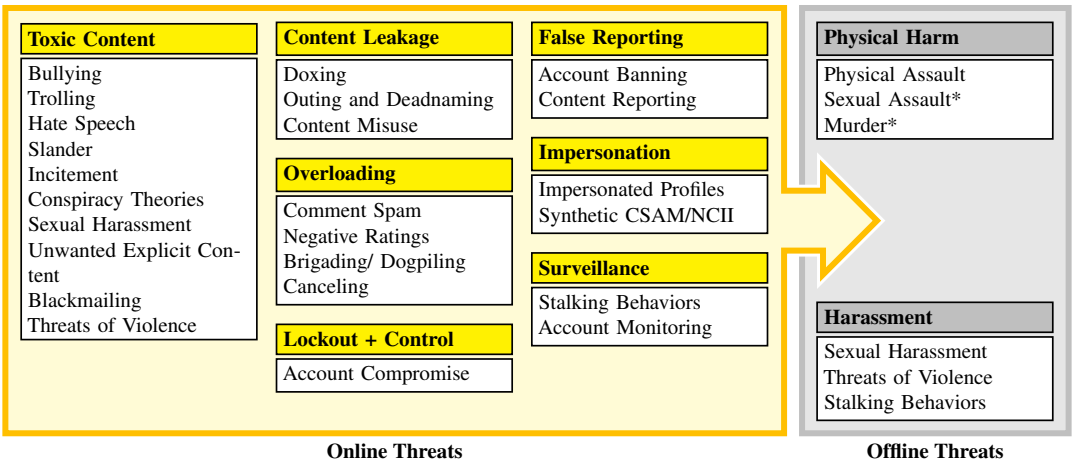

Figure 1: Spillover effect from online (yellow) to offline threats (gray) for content creation in Pakistan, grouped by categories. A * denotes threats explicitly mentioned (but not experienced); all other threats were experienced by at least one participant.

overview of online and offline threat categories, including associated attacks that emerged in the interviews. Throughout our results, we compare our findings to prior work on U.S. creators [79, 87] and the general Pakistani population [15], however, we do not provide comparative statements for all findings due to differences in analysis lense.

4.2.1 Toxic Content

This category includes online attacks that are intended to be seen by the creator, often with the goal to intimidate, harass, influence, or silence them. In line with our work, research on U.S. creators found these attacks impact emotional safety [79].

Bullying and Trolling. The boundaries between *bullying* and *trolling* are fluid, and attacks are prevalent across genders [w:11 | m:6 | k:2]. Women report instances of body (W7, W8, W11) and skin tone (W1) shaming, with one participant developing an eating disorder as an outcome (W7). Similarly, men are also shamed for their bodies, such as for not growing facial hair (M3). Participants are also targeted if they are in positions of power; for instance, W9 works as a civil servant and experienced bullying and harassment due to attackers' belief that she does not deserve a good life while the local population lives in poverty. In general, *marginalization* is a risk factor [94] that applies to U.S. creators, too. Attackers focus on identity characteristics such as gender or religion, with intersectional identities (e.g., woman, overweight) compounding risk [79].

Hate Speech. In our sample, all genders report experiencing *hate speech*, although it was most prevalent among woman and Khwaja Sira [w:10 | m:3 | k:2]. Extreme cases of hate speech and harassment are associated with specific content topics, such as religion (W9), politics (K1), and discussion of social issues such as rights for women and/or transgender people (W9, K1, K2). As W9 puts it, activism can be dangerous, as people start targeting you based on the minorities you speak up for: *"When I talk about Ahmadis [religious sect] or religious minorities in general, people would just tag me an Ahmadi. [...] If I talk about women, they would just spew slurs at me just randomly. They would curse me, saying I'm a feminist. It's a curse word itself. March is just upon us and Women's Day is here. The whole cycle just starts every single year. The other kind of harassment that I face is when I talk about transgenders. They would just say that I am promoting gay marriages just because I'm talking about the protection of transgenders"* (W9). Outside of religious, political, and social issues-centric content, even pictures with friends in a mixed-gendered group can lead to harassment: *"I remember posting a group picture with my friends and they were, five or six girls and there were two guys in the picture [...]. For some reason, I received a lot of backlash on that post. 'Why are you promoting Western ideals in Pakistan? [...] She's Western propaganda.' Even though it was literally just a couple of friends standing together and there was nothing problematic about it"* (W3).

Slander, Incitement, and Conspiracy Theories. *Slander* is a distinct type of toxic content meant to harm the reputation of creators. It is particularly dangerous in Pakistan, as people – especially women – need to comply with strict cultural norms of *pardah/modesty* and *honor* [88]. Among our participants, all genders [w:3 | m:3 | k:2] fell victim to the spread of disinformation – not only those who produce content associated with heightened risk (e.g., religion, politics, social issues), but also those of non-sensitive content types. For instance, W12 owns a pet account, and a competing account that did not grow as quickly as hers spread disinformation that she abused her animals. Slander may lead to *incitement* [w:2 | m:3 | k:2], either when creators decide to remain associated with someone who got canceled (W2), or if their content is controversial or sensitive (K1). *Incitement*

can easily lead to online (W2, W9, M4, M5) and offline (K1) harassment, for instance K1 reports of online articles speculating about her gender that led to religious sects showing up at her home (K1). The boundaries between *slander* and *conspiracy theories* are fluid, and the potential for harm is great if the creators reputation is damaged. We identified four instances of *conspiracy theories* [w:1 | m:1 | k:2] that were all trying to explain creators behavior with hidden political agendas: *"If these individuals are sitting and saying that I'm an American agent working on a conspiracy to promote LGBT rights in Pakistan"* (K1).

Sexual Harassment. Among our participants, only women and Khwaja Sira reported experiencing online sexual harassment [w:2 | m:0 | k:2]. One prominent experience involved religious figures leaving sexualized comments on the (non-sexual) picture of a woman creator. Being targeted by religious figures can escalate the incident into the religious realm where the victim becomes a target for a broader audience that often is willing to resort to offline violence [55].

Unwanted Explicit Content. The majority of women in our sample receive unsolicited messages [w:9 | m:2 | k:1], often containing explicit content. In one incident, a friend sent the unwanted content (M2), but usually attackers are unknown as the messages are sent from throw-away accounts: *"This guy was DMing me really sexual stuff, pictures, unsolicited pictures of himself [...]. I blocked him from so many different accounts and he would just keep making new accounts to do that"* (W7).

Blackmailing. Few participants report being *blackmailed* with the majority identifying as men [w:1 | m:3 | k:1]. Attackers can come from within their inner circle; for example, former friends threatened to leak a participant's identity (W11). Other attackers are anonymous and threaten participants for attention or personal gain: *"The most extreme level of it is when someone reaches out to you and basically asks you for [...] a shout out, [...] a reply, it can be anything like that and if you don't do it, then they threaten people that you love"* (M9).

Threats of Violence. Some participants report receiving *threats of violence* [w:2 | m:2 | k:2] through comments and direct messages (M2). Attackers are usually anonymous, although participants sometimes link them to known opponents of them (e.g., influential industrialists (K1)). Threats can also be a result of *hate* and *canceling* campaigns (M4). The goal is to intimidate or silence the creator, and may target the creator's family and close ones as well (M4). Participants report that content on topics like religion and gender lead to more extreme (sexualized) threats of violence (K1,K2, M2, W9). In general, creators think that *"mostly girls [...] receive perverted comments"*(M2). In our data set only woman and Khwaja Sira received comments regarding sexualized violence: *"Whenever I talk about religion, these kinds of things happen. I believe that [...] the concept of religion that has been propagated by different Madrasas is the ultraconservative version of a religion that I'm talking about. They do not see women or random creators talking about religion. They believe that religion is something that only they can talk about. Just because a person [...] is in westernized attire, [they are] not Muslim enough to talk about the religion. That's when I got some rape threats"* (W9). Participants did not report any threats of violence that were realized offline.

4.2.2 Content Leakage

Content leakage summarizes attacks in which private data or content is leaked by a third party. This was done to threaten, embarrass, or shame our participants.

