Examining the Unique Online Risk Experiences and Mental Health Outcomes of LGBTQ+ versus Heterosexual Youth

Tangila Islam Tanni
TangilaIslam.Tanni@ucf.edu
University of Central Florida
Orlando, Florida, USA

Mamtaj Akter
Mamtaj.Akter@vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

Joshua Anderson
Jo178841@ucf.edu
University of Central Florida
Orlando, Florida, USA

Mary J. Amon
MJAmon@ucf.edu
University of Central Florida
Orlando, Florida, USA

Pamela J. Wisniewski
Pamela.Wisniewski@vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

Abstract

We collected and analyzed Instagram direct messages (DMs) from 173 youth aged 13–21 (including 86 LGBTQ+ youth). We examined youth's risk-flagged social media trace data with their self-reported mental health outcomes to examine how the differing online experiences of LGBTQ+ youth compare with their heterosexual counterparts. We found that LGBTQ+ youth experienced significantly more high-risk online interactions compared to heterosexual youth. LGBTQ+ youth reported overall poorer mental health, with online harassment specifically amplifying Self-Harm and Injury. LGBTQ+ youth's mental well-being linked positively to sexual messages, unlike heterosexual youth. Qualitatively, we found that most of the risk-flagged messages of LGBTQ+ youth were sexually motivated; however, a silver lining was that they sought support for their sexual identity from peers on the platform. The study highlights the importance of tailored online safety and inclusive design for LGBTQ+ youth, with implications for CHI community advancements in fostering a supportive online environments.

Content Warning: This research discusses sensitive topics, including explicit sexual content, abusive language, and homophobic slurs, which may cause discomfort. Reader discretion is advised.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Empirical studies in HCI**;

Keywords

LGBTQ; Online Risks; Youth; Social Media; Self-Harm; Instagram; Mental Health; Online Safefy; Sexual Risk

ACM Reference Format:
Tangila Islam Tanni, Mamtaj Akter, Joshua Anderson, Mary J. Amon, and Pamela J. Wisniewski. 2024. Examining the Unique Online Risk Experiences and Mental Health Outcomes of LGBTQ+ versus Heterosexual Youth. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642509

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642509

1 Introduction

Social media platforms like Instagram, TikTok, and Snapchat are popular among teens, especially those who feel isolated or marginalized, such as youth from the lesbian, gay, bisexual, transgender, queer or questioning communities (LGBTQ+) [4, 52]. As of 2020, an estimated 9.5% of the youth population in the United States is comprised of LGBTQ+ youth (ages 13-17) [35], which represents a growing demographic of youth with unique needs related to their sexual and gender identities [131]. Consequently, many LGBTQ+ youth turn to social media platforms as a means to seek emotional and social support, foster friendships, and find validation [95]. Community engagement within the LGBTQ+ community can aid these youth in their identity exploration and foster confidence in themselves and their sexuality [37, 96]. However, other researchers have highlighted the negative effects of online engagement for this vulnerable community, as LGBTQ+ youth also face unique risks when it comes to navigating online spaces [2, 66, 127, 151]. For instance, research has shown that they are disproportionately a target of harassment and other forms of cyberbullying, compared to their heterosexual counterparts [2]. These negative experiences have been shown to cause mental health issues for LGBTQ+ youth and, at the extreme, may even lead to suicidal tendencies or self-harm [88].

Recently, social media interactions have been trending towards private message communication instead of public postings due to several reasons, including privacy concerns [45, 75, 148], to escape public scrutiny, and to reduce data visibility [125]. Private forms of communication have unique benefits (e.g., social support, advice from peers, etc.) but may also subject users to new harms (e.g., targeted harassment, hate speech, sexual grooming). Private conversations allow increased anonymity, sometimes facilitating inappropriate or uncharacteristic behaviors online [149]. For instance, prior research suggests that most of the abuse on social media platforms comes in the form of Direct Messages (DMs) from strangers [80]. As DMs are considered private conversations, the majority of social media platforms do not address hate speech or bullying experienced through these channels [76]. Therefore, these private chats represent a potentially unique threat yet to be examined, particularly

in the context of LGBTQ+ youth in comparison with heterosexual youth. Yet, a common theme among existing research studies is that they have historically relied heavily on youth self-reports [12] or publicly scraped social media data [121, 124], which constrains what we know about LGBTQ+ youth's experiences in private spaces on social media. Therefore, to better understand the online risks encountered by LGBTQ+ youth in private social media spaces and how this may influence their overall mental health, the current study poses the following high-level research questions:

- **RQ1:** *Do LGBTQ+ youth experience more frequent and/or severe online risks in their private social media conversations compared to heterosexual youth?*
- **RQ2:** *Do LGBTQ+ youth who experience negative online experiences within their private conversations have more mental health challenges compared to heterosexual youth?*
- **RQ3:** *What is the unique nature of the online risks LGBTQ+ youth experience in private online chats?*

To answer our research questions, we conducted a user study with 173 youth (ages 13-21), including 86 LGBTQ+ participants. We measured several pre-validated constructs regarding the youths' mental health and had them upload and flag their Instagram DMs that they felt uncomfortable or unsafe about. To answer RQ1, we first conducted between-group tests to evaluate LGBTQ+ youth versus heterosexual youth's self-reported online risk experiences and mental health challenges. For RQ2, we also applied regression models to analyze the relationship between online risk experiences and the mental health outcomes of youth participants. Finally, for RQ3, we conducted a qualitative analysis of the flagged messages for deeper insights into LGBTQ+ youth's online risk experiences.

Overall, we found that the private online experiences of LGBTQ+ youth differ significantly from the experiences of heterosexual youth in several key ways. For RQ1, we found that while LGBTQ+ youth do not report more online risk experiences overall, the experiences that they report are of higher risk severity and are more likely to involve messages that contain sexual or self-injurious content. For RQ2, we confirmed findings from prior literature that found LGBTQ+ youth report significantly worse mental health outcomes (e.g., Self-Harm and Injury, Depression, and decreased Mental Well-being) compared to heterosexual youth. In addition, we uncovered two significant interaction effects; LGBTQ+ youth who receiving more online harassment report significantly more self-harm and self-injurious behavior, while this effect was not seen among heterosexual youth. In contrast, LGBTQ+ youth who received more sexually risky messages reported higher levels of mental well-being, while this effect was in the opposite direction for heterosexual youth. Our RQ3 qualitative analyses revealed additional insights into our quantitative findings. The conversations flagged as risky by LGBTQ+ youth often contained sexual undertones, even when the intent was to harass. However, some of these risky messages gave LGTBQ+ youth the opportunity to explore their sexual and gender identities. Importantly, Instagram DMs gave them a channel to share about their struggles with others who provided needed support. Our results provide useful insights for supporting the unique online experiences of LGBTQ+ youth, which we unpack in our discussion.

Our research makes a distinct contribution to the CHI community, particularly in the context of inclusive design and research advocating for LGBTQ+ populations. Building upon the LGBTQ+ advocacy research within the CHI community (c.f., [5, 26, 58]), we advance this body of knowledge by delving into youths' private social media experiences through the use of their Instagram Direct Messages. Focusing on this timely subject, our research furthers understandings of the challenges faced by a particularly vulnerable demographic, LGBTQ+ youth, in the realm of online threats and bullying experiences but also critically evaluates the efficacy of existing filtering tools and protective measures for vulnerable users in the face of cyberbullying, urging a call-to-action for risk prevention and mitigation program. The broader implications of the research extend to the advancement of Human-Centric Machine Learning (HCML) algorithms designed to detect and address online sexual risks, thus contributing significantly to the ongoing efforts to enhance LGBTQ+ youth's digital safety and well-being. Specifically, we make the following unique research contributions:

- Through a mixed method study, we disentangled how LGBTQ+ and heterogeneous youth's online risk experiences differed, finding that LGBTQ+ youth report higher risk conversations containing sexual and self-harm content.
- Our research highlighted a range of online risks experienced by LGBTQ+ youth and demonstrates how the multi-faceted risk factors go beyond online public spaces and occur in private contexts.
- We examined social media direct message data to underscore the association between receiving harassing messages and the self-harming behaviors of LGBTQ+ youth.
- We gained deeper understanding of how LGBTQ+ youth did not only experience sexual harassment and unsolicited sexual content from strangers and peers, they also sought support from their peers.

Content Warning: This study discusses delicate subjects such as sexual content, offensive language, and derogatory remarks aimed at LGBTQ+ individuals, which might evoke unease. Caution is advised for readers.

2 Related Work

In this section, we synthesize related work regarding LGBTQ+ youth and their social media usage. Next, we introduce literature highlighting how LGBTQ+ youth are more vulnerable to online risks and how the negative online experiences impact mental health.

2.1 LGBTQ+ Youth, Social Media, and Support

LGTBQ+ turn to online platforms to seek information related to gender, sexuality, and romantic relationships, as well as to seek support for physical, mental, and sexual health needs [38, 41, 64, 91, 108, 113]. Yet, LGBTQ+ youth reported having smaller online social networks compared to heterosexual youth [29]. In contrast, they reported joining online communities or groups more frequently to feel less isolated or lessen social isolation compared to heterosexual youth [29]. Furthermore, the likelihood of LGBTQ+ teenagers having friends they only know online is higher as they often view online friends as being more supportive than their in-person peers [144].

Past research highlighted the role of online social networks as safe havens for LGBTQ+ youth who live in hostile environments. They found that the anonymity and lack of geographic restrictions in digital spaces offer an ideal platform for coming out, engaging with a communal gay culture, experimenting with non-heterosexual intimacy, and socializing with other LGBTQ+ youth [71]. Furthermore, LGBTQ+ youth also interact and build connections with a specific community and engage themselves with content that validates and recognizes their unique experiences with being an LGBTQ+ individual. However, seeking social support in online communities often requires disclosing personally sensitive information, such as one's gender identity and personal struggles [63, 71]. In the early phases of developing their LGBTQ+ identities, LGBTQ+ individuals may use social media platforms as informal learning environments [55]. For example, previous research findings identified three educational purposes associated with online information seeking on social media: traditional learning about LGBTQ+-related issues, social learning involving mirroring role models or other LGBTQ+ people's behavior, and experiential learning with online dating sites and dating apps. These learning behaviours were common during the coming-out process [55].

Decisions of coming out are further complicated by the potential risks and negative consequences of disclosing one's LGBTQ+ identity in unsupportive environments [43, 63]. Violations of privacy [59], being disowned by family members [31], and rejection from society are potential consequences of coming out. Regardless of these fears, youth often choose to disclose their sexual identity online to obtain needed support and social connections [56, 59]. Coming out about one's sexual identity is positively associated with relationship satisfaction, self-efficacy, self-esteem, lower anxiety, and lower levels of depression [13, 99, 101]. Thus, LGBTQ+ youth have multiple co-acting motives for seeking online support and also experience a range of benefits and risks within online spaces. However, most prior research in this area has focused on how LGBTQ+ youth self-report on their experiences in reference to online public spaces (e.g., via surveys, focus group studies, interviews, or analyzing public social media data). Further research is required to understand the online risks in private one-to-one communication, particularly through the analysis of real-world direct messages.

2.2 LGBTQ+ Youth Online Risks and Well-being

Although the Internet provides LGBTQ+ youth numerous opportunities to form new relationships and explore their sexual identities [56, 59], it can pose a range of online risks, such as privacy violations [93, 102, 129], cyberbullying [89, 127, 144], sexual risks [103], and other types of abuse [106]. Social media platforms provide social visibility, connectivity, feedback, and ease of accessibility, which can increase cyberbullying compared to traditional bullying [54]. In particular, LGBTQ+ youth experience higher rates of cyberbullying ([127, 144]) and are often targeted due to their nonconforming sexual identities [2, 15]. Moreover, a greater number of polyvictimization incidents, where youth who experience one type of victimization are more at risk of experiencing other types as well, are prevalent among LGBTQ+ youth [49, 50, 132, 139]. LGBTQ+ youth report being bullied or harassed online nearly three times as frequently as non-LGBTQ+ youth (42% vs. 15%) [61]. Other forms of online risks, such as hate-based abuse, violence, and discrimination are also prevalent among sexual minority youth [62, 70].

As a result, LGBTQ+ youth experience "minority stress," a chronic form of stress caused by negative social experiences associated with identifying with a minority group [15, 98, 124], which can be amplified in social media contexts [124]. According to the minority stress theory, sexual and gender minority health disparities may be traced to stressors brought on by hostile, homophobic, and transphobic cultures. These stressors frequently result in harassment, abuse, and victimization and may ultimately affect access to care [42, 57, 69, 124]. In particular, stigma, prejudice, and discrimination create a hostile social environment that can lead to mental health difficulties [97]. LGBTQ+ youth reported higher rates of suicidal thoughts and depressive symptoms than their straight peers, which are influenced by negative experiences, including discrimination and victimization [92]. These mental health issues, which result from bullying and other online risk factors, can significantly impact well-being and relationships with others [78].