Doxing. Most woman creators were concerned about or experienced *doxing*, where personal information is exposed to a greater audience [w:5 | m:2 | k:2]. Attackers can come from the inner circle (W11), or engage in social engineering to obtain personal information such as phone numbers (W4). Through *doxing* creators and their loved ones become targets for offline attacks (e.g., *stalking behaviors*, *physical harm*), but also because people surrounding them might feel threatened or dishonoured. In the context of Pakistan, leakage of personal information such as family details are therefore severe threats, and deanonymization has led to the murder of a woman creator in the past [55]. Brands can also be a threat as they obtain participants' information to send PR packages; one participant had a brand give out her home address to a fan who was a relative of the brand's owner (W7).

Deadnaming. Another form of leaking personal information and disrespecting the victim is deadnaming a Khwaja Sira, which was reported in one instance: *"Dead naming me is a line for me [...]. When people [call me by my old name] to belittle me that I think is a line that I do not like being passed because <old name> is a very personal name."* (K1). Here, *deadnaming* was used to harass the creator, however this practice reveals sensitive gender information that might not be intended for a broader audience. In case of K1, gender-identity is part of her content.

Content Misuse. We expanded this category to also include unauthorized content use, which was a prominent concern across participants [w:5 | m:3 | k:1]. It is different from content leakage, as in this case the content was previously posted on a public profile. Unauthorized content use by third parties primarily leads to financial harm [38], but may also be used for *impersonation* that in some cases can result in *physical harm*. In Pakistan this is a severe threat to woman, as it damages their reputation: *"Being a Pakistani girl with a public account, there is always a little thing in your heart that anything can happen anytime, your pictures can be edited and transformed and a lot of other things can happen so social media is one of the most unsafe places, which we all know but still we use it"* (W1).

4.2.3 Overloading

Overloading refers to online attacks or interactions that overwhelm the creator by spamming communication channels

(e.g., comments or notifications), usually with the aim to silence or influence them. Non-malicious intentions, such as getting the creator's attention or showing fan affection were also reported (W12), but can equally be perceived as harassment. Attackers are groups of people, often entire communities, that coordinate to target a victim.

Comment Spam and Negative Ratings. Especially when it comes to religious content (M5), our participants face excessive negative engagement like hate comments and messages [w:7 | m:3 | k:1]. Comment spam can also be a way for people to get the attention of creators. Complementary, W12 describes out of the norm positive attention from fans as a form of harassment.

Canceling, Brigading, and Dogpiling. These attacks [w:3 | m:2 | k:1] are similar because they are all carried out by coordinated online communities, e.g., through comment spams. The goal is what makes them different: *Canceling* aims to silence and de-plotform creators (W2, M4, M7), *brigading* tries to disrupt the discourse (W2, W8, W9), and *dogpiling* aims to get victims to recant their views (K2).

4.2.4 Lockout and Control

These attacks aim to silence creators or de-platform them alltoghether by breaking into or leveraging privileged access to a targets' accounts or devices [87]. In our sample, attackers were never identified, but third-party reports highlight the risk of attackers from the inner circle of the victim [60, 78].

Account compromise. The threat of getting hacked is known to the majority of participants, and they deploy Two-Factor-Authentication (2FA) as a defense [w:12 | m:7 | k:2]. However, the concepts of attack vectors and recovery seem to be obscure: *"I couldn't access my account and I don't know why that happened"* (W12). M7 told us about the account deletion of a fellow creator, allegedly because the attacker did not like her content: *"Her account got deleted recently. And the kid who hacked it said that because he didn't like the content [...] he deleted it. If it's so simple for a kid like that to do something, I would say Instagram is not really a safe account. And you can't even appeal to anyone to help"* (M7). In our sample, three creators (W12, M9, K2) have fallen victim to hacks that resulted in account lockouts. None named details on who the attacker was or how their account got compromised. All were able to recover their account with the help of the platform's support system and external digital rights organization. Only after the hacks and on the platform's request, two of them switched to 2FA. Participants also expressed their concerns of account or content deletion, although nobody in our sample experienced it firsthand.

4.2.5 False Reporting

A platform's reporting function is intended to protect individuals and the community at large. However, due to its semi-automated handling, reporting can be misused to silence, de-platform, and financially harm creators or, more generally, demonstrate power. Attackers usually stay anonymous. We extended the taxonomy [86] to fit creators.

Account Banning and Content Reporting Participants were as concerned about attacks in this scenario as they were about hacking attempts. However, they were dissatisfied with the lack of defensive options. One participant experienced account loss through banning twice on TikTok (W8) but was able to recover from it. Another participant reported a similar incident of a fellow creator: *"Some random person [...] reported her account and got her banned from Instagram. There's nothing that she could do to get her account back. She had around 20,000 followers. The person who reported her account also told her, 'I don't agree with your content, and I think that you're annoying [so I took your platform], and I want you to apologize to me, then I'll give it back to you.' "* (W3). Similarly, some participants had their content wrongly taken down from platforms [w:1 | m:1 | k:0]. Attackers coordinated malicious reporting attacks to silence creators: *"Even though my videos did not have any sort of explicit content [dance], due to people reporting it, a lot of my videos got taken down. They still do to this day"* (W8). Generally, participants wished for transparency on how banning works from the platforms (W1, W3, W12).

4.2.6 Impersonation

This category involves using, altering, or artificially recreating content to impersonate a creator online. These kinds of attacks can do severe reputational damage. As a result, M9 reported offline harm such as physical assault and lethal threats.

Impersonated Profiles. Half of the woman creators in our sample were impersonated online [w:6 | m:1 | k:0]. Common platforms are dating profiles (W2, M9), and social media platforms (W1, W10). Attackers are unknown to participants, and motives are often unclear. Attackers deceived people to send them money (W2) or extract information from family members (W10). There was one case that led to offline harm: *"People started making fake Tinder and Bumble profiles. I don't know if it was something personal against me."* (M9). Even several years after the incident, the participant faces severe threats, as family members of the matches track him down and threaten to kill him. They are motivated by violations of strict cultural norms of *pardah/modesty* and *honor* [88].

Synthetic CSAM and NCII. Another form of impersonation, that was experienced by one woman in our sample, is synthetic non-consensual intimate imagery (NCII) or synthetic child sexual abuse material (CSAM). Participant W7 had synthetically-created pictures of her shared in an online community forum: *"I think to me the most unpleasant experience was the Reddit picture [...] where someone did something really weird on my picture. That was uncomfortable for me because that was more of a sexual nature. [...] I was 16. [...] someone sent me a screenshot of [the image]. I reported that and it got taken down because I [was] a minor "* (W7). Further, technologies such as Deepfakes generate content that is

harder to distinguish from reality, opening the door to harass the victim based on societal norms of decency.

4.2.7 Surveillance

Participants identified two types of attackers: (1) online contacts, such as fans or opponents, and (2) offline contacts, such as friends, family, or work-related acquaintances.

Stalking Behaviors. Stalking is a common experience among creators [w:4 | m:2 | k:2]. One participant reported being digitally stalked: *"There was a case with a house help of mine that got access to my accounts and then ended up stalking me from different accounts. That was a very big concern for me and a very unpleasant experience"* (W6). Stalking, especially if exerted by people close to the victim, directly or indirectly reduces the victim's physical or psychological integrity. Incidents of stalking may co-occur with physical threats, as was the case for this participant: *"I've had physical concerns in the sense that randomly, people 1-2 times have sent me messages: 'This was your car and this was the numberplate. We saw you take a u-turn from here.' "* (W5). Thereby stalking of creators is an example of a post-digital security issue [27], as it blurs the line between the online and offline worlds.

Account Monitoring. Similarly, female participants reported that they are concerned and annoyed that parents and other relatives monitor their account (W2, W3, W6, W9, W11, W12). This is in part due to socio-cultural norms of decency, as W9 puts it: *"People in our society are not accustomed to looking at pictures or videos of a woman on social media."* and W3: *"I know that a lot of people from my family have approached my mom in a way, saying you have absolutely no control over her daughter and she should be ashamed of what I was doing."*

4.2.8 [Offline] Physical Harm

The various online threats above can manifest into offline attacks causing physical harm to creators. Participants are especially concerned about influential Pakistani figures abusing their power and network to harm them. For U.S. creators physical-world harm was a top concern, however only few had personal experiences [79]. In our study 11 participants reported negative offline experiences, suggesting that they might be more prominent in the context of Pakistan.

Physical and Sexual Assault, Murder. The majority of participants were aware and concerned about the possibility of becoming targets in the real world. Three participants experienced physical assault [w:1 | m:1 | k:1]. M9 went through an extreme case of assault as a consequence of being impersonated on a dating profile: *"I was getting into my car and behind my car two cars came and parked. A person came out from one of the car and he held me by the collar pushed me against the wall and he pointed a gun to my head and he was you have done this and that to my sister."* K2 reported getting kidnapped. Mostly Khwaja Sira were concerned about sexual assault; none of our participants experienced it. Murder is real threat as is exemplified by M9 and past events [55].

4.2.9 [Offline] Harassment.

This category contains offline attacks that cause the victim to feel intimidated, dehumanized, or belittled.

Sexual Harassment. Sexual harassment like catcalling (W9) was not experienced by men [w:1 | m:0 | k:2]. Khwaja Sira faced severe instances of offline sexual harassment through cis-woman: *"As a trans woman interacting with a cis woman, the privilege lies with the cis women. Harassment at the hands of cisgender women, I have faced a lot, which can both be what I call sexual harassment and then harassment of a private nature where they ask me very, very intrusive and disgusting questions in the presence of other people"* (K1). She explains that in Pakistani patriarchy, women cannot harass cis-men, so they harass transgender people.

Threats of Violence. Another form of intimidation is threats of violence, such as threatening sexual assault (W9) or death (W9, M9, K2). Many participants are concerned about the possibility of somebody showing up at their address to threaten them. Creators are sometimes threatened to influence their posting behavior or delete content: *"Lucky for me it was not someone influential otherwise they would have forced me to delete it rather than asking politely and might have reached my house and telling me are you deleting it or we make you delete it"* (M1). Others will threaten to kill creators to restore honor (M9).

Stalking Behaviors. All genders experienced offline stalking behaviors [w:4 | m:3 | k:2]. While participants are generally open to take pictures and meet fans in public, some are concerned about having pictures and videos taken without their consent because of being public figures (W3). Similarly, creators reported instances where fans went to their house or current location to take pictures or meet them (W7). In contrast to the previous category, here the intention is not to threaten the creator, but to be near them. The close-knit communities within Pakistan simplify the process of obtaining creators' personal information for attackers (K1). This is especially the case in rural areas, and when the creator has great visibility [87], e.g., because of being Khwaja Sira (K2).

4.3 Coping with Threats

We categorize *defensive mechanisms* and *mindsets* of creators when handling threats described above in Section 4.2 and the *external support* they rely on. The results of this section contain a thematic analysis of the axial category *defense practices* and *support*. Table 1 relates the findings of this section to the previously described *threat categories*.

4.3.1 Defensive Mechanisms

Technical Defenses. All participants rely on **technical** defense mechanisms to protect their accounts, content, and

Table 1: Mapping of defensive mechanisms, mindsets, and external support to threat categories (c.f. Figure 1).

Threat Category	Defenses				Mindsets				External Support				
	Technical	Data Policies	Self-Censorship	Offline	Ignore	Comply	Fight	Fatalism	Information Sources	Platforms	Authorities	Social Suppport	Therapy
Toxic Content	●		●		●	●	●	●	●	●	●	●	●
Content Leakage + Theft	●	●					●					●	
Overloading	●					●	●					●	●
Lockout + Control	●								●	●	●	●	
False Reporting								●	●	●		●	
Impersonation					●		●		●	●	●	●	
Surveillance		●						●	●		●	●	
Physical Harm		●	●	●			●	●	●			●	
Offline Harassment		●	●	●			●	●	●			●	

●: observed with one or more participants

manage their community. To battle *account lockout*, few participants use *authentication-related* defenses, such as password complexity rules (W1, M7), password managers (W1, W9), and login monitoring (F6). While some participants follow outdated security practices, such as changing passwords regularly (W1, W5, W6, M9), the majority use enhanced authentication schemes such as 2FA [w:12 | m:7 | k:2]. 2FA in particular is perceived as a strong security mechanism, however some participants only adopted it when the platform pushed it on them after account compromise (M9, W12), or they saw suspicious login attempts (W5). Similarly, U.S. creators adopt protective practices in response to attacks [79]. Moreover, most participants [w:12 | m:6 | k:2] rely on *community management* tools to prevent and *combat toxic content* and *overloading*. Mechanisms that restrict who can send messages or reply to content (W7) are used to silence attackers and perceived as effective. Yet, the majority of participants were not satisfied with the available comment moderation, blocking, and reporting tools platforms provide to prevent hate and harassment and de-platform attackers. Similar to how young Pakistanis collaborate to counteract impersonated profiles [15], creators rely on their community to mass-report attackers in the case of hate speech and impersonation. Some participants [w:5 | m:1 | k:0] use *content control* methods to protect against *content theft*. For example, creators sometimes use watermarks on their content (W11). However, these do not stop the content theft itself, but rather ensures that the source will be known (if the watermark is not removed entirely). Other creators completely separate content into private accounts to prevent it from being misused (W1, W2, W3, W12, M9). To prevent content being deleted after a hack, one creator curates backup accounts to which they can switch in case of a hack (W1). This purpose is also communicated to fans and they are encouraged to follow the backup account.

Data Policies. Most participants [w:12 | m:6 | k:2] report following personal data control policies to combat *content leakage*, *surveillance*, and offline threats. The underlying principle involves keeping personal data points private, such as details about family and friends, location data, daily activities, or means of transport. Some go as far as hiding their identity online or in real life (W9, W11, M9), or leaking false information to distract people (K1). These practices are universally perceived as effective strategies to preserve online and offline privacy among our participants.

Self-Censorship. Similarly to a study on U.S. creators [79], the majority [w:9 | m:9 | k:2] of our participants engage in self-censorship to protect from becoming the targets of online and offline hate and harassment. Self-censorship is a learned pattern of behavior that results from one's own or other people's experiences. Thereby, implicit or explicit social norms influence what may be said and done. While one participant acknowledges self-censorship as an effective strategy to prevent hate (W2), our data suggests that it does not prevent creators from harassment, particularly related to religion. Against this background, self-censorship is a fragile precaution: *"I also avoid political and religious discussions online because I know for a fact that will never end well. [...] The most precautionary thing is to extremely filter out everything that I say and do"* (W3).

Offline Protections. Several participants [w:1 | m:2 | k:2] report relying on offline defenses to protect from *physical harm* and *harassment*. They carry pepper spray and rely on physical protection from other people (W9). Khwaja Sira face increased risk of offline harassment and report being selective about the people they meet (K2) and avoid attending parties in response to their digital visibility: *"I'm a digital creator who was getting very famous, and if I'm inviting 20 people and 25 show up, at least eight of them are going to show off saying [...] we ran into her. [...] They will go post my pictures and then that becomes a whole shitstorm. Like they would say this person talks about Islam or Sufism and now you see them in short cloths"* (K1). In that way, for creators offline privacy invasions can feed into online harassment.

4.3.2 Coping Mindsets

Based on how participants reacted to negative experiences we derived four *coping mindsets*:

Ignore. When coping with online harassment participants had the sentiment that this is something that one simply has to *accept* and *ignore* [w:12 | m:8 | k:2]. A similar theme was identified by Samermit et al. when U.S. creators claimed to develop a "thick skin" [79]. In line with their work, we find

that creators do not want to give haters and trolls attention, hoping that this way the attack will die down. We identified an additional theme of Pakistani creators being concerned that by standing up for themselves they might confront someone influential who could target them in the real-world (M1). *Ignoring* is often informed by a sense of helplessness when creators feel they cannot defend against a threat. This resetting of safety expectations when facing elevated risk applies to U.S. creators, too [79]. We find this theme especially for *toxic content* type threats when creator's way of coping is to *"just get used to it"* (W7). Similarly, most could not imagine a world without online hate and harassment attacks. However, this does not mean that they do not deploy defensive mechanisms. Acceptance of risks was felt throughout our participants; for instance regarding impersonation, one participant noted they lacked adequate tools to deal with such attacks and therefore became desensitized: *"Impersonation happens a lot. Initially, I would get very scared about it. I would report the account and I would ask my friends [to report] them also. But then slowly, as the number of accounts increased, I got desensitized."* (W3).