Whereas much of the current research is focused on young LGBTQ+ adults, more research is needed to understand how younger LGBTQ+ adolescents (ages 13-17) experience these online risk factors. This is especially true given that the permissible age for having a social media account (e.g., for Facebook or Instagram) is 13 [72, 100], a time during which LGBTQ+ youth may be particularly susceptible to online dangers and associated mental health consequences. Moreover, the risk factors associated with online engagement highlight the critical need to go beyond just examining public online communities, but also understanding the multifaceted risk factors that occur in private contexts as well. As such, our research complements and extends beyond previous work by examining the online risk experiences and associated mental health outcomes of adolescent LGBTQ+ youth between the ages of 13 to 21. Our study adds depth by adopting a mixed-method approach and pairing the self-reported mental health data of LGBTQ+ youth with their real-world social media private conversations that they flagged as unsafe or risky. By triangulating these two data sources, our analysis is one of the first to deeply examine the association between these two facets of LGBTQ+ youths' personal and social experiences, contributing to a more holistic representation of LGBTQ+ youth's online experiences.

3 Methods

In this section, we provide a detailed description of our study, including the survey design, Instagram data collection procedures, and the data analysis approach we used to answer each of our research questions.

3.1 Study Overview

We conducted a user study of youth (ages 13-21), who first completed a web-based survey, then were asked to upload their Instagram data and subsequently flag their DM conversations for interactions that made them feel uncomfortable or unsafe. We chose Instagram as our social media platform of choice because, according to Pew Research, 72% of teens use Instagram, making it one of the most popular social media platforms among youth [14]. We recruited participants, who met the following eligibility criteria:

1) English speakers based in the United States between the ages of 13-21, 2) Had an active Instagram account at while they were a teen (ages 13-17), 3) Communicated with at least 15 people through Instagram direct messaging (DMs), and 4) Had at least two conversations with other users that made them feel unsafe or uncomfortable. If participants met this eligibility criteria, parental consent was obtained for those under 18; otherwise, participants consented to participate themselves. Participants were compensated with a $50 Amazon gift card for their time and data. This study was approved by our university's Institutional Review Board (IRB).

3.2 Ethical Considerations

We took the utmost care to protect the participants' anonymity and privacy due to the complicated and sensitive nature of the data collected. First, we obtained IRB approval for our work, declared our position as mandated child abuse reporters within the statement of informed consent in the case of an impending danger posed to a minor, and followed the guidance of Badillo-Urquiola et al. [18] on conducting risky research with minors. For instance, we provided explicit instructions NOT to share digital images that depicted the nudity of a minor and outlined our statutory responsibility to report child pornography to the proper authorities. We also provided clear directions to delete such materials prior to data sharing. Additionally, we secured a Certificate of Confidentiality from the National Institute of Health, which further protected participants by precluding the subpoenaing of the data during legal discovery. To further protect the privacy of our participants and everyone else who participated in direct message conversations, we removed all personally identifiable information from any textual data reported in our results and paraphrased all quotations. To prevent sharing the data to outside parties, we also did not analyze our data using any cloud-based services. Researchers were prohibited from downloading the data onto personal devices and were required to complete IRB Human Subjects CITI training. We also gave research assistants, who assisted in verifying and qualitatively analyzing the data, mental health support, and adequate breaks.

3.3 Participant Recruitment and Demographics

A total of (N = 173) participants completed the study. In the survey, participants were asked to report several demographic characteristics, including their gender identity and sexual identity (e.g., Heterosexual or straight, Homosexual or gay, Bisexual, or Prefer to self-identify). Details of participant demographics based on age, gender identity, race, and sexual identity can be found in Table 1. Since our participants could select multiple races, the total percentages of all categories can be greater than 100%. We grouped homosexual/gay, bisexual, and individuals who preferred to self-identify as LGBTQ+ youth, as the purpose of the term "LGBTQ+" is to be inclusive of all same-gender attracted and trans people [30]. Among all participants who completed the study, 50.29% (n = 87) of youth identified as heterosexual, whereas 49.71% (n = 86) were LGBTQ+. Participants also responded to the social media usage questions, which helped us better understand their Instagram use. For almost half (49.71%) of our participants, Instagram was their most used social media platform with 64.16% (n = 111) participants having more than one Instagram account. 47.97% (n = 83) of the participants used Instagram several times a day, 24.27% (n = 42) used every day or almost every day, 21.96% (n = 38) used several times an hour, 4.04% (n = 7) used once or twice a week, 1.15% (n = 2) used less than once a month, and 0.57% (n = 1) used less than once a week. 97.68% (n = 169) participants never met their Instagram followers in person.

Table 1: Summary of participants demographics

	n (%)
Total (N)	173 (100%)
Age (M (SD))	17 (2.14)
Gender Identity	
Female	117 (67.63%)
Male	39 (22.54%)
Non-binary	14 (8.09%)
Prefer to self-identify	3 (1.73%)
Race	
White/Caucasian	92 (53.18%)
Black/African-American	45 (26.01%)
Asian or Pacific Islander	36 (20.81%)
Hispanic/Latino	26 (15.02%)
American Indian/Alaska Native	7 (4.04%)
Prefer to self-identify	5 (2.89%)
Two or more races or ethnicity	32 (18.49%)
Sexual Identity	
Heterosexual/straight	87 (50.29%)
Bisexual	47 (27.16%)
Homosexual/gay	17 (9.82%)
Self-identify	22 (12.71%)
Pansexual	10 (45.45%)
Omnisexual	2 (9.00%)
Asexual	2 (9.00%)
Biromantic asexual	1 (4.55%)
Neptunic	1 (4.55%)
Panoromantic and demisexual	1 (4.55%)
Fluid	1 (4.55%)
Curious	1 (4.55%)
I don't know	3 (13.64%)

3.4 Data Collection and Risk-Flagging

After completing the web-based survey, participants uploaded their Instagram data to a secure web-based system developed by Razi et al. [111]. This web-based system was developed using PHP with an Amazon Web Services (AWS) back-end infrastructure to encrypt and store the data. Participants requested their data from Instagram and uploaded it in the form of JSON files that were translated and stored on our AWS server. Then, participant's DMs were presented through our web-based interface in reverse chronological order for them to review and flag the uncomfortable or unsafe (i.e., "risky") messages of each conversation. Participants marked individual messages by risk type (e.g., sexual content, harassment, spam, hate speech, violence, illicit actions, and self-injury) and risk level (low, medium, high). While aligned with Instagram's reporting

features [46], participants could denote their own risk type. We provided a benchmark for risk level: low for discomfort without harm, medium for potential emotional or physical harm, and high for actual harm. The dataset included 32,055 DM conversations with over 6 million (6,863,161) messages. Participants marked 2,515 conversations as 'risky,' flagging a total of 3,023 messages by risk type and level. High-level risks comprised 402 messages, while medium and low risks were 821 and 1,813, respectively. Table 2 displays message counts based on risk types and severity. Past studies (as synthesized in section 2.2) suggested that LGBTQ+ youths face bullying and harassment more on both offline and online, including on social media platforms. Based on prior findings from the literature, we hypothesized that:

H1: *LGBTQ+ youth will report encountering more online risks than heterosexual youth.*

H2: *LGBTQ+ youth will report encountering more high-risk online experiences than heterosexual youth.*

Given that the literature was not conclusive regarding the specific types of online risks LGBTQ+ youth encountered compared to their heterosexual counterparts, we examined these relationships in our results without formalizing hypotheses based on risk types.

3.5 Mental Health and Well-being Measures

In the survey, participants were also asked to report on several prevalidated constructs to assess their mental health and well-being. These survey constructs included: The Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) [135], the Patient Health Questionnaire (PHQ)-9, a commonly used measure for Depression [83], and the Inventory of Statements about Self-injury (ISAS) [82]. These constructs were measured on a 5-point Likert scale. To ensure alignment with other survey measures, we modified the (PHQ)-9 scale from a 4 to a 5-point scale. This adjustment was validated through factor analyses, incorporating Horn's parallel analysis in exploratory factor analysis ([73]) and taking into account prior research on the 4-point scale [141]. The subsequent confirmatory factor analysis produced favorable fit indices, surpassing the 0.9 threshold. Notably, both RMSEA and SRMR scores exceeded the recommended cutoff of 0.05, confirming the validity of our refined 5-point (PHQ)-9 [60, 81, 140]. We tested for internal consistency using Cronbach's alpha and all values were above the acceptable threshold (0.7) [32]. Table 4 shows the scale reliability and statistical description of the mental health constructs for both LGBTQ+ and heterosexual youth who participated in our study. As discussed in the prior literature, many studies have found evidence that LGBTQ+ youth showed greater depressive symptoms and, in some cases, suicidal ideation as a result of the negative consequences of stigma, prejudices, discrimination, and abusive attitudes held by the predominately heterosexual society [33, 133, 145]. Generally, LGBTQ+ youth are found to have poor mental health conditions due to the negative experiences faced in both offline and online settings. Based on this literature, we formulated the following hypotheses related to the well-being and mental health of LGBTQ+ versus hetersexual youth:

H3: *LGBTQ+ youth will report worse mental health outcomes, including a) lower Well-Being, b) higher Depression, c) and higher Self-Harm and Injury than heterosexual youth.*

H4: *Youth who experience more high-risk online experiences will report worse mental health outcomes than those who report experiencing more low-level online risk.*

H5: *The negative mental health outcomes associated with online risk experiences will be moderated by sexual identity, such that the negative effect will be stronger for LGBTQ+ youth than for heterosexual youth.*

In the next section, we describe our statistical methods for testing our research hypotheses, as well as our qualitative approach for gaining additional insights.

3.6 Data Analysis Approach

In this section, we first describe how we conducted our statistical analyses to answer our hypotheses related to RQ1 and RQ2. Then, we describe our qualitative analysis approach for investigating RQ3.

3.6.1 Examining the Differing Online Risk Experiences of LGBTQ+ Versus Heterosexual Youth (RQ1). To investigate differences in online risk experiences encountered by LGBTQ+ and heterosexual youth, we first analyzed the risk-flagged messages provided by youth. Among the total (*n=3,023*) messages, 48.56% (*n=1468*) were flagged by LGBTQ+ youth, while 51.44% (*n=1555*) were flagged by heterosexual youth. Table 2 summarizes the number and percent of risk-flagged messages by risk type and level for both LGBTQ+ and heterosexual youth. We conducted a between-group chi-square test (χ^2) to compare the difference in total number of online risks experienced by these two youth groups (H1). The χ^2 test of independence is employed for between-group comparisons involving two or more groups [94], and previous studies contrasting sexual minority groups with heterosexual individuals have successfully employed this approach [44].

Furthermore, we explored significant between-group differences between LGBTQ+ and heterosexual youth based on risk level to test H2. Specifically, we utilized standardized residuals to provide insights into cells significantly impacting the chi-square value. Cells with standardized residuals exceeding +2 are considered major contributors, while those surpassing -2 weakly contribute to the overall chi-square calculation [142]. We performed another χ^2 test to identify significant differences in the total number of messages falling into each of the six risk types, as self-annotated by the youth. Similar to the approach with standardized residuals described above, we employed this analysis to gain insights into risk type disparities within the dataset.