Comply. One creator reports *complying with* attackers to stop hate and harassment, especially when facing cancelling. M4 apologized and changed his behavior and content. *"I uploaded three apology videos. [...] Did it once and deleted that account because I was getting too many death threats."* In that way, this coping mindset is connected to *self-censorship*.

Fight. Many creators [w:4 | m:3 | k:2] report fighting online hate and harassment, especially if other coping mechanisms like *ignoring* or *complying* fail: *"I try to ignore it as much as I can, but if I can't and if I lose [my mind], then I'll have to answer to that person back"* (M3). They speak up for themselves (M3, K2), call people out (W2, W5), respond to comments (K2), and sometimes harass people back - the latter also offline (K1). We found that fighting back is a coping strategy that all participants who create sensitive or controversial topics rely upon (K1, K2, W9). We theorize that creators who actively decide to put out controversial content are willing to fight for their voice. But also creators on seemingly uncritical topics easily become targets. When asked whether she ever considered leaving the platform due to the hate she faces, W12 responded: *"I think that would be the weakest I can do. I chose to fight, be brave, and to deal with it."*

Fatalism. Across genders [w:7 | m:7 | k:1] participants developed a fatalistic mindset towards online and offline threats: *"If it's harassment, that's something that I still haven't figured out how to deal with [...] Though, I don't know how I will [ever] figure out how to deal with that"* (W3). In response to online hate participants left platforms (W5, W9), or even fled the country (M9). The attacker who almost killed M9 advised him to leave the country even after being convinced that M9 was innocent: *"He's like, 'Today was me tomorrow it could be some other guy from that group who tries to do something like this.' I was like, 'What do you suggest I should do?' He's like 'I would honestly say, stay off the grid. If you can [...] leave the country for a while.' "* (M9). W8 stopped creating content on TikTok because her account kept getting reported.

4.3.3 External Support

Information Resources. Two participants (W9, K2) explicitly mentioned education as a critical step to maintaining their security and privacy: *"Everything is just changing so much and technology is advancing at a speed of light. I'm just trying to understand it first, and then I'll take some measures about it."* (W9). Participants got their information about how to defend against threats from fellow creators [w:7 | m:7 | k:0], people in their personal lives [w:10 | m:7 | k:2], search engines [w:2 | m:1 | k:1], platform guidelines [w:3 | m:2 | k:1], advertisements (M8), and organizations (e.g., Digital Rights Foundation) (K1, K2).

Platforms. Creators appreciate platforms having physical offices (M8) and help centers (W12). They rely on *platform support* to combat *impersonation* (W4), and recover from *account compromise* after being hacked (W12). Also, platforms are the only way to recover from account loss due to *false reporting* attacks. In the case of *impersonation* participants stress that they *"don't know any other way to deal with it"* (W4). Creators criticize the slow response time of some platforms (Instagram) when taking down *content misuse*, and highlight that others (YouTube) are doing a good job (W9). Wanted features are tools to check for re-posted and impersonated content (W2, W3, W4, W7), controls over who can view and screenshot content (W1, W3, M9), measures to counter spread of *slander* (W2), and involving real people to resolve conflict (W2, W6, W12, M2, M9), e.g. in the form of a help line (W2, W4, W6, W7, W9, M2). Moreover, they would appreciate transparency over how banning works (W1, W3, W12) and the formation of support communities (W5, W7, W9, W11, M7).

Authorities. Few participants reported *involving authorities* when dealing with online or offline threats. Creators rarely reached out to *government agencies*, and if they did it was not very effective (e.g., they got no response (W5)). W2 summarizes why: *"I think in Pakistan especially, like nobody really directly goes to authorities. It's usually through contacts. [...] I don't recall there being sort of easy access to authorities where it's like a helpline or something. Maybe there is one, but [...] it's not being advertised enough. Because I think there's also a general distrust in society, like in terms of authority. We don't really trust them because it's very hard to find authority that takes your word and actually brings justice"* (W2). Even creators who generally fight back against hate and harassment are reluctant to report attackers officially: *"[If] I lodge cases against them, I'm pretty certain things against me are going to go ballistic. It's a double edged sword. I will take an action, but that action is also probably going to cost me my life"* (K1). None of our participants mentioned involving authorities in response to offline threats. Other

participants reported being in contact with activist agencies (K1), or reaching out to lawyers (W2, K2) and experts (K2).

Social Support. To cope with negative experiences, especially hate and harassment, participant rely on *social support*. They commonly refer to family and friends for support and to get a reality check (M3), although W3 tends to hide negative experiences from her close family because *"they already won't approve of <content creation>, and they'll tell me to stop"*. Another popular form of support is to talk to fellow creators and receive advice on how they navigate negative experiences (W4).

Therapy. Last but not least, a few participants go to therapy to cope with the outcome of online hate and harassment. The vast majority of our participants report on how content creation and the threats they face negatively impacts their mental health. They question their self-worth (M4), and face depression (M4) and burnout (W12, K1). One participant reported becoming suicidal after facing severe online and offline harassment (M4).

5 Discussion

5.1 Risk Factors (RQ1)

We discuss factors that correlate with experiencing heightened risk based on the relative frequency of reported threat types (compare Table 4 in Appendix C), and the reasoning participants offered. These factors have to be quantitatively validated in future work.

Gender. Our data suggest that women and Khwaja Sira face heightened threats and are more likely to experience sexual harassment than men (average # of reported negative experiences: [w:7.33 | m:5 | k:17]); future work needs to validate this hypothesis. Women are pressured to adhere to social norms and face heightened risks in online spaces when identifiable as women (W12). Thomas et al. [87] found that woman creators in the U.S. statistically faced more threats than other genders; they did not find effects for transgender persons. In this study the threats Khwaja Sira faced were severe: Among others, they got kidnapped (K2), were sexually harassed online and offline (K1, K2), and experienced human-trafficing attempts (K2). Sexual harassment is an extremely sensitive topic. Victims are reluctant to speak up because of getting stigmatized, i.e., tainting their honor. Also, victims need to be cautious when acting against attackers. Confrontational responses like calling out people, especially if they are powerful (e.g., religious leaders) can become lethal [55]. Thus, technical solutions need to focus on victim protections; solutions such as blocking that draw attention on the actions of victims might not be suitable as it can provoke attackers.

Content Topic. Participants who create content on topics that are controversial face more negative experiences than those who do not. In contrast, in the U.S. content topics were not correlated with higher risk [87]. However, recent research identified content on sex work and nudity leads to de-platforming [13, 14] or shadowbanning [12]. In our study, examples of critical content topics are social issues and activism (W9), women's rights (W9), transgender rights (K1, K2), and sex and sexuality (K1, K2). Moreover, many participants report avoiding content on religion (W2), politics (M2), and influential people (W9) to avoid harm. On the contrary, creators in our sample whose content centers on topics like music (W4) and pets (W12) face fewer threats. Yet, seemingly harmless content such as pictures in a mixed-gendered group (W3), dance videos (W8), or wearing shorts (K1) can be *"stigmatized in [Pakistan's] society"* (W8), and bear the potential to be weaponized, especially against woman and Khwaja Sira. Thus, platforms need to take local cultural norms into consideration when deciding what classifies as "harmful" content and must not make assumptions based on Western views.

Platform and Audience. Some participants see a connection between the platform and the hate they receive. They suspect that the audiences differ between platforms, and that platforms with more low-literacy users (they name TikTok) harbor more potential for hate and harassment. Moreover, they think certain interaction features and affordances that allow direct contact with creators enable harassment. However, results are inconclusive, and this factor must be explored further in future work. Some creators reported that they started attracting hate only after reaching a bigger audience (W8). This is in line with Thomas et al. [87], who reports a correlation between hate and audience size. Moreover, Samermit et al. [79] describe *prominence* (in accordance to Warford et al. [94]) as a risk factor for creators in the US, e.g. when massive popularity and virality further amplify risk.