3.6.2 Exploring Associations Between Mental Health and Negative Online Experiences (RQ2). To operationalize the frequency and levels of risk for each participant, we utilized a weighted model, with higher severity risks having a higher weight. Our methodology involved multiplying the number of messages within each risk type by the corresponding risk weight. This resulted in a comprehensive risk score for each risk type and user, providing a nuanced representation of the interplay between message volume and risk severity. Next, we averaged items associated with the three mental health

Table 2: Counts and percentage of risks based on risk type and severity for LGBTQ+ and heterosexual participants (*N* = 173)

Risk Type	LGBTQ+ (*N* = 86)			Heterosexual (*N* = 87)		
	Low	Medium	High	Low	Medium	High
Sexual messages *N* = 1157	*n* = 286 35.18%	*n* = 171 43.85%	*n* = 141 53.21%	*n* = 339 34.35%	*n* = 151 35.03%	*n* = 69 50.36%
Harassment *N* = 963	*n* = 276 33.94%	*n* = 101 25.89%	*n* = 64 24.15%	*n* = 327 33.13%	*n* = 159 36.89%	*n* = 36 26.27%
Spam and others *N* = 456	*n* = 158 19.43%	*n* = 39 10.00%	*n* = 9 3.40%	*n* = 185 18.74%	*n* = 52 12.06%	*n* = 13 9.49%
Hate Speech and violence *N* = 244	*n* = 48 5.90%	*n* = 47 12.05%	*n* = 28 10.57%	*n* = 68 6.89%	*n* = 41 9.51%	*n* = 12 8.76%
Illicit actions *N* = 147	*n* =30 3.69%	*n* = 21 5.38%	*n* = 13 4.90%	*n* = 55 5.57%	*n* = 21 4.87%	*n* = 7 5.10%
Self-injury *N* = 56	*n* = 15 1.84%	*n* = 11 2.82%	*n* = 10 3.77%	*n* = 13 1.31%	*n* = 7 1.62%	*n* = 0 0.00%
Total	*N* = 813 (100%)	*N* = 390 (100%)	*N* = 265 (100%)	*N* = 987 (100%)	*N* = 431 (100%)	*N* = 137 (100%)

constructs and calculated the mental health scores for each participant. After calculating their Self-Harm and Injury, Depression, and Mental Well-Being scores for both LGBTQ+ and heterosexual youth, we examined the distribution and variance of the two groups for each construct. As the two populations were normally distributed, and the variances for the constructs were not equal for our independent populations, we performed Welch's two-sample *t*-test to compare the between-group differences in their mental health scores [90].

Next, we investigated the relationship between the risky messages received via DMs and the mental health of these youth via multiple linear regressions, and we also tested whether this relationship is modulated by the self-reported sexual identity of these youth. We started by fitting our data into a linear regression model with each youth's risk type scores and sexual identity as predictor variables and mental health constructs as our outcome variables. Six risk-type scores (e.g., sexual message, harassment, spam and others, hate speech and violence, illicit actions, and self-injury) were entered as predictor variables. Three mental health scores (Depression, Self-Harm and Injury, Mental Well-Being) were the outcome variables. Furthermore, the sexual identity of these youth was used as an interaction term in each regression model. Here, messages categorized as self-injury are not synonymous with the mental health construct known as Self-Harm and Injury.

3.6.3 Qualitative Analysis of Instagram Private Message Conversations (RQ3). To further explore the types of risks that LGBTQ+ youth faced in private messaging contexts, we performed a thematic analysis [23] of the private message conversations risk-flagged by LGBTQ+ youth. Each conversation represented a messaging thread where the participant and other person/people messaged back and forth intermittently. Therefore, each conversation could have either a single message or a series of messages that were exchanged over a long period of time. There were a total 1,468 unsafe messages (of 808 conversations) that were flagged by our LGBTQ+ youth participants. However, we analyzed all messages (N = 216,332) of these 808 conversations. All messages were in textual format. When participants or their conversation partners shared media, our data showed the message as "Instagram User sent an attachment." or "[Name] sent an attachment" instead of showing the actual file. In such cases, we used contextual cues from the larger conversation to interpret the risk.

To complete the qualitative analysis, the second author initially familiarized themselves with the data by reading through the conversations and creating initial codes, which were then discussed among all co-authors. Subsequently, the second and last author worked together to iteratively establish consensus and incorporate codes as they emerged. They jointly collaborated to conceptually group the codes into cohesive subthemes and overarching themes. Our thematic analysis, as presented in Table 3, aimed to identify key characteristics of the risk-flagged conversations LGBTQ+ youth had with others. For each theme, the total count and percentages of codes can be greater than the total number of conversations as we double-coded the conversations for different risk types. For example, there were instances when sexual message content, sexual solicitations, and bullying were all present in the same conversations.

4 Results

In the following sections, we highlight distinctions in the total number of online risks and their severity between LGBTQ+ and heterosexual youth by presenting the outcomes of chi-square tests. Subsequently, we delve into the results of linear regression models, investigating the correlation between online risk experiences and mental health outcomes for both groups. Lastly, we provide qualitative insights to enhance the overall understanding of online risks in private conversations among LGBTQ+ youth.

4.1 LGBTQ+ Youth at Higher Risk for Sexual and Self-Injury Messages (RQ1)

4.1.1 Flagged Messages by Risk Level (H1 & H2). We first tested the degree to which LGBTQ+ youth flagged riskier messages compared to heterosexual youth. Overall, LGBTQ+ youth flagged (*N = 1468*) messages, whereas heterosexual youth flagged (*N = 1555*) as risky. Our chi-square test indicated significant variation in the total number of online risks experienced by LGBTQ+ youth versus heterosexual youth, χ^2 (2, *N* = 3023) = 57.167, $p < .001$. However, this effect was in the opposite direction than hypothesized. Hence, the result did not support our first hypothesis (H1) that LGBTQ+ youth would report more online risks than heterosexual youth. In contrast, heterosexual youth reported significantly more online risks than LGBTQ+ youth.

For H2, Figure 1 (a) reveals that LGBTQ+ youth experienced a significantly higher frequency of high severity online risks. Moreover,

Table 3: Codebook for Risk-Flagged Conversations of LGBTQ+ Youth (RQ3)

Themes	Subthemes (Codes)	Illustrative Quotations
Most risky conversations had sexual undertones	**Exchanged sexual message content** (Sexual content from strangers; from peers; Sent to others; Added to porn groupchats by strangers)	**OP*:** *hurry up bitch+* **OP:** *u wasting my time* **P*:** *[User ID] sent an attachment* **OP:** *oh girl, u are a good submissive. I fucking love u*
	Received sexual solicitations (Sexual solicitations from strangers; From peers; With monetary offers; Received persistent messages; Led to harassment)	**OP:** *Can I get my duck sucked* **P:** *I don't know you do I? And not by me*
	Got harassed for sexual identity (Harassed for LGBTQ+ identity by peers; by strangers)	**OP:** *Faggots like you are in this world will take us in hell* **P:** *leave me alone please*
	Received spam targeted to sexual identity (Advertisements; Scams targeted to LGBTQ+)	**OP:** *Hey, i'm from [Organization name]! I hope you're having a great January. We are looking for brand ambassadors and reps. Please send us a message at our main account [UserID] ASAP! Have a wonderful day*
LGBTQ+ youth encountered general harassment and scams	**Got harassed** (Harassed by strangers; By peers)	**OP:** *Instagram User sent an attachment* **OP:** *can we be friends* **OP:** *Instagram User sent an attachment* **OP:** *bitch talk with me*
	Received spam/scams (Received general advertisements; scams; ads for illegal substances)	**OP:** *you won't believe this, I just got sent a $1,000 GiftCardVisa yesterday and all I had to do was participate a little and they sent me it. I used it to pick up a new gaming computer Ive been eyeing LOL. I thought u would want one since your my follower. Hurry thought there is just a few left!! Go*
LGBTQ+ youth received support from peers regarding negative experiences	**Shared negative experiences related to sexual identity** (Received support for negative experiences with transphobic people; Fear of coming out; Urge to self-harm)	**P:** *I am tired of my dad. He gives me pills to cure my homosexuality. Its as if a disease my parents want me to get well. At this point I don't think I should be living anymore.* **OP:** *I am sorry your going through these. It suckz when family is liek them. Can you move out*
	Shared mental health concerns (Received support for sad feelings; Depressive thoughts)	**P:** *I ve bn having break up after break up. this time her problem was my bisexuality. I feel like cuting, I tried so many times* **OP:** *Honestly, she doesn't deserve you and your such a great guy and any girl would be lucky to have you*
	Received other support and advice (Received advice for general negative experiences; Drug abuse)	**OP:** *Oh so u out for smoking blunt again?* **P:** *a little* **OP:** *u know u shouldn't be smoking*

*P: Participant; *OP: Other Person
+While the authors recognize the significance of portraying the real-life experiences of LGBTQ+ youth, they firmly denounce the utilization and spread of such language.

a significant difference exists between heterosexual and LGBTQ+ youth regarding the number of low-level risk conversations, with heterosexual youth reporting more low-level risk conversations. However, there was no significant difference in medium-level risks based on sexual identity. Thus, these results support our hypothesis (H2).

4.1.2 Flagged Messages by Risk Type. We conducted an additional chi-square test to identify differences between LGBTQ+ and heterosexual youth groups based on risk types. The chi-square test revealed significant differences between the two groups concerning risk types, χ^2 (5, N = 3023) = 16.927, p = .004. Standardized residuals indicated a significant difference between LGBTQ+ and heterosexual individuals, particularly when examining messages containing sexual content, harassment, and self-injury. As shown in Figure 1 (b), LGBTQ+ youth experienced a significantly higher frequency of sexual messages and messages containing self-injury language compared to their heterosexual counterparts. In contrast, heterosexual youth reported a significantly higher number of harassment messages than LGBTQ+ youth. In the next section, we examine the association between online risks and mental health for youth.

4.2 Associations Between Online Risks and Mental Health (RQ2)

4.2.1 Investigating the Differences in Mental Health Outcomes based on Sexual Identity (H3). Next, we investigated whether there were significant differences in mental health challenges reported by LGBTQ+ versus heterosexual youth. Our Welch's Two-Sample t-test compared between-group differences based on mental health scores and revealed a statistically significant difference in self-reported Depression, Self-Harm and Injury, and Mental Well-Being scores $p <$ .001, as illustrated in Table 4. LGBTQ+ youth reported higher levels

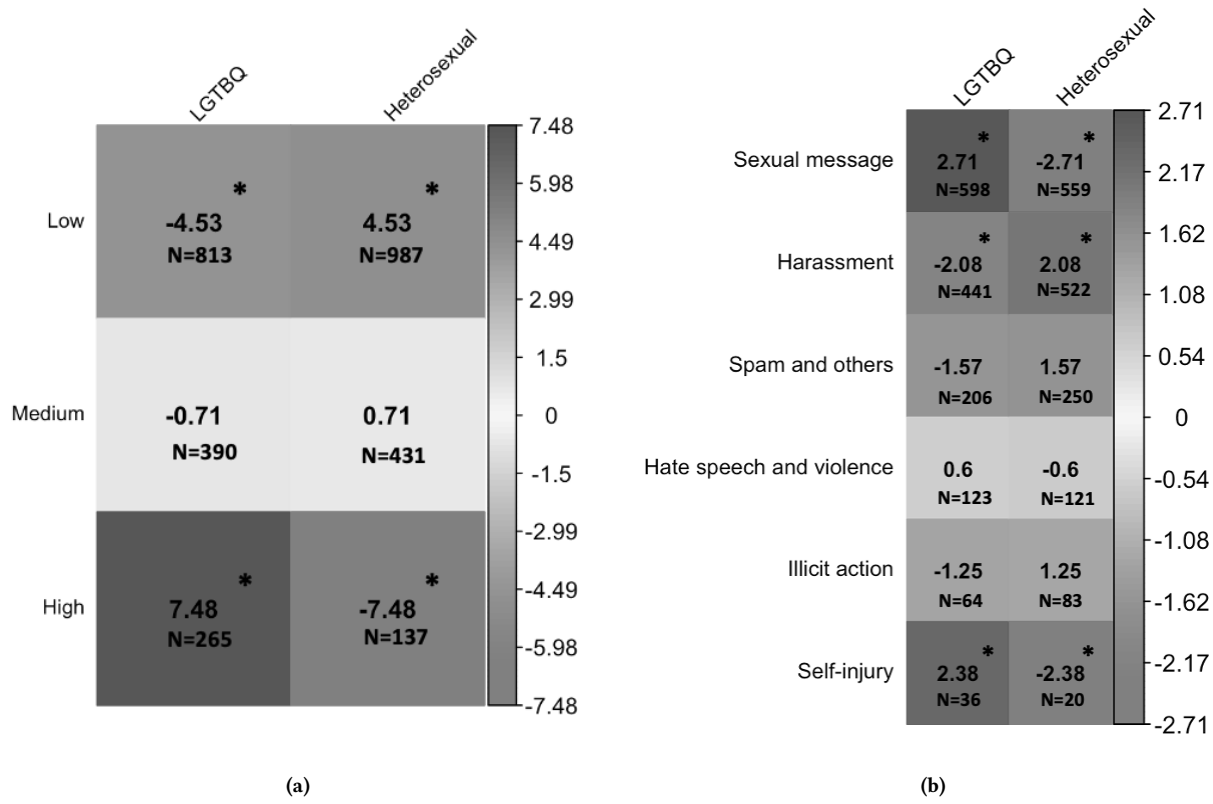


Figure 1: Results (Standardized residuals) from chi-square test showing the between-group analysis of online risk messages encountered by LGBTQ+ and heterosexual youth based on (a) risk level and (b) risk types (N = 3023). (*) indicates significant association. Note that green indicates a positive association, while red indicates a negative one.

of depressive thoughts, Self-Harm, and Injury behavior than heterosexual youth. LGBTQ+ youth also scored significantly lower in Mental Well-Being than heterosexual youth. Therefore, the results confirmed our hypothesis (H3) that LGBTQ+ individuals report worsened mental health outcomes compared to heterosexual youth. These results serve to confirm past studies reporting similar results [39, 68].

4.2.2 Examining the Associations between Online Risk and Mental Health Outcomes (H4). The results of the second chi-square test for RQ1 (Figure 1 (b)), which was conducted to examine the variation in risk types, guided our subsequent analyses for H4. Specifically, we focused our analysis on sexual messages, harassment, and self-injury-related online interactions due to the significant differences found between LGBTQ+ and heterosexual youth in relation to these distinct categories of risky online messages received. Therefore, we investigated to what extent the online risk experiences encountered by youth in private online spaces were associated with mental health outcomes (i.e., Self-Harm and Injury, Depression, and Mental Well-Being), and how this relationship was moderated by participants' sexual identity.

Self-Harm and Injury: Our first multiple regression model (M1) identified a significant and positive association between messages containing sexual content and youths' Self-Harm and Injury scores (see Table 5). Results indicated that youths more frequently engaged in self-harming behaviors when they also received more sexual messages through private messages, regardless of sexual identity (t = 1.991, df = 165, p = .048, 95% confidence interval: [0.000, 0.020]). In contrast, harassing messages and those having to do with self-injury did not predict Self-Harm and Injury mental health scores (p > .05). Additionally, we conducted a detailed examination of individual Self-Harm and Injury behaviors by implementing a multivariate regression model (M4) to discern the specific self-harm behaviors that significantly impact youth's mental health. The outcomes of our M4 model align with those of the M1 model, reinforcing earlier findings. Notably, we observed a positive and statistically significant main effect between messages containing sexual content and particular self-harm behaviors, such as carving and pulling hair, aligning with M1. Additionally, the model (M4) reveals a significant interaction effect between messages containing harassment and behaviors such as cutting, carving, and rubbing skin against a rough surface, specifically for LGBTQ+ youth. However, no significant effects were identified for other risky messages and their association with Self-Harm and Injury behaviors (p > .05). Appendix A provides comprehensive details of this model.

Table 4: *T*-test results comparing LGBTQ+ and heterosexual on mental health constructs

Measure	*Cronbach's α*	Sexual Identity	*Mean*	*SD*	*t*	*df*	*Std. CI*	*p*
Depression	0.918	LGBTQ+	3.598	0.881	-5.788	170.94	-1.036 - -0.509	**<.001**
		Heterosexual	2.825	0.876				
Self-Harm and Injury	0.878	LGBTQ+	1.858	0.773	-4.908	142.47	-0.676 - -0.287	**<.001**
		Heterosexual	1.375	0.484				
Mental Well-Being	0.897	LGBTQ+	4.279	1.102	4.248	170.82	0.377 - 1.032	**<.001**
		Heterosexual	4.984	1.079				

Table 5: Unstandardized estimates for linear regression models examining the relationship between online risky messages (e.g., sexual messages, harassment, and self-injury) and mental health constructs (e.g., Self-Harm and Injury, Depression, and Mental Well-Being)

	Self-Harm and Injury (M1)			Depression (M2)			Mental Well-Being (M3)		
Predictors	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*
(Intercept)	1.284	1.100 - 1.467	<0.00***	2.639	2.382 - 2.895	<0.00***	5.175	4.858 - 5.492	<0.00***
Sexual Messages	**0.010**	**0.000 - 0.020**	**0.048***	**0.014**	**0.000 - 0.028**	**0.048***	**-0.020**	**-0.037 - -0.002**	**0.026***
Harassment	-0.000	-0.008 - 0.007	0.865	0.002	-0.009 - 0.013	0.717	-0.001	-0.015 - 0.012	0.803
Self-Injury	-0.011	-0.167 - 0.144	0.885	0.082	-0.136 - 0.301	0.457	0.064	-0.205 - 0.334	0.636
Sexual Identity (LGBTQ+)	**0.328**	**0.052 - 0.605**	**0.020***	**0.891**	**0.504 - 1.278**	**< 0.001*****	**-0.994**	**-1.472 - -0.517**	**< 0.001*****
Sexual Messages x Sexual Identity (LGBTQ+)	-0.006	-0.019 - 0.007	0.379	-0.012	-0.031 - 0.006	0.200	0.028	0.004. - 0.051	**0.018***
Harassment x Sexual Identity (LGBTQ+)	0.020	0.003 - 0.037	**0.018***	-0.003	-0.020 - 0.027	0.749	0.001	-0.028 - 0.031	0.919
Self-injury x Sexual Identity (LGBTQ+)	0.058	-0.108 - 0.225	0.489	-0.088	-0.322 - 0.145	0.458	-0.063	-0.352 - 0.225	0.663
Observations	173			173			173		
R^2 / Adjusted R^2	0.195 / 0.161			0.190 / 0.156			0.135 / 0.098		

Note. *p < .05; **p<.01; ***p<.001

Depression: In our second regression model (M2), a significant correlation emerged between messages featuring sexual content and the self-reported Depression scores of young individuals (see Table 5). The positive direction suggested that regardless of their sexual identity, young individuals reported more frequent experiences of depressive thoughts when they received a higher volume of sexual messages through private messaging (t = 1.992, df = 165, p = .048, 95% confidence interval: [0.000, 0.028]). Conversely, messages that included harassment or language likely to induce self-injury did not forecast self-reported Depression scores in young individuals (p > .05).

Mental Well-Being: In our third regression model (M3), we observed a significant negative association between messages containing sexual content and the Mental Well-Being scores of young individuals (see Table 5). The findings revealed that irrespective of their sexual identity, young individuals reported lower scores, indicating a decline in Mental Well-Being, when they received an increased number of sexual messages via private messaging (t = -2.244, df = 165, p = .026, 95% confidence interval: [-0.037, -0.002]). On the other hand, messages that contained harassment or language that might prompt self-injury did not predict the Mental Well-Being scores of young individuals (p > .05). According to established standards in social science research an R^2 value of 10% or above is considered acceptable [47, 104]. Our models, across all three mental health measures (Mental Well-being marginally close to the acceptable threshold), meet this threshold, signifying that they provide meaningful insights into the relationships between the predictors and mental health outcomes.

In summary, our results (M1, M2) reveal that a substantial correlation exists between youth encountering a significant number of messages containing sexual content and reporting more pronounced mental health challenges (e.g., higher Depression scores, increased Self-Harm and Injury), and lower Mental Well-Being scores (M3). This association stands out, as other types of risky encounters, such as messages containing harassment and self-injury, did not demonstrate statistical significance with any of the mental health constructs for the youth as a whole. Consequently, our multiple linear regression models provide insightful findings, partially confirming our hypothesis (H4), emphasizing that youth with more high-risk online experiences tend to report more adverse mental health outcomes compared to those with lower levels of online risk. Next, we report on the moderating effects of our models.

4.2.3 Examining the Moderating Effect of Online Risk and Sexual Identity on Mental Health (H5). We examined the moderating effect of sexual identity on the relationship between risky online messages and youth's mental health outcomes. In doing so, our multiple linear regression model (M1) identified a significant interaction effect between harassing messages and Self-Harm and Injury such that

the non-significant main effect held for heterosexual youth; yet, the correlation between harassing messages and Self-Harm and Injury for LGBTQ+ youth became significant and positive. Figure 2 (a) shows the significant interaction effect between harassment messages and Self-Harm and Injury behavior of youth. Results indicated that harassment within direct messages represents a unique risk factor for Self-Harm among LGBTQ+ youth (t = 2.386, df = 165, p = .018, 95% confidence interval: [0.003, 0.037]). In other words, harassment is associated with an increase in Self-Harm for young people who identify as LGBTQ+, whereas heterosexual youth are less affected. Hence, this finding supports our hypothesis (H5), which is the relationship between online risks (e.g., harassment) and adverse mental health outcome (e.g., Self-Harm and Injury behavior in this instance) will be stronger for LGBTQ+ youth. In contrast, the non-significant interaction effect between sexual messages and sexual identity indicates that sexual messages are positively associated with Self-Harm and Injury behaviors for LGBTQ+ and heterosexual youth alike. Similarly, there was no significant interaction effect found between messages containing self-injury language and sexual identity in relation to Self-Harm and Injury behaviors (p > .05).

For our model (M2) with Depression score as the outcome variable, the interaction effect between risky messages (e.g., sexual messages, harassment, and self-injury-containing messages) and Depression scores was not statistically significant (p > .05). The lack of an interaction effect between sexual messages and sexual identity suggests that sexual messages can be concerning for both LGBTQ+ and heterosexual youths when it comes to their depressive thoughts.

On the other hand, our model (M3) identified a significant relationship between messages containing sexual content toward LGBTQ+ youth and Mental Well-Being scores. Figure 2 (b) shows the significant interaction effect between sexual messages (X-axis) and the Mental Well-Being score (Y-axis) of youths. The results indicated that sexual messages are uniquely associated with positive Mental Well-Being for LGBTQ+ youth only (t = 2.386, df = 165, p = .018, 95% confidence interval: [0.004, 0.051]). In other words, sexual messages are associated with a decrease in Mental Well-Being scores for heterosexual young people (blue straight line), while for LGBTQ+ youth (orange dotted line) the association is in the opposite direction. Consequently, this finding contradicts our hypothesis (H5), which stated that the relationship between online risks (e.g., sexual messages) and negative mental health outcomes (e.g., Mental Well-Being scores) would be more pronounced for LGBTQ+ youth compared to heterosexual youth. However, there was no significant interaction effect when examining the relationship between messages containing harassment or self-injury inducing language and sexual identity in relation to youth's self-reported Mental Well-Being scores (p > .05).

In summary, our investigation into how sexual identity moderates the relationship between negative mental health outcomes and online risk experiences yields novel insights. Our findings illuminate the distinctive role of harassment as a risk factor for Self-Harm and Injury, while also uncovering an unexpected positive association between sexual messages and the Mental Well-Being of LGBTQ+ youth. Therefore, our results contribute to a nuanced understanding and partially support our hypothesis (H5) that the link between online risks and adverse mental health outcomes will be more pronounced for LGBTQ+ youth. Table 6 provides a summary of the hypothesis testing results from our statistical model.

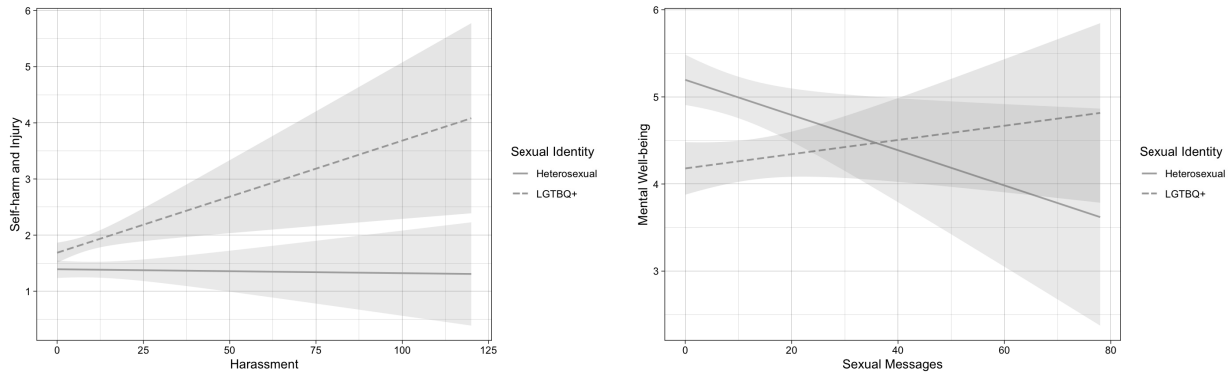
4.3 Nature of Risks Experienced by LGBTQ+ Youth in Private Online Spaces (RQ3)

To further unpack the statistically significant effects found in RQ1 and RQ2, this section qualitatively examines risk-flagged Instagram conversations donated by the LGBTQ+ youth to shed light on the nature of online risks. LGBTQ+ youth participants flagged a total of 1,468 private messages of 808 conversations as "risky", and these conversations serve as the unit of analysis for the results presented below. Table 3 presents the themes and subthemes for RQ3, as well as their corresponding codes and illustrative quotations. We used illustrative quotations to describe each of the themes that emerged from our thematic analysis. In the illustrative conversations, we replaced all referenced names with the letter "X" in the messages and removed the Unicode characters (emojis). Overall, we found that the risky experiences reported by LGBTQ+ youth were often related to their sexual identity.