5.2 Gaps in Defenses and Support (RQ2+3)

Flawed Defenses. *Toxic content* is a prominent threat that a majority of our participants experience. It is also ranked a top priority by experts, and they advise *blocking*, *muting*, *reporting*, and *moderating* [95] which our participants also rely on. Platforms invest great effort into combating *toxic content*, and we find that participants use and appreciate moderation tools that help filter comments. However, they report that it is especially difficult to de-platform malicious actors, and criticise the lack in transparency and efficacy of reporting mechanisms. In line with this, other research discusses how reporting in Urdu is harder than reporting in English [88]. Work on U.S. creators found that involving platforms can be slow and opaque if the creator does not have a human contact at the platform [79]. In our sample creators often felt left alone with the responsibility to defend themselves. However, some active forms of defense, like speaking up and addressing hate, can have lethal consequences in Pakistan's tight social net when perpetrators in turn feel attacked and use their network to reach creators offline. Creators deliberate this when speaking up against hate; some may have a cause that they deem worth fighting for, e.g., activism for minorities, women and transgender rights. However, a policy brief by

UNESCO and the Digital Rights Foundation [88] discusses that posts on sensitive topics lead to less engagement, either because audiences self-censor too or because of algorithmic biases [10]. In general, all creators were aware of the risks associated with sensitive topics such as religion and politics, and thus most of our participants refer to self-censorship.

Missing Protections. Participants lacked any defenses against threats of *false reporting* and *impersonation*. A recent study by Wei et al. [95] in which experts ranked threat categories internet users should prioritize, threats that fall under the category *impersonation* and *false reporting* were assigned medium ranking and rarely identified as top threats. In contrast, we found that *impersonation* attacks likely lead to offline threats and are thus highly critical in Pakistan. In our data set, we had one participant who almost got killed due to being impersonated on an online dating app (M9). Regarding Pakistan, *impersonation* poses a high risk for (1) the reputation of the victim, but also (2) for the reputation of the person who falls for the scam, especially in the context of impersonated dating profiles. M9 still receives online and offline threats from family members of the scam victim, even several years after the incident. Threats are so severe, that even former attackers advise him to leave the country. This points out the potential for geographic biases when selecting generalized security advice [95] and highlights the need for targeted advice, especially for populations facing elevated threats. Critically, even for the general internet user experts identified a lack of effective advice to defend against *impersonation* [95]. Future research needs to work towards effective and feasible defensive mechanisms taking into account the needs of marginalized populations such as creators in Pakistan. Moreover, creators are lacking effective protections against *stalking behaviors*; they rely on *data policies* that are implicit rules to control which data they release. However, it is difficult to foresee which data points might result in *stalking behaviors* and retracting data that is published once, is hard or impossible. In this regard creators are different from young adult internet users in Pakistan [15], as they cannot rely on the same affordances (e.g. having private profiles).

Lacking Support. The socio-cultural context of Pakistan impacts availability of support structures, especially as content creation is unregulated and lacks protective legislation [15, 88]. First, there are gaps when it comes to official support structures for online harm. Participants were reluctant to approach authorities, or unaware of agencies that deal with online hate and harassment. Underscoring the problem, even the one participant in our sample who filed a report never heard back from the authorities. This issue is similar for marginalized content creators in the West such as sexual content creators who similarly report being unable to turn to authorities due to stigma and lack of respect [38]. For U.S. content creators, law enforcement can be helpful, but they are not always taken seriously [79]. Second, participants report a stigma that is attached to content creation in Pakistan, especially for women: in our sample they were concerned about their parents and close relatives monitoring their accounts. This in turn might constrain the *social support* resources women can rely on, if they do not feel they can approach their parents when being attacked. Similarly, young adults in Pakistan are reluctant to report cybercrimes because they do not want their families to worry, or women want to avoid being victim blamed [15].

5.3 Towards Solutions

Threat Modeling. There is a need to explore the design for flexible threat priorities where socio-economic context determines what are severe threats, as opposed to viewing threats as static across social contexts. Comparing our results (e.g. impersonation attacks leading to offline harm) to previous work focusing on Western populations where experts ranked threat categories that internet users should prioritize [95], we suggest that different socio-economic contexts influence the prioritisation of threats.

Technical Defenses. Creators in Pakistan are especially dependent on well-designed technological defenses because of missing social and societal protections (e.g. reluctance to approach authorities, involve families). Defenses need to take cultural norms into account when (1) detecting harmful content (e.g. pictures of mixed-gendered groups, comments by religious figures), (2) protecting victims (defensive mechanisms should not leave traces), as Pakistan's tightly-woven social-net allows attackers to reach creators offline, or mobilize a crowd to attack victims based on false claims (e.g. religious offense).

Tailored Information Sources. Creators expressed a desire for contextualized information sources. We suggest to consider factors that impact creators' threat experiences when generating advice. Based on our findings we identified gender, content topic, platform, (size of) audience, and broader social embeddedness including cultural background.

6 Conclusion

Creators in Pakistan and the US [87] face online threats in the same categories, however there are differences in risk surface and expected harms: (1) what classifies as harmful content depends on the cultural context, and (2) we observe that online threats can be expected to lead to offline harm, which was not reported for U.S. creators [79]. Here, threats like impersonation that are rated "moderate" by experts [95] lead to potentially lethal attacks. Hence, we hypothesize that although creators across the world face the same threat categories, the prioritization of threats changes across cultures. Finally, compared to young adults in Pakistan [15], the threats creators face are even amplified. Furthermore, countermeasures such as taking their profiles offline do not work for creators which suggests that they need protection mechanims for their specific needs.

Acknowledgements

The fifth author conducted a portion of this work while supported by the Max Planck Institute for Software Systems and additionally acknowledges support from a Google Research award focused on "at-risk users and creator safety".

References

[1] Pakistan labour force survey 2020-21. https://www.pbs.gov.pk/sites/default/files/labour_force/publications/lfs2020_21/LFS_2020-21_Report.pdf, last accessed: 20.09.2023.

[2] Pakistan's third gender recognized but still discriminated. DW, https://www.dw.com/en/pakistans-third-gender-recognized-but-still-discriminated/audio-16958385, last accessed: 12.04.2023.

[3] 24 NEWS HD. Jannat Mirza returns to Pakistan after long stay in Tokyo. 24 NEWS HD, https://www.24newshd.tv/10-Nov-2020/jannat-mirza-returns-to-pakistan-after-long-stay-in-tokyo, last accessed: 05.04.2023, 2020.

[4] ABBAS, S. K., HASSAN, H. A., ASIF, J., AND ZAINAB, F. How income level distribution responds to poverty: Empirical evidence from pakistan. *Global Scientific Journals 6*, 3 (2018), 131–142.

[5] ABOKHODAIR, N., AND VIEWEG, S. Privacy & social media in the context of the Arab Gulf. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (2016), pp. 672–683.

[6] AHMED, S. I., HAQUE, M. R., CHEN, J., AND DELL, N. Digital privacy challenges with shared mobile phone use in Bangladesh. *Proceedings of the ACM Conference on Human-Computer Interaction 1*, CSCW (dec 2017).

[7] AHMED, S. I., HAQUE, M. R., CHEN, J., AND DELL, N. Digital privacy challenges with shared mobile phone use in bangladesh. *Proceedings of the ACM on Human-Computer Interaction 1*, CSCW (2017), 1–20.

[8] AHMED, T., HOYLE, R., CONNELLY, K., CRANDALL, D., AND KAPADIA, A. Privacy concerns and behaviors of people with visual impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 3523–3532.

[9] ALI, F. Z. Gender discourse and transphobia in pakistan's digital sphere. https://blogs.lse.ac.uk/southasia/2022/12/19/gender-discourse-and-transphobia-in-pakistans-digital-sphere/, 2022.

[10] ALLISON-HOPE, D. Human rights due diligence of Meta's impacts in Israel and Palestine in May 2021. BSR, https://www.bsr.org/en/blog/human-rights-due-diligence-of-meta-impacts-in-israel-and-palestine-may-2021, last accessed: 31.05.2023, 2022.

[11] AMMARI, T., NOFAL, M., NASEEM, M., AND MUSTAFA, M. Moderation as empowerment: Creating and managing women-only digital safe spaces. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 1–36.