4.3.1 Most Risky Conversations had Sexual Undertones. Overall, we found that most of the conversations LGBTQ+ youth flagged as making them feel unsafe or uncomfortable contained sexual content, either related to their gender or sexual identity and/or made direct sexual solicitations. For instance, more than one-third of the conversations contained *messages with sexual content*, and one-fourth of these conversations took place *with strangers*. Sexual messages often contained explicit adult content (e.g., photos and videos), porn website URLs, and sexual texts. For example, LGBTQ+ youth were often added to porn group chats by strangers, where they also received inappropriate content and/or porn site URLs. Interestingly, we often found LGTBQ+ youth were interested in participating in such conversations as a way to explore their sexual identities but wanted to do so slowly. For example, a 21-year-old homosexual woman had the following conversation with a stranger:

> ***Other person:*** *You are so incredibly beautiful. Its an absolute pleasure to meet you. What exactly is it you are looking for? I'm looking for a special friend to share, maybe explore things together. We can be friends and forward funny videos. You feel you might be interested in something casual like this?*
> **Participant:** *Uh sure, Haha, Omg! I think I probably seen you on a dating app lol*
> **Other person:** *Instagram User sent an attachment [flagged as sexual content]*
> **Participant:** *whoa, thats just downright fast*

While the LGBTQ+ youth mostly received sexual message content from strangers, we found some conversations where participants *received sexual messages from their peers* and the *youth themselves also sent sexual messages to others*. When asked to describe the situation, they explained that they sent these sexual photos either to experiment or to earn money, but often later they regretted it. Below is an example conversation that a 21-year-old bisexual woman flagged:



(a) (b)

Figure 2: Moderating effect of self-reported sexual identity on the association between online risk experiences and mental health outcomes: (a) Shows a positive relationship between harassment (X-axis) and Self-Harm/Injury (Y-axis); (b) Demonstrates a positive relationship between sexual messages (X-axis) and Mental Well-being (Y-axis) for LGBTQ+ youth. Orange and blue lines depict the moderating effect for LGBTQ+ and heterosexual youth, respectively.

Table 6: Hypotheses testing results

Hypotheses	Supported
H1: LGBTQ+ youth will report more online risks than heterosexual youth	No
H2: LGBTQ+ youth will report more high risk online experiences than heterosexual youth	Yes
H3: LGBTQ+ youth will report worse mental health outcomes than heterosexual youth	Yes
H4: Youth with more high risk online experiences will report worse mental health outcomes	Partially
H5: LGBTQ+ youth will exhibit a strong relationship between online risks and negative mental health outcomes	Partially

Other Person: *Do u have haire on pussy* [1]
Participant: *A lil*
Other Person: *Can I see*
Participant: *X sent an attachment*
[flagged as sexual content]
Other Person: *and boobs*

Almost a quarter of the conversations consisted of text messages with *sexual solicitations*. These conversations frequently included adult content; therefore, they were double-coded for sexual content and solicitations. Similar to the sexual message content, LGBTQ+ youth mostly received these *sexual solicitations from strangers*. In some cases, participants were offered different kinds of incentives, e.g., *monetary and/or other benefits*, in exchange for a sexual relationship with these strangers. For instance, a 19-year-old bisexual woman received the following messages:

Other Person: *Hello beautiful, can you be my sugar baby? I'm ready to help you and you gonna get paid weekly. Let me know when you're ready. Just chat with me everyday and you get your paid for that* [flagged as harassment]
Participant: *can u buy my amazon wish list*
Other Person: *Money has never been my problem,How much is your amazon wishlist*
Participant: *like $300ish*

[1]While the authors acknowledge the importance of reflecting the lived experiences of LGBTQ+ youth, they explicitly condemn the use and propagation of such language.

Similar to the sexual content, LGBTQ+ youth did not just receive these sexual solicitations from strangers, but some of the sexual solicitations came *from their peers* as well. While in majority of the conversations we saw youth to participate or interact with the other person, there were a few other conversations where we found our participants *received persistent unwanted messages* that they flagged as "harassment". These messages were received from strangers and peers both. Although youth repeatedly requested to stop texting, they kept sending these messages. Interestingly, we noticed there were a small portion of conversations where they started with sexual content and/or sexual solicitations, but later they *moved toward harassment*. This is because participants often refused to advance with the sexual solicitations, which caused the other person to become aggressive, sometimes even threatening the youth.

Along with the above sexual solicitations, we found conversations where participants were *harassed because of their sexual identity*. Participants often labeled these messages as "harassment" or "unwanted messages". They mostly received these messages from their peers. Some of these conversations also had sexual slurs mentioned within instances of *sexual harassment*. For example, a 20-year-old transgendered man flagged the following conversation that they had with someone they knew:

Other Person: *Call me.*
[Video call started.] [Video call ended.]
Participant: *Wtf what's wrong with you*
Other Person: *I didn't mean to do that*
Participant: *Oh really?*

[Video call started.] [Video call ended.]
Other Person: *Answer the phone fag* [flagged as harassment]
Participant: *Stop texting me please*

LGBTQ+ youth also received similar messages from strangers where they were harassed for their sexual identities. Interestingly, most of these conversations took place in large group chats created for makeups, games, travels, electronics troubleshooting, etc, where people may not know each other. Youth were often harassed during casual conversations related to technical problems with their games and/or computers. For instance, a 21-year-old gender-fluid woman had the following conversation in a group chat for gamers:

Participant: *My fucking laptop is just burning up*
Other Person: *thats because you're a faggot and your gonna burn in hell* [flagged as harassment]
Participant: *well that was impolite because you dont know anythin abt me*

LGBTQ+ youth also often *received different advertisements to promote sexual products and services that were specifically targeted to their sexual identities*. Participants mostly flagged these messages as spam or harassment. In these cases, it seemed apparent that these advertisements and/or scams were aware of their sexual identities, possibly disclosed through their profile descriptions or public posts. LGBTQ+ youth also *received scam messages* that often contained a monetary offer for signing up on a porn website, sending body photos, or working as a sex worker. Participants mostly did not reply to these messages; therefore, we did not find any conversations where they stepped into these monetary traps. In fact, many of the youth seemed savvy to avoid or exploit these solicitations to their own advantage. Below is an example conversation that a 21-year-old transgender (bisexual) woman had with a stranger:

Other Person: *Hi, My ex trans baby left me. Will you be my sugar baby* [flagged as spam]
Participant: *pay me first ? bc i've been in these situations b4*
Other Person: *I'm a legit sugar daddy*
Participant: *send me $100 den*

As shown above, LGBTQ+ youth often received sexual content and solicitations from strangers, but they often took these interactions in stride until they became threatening. However, when they were harassed because of their sexual identities, they did not appreciate such messages. In the next section, we present other general types of online risks that LGBTQ+ youth often reported.

4.3.2 LGBTQ+ Youth Encountered Harassment and Scams. LGBTQ+ youth also received other types of risky messages that were not sexual in nature. For example, we found one-third of the conversations had some form of harassment, e.g., namecalling, threats, abusive words. Participants received these messages not just from strangers but from their peers also. We noticed that youths were often bullied (with threats and abusive words) by peers because of arguments over some physical incidents that happened at the their schools or workplaces. Also, they were often body-shamed by their peers. For instance, a 15 years old non-binary (bisexual) woman had the following conversation with a peer:

Other Person: *hey, your kind fat. you should lose weight. I remember you lost weight... what happened* [flagged as harassment]
Participant: *um it was hard for me. I've been suffering... Stop sending me hate messages*

Besides the harassing messages, participants also received *different advertisements* and *scam messages* that were not specifically targeted to their sexual identity. LGBTQ+ youth flagged these unsolicited messages as unsafe because they often promoted different products or websites and/or intended to cause financial harm. Some of these spam messages contained links to the websites of *illegal products*, e.g., tobacco or recreational drugs. Participants rarely responded to these messages, and therefore, these conversations were mostly one-way. Because these risk experiences seemed relatively typical to the risks generally experienced by youth online [7, 65, 112, 115], we chose not to examine these conversations in more depth.

4.3.3 LGBTQ+ Youth Received Support from Peers Regarding Negative Experiences. Although our conversation dataset focused on risky messages flagged by the LGBTQ+ youth, we often found messages in these conversations where the LGBTQ+ youth *shared negative experiences relating to their sexual identities with peers*. Most of these conversations had messages that they flagged as harassment or bullying. This is because, in the same messaging thread, the youth had arguments with their peers over an issue that led to harassment, and therefore, they flagged the messages as unsafe. However, the youth also shared their negative emotions with the same peers, receiving compassion and moral support. Participants were also seen to share *experiences that they had with homophobic people around them*, mostly in their families. In some of these conversations, they expressed that they are *afraid of coming out* because of their LGBTQ+ intolerant families. Participants also often discussed their *urge to self-harm* as they became depressed about their daily struggle for their gender identities. Interestingly, youth often mentioned about the negative experiences they faced when they publicly shared their photos on different social media platforms. Below is an example conversation that a 19-year-old gender-fluid man had with a peer:

Participant: *i wish things were going ok i have been having thoughts of wanting to cut and i freaking hate it so much i just want to scream and cry for hours*
Other Person: *I'm sorry to hear it, self-harm urges are awful*
Participant: *I hate that i am still afraid to posting wearing a dress or skirt it is making me dysphoric as fuck. People are good for judging me for showing my pics on Instagram and This is the 5th time I was treated badly online And im so fucking tired of it.*

Participants did not just share their struggles related to gender identity; they also often discussed their mental health issues with peers. LGBTQ+ youth expressed their sad feelings about different negative incidents that occurred with their friends and family. Besides these, participants also *shared their depressive thoughts*, e.g., self-harm, suicide, with their peers. In all these messages, we saw their peers being compassionate and providing advice to cope. For

instance, a 15-year-old homosexual woman and their peer had the following conversations:

> **Participant:** *And maybe I'd been lucky to die*
> **Other Person:** *Umm X, don't die on me*
> **Participant:** *You know almost everytime I smile or something of that sort is fake, because I've been broken for so long I've become the person I thought I was when I was 9-12. I've become the girl who literally lost it from grief, and wants to just end it all*
> **Other Person:** *Wait X plz no crying...imma bout to be sad now*

Alongside the depressive feelings, we also often found the LGBTQ+ youth sharing other feelings about the *negative experiences* that they had with their families, friends, and others. We observed that these feelings were not sad in particular, but more related to fear, anger, disgust, and/or anxiety. Participants also occasionally *exchanged advice on their drug usage.* A few conversations contained messages where one person used illegal tobacco or drugs and the other person gave advice or showed concern for their risky behavior. In summary, LGBTQ+ youth did not just receive risky messages from others; they also received support and advice from their peers about their negative experiences with their family and friends. More importantly, they often sought support for the struggles that they dealt with in their families regarding their sexual identity.

5 Discussion

In this section, we briefly summarize our key findings, unpack the implications of these results, and provide recommendations for empowering and protecting LGBTQ+ youth in social media spaces.

5.1 LGBTQ+ Young Experience More Severe Online Risks on Social Media (RQ1)

While LGBTQ+ youth in our study flagged fewer Instagram DMs as risky (48.56%) compared to heterosexual participants (H1), they encountered significantly more high-risk situations (H2), involving increased instances of sexual messages and self-injury. Conversely, heterosexual youth reported more low-risk conversations, primarily centered around online harassment. Despite conflicting with previous research that positioned LGBTQ+ youth as highly vulnerable to cyberbullying [1, 150, 151], our qualitative findings elucidate this apparent contradiction. Many instances of online harassment experienced by LGBTQ+ participants were sexually motivated, leading to categorization as sexual rather than general harassment. This insight reveals the unique challenges faced by LGBTQ+ youth, who often endure online harassment targeting their sexual and gender identities, arguably more detrimental than bullying experienced by heterosexual counterparts. Consequently, our study indicates that LGBTQ+ youth reported significantly more risky conversations involving self-injury and mental health struggles compared to their heterosexual counterparts.

These significant findings underscore the need for urgent action in implementing targeted risk prevention and mitigation programs addressing sexually motivated online harassment and supporting the mental health of LGBTQ+ youth both online and offline. Existing national programs like "Netsmartz" in the United States [51] and "ThinkUKnow" in the United Kingdom [137] offer commendable educational resources but primarily target broader audiences. To effectively safeguard vulnerable communities, such as LGBTQ+ youth, there's a growing demand for specialized education programs tailored to their unique risks and challenges. While organizations like The Trevor Project provide vital crisis intervention and suicide prevention services for LGBTQ+ youth [107], recent studies highlight limitations, including resource availability and operational challenges [87]. Establishing such platforms alone is insufficient; continuous monitoring and timely support are crucial for their effectiveness. Our research emphasizes the need for more focused studies on vulnerable youth, including LGBTQ+ [58, 124], neurodiverse [19, 105], and foster youth [17, 53], concerning their online safety. Recent discourse on youth empowerment and resilience in online spaces [6, 8–11, 84, 143] may not be inclusive of the needs and struggles of more vulnerable youth populations, necessitating a more comprehensive approach.

5.2 Social Media's Dual Impact on LGBTQ+ Youth Mental Health (RQ2)

Our H3 results confirmed existing research indicating that depression disproportionately affects LGBTQ+ minority youths compared to their heterosexual counterparts [109]. Furthermore, LGBTQ+ youth exhibit higher rates of self-harm and diminished mental well-being [12, 29] when compared to heterosexual peers [122, 123] (refer to Table 4 showing the differences in mental health constructs). In addition, our H4 and H5 results introduced novel insights by linking online risk experiences to mental health outcomes for both LGBTQ+ and heterosexual youth, addressing a crucial gap in the literature. Through our models M1-M3 (refer to Table 5), we observed that sexual messages are significantly and positively associated with increased levels of self-harm, injury, and depression, while negatively impacting mental well-being for all youth, regardless of sexual identity. This supports our H4 that more high-risk online experiences correlate with worse mental health outcomes. However, examining interaction effects (H5) revealed nuances. Online harassment was significantly and positively linked to self-harm and injury (Figure 2) for LGBTQ+ youth, not evident in the main effects, while an opposite effect emerged for sexual messages: more messages were associated with reduced mental well-being for heterosexual youth but positively correlated with mental well-being for LGBTQ+ youth.

Careful consideration of these findings is essential due to the conflicting mental health outcomes associated with receiving sexual messages. Both LGBTQ+ and heterosexual youth showed higher rates of Depression and Self-Harm/Injury in association with sexual messages, yet for LGBTQ+ youth, an increase in sexual messages was also linked to improved Mental Well-Being. Our exploration of qualitative insights in response to RQ3 offers a potential explanation for this paradox. While LGBTQ+ youth faced sexual messages harassing their identities, they also received messages facilitating exploration without judgment. Future studies should distinguish between *harmful* and *helpful* sexual messages sent to and by LGBTQ+ youth. This holds crucial implications for Human-Centered Machine Learning (HCML) researchers developing algorithms to detect and mitigate online sexual risks [110, 114, 116]. If these algorithms

inadvertently restrict or censor beneficial sexual conversations of LGBTQ+ youth, the technologies designed to protect them may disproportionately harm them. Our research underscores the need to provide LGBTQ+ youth a safe space for exploring their sexual and gender identities. Future HCML research should strive to differentiate between sexually motivated harassment and violence online and healthy sexual exploration during adolescent developmental growth.

5.3 Social Media Amplifies Risks But Provides Needed Support for LGBTQ+ Youth (RQ3)

Our qualitative findings played a crucial role in elucidating unexpected and occasionally conflicting results from our statistical analyses, offering deeper insights into private online interactions that made LGBTQ+ youth feel uncomfortable or unsafe. One key theme revealed in risk-flagged messages was the prevalence of sexual undertones, even in harassing messages, intensifying the interconnectedness between online risk and the sexual identities of LGBTQ+ youth. This underscores the need for spaces where LGBTQ+ youth can interact online without their sexual identities being the focal point, a benefit often naturally afforded to heterosexual youth. Our results emphasize the urgent societal need for increased awareness and education about the LGBTQ+ community to foster acceptance and inclusion. Additionally, our research highlights the normalcy of some level of risk-seeking behavior in youth development [21], as evidenced by LGBTQ+ youth's willingness to engage with flagged sexual messages, sometimes benefiting from these interactions. However, a concern arises as these conversations often occur with strangers, posing potential risks, including sexual predation. To address this, it is crucial to provide safe outlets for LGBTQ+ youth [25], allowing them to question, explore, and discuss their sexual identities online.

Our research also revealed that LGBTQ+ youth flagged messages as risky even when sharing struggles with peers to seek support, highlighting the double-edged nature of social media. This aligns with existing literature indicating that LGBTQ+ individuals often utilize social media as a support group to share unique experiences both online and offline [38, 91]. Notably, a proposed Kids Online Safety Act (KOSA) [22] aims to enhance minors' online safety but raises privacy concerns due to increased surveillance and potential content filtering. Legislative efforts to ban certain platforms could inadvertently harm marginalized groups, such as LGBTQ+ teens who rely on social media to connect to peers [24, 79], to question, explore, and understand their sexual identities [48]. Restricting Internet use [119] for LGBTQ+ youth might diminish their support networks, necessitating practical solutions that reduce harmful risk exposure while providing opportunities for online support.

5.4 Implications for Practice and Design

We provide several actionable recommendations towards education and design in promoting the online safety and digital well-being of LGBTQ+ youth.

5.4.1 Establishing Stronger Community Guidelines and Norms to Protect LGBTQ+ Youth Online. Recent research highlights significant safety risks for LGBTQ+ users on social media platforms [128], with claims that these platforms prioritize profit over LGBTQ+ safety and lives [146]. Moreover, researchers have raised doubts about the effectiveness of existing community guidelines [34, 146], emphasizing their inadequacy in protecting the LGBTQ+ community. Our study advocates for stronger community-based guidelines, specifically focusing on safeguarding sexual minority social media users, especially LGBTQ+ youth. These guidelines provide a crucial foundation for curtailing harmful content targeting sexual minority youths, setting clear boundaries for acceptable behavior, and discouraging harassment and discrimination. They convey a powerful message affirming the right to a safe online environment for all users, irrespective of sexual identity. Additionally, guidelines offer a structured mechanism for reporting unacceptable behavior, empowering victims to seek assistance. Violators can face consequences, including warnings, suspensions, or bans from the platform [118]. In summary, when combined with effective enforcement, these guidelines contribute to positive community norms by setting clear expectations, reinforcing social norms, fostering trust, and promoting responsible behavior [117, 126].

5.4.2 Providing Resources for Cyberabuse Prevention and Support. To this end, providing help resources and establishing peer support networks for LGBTQ+ youth who are victims of cyberabuse is crucial. These resources empower youth to combat the effects of online abuse and seek help when needed. Equipping them with tools to address cyberabuse is paramount in giving them control over their online experiences. Previous research delved into the effectiveness of diverse support systems, encompassing family, curriculum, peer networks, school policies, Gay-Straight Alliances, etc. and confirmed that these elements are positively linked to the enhancement of positive socioemotional, behavioral, and educational outcomes for LGBTQ+ youth [86]. Additionally, we must establish mechanisms for bystander intervention to create a safer online environment. Encouraging friends, family members, and online community members to step in and advocate for LGBTQ+ youth ensures that they do not bear the burden of responsibility alone. In sum, our research serves as a clarion for immediate action, emphasizing the necessity of creating safer and more supportive digital spaces for all youth, with particular attention to the unique needs and challenges faced by LGBTQ+ individuals.

5.4.3 Creating Safe Online Spaces for LGBTQ+ Youth. Our research underscores the need for safe online spaces explicitly designed for LGBTQ+ youth. These spaces should serve as shelter, where they can explore their sexual identities and access accurate sexual health information without fear of judgment. The absence of such secure environments often turn LGBTQ+ youth to seek information and support from fringe or unregulated online communities [16, 136]. Unfortunately, these spaces may lack adequate moderation, exposing youth to harmful ideologies and potential safety risks. Furthermore, our findings highlight the significant mental health challenges faced by LGBTQ+ youth, who reported a higher prevalence of such challenges compared to heterosexual youth. This highlights the urgency of ensuring easy access to professional support resources in online settings to mitigate these adverse health outcomes. Such initiatives can act as protective measures, reducing the need for LGBTQ+ youth to engage with potentially harmful strangers online. In essence, our research serves as a call to action

to create safer and more supportive digital environments for all youth, with a particular focus on the unique needs and challenges faced by LGBTQ+ individuals [27, 74, 107].

5.4.4 Developing Automated Risk Detection Differentiating between Sexual Harassment and Exploration. In contrast to our previous point, we also advocate for existing social media spaces to be safer for LGBTQ+ youth, so that they do not need to segregate themselves. Our findings shed light on the heightened vulnerability of LGBTQ+ youth to unsolicited sexual messages from strangers. Consequently, there is a pressing need for the development and implementation of automated risk detection systems capable of identifying cyberabuse directed at the LGBTQ+ community and enforcing appropriate penalties. Navigating this challenging landscape reveals algorithmic bias as a significant threat to the well-being of LGBTQ+ individuals [40]. Past studies have shown that automated content moderation often restricts LGBTQ+ content, under the guise of "preserving decency" and "protecting the youth" [40, 67, 77]. The challenges associated with bias detection involve anomaly identification and assessing error rate equality [36, 85, 130, 147]. HCML emerges as a pivotal solution, emphasizing fair language models and comprehensive training in gender-neutral pronouns [28, 138]. HCML strategies include expanding nondiscrimination laws, regulatory sandboxes, safe harbors, and self-regulatory practices with bias impact statements and inclusive design. Efforts also focus on algorithmic literacy and collaborative mitigation through formal feedback mechanisms [85]. However, a significant challenge lies in ensuring that these penalties do not inadvertently infringe upon or harm the LGBTQ+ population. Striking a delicate balance is essential to mitigate potentially harmful interactions while preserving the rights of LGBTQ+ individuals to freely express their sexual identity. These automated systems should demonstrate sophistication in distinguishing between different types of content. For instance, they should distinguish between hate speech directed towards these communities and content that addresses transgender or homosexual issues.

5.4.5 Raising Awareness and Inclusivity. Our study underscores the need to foster greater acceptance and understanding of the LGBTQ+ community. Achieving this goal requires a concerted effort to expand online awareness and educational programs. A diverse array of strategies may be employed to cultivate more inclusive and empathetic environments for LGBTQ+ youth. One strategy may involve educational campaigns for disseminating accurate and up-to-date information regarding LGBTQ+ issues. Previous research highlights how pervasive myths can undermine and dehumanize LGBTQ+ individuals, often unfairly labeling these young individuals as “confused” or “misguided” [3]. Therefore, we suggest that these campaigns adopt a conservative approach that incorporates storytelling and representation supported by facts from evidence-based research. Such personal narratives, combined with objective research, have the potential to put faces and voices to humanize LGBTQ+ experiences, making it easier for others to empathize and relate [20, 134]. A social media campaign explaining ‘deadnaming’ and how it is harmful to transgender people would be a specific example [120]. This approach has the potential to address common misconceptions, ultimately enlightening the public on the complexities of LGBTQ+ identities.

5.5 Limitations and Future Research

We would like to highlight some limitations of our work, which inform future research directions. First, since youth participants were asked to flag at least two conversations that made them feel unsafe or uncomfortable, it is possible that they did not flag all conversations that met this criteria. Further, because we allowed youth to self-annotate their risk flagged messages by risk level and type, we do not account for individual differences in their perceptions of risk. At the same time, we adopt a victim-centered lens that believes participants' lived experiences, rather than questioning these experiences. Second, to have enough power to detect significant differences between heterosexual youth and LGBTQ+ youth, we had to group LGBTQ+ youth allies (e.g., lesbian, gay, bisexual, transgender, queer or questioning, intersex, and asexual). As different LGBTQ+ youth likely have unique experiences concerning their sexual and gender identities, as well as their online risk experiences, studying specific subgroups of the LGBTQ+ community might yield important insights not uncovered in our study. We suggest that future studies oversample from LGBTQ+ communities to achieve large enough sample sizes to detect medium to large size effects between subgroups within the LGBTQ+ community. While a strength of our research is examining the private online communications and risk experiences of LGBTQ+ and heterosexual youth, this means that our results should not be generalized to their public social media communications. Further, private communications via Instagram may also differ from those of other social media platforms. Therefore, we encourage future research to take into account different private versus public contexts, as well as diversify to study youths' experiences on other popular social media platforms. Ultimately, the landscape of adolescent mental health is likely influenced by a multitude of factors beyond private conversations. Future research should explore additional factors that play a role in shaping the mental well-being of LGBTQ+ youth, extending beyond the scope of their online risk experiences.