[12] ARE, C. The shadowban cycle: an autoethnography of pole dancing, nudity and censorship on instagram. *Feminist Media Studies 22*, 8 (2022), 2002–2019.

[13] ARE, C. An autoethnography of automated powerlessness: lacking platform affordances in instagram and tiktok account deletions. *Media, Culture & Society 45*, 4 (2023), 822–840.

[14] ARE, C., AND BRIGGS, P. The emotional and financial impact of de-platforming on creators at the margins. *Social Media+ Society 9*, 1 (2023), 20563051231155103.

[15] ASHRAF, A., KÖNIG, C. J., JAVED, M., MUSTAFA, M., ET AL. " stalking is immoral but not illegal": Understanding security, cyber crimes and threats in pakistan. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)* (2023), pp. 37–56.

[16] BARAKAT, H., AND REDMILES, E. M. Community under surveillance: Impacts of marginalization on an online labor forum. In *Proceedings of the International AAAI Conference on Web and Social Media* (2022), vol. 16, pp. 12–21.

[17] BARROSO DOMÍNGUEZ, A., ET AL. ¿ Redes de apoyo, comunicación y sororidad entre las mujeres creadoras de contenido erótico en OnlyFans? http://hdl.handle.net/10651/64197, 2022.

[18] BARWULOR, C., MCDONALD, A., HARGITTAI, E., AND REDMILES, E. M. "Disadvantaged in the American-dominated internet": Sex, work, and technology. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), pp. 1–16.

[19] BEGUM SADAQUAT, M., AND SHEIKH, Q.-T.-A. A. Employment situation of women in Pakistan. *International Journal of Social Economics 38*, 2 (2011), 98–113.

[20] BLACKWELL, L., HARDY, J., AMMARI, T., VEINOT, T., LAMPE, C., AND SCHOENEBECK, S. LGBT parents and social media: Advocacy, privacy, and disclosure during shifting social movements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 610–622.

[21] BOL NEWS. TikTok star Jannat Mirza does not want to hear criticism from Bushra Ansari. BOL NEWS, https://www.bolnews.com/entertainment/2021/06/tiktok-star-jannat-mirza-does-not-want-to-hear-criticism-from-bushra-ansari, last accessed: 05.04.2023, 2021.

[22] BUF, D.-M., AND ȘTEFĂNIȚĂ, O. Uses and gratifications of YouTube: A comparative analysis of users and content creators. *Romanian Journal of Communication and Public Relations 22*, 2 (2020), 75–89.

[23] CHARMAZ, K. C. *Constructing grounded theory*. Sage, 2014.

[24] CHATTERJEE, R., DOERFLER, P., ORGAD, H., HAVRON, S., PALMER, J., FREED, D., LEVY, K., DELL, N., MCCOY, D., AND RISTENPART, T. The spyware used in intimate partner violence. In *2018 IEEE Symposium on Security and Privacy (SP)* (2018), IEEE, pp. 441–458.

[25] CHAUDHARY, S., ZHAO, Y., BERKI, E., VALTANEN, J., LI, L., HELENIUS, M., AND MYSTAKIDIS, S. A cross-cultural and gender-based perspective for online security: Exploring knowledge, skills and attitudes of higher education students. *IADIS International Journal on WWW/Internet 13*, 1 (2015).

[26] CHOI, D., LEE, U., AND HONG, H. "it's not wrong, but i'm quite disappointed": Toward an inclusive algorithmic experience for content creators with disabilities. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022), pp. 1–19.

[27] COLES-KEMP, L., JENSEN, R. B., AND HEATH, C. P. Too much information: Questioning security in a post-digital society. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–14.

[28] CONSOLVO, S., KELLEY, P. G., MATTHEWS, T., THOMAS, K., DUNN, L. C., AND BURSZTEIN, E. "Why wouldn't someone think of democracy as a target?": Security practices & challenges of people involved with US political campaigns. In *Proceedings of the 30th USENIX Security Symposium* (2021), USENIX, pp. 1181–1198.

[29] CORBIN, J. M., AND STRAUSS, A. L. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology 19*, 6 (1990), 418–427.

[30] DAFFALLA, A., SIMKO, L., KOHNO, T., AND BARDAS, A. G. Defensive technology use by political activists during the Sudanese revolution. In *2021 IEEE Symposium on Security and Privacy (SP)* (2021), IEEE, pp. 372–390.

[31] DAILY PAKISTAN. TikToker assaulted by 400 men at Minar-e-Pakistan shares her ordeal (DP Exclusive). Daily Pakistan, https://en.dailypakistan.com.pk/18-Aug-2021/tiktoker-assaulted-by-400-men-at-minar-e-pakistan-shares-her-ordeal-dp-exclusive, last accessed: 05.04.2023, 2021.

[32] DEV, J., MORIANO SALAZAR, P., AND CAMP, J. Lessons learnt from comparing WhatsApp privacy concerns across Saudi and Indian populations. In *Proceedings of the Sixteenth Symposium on Usable Privacy and Security (SOUPS)* (2020), USENIX Association, pp. 81–97.

[33] DEVITO, M. A. How transfeminine tiktok creators navigate the algorithmic trap of visibility via folk theorization. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 1–31.

[34] EBERSOLE, C. Online sex work in the time of Covid-19. *Illinois State University Research Symposium 403* (2022).

[35] FORUM, W. E. Global gender gap report 2022. https://www3.weforum.org/docs/WEF_GGGR_2022.pdf, last accessed: 31.05.2023, 2022.

[36] FREED, D., PALMER, J., MINCHALA, D., LEVY, K., RISTENPART, T., AND DELL, N. "A Stalker's Paradise": How intimate partner abusers exploit technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–13.

[37] HADI, A. Patriarchy and gender-based violence in Pakistan. *European Journal of Social Science Education and Research 4*, 4 (2017), 289–296.

[38] HAMILTON, V., BARAKAT, H., AND REDMILES, E. M. Risk, resilience and reward: Impacts of shifting to digital sex work. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 1–37.

[39] HAMILTON, V., SONEJI, A., MCDONALD, A., AND REDMILES, E. M. "Nudes? Shouldn't I charge for these?": Motivations of new sexual content creators on OnlyFans. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 1–14.

[40] HAQUE, R. *Purdah of the heart and eyes: An examination of purdah as an institution in Pakistan*. University of New South Wales, 2003.

[41] HAVRON, S., FREED, D., CHATTERJEE, R., MCCOY, D., DELL, N., AND RISTENPART, T. Clinical computer security for victims of intimate partner violence. In *USENIX Security Symposium* (2019), pp. 105–122.

[42] HENNINK, M., AND KAISER, B. N. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine 292* (2022), 114523.

[43] HENNINK, M. M., KAISER, B. N., AND MARCONI, V. C. Code saturation versus meaning saturation: how many interviews are enough? *Qualitative health research 27*, 4 (2017), 591–608.

[44] HENRICH, J., HEINE, S. J., AND NORENZAYAN, A. The weirdest people in the world? *Behavioral and brain sciences 33*, 2-3 (2010), 61–83.

[45] IRFAN, A. Gruesome murder prompts Irfan Junejo to break his non-political stance. Lens, https://propakistani.pk/lens/gruesome-murder-prompts-irfan-junejo-to-break-his-non-political-stance/, last accessed: 05.04.2023, 2020.

[46] JACK, M. C., SOVANNAROTH, P., AND DELL, N. "Privacy is not a concept, but a way of dealing with life": Localization of transnational technology platforms and liminal privacy practices in Cambodia. *Proceedings of the ACM on Human-Computer Interaction 3*, CSCW (2019), 1–19.

[47] JAMAL, S. Social media users demand justice for Pakistani blogger stabbed to death. Gulf News, https://gulfnews.com/world/asia/pakistan/social-media-users-demand-justice-for-pakistani-blogger-stabbed-to-death-1.64684353, last accessed: 05.04.2023, 2019.

[48] KHAN, F. A. *Khwaja sira: Culture, identity politics, and" transgender" activism in Pakistan*. PhD thesis, Syracuse University, 2014.

[49] KINGSLEY, S., SINHA, P., WANG, C., ESLAMI, M., AND HONG, J. I. " give everybody [..] a little bit more equity": Content creator perspectives and responses to the algorithmic demonetization of content associated with disadvantaged groups. *Proceedings of the ACM on Human-Computer Interaction 6*, CSCW2 (2022), 1–37.

[50] KOMOK, H. Does the pay gap exist on Instagram? Remuneration of male vs female creators. HypeAuditor, https://hypeauditor.com/blog/does-the-pay-gap-exist-on-instagram-remuneration-of-male-vs-female-creators/, last accessed: 19.04.2023, 2020.

[51] KOW, Y. M., KOU, Y., SEMAAN, B., AND CHENG, W. Mediating the undercurrents: Using social media to sustain a social movement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 3883–3894.

[52] KUCKARTZ, U. Qualitative content analysis: From Kracauer's beginnings to today's challenges. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research 20*, 3 (2019), Art. 12.

[53] LERNER, A., HE, H. Y., KAWAKAMI, A., ZEAMER, S. C., AND HOYLE, R. Privacy and activism in the transgender community. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–13.

[54] LEUNG, L. User-generated content on the internet: An examination of gratifications, civic engagement and psychological empowerment. *New Media & Society 11*, 8 (2009), 1327–1347.

[55] MAHER, S. Viewpoint: Qandeel Baloch was killed for making lives 'difficult'. The British Broadcasting Corporation (BBC), https://www.bbc.com/news/world-asia-49874994, last accessed: 05.04.2023, 2019.

[56] MAJID, H., AND MUSTAFA, M. Transformative digital spaces? Investigating women's digital mobilities in Pakistan. *Gender & Development 30*, 3 (2022), 497–516.

[57] MARCZAK, W. R., AND PAXSON, V. Social engineering attacks on government opponents: Target perspectives. In *Proceedings on Privacy Enhancing Technologies* (2017), pp. 152–164.

[58] MARNE, S. T., AL-AMEEN, M. N., AND WRIGHT, M. K. Learning system-assigned passwords: A preliminary study on the people with learning disabilities. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS)* (2017).

[59] MASOOD, S. Pakistan strengthens already harsh laws against blasphemy. New York Times, https://www.nytimes.com/2023/01/21/world/asia/pakistan-blasphemy-laws.html, last accessed: 19.04.2023, 2023.

[60] MATTHEWS, T., O'LEARY, K., TURNER, A., SLEEPER, M., WOELFER, J. P., SHELTON, M., MANTHORNE, C., CHURCHILL, E. F., AND CONSOLVO, S. 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* (2017), pp. 2189–2201.

[61] MAYRING, P. Qualitative content analysis. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research 1*, 2 (2000), Art. 20.

[62] MCDONALD, A., BARWULOR, C., MAZUREK, M. L., SCHAUB, F., AND REDMILES, E. M. "It's stressful having all these phones": Investigating sex workers' safety goals, risks, and practices online. In *30th USENIX Security Symposium* (2021), USENIX, pp. 375–392.

[63] MCDONALD, N., BADILLO-URQUIOLA, K., AMES, M. G., DELL, N., KENESKI, E., SLEEPER, M., AND WISNIEWSKI, P. J. Privacy and power: Acknowledging the importance of privacy research and design for vulnerable populations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–8.

[64] MCGREGOR, S. E., CHARTERS, P., HOLLIDAY, T., AND ROESNER, F. Investigating the computer security practices and needs of journalists. In *24th USENIX Security Symposium (USENIX Security 15)* (2015), pp. 399–414.

[65] MCGREGOR, S. E., ROESNER, F., AND CAINE, K. Individual versus organizational computer security and privacy concerns in journalism. *Proceedings on Privacy Enhancing Technologies 2016*, 4 (2016), 1–18.

[66] MCGREGOR, S. E., WATKINS, E. A., AL-AMEEN, M. N., CAINE, K., AND ROESNER, F. When the weakest link is strong: Secure collaboration in the case of the Panama Papers. In *USENIX Security Symposium* (2017), pp. 505–522.

[67] MEHMOOD, H., AHMAD, T., RAZAQ, L., MARE, S., USMANI, M. Z., ANDERSON, R., AND RAZA, A. A. Towards digitization of collaborative savings among low-income groups. *Proceedings of the ACM on Human-Computer Interaction 3*, CSCW (2019), 1–30.

[68] MUMTAZ, Z., AND SALWAY, S. 'I never go anywhere': Extricating the links between women's mobility and uptake of reproductive health services in Pakistan. *Social Science & Medicine 60*, 8 (2005), 1751–1765.

[69] MUSTAFA, M. Negotiating borders through feminisms. *Interactions 27*, 6 (2020), 49–51.

[70] NAVEED, S., NAVEED, H., JAVED, M., AND MUSTAFA, M. "Ask this from the person who has private stuff": Privacy perceptions, behaviours and beliefs beyond WEIRD. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022), pp. 1–17.

[71] NISAR, M. A. (Un)becoming a man: Legal consciousness of the third gender category in Pakistan. *Gender & Society 32*, 1 (2018), 59–81.

[72] REDMILES, E. M., KROSS, S., AND MAZUREK, M. L. How I learned to be secure: A census-representative survey of security advice sources and behavior. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (2016), pp. 666–677.

[73] REDMILES, E. M., KROSS, S., AND MAZUREK, M. L. Where is the digital divide? A survey of security, privacy, and socioeconomics. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), pp. 931–936.

[74] REICHEL, J., PECK, F., INABA, M., MOGES, B., CHAWLA, B. S., AND CHETTY, M. 'I have too much respect for my elders': Understanding South African mobile users' perceptions of privacy and current behaviors on Facebook and WhatsApp. In *Proceedings of the 29th USENIX Conference on Security Symposium* (2020), pp. 1949–1966.

[75] ROLLER, M. R. A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research 20*, 3 (2019), Art. 31.

[76] SAMBASIVAN, N., AHMED, N., BATOOL, A., BURSZTEIN, E., CHURCHILL, E., GAYTAN-LUGO, L. S., MATTHEWS, T., NEMAR, D., THOMAS, K., AND CONSOLVO, S. Toward gender-equitable privacy and security in South Asia. *IEEE Security & Privacy 17*, 4 (2019), 71–77.

[77] SAMBASIVAN, N., BATOOL, A., AHMED, N., MATTHEWS, T., THOMAS, K., GAYTÁN-LUGO, L. S., NEMER, D., BURSZTEIN, E., CHURCHILL, E., AND CONSOLVO, S. "They don't leave us alone anywhere we go": Gender and digital abuse in South Asia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–14.

[78] SAMBASIVAN, N., CHECKLEY, G., BATOOL, A., AHMED, N., NEMER, D., GAYTÁN-LUGO, L. S., MATTHEWS, T., CONSOLVO, S., AND CHURCHILL, E. "Privacy is not for me, it's for those rich women": Performative privacy practices on mobile phones by women in South Asia. In *Proceedings of the Fourteenth Symposium on Usable Privacy and Security (SOUPS)* (2018), USENIX Association, pp. 127–142.

[79] SAMERMIT, P., TURNER, A., KELLEY, P. G., MATTHEWS, T., WU, V., CONSOLVO, S., AND THOMAS, K. {"Millions} of people are watching {you"}: Understanding the {Digital-Safety} needs and practices of creators. In *32nd USENIX Security Symposium (USENIX Security 23)* (2023), pp. 5629–5645.

[80] SCHEUERMAN, M. K., BRANHAM, S. M., AND HAMIDI, F. Safe spaces and safe places: Unpacking technology-mediated experiences of safety and harm with transgender people. In *Proceedings of the ACM on Human-Computer Interaction* (2018), vol. 2, ACM New York, NY, USA, p. Article 155.

[81] SCHREIER, M. *Qualitative content analysis in practice*. Sage, 2012.

[82] SHAH, Z. Muskan Sheikh and Rehan Shah: The TikTok stars gunned down in Karachi. Geo News, https://www.geo.tv/latest/332999-muskan-sheikh-and-rehan-shah-the-tiktok-stars-gunned-down-in-karachi, last accessed: 05.04.2023, 2021.

[83] STRAUSS, A. L., AND CORBIN, J. M. *Grounded theory in practice*. Sage, 1997.

[84] STROHMAYER, A., CLAMEN, J., AND LAING, M. 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* (2019), pp. 1–14.