6 Conclusion

The present study offers empirical evidence regarding the online risk experiences encountered by LGBTQ+ youth in private social media spaces. Furthermore, we examined how these interactions may impact the overall mental health of young adults. We conducted a mixed-method study to understand the severity and types of risks within online private messages experienced by LGBTQ+ youth. In doing so, we differentiate various types of risky and uncomfortable experiences of LGBTQ+ youth and show that LGBTQ+ youth face more severe risks in private spaces than their heterosexual peers. In addition, our findings highlight the relationship between sexual messages and Self-Harm behavior among both LGBTQ+ and heterosexual youth. In contrast, young people identifying as LGBTQ+ are more likely to Self-Harm due to harassment, whereas heterosexual youth are less affected. At the same time, an increased number of sexual messages corresponded to positive mental Well-Being among LGBTQ+ youth. Thematic analysis of DMs show that LGBTQ+ youth engage in risky conversations with strangers and frequently receive sexual messages and solicitations with inappropriate content. Our findings highlight the critical need for social media tools and support mechanisms to combat risks that

take place within private online spaces and are particularly likely to negatively effect LGBTQ+ youth.

Acknowledgments

This research is supported in part by the U.S. National Science Foundation under grants #IIP-2329976, #IIS-2333207 and by the William T. Grant Foundation grant #187941. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the research sponsors. We would also like to thank all the participants who donated their data and contributed towards our research.

References

[1] Elias Aboujaoude, Matthew W Savage, Vladan Starcevic, and Wael O Salame. 2015. Cyberbullying: Review of an old problem gone viral. *Journal of adolescent health* 57, 1 (2015), 10–18. https://doi.org/10.1016/j.jadohealth.2015.04.011

[2] Roberto L Abreu and Maureen C Kenny. 2018. Cyberbullying and LGBTQ youth: A systematic literature review and recommendations for prevention and intervention. *Journal of Child & Adolescent Trauma* 11, 1 (2018), 81–97. https://doi.org/10.1007/s40653-017-0175-7

[3] Roberto L Abreu, Adriana G McEachern, and Maureen C Kenny. 2017. Myths and Misconceptions about LGBTQ Youth: School Counselors' Role in Advocacy. *Journal of School Counseling* 15, 8 (2017), n8. https://eric.ed.gov/?id=EJ1146191

[4] Newport Academy. 2021. *How Does Social Media Affect Teenagers?* https://www.newportacademy.com/resources/well-being/effect-of-social-media-on-teenagers/.

[5] Dane Acena and Guo Freeman. 2021. "In My Safe Space": Social Support for LGBTQ Users in Social Virtual Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6. https://doi.org/10.1145/3411763.3451673

[6] Zainab Agha, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2023. "Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–32. https://doi.org/10.1145/3579625

[7] Zainab Agha, Neeraj Chatlani, Afsaneh Razi, and Pamela Wisniewski. 2020. Towards Conducting Responsible Research with Teens and Parents Regarding Online Risks. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383073

[8] Mamtaj Akter, Leena Alghamdi, Dylan Gillespie, Nazmus Sakib Miazi, Jess Kropczynski, Heather Lipford, and Pamela J. Wisniewski. 2022. CO-oPS: A Mobile App for Community Oversight of Privacy and Security. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) *(CSCW'22 Companion)*. Association for Computing Machinery, New York, NY, USA, 179–183. https://doi.org/10.1145/3500868.3559706

[9] Mamtaj Akter, Leena Alghamdi, Jess Kropczynski, Heather Richter Lipford, and Pamela J. Wisniewski. 2023. It Takes a Village: A Case for Including Extended Family Members in the Joint Oversight of Family-Based Privacy and Security for Mobile Smartphones. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 194, 7 pages. https://doi.org/10.1145/3544549.3585904

[10] Mamtaj Akter, Amy J. Godfrey, Jess Kropczynski, Heather R. Lipford, and Pamela J. Wisniewski. 2022. From Parental Control to Joint Family Oversight: Can Parents and Teens Manage Mobile Online Safety and Privacy as Equals? *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 57 (apr 2022), 28 pages. https://doi.org/10.1145/3512904

[11] Mamtaj Akter, Madiha Tabassum, Nazmus Sakib Miazi, Leena Alghamdi, Jess Kropczynski, Pamela J. Wisniewski, and Heather Lipford. 2023. Evaluating the Impact of Community Oversight for Managing Mobile Privacy and Security. In *Nineteenth Symposium on Usable Privacy and Security (SOUPS 2023)*. USENIX Association, Anaheim, CA, 437–456. https://www.usenix.org/conference/soups2023/presentation/akter

[12] Joanna Almeida, Renee M Johnson, Heather L Corliss, Beth E Molnar, and Deborah Azrael. 2009. Emotional distress among LGBT youth: The influence of perceived discrimination based on sexual orientation. *Journal of youth and adolescence* 38, 7 (2009), 1001–1014. https://doi.org/10.1007/s10964-009-9397-9

[13] Nazanin Andalibi. 2019. What happens after disclosing stigmatized experiences on identified social media: Individual, dyadic, and social/network outcomes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15. https://doi.org/10.1145/3290605.3300367

[14] Monica Anderson, Jingjing Jiang, et al. 2018. Teens, social media & technology 2018. *Pew Research Center* 31, 2018 (2018), 1673–1689.

[15] Ikuko Aoyama, Lucy Barnard-Brak, and Tony L Talbert. 2011. Cyberbullying among high school students: Cluster analysis of sex and age differences and the level of parental monitoring. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)* 1, 1 (2011), 25–35. https://doi.org/10.4018/ijcbpl.2011010103

[16] Rachel Badham. 2021. *64an alternative educational resource.* https://www.gscene.com/news/durex-mysexmyway-lgbtq-study-uk-sex-ed/

[17] Karla Badillo-Urquiola, Xinru Page, and Pamela J Wisniewski. 2019. Risk vs. restriction: The tension between providing a sense of normalcy and keeping foster teens safe online. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3290605.3300497

[18] Karla Badillo-Urquiola, Zachary Shea, Zainab Agha, Irina Lediaeva, and Pamela Wisniewski. 2021. Conducting risky research with teens: co-designing for the ethical treatment and protection of adolescents. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–46. https://doi.org/10.1145/3432930

[19] Belén Barros Pena, Nelya Koteyko, Martine Van Driel, Andrea Delgado, and John Vines. 2023. " My Perfect Platform Would Be Telepathy"-Reimagining the Design of Social Media with Autistic Adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16. https://doi.org/10.1145/3544548.3580673

[20] Sue R Bauer. 2003. The power of story. In *Proceedings of the 31st annual ACM SIGUCCS fall conference*. 151–153. https://doi.org/10.1145/947469.947510

[21] Diana Baumrind. 1987. A developmental perspective on adolescent risk taking in contemporary America. *New Directions for Child Development* 37 (1987), 93–125. https://doi.org/10.1002/cd.23219873706

[22] Richard Blumenthal. 2022. *Blumenthal & Blackburn Introduce Comprehensive Kids' Online Safety Legislation.* Retrieved Jan. 31, 2023 from https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-blackburn-introduce-comprehensive-kids-online-safety-legislation

[23] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association. https://doi.org/10.1037/13620-004

[24] Christian Brown. 2023. Clarksville-Montgomery County Schools sues Facebook, Tik Tok, others over student mental health crisis. https://clarksvillenow.com/local/clarksville-montgomery-county-schools-sues-facebook-tik-tok-others-over-student-mental-health-crisis/

[25] Dylan Buckley. 2022. *A Deep Dive: Online Anonymous Chat Rooms.* https://www.betterhelp.com/advice/chat/pros-and-cons-of-an-anonymous-chat-room/.

[26] Matthew Carrasco and Andruid Kerne. 2018. Queer visibility: Supporting LGBT+ selective visibility on social media. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12. https://doi.org/10.1145/3173574.3173824

[27] The LGBT National Help Center. 2023. *The LGBT National Help Center.* https://www.lgbthotline.org/youthchatrooms.

[28] Stevie Chancellor. 2023. Toward Practices for Human-Centered Machine Learning. *Commun. ACM* 66, 3 (2023), 78–85. https://doi.org/10.1145/3530987

[29] Linda Charmaraman, Rachel Hodes, Amanda M Richer, et al. 2021. Young Sexual Minority Adolescent Experiences of Self-expression and Isolation on Social Media: Cross-sectional Survey Study. *JMIR mental health* 8, 9 (2021), e26207. https://doi.org/10.2196/26207

[30] Kendra Cherry. 2020. *What Does LGBTQ+ Mean?* https://www.verywellmind.com/what-does-lgbtq-mean-5069804

[31] Alexander Cho. 2018. Default publicness: Queer youth of color, social media, and being outed by the machine. *New Media & Society* 20, 9 (2018), 3183–3200. https://doi.org/10.1177/1461444817744784

[32] Eunseong Cho and Seonghoon Kim. 2015. Cronbach's coefficient alpha: Well known but poorly understood. *Organizational research methods* 18, 2 (2015), 207–230. https://doi.org/10.1177/1094428114555994

[33] Victoria Clarke, Sonja J Ellis, Elizabeth Peel, and Damien W Riggs. 2010. *Lesbian, gay, bisexual, trans and queer psychology: An introduction.* Cambridge University Press. https://doi.org/10.1017/CBO9780511810121

[34] Shannon Connellan. 2022. *Social media giants aren't protecting LGBTQ users enough, GLAAD says.* https://mashable.com/article/social-media-lgbtq-safety-index-glaad

[35] Kerith J. Conron. 2020. LGBT Youth Population in the United States. (2020).

[36] Sam Corbett-Davies, Emma Peirson, Avi Feller, and Sharad Goel. 2016. *A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear.* https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.

[37] Shelley L. Craig, Andrew D. Eaton, Lauren B. McInroy, Vivian W. Y. Leung, and Sreedevi Krishnan. 2021. Can Social Media Participation Enhance LGBTQ+ Youth Well-Being? Development of the Social Media Benefits Scale. *Social Media + Society* 7, 1 (2021), 2056305121988931. https://doi.org/10.1177/2056305121988931 arXiv:https://doi.org/10.1177/2056305121988931

[38] Shelley L Craig and Lauren McInroy. 2014. You can form a part of yourself online: The influence of new media on identity development and coming out for LGBTQ youth. *Journal of Gay & Lesbian Mental Health* 18, 1 (2014), 95–109. https://doi.org/0.1080/19359705.2013.777007

[39] Lindsey Dawson, Brittni Frederiksen, and Michelle Long. 2023. *Mental Health Care Needs and Experiences Among LGBT+ People.* https://www.kff.org/mental-health/issue-brief/mental-health-care-needs-and-experiences-among-lgbt-people/#:~:text=Experiences%20with%20Mental%20Health%20Problems,a%20recent%20KFF%2FCNN%20poll.

[40] Scott DeGeest. 2022. *Let's Talk about Bias: LGBTQ+ People and AI.* https://www.correlation-one.com/blog/llgbtq-people-ai-algorithmic-bias-harm.

[41] Samantha DeHaan, Laura E Kuper, Joshua C Magee, Lou Bigelow, and Brian S Mustanski. 2013. The interplay between online and offline explorations of identity, relationships, and sex: A mixed-methods study with LGBT youth. *Journal of sex research* 50, 5 (2013), 421–434. https://doi.org/10.1080/00224499.2012.661489

[42] Nicholas F Denton, Sharon Scales Rostosky, and Fred Danner. 2014. Stigma-related stressors, coping self-efficacy, and physical health in lesbian, gay, and bisexual individuals. *Journal of counseling psychology* 61, 3 (2014), 383. https://doi.org/10.1037/a0036707

[43] Michael A DeVito, Ashley Marie Walker, and Jeremy Birnholtz. 2018. 'Too Gay for Facebook' Presenting LGBTQ+ Identity Throughout the Personal Social Media Ecosystem. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23. https://doi.org/10.1145/3274313

[44] Elizabeth W. Diemer, Julia D. Grant, Melissa A. Munn-Chernoff, David A. Patterson, and Alexis E. Duncan. 2015. Gender Identity, Sexual Orientation, and Eating-Related Pathology in a National Sample of College Students. *Journal of Adolescent Health* 57, 2 (2015), 144–149. https://doi.org/10.1016/j.jadohealth.2015.03.003

[45] Harry Domanski. 2019. *Facebook says it's switching focus from public posts to private messages.* https://www.techradar.com/news/facebook-says-its-switching-focus-from-public-posts-to-private-messages.

[46] Facebook. 2022. *How do I report a post or profile on Instagram?* https://www.facebook.com/help/instagram/192435014247952.