[85] TADIC, B., ROHDE, M., WULF, V., AND RANDALL, D. ICT use by prominent activists in Republika Srpska. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 3364–3377.

[86] THOMAS, K., AKHAWE, D., BAILEY, M., BONEH, D., BURSZTEIN, E., CONSOLVO, S., DELL, N., DURUMERIC, Z., KELLEY, P. G., KUMAR, D., ET AL. Sok: Hate, harassment, and the changing landscape of online abuse. In *2021 IEEE Symposium on Security and Privacy (SP)* (2021), IEEE, pp. 247–267.

[87] THOMAS, K., KELLEY, P. G., CONSOLVO, S., SAMERMIT, P., AND BURSZTEIN, E. "It's common and a part of being a content creator": Understanding how creators experience and cope with hate and harassment online. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (2022), pp. 1–15.

[88] UNSECO, D. R. F. Policy brief: Centering Pakistani digital content and creators. https://digitalrightsfoundation.pk/wp-content/uploads/2022/12/UNESCO-Policy-Paper_FINAL-December-2022.pdf, last accessed: 31.05.2023, 2022.

[89] UTTARAPONG, J., BONIFACIO, R., JEREZA, R., AND WOHN, D. Y. Social support in digital patronage: OnlyFans adult content creators as an online community. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (2022), pp. 1–7.

[90] VAISMORADI, M., AND SNELGROVE, S. Theme in qualitative content analysis and thematic analysis. *Forum: Qualitative Sozialforschung / Forum: Qualitative Social Research 20*, 3 (2019), Art. 23.

[91] VASHISTHA, A., ANDERSON, R., AND MARE, S. Examining security and privacy research in developing regions. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18)* (2018), pp. 1–14.

[92] VASHISTHA, A., GARG, A., ANDERSON, R., AND RAZA, A. A. Threats, abuses, flirting, and blackmail: Gender inequity in social media voice forums. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–13.

[93] VITAK, J., LIAO, Y., SUBRAMANIAM, M., AND KUMAR, P. "I knew it was too good to be true": The challenges economically disadvantaged internet users face in assessing trustworthiness, avoiding scams, and developing self-efficacy online. *Proceedings of the ACM on Human-Computer Interaction 2*, CSCW (2018), 1–25.

[94] WARFORD, N., MATTHEWS, T., YANG, K., AKGUL, O., CONSOLVO, S., KELLEY, P. G., MALKIN, N., MAZUREK, M. L., SLEEPER, M., AND THOMAS, K. Sok: A framework for unifying at-risk user research. In *2022 IEEE Symposium on Security and Privacy (SP)* (2022), IEEE, pp. 2344–2360.

[95] WEI, M., CONSOLVO, S., KELLEY, P. G., KOHNO, T., ROESNER, F., AND THOMAS, K. "There's so much responsibility on users right now": Expert advice for staying safer from hate and harassment. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 1–17.

[96] WENDT, N., JENSEN, R. B., AND COLES-KEMP, L. Civic empowerment through digitalisation: The case of greenlandic women. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–13.

[97] YOUNAS, F., NASEEM, M., AND MUSTAFA, M. Patriarchy and social media: Women only Facebook groups as safe spaces for support seeking in Pakistan. In *Proceedings of the 2020 International Conference on Information and Communication Technologies and Development* (2020), pp. 1–11.

Table 2: The demographic of the 23 creators of this study.

Demographic	**Group**	**N**	**%**
Gender	Woman	12	52%
	Man	9	39%
	Non-Binary	0	0%
	Khwaja Sira	2	9%
	Prefer not to say	0	0%
Age	18-24	19	83%
	25-34	4	17%
Income from Creation in PKR	<50k	1	4%
	50k - 100k	1	4%
	100k - 150k	2	9%
	150k -200k	2	9%
	>200k	7	30%
	Prefer not to say	10	44%
Education	Below high school	1	4%
	High school	11	48%
	University	11	48%
Total Followers	10k - 50k	11	48%
	50k - 100k	3	13%
	100k - 200k	5	22%
	200k - 300k	3	13%
	>300k	1	4%

Table 3: Platforms where participants create content.

Online Community	**N**	**%**
Instagram	23	100%
TikTok	15	65%
YouTube	14	61%
Facebook	6	26%
Twitter	2	9%
Snapchat	1	4%
Spotify	1	4%

Appendix

A Sample Descriptives

Table 2 summarizes the demographics of the study participants. Table 3 provides an overview of the platforms on which participants create content.

B Interview Guideline

Figure 2 illustrates the structure of the interviews. For the full interview guideline, as well as details on the screening survey, refer to the supplementary material.

C Overview of Negative Experiences

Table 4 provides an overview of the negative experiences participants reported on during the interview.

Table 4: Overview of threats experienced by participants according to the coded interview data. "E" denotes that the participant reported a personal negative experience in this category. "T" denotes that they did not report a negative experience, but explicitly stated that they are concerned about an attack in this threat category.

Threat Category		W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	M1	M2	M3	M4	M5	M6	M7	M8	M9	K1	K2
Toxic Content	Bullying	E	E	E		E	E	E	E	E	E	E	E	E	E	E	E			E			E	E
	Trolling	E	E	E													E		E				E	
	Hate Speech	E	E	E	T	E	E	T	E	E	E	E	E	E			E	E					E	E
	Sexual Harassment							E	T	E													E	E
	Slander		E							E			E	T			E			E		E	E	E
	Conspiracy Theories		E														E						E	E
	Unwanted Explicit Content			E	E	E		E	E	E	E	E	E		E					E				E
	Blackmailing											E					E				E	E		E
	Threats of Violence			E						E				T	T		E	E					E	E
	Incitement		E							E							E	E				E	E	E
Content Leakage	Doxing	T		T	E	T	T	E	E	E	T	E	T				E	E					E	E
	Outing and Deadnaming																						E	
	Content Misuse	T	E	E			T	E		E		E	T	E						E		E		E
Overloading	Comment Spam	E	E			E			E		E		E				E	E		E				E
	Negative Ratings								E								E							
	Brigading/ Dogpiling		E						E	E														E
	Canceling		E														E			E				
False Reporting	Account Banning	E		T					E				T		T									E
	Content Reporting	T							E													E		
Impersonation	Impersonation	T	E	E	E	E		E			E											E		
	Synthetic CSAM/NCII							E					T											
Surveillance	Stalking Behaviors		E			E	E				E						E	T				E	E	E
	Account monitoring		T	T			T			T		T	T											E
Lockout + Control	Account compromise			T		T			E				E			T		T	T	T		E		E
Physical Harm	Physical Assault	E		T		T		T		T				T								E	T	E
	Sexual Assault									T							T						T	T
	Murder					T				T							T	T				T	T	T
Offline Harassment	Sexual Harassment							T		E													E	E
	Stalking Behaviors			E		E		E			E		T	T			T	T	E		E	E	E	E
	Threats of Violence									E				T			T	T				E	T	E

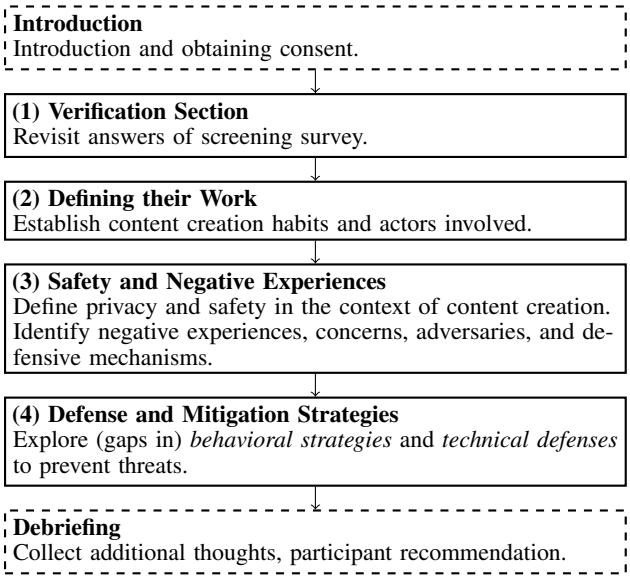

Figure 2: Overview of the semi-structured interview guideline.