[47] R Frank Falk and Nancy B Miller. 1992. *A primer for soft modeling.* University of Akron Press. https://psycnet.apa.org/record/1992-98610-000

[48] Lauren Feiner. 2022. *Kids Online Safety Act may harm minors, civil society groups warn lawmakers.* Retrieved Jan. 31, 2023 from https://www.cnbc.com/2022/11/28/kids-online-safety-act-may-harm-minors-civil-society-groups-warn.html

[49] David Finkelhor, Richard Ormrod, Heather Turner, and Melissa Holt. 2009. Pathways to poly-victimization. *Child maltreatment* 14, 4 (2009), 316–329. https://doi.org/10.1177/1077559509347012

[50] David Finkelhor, Richard K Ormrod, and Heather A Turner. 2007. Polyvictimization and trauma in a national longitudinal cohort. *Development and psychopathology* 19, 1 (2007), 149–166. https://doi.org/10.1017/S0954579407070083

[51] The National Center for Missing and Exploited Children. 2023. *Netsmartz.* https://www.missingkids.org/netsmartz/home.

[52] Annie E. Casey Foundation. 2021. *Definitions of Common LGBTQ Concepts and Terms.* https://www.aecf.org/blog/lgbtq-definitions.

[53] John Fowler, Mark Zachry, and David W McDonald. 2022. Fostering communication: Characterizing the concerns of former foster youth in an online community. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–23. https://doi.org/10.1145/3492834

[54] Jesse Fox and Jennifer J Moreland. 2015. The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in human behavior* 45 (2015), 168–176. https://doi.org/10.1016/j.chb.2014.11.083

[55] Jesse Fox and Rachel Ralston. 2016. Queer identity online: Informal learning and teaching experiences of LGBTQ individuals on social media. *Computers in Human Behavior* 65 (2016), 635–642. https://doi.org/10.1016/j.chb.2016.06.009

[56] Jesse Fox and Katie M Warber. 2015. Queer identity management and political self-expression on social networking sites: A co-cultural approach to the spiral of silence. *Journal of Communication* 65, 1 (2015), 79–100. https://doi.org/10.1111/jcom.12137

[57] David M Frost, Keren Lehavot, and Ilan H Meyer. 2015. Minority stress and physical health among sexual minority individuals. *Journal of behavioral medicine* 38, 1 (2015), 1–8. https://doi.org/10.1177/1745691613497965

[58] Cally Gatehouse, Matthew Wood, Jo Briggs, James Pickles, and Shaun Lawson. 2018. Troubling vulnerability: Designing with LGBT young people's ambivalence towards hate crime reporting. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13. https://doi.org/10.1145/3173574.3173683

[59] Christine Geeng and Alexis Hiniker. 2021. LGBTQ privacy concerns on social media. *arXiv preprint arXiv:2112.00107* (2021). https://doi.org/10.48550/arXiv.2112.00107

[60] Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment* 4, 1 (1992), 26. https://doi.org/10.1037/1040-3590.4.1.26

[61] Emily A. Greytak, Josephine Korchmaros, Joseph G. Kosciw, Kimberly J. Mitchell, Neal A. Palmer, and Michele L. Ybarra. 2013. Out online: The experiences of lesbian, gay, bisexual and transgender youth on the Internet. *New York, NY* (2013). https://search.issuelab.org/resource/out-online-the-experiences-of-lesbian-gay-bisexual-and-transgender-youth-on-the-internet.html

[62] Amy Guasp, Anne Gammon, and Gavin Ellison. 2013. Homophobic hate crime: The gay British crime survey 2013. *London: Stonewall* (2013).

[63] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. 2015. Disclosure, stress, and support during gender transition on Facebook. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing.* 1176–1190. https://doi.org/10.1145/2675133.2675152

[64] Gary W Harper, Douglas Bruce, Pedro Serrano, and Omar B Jamil. 2009. The role of the Internet in the sexual identity development of gay and bisexual male adolescents. *The story of sexual identity: Narrative perspectives on the gay and lesbian life course* (2009), 297–326. https://doi.org/10.1093/acprof:oso/9780195326789.003.0013

[65] Heidi Hartikainen, Afsaneh Razi, and Pamela Wisniewski. 2021. Safe Sexting: The Advice and Support Adolescents Receive from Peers Regarding Online Sexual Risks. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 42 (apr 2021), 31 pages. https://doi.org/10.1145/3449116

[66] Tyler Hatchel, Cagil Torgal, America J. El Sheikh, Luz E. Robinson, Alberto Valido, and Dorothy L. Espelage. 2021. LGBTQ youth and digital media: online risks. https://doi.org/10.1016/B978-0-12-817499-9.00015-6

[67] Mark Hay. 2021. *How AI lets bigots and trolls flourish while censoring LGBTQ+ voices.* https://www.inkl.com/news/how-ai-lets-bigots-and-trolls-flourish-while-censoring-lgbtq-voices.

[68] healthcare.utah.edu. 2021. *WHY DOES THE LGBTQIA+ COMMUNITY SUFFER FROM POOR MENTAL HEALTH AT HIGHER RATES?* https://healthcare.utah.edu/healthfeed/2021/07/why-does-lgbtqia-community-suffer-poor-mental-health-higher-rates

[69] Michael L Hendricks and Rylan J Testa. 2012. A conceptual framework for clinical work with transgender and gender nonconforming clients: An adaptation of the Minority Stress Model. *Professional Psychology: Research and Practice* 43, 5 (2012), 460. https://doi.org/10.1037/a0029597

[70] Lynne Hiller, Alina Turner, and Anne Mitchell. 2005. Writing Themselves in Again: 6 Years on. The 2nd National Report on the Sexuality, Health & Well-being of Same Sex Attracted Young People in Australia. Melbourne: Australian Research Centre in Sex. *Health & Society* (2005). https://rainbowhealthaustralia.org.au/media/pages/research-resources/writing-themselves-in-again/2614754494-1650953507/writing_themselves_in_again.pdf

[71] Lynne Hillier and Lyn Harrison. 2007. Building realities less limited than their own: Young people practising same-sex attraction on the internet. *Sexualities* 10, 1 (2007), 82–100. https://doi.org/10.1177/136346070707295

[72] Donell Holloway and Lelia Green. 2016. The internet of toys. *Communication Research and Practice* 2, 4 (2016), 506–519. https://link.springer.com/book/10.1007/978-3-030-10898-4

[73] Darren Homrighausen. 2015. *Factor Analysis -Applied Multivariate Analysis.* https://darrenho.github.io/AMA/factorAnalysis.pdf.

[74] Trans Lifeline Hotline. 2023. *TransLine.org.* https://translifeline.org/hotline/

[75] Jane C. Hu. 2020. *We Want a More Private Internet, but We Want to Screenshot It Too.* https://slate.com/technology/2020/02/screenshots-text-conversations-privacy-social-media.html.

[76] Instagram. 2021. *Introducing new tools to protect our community from abuse.* https://about.instagram.com/blog/announcements/introducing-new-tools-to-protect-our-community-from-abuse.

[77] Nicolas Kayser-Bril. 2020. *Automated moderation tool from Google rates People of Color and gays as "toxic".* https://algorithmwatch.org/en/automated-moderation-perspective-bias/.

[78] Rachel Keighley. 2022. Hate Hurts: Exploring the Impact of Online Hate on LGBTQ+ Young People. *Women & Criminal Justice* 32, 1-2 (2022), 29–48. https://doi.org/10.1080/08974454.2021.1988034

[79] Mark Kelly. 2023. Tennessee school district joins mass action lawsuit against Big Tech. https://www.wkrn.com/special-reports/back-to-school/tennessee-school-district-joins-mass-action-lawsuit-against-big-tech/

[80] Amrita Khalid. 2019. *Twitter is testing a filter for potentially offensive messages.* https://www.engadget.com/2019-08-15-twitter-is-testing-a-filter-for-potentially-offensive-messages.html.

[81] Rex B Kline. 2023. *Principles and practice of structural equation modeling.* Guilford publications. https://www.guilford.com/books/Principles-and-Practice-of-Structural-Equation-Modeling/Rex-Kline/9781462551910

[82] E. David Klonsky and Catherine R. Glenn. 2009. Assessing the Functions of Non-suicidal Self-injury: Psychometric Properties of the Inventory of Statements About Self-injury (ISAS). *Journal of Psychopathology and Behavioral Assessment* 31, 3 (Sept. 2009), 215–219. https://doi.org/10.1007/s10862-008-9107-z

[83] Kurt Kroenke, Robert L. Spitzer, and Janet B. Williams. 2001. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine* 16, 9 (Sept. 2001), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Publisher: Blackwell Science Inc.
[84] Priya Kumar. 2022. *TikTok spat between Kanye West and Kim Kardashian has lessons for all parents.* https://www.nbcnews.com/think/opinion/tiktok-spat-between-kanye-west-kim-kardashian-has-lessons-all-ncna1289172
[85] Nicol Turner Lee, Paul Resnick, and Genie Barton. 2019. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings Institute: Washington, DC, USA* 2 (2019). https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/
[86] Enoch Leung, Gabriela Kassel-Gomez, Samantha Sullivan, Flavio Murahara, and Tara Flanagan. 2022. Social support in schools and related outcomes for LGBTQ youth: a scoping review. *Discover education* 1, 1 (2022), 18. https://doi.org/10.1007/s44217-022-00016-9
[87] Joel Levtov. 2023. *Trevor Project in crisis amid financial woes, staff dissension, 'union busting': sources.* https://www.washingtonblade.com/2023/08/10/trevor-project-crisis/
[88] Stephen P Lewis, Nancy L Heath, Jill M St Denis, and Rick Noble. 2011. The scope of nonsuicidal self-injury on YouTube. *Pediatrics* 127, 3 (2011), e552–e557. https://doi.org/10.1542/peds.2010-2317
[89] Ruili Li, Qiguo Lian, Qiru Su, Luhai Li, Meixian Xie, and Jun Hu. 2020. Trends and sex disparities in school bullying victimization among US youth, 2011–2019. *BMC public health* 20, 1 (2020), 1–6. https://doi.org/10.1186/s12889-020-09677-3
[90] Zhenqiu Lu and Ke-Hai Yuan. 2010. *Welch's t test.* 1620–1623. https://doi.org/10.13140/RG.2.1.3057.9607
[91] Leanna Lucero. 2017. Safe spaces in online places: Social media and LGBTQ youth. *Multicultural Education Review* 9, 2 (2017), 117–128. https://eric.ed.gov/?id=EJ1140031
[92] Michael P Marshal, Laura J Dietz, Mark S Friedman, Ron Stall, Helen A Smith, James McGinley, Brian C Thoma, Pamela J Murray, Anthony R D'Augelli, and David A Brent. 2011. Suicidality and depression disparities between sexual minority and heterosexual youth: A meta-analytic review. *Journal of adolescent health* 49, 2 (2011), 115–123. https://doi.org/10.1016/j.jadohealth.2011.02.005
[93] Emma Powys Maurice. 2021. *Hackers demand $1 million to stop leak of private user info from Israeli LGBT+ dating site.* https://www.pinknews.co.uk/2021/11/01/atraf-cyberserve-hack-lgbt-data-leak-israel/.
[94] Mary L McHugh. 2013. The chi-square test of independence. *Biochemia medica* 23, 2 (2013), 143–149. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900058/
[95] Lauren B McInroy and Shelley L Craig. 2020. "It's like a safe haven fantasy world": Online fandom communities and the identity development activities of sexual and gender minority youth. *Psychology of Popular Media* 9, 2 (2020), 236. https://doi.org/10.1037/ppm0000234
[96] Lauren B McInroy, Rebecca J McCloskey, Shelley L Craig, and Andrew D Eaton. 2019. LGBTQ+ youths' community engagement and resource seeking online versus offline. *Journal of Technology in Human Services* 37, 4 (2019), 315–333. https://doi.org/10.1080/15228835.2019.1617823
[97] Ilan H Meyer. 1995. Minority stress and mental health in gay men. *Journal of health and social behavior* (1995), 38–56.
[98] Ilan H Meyer. 2003. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychological bulletin* 129, 5 (2003), 674. https://doi.org/10.1037/0033-2909.129.5.674
[99] Michele Meyer. 2019. The impact of social media on non-monosexuals' responses to discrimination: A co-cultural approach. *Social Media+ Society* 5, 1 (2019), 2056305119826120. https://doi.org/10.1177/20563051198261
[100] Kathryn C Montgomery, Jeff Chester, and Tijana Milosevic. 2017. Ensuring young people's digital privacy as a fundamental right. In *International handbook of media literacy education.* Routledge, 85–102. https://doi.org/10.4324/9781315628110-9
[101] Brian Mustanski, Michael E Newcomb, and Robert Garofalo. 2011. Mental health of lesbian, gay, and bisexual youths: A developmental resiliency perspective. *Journal of gay & lesbian social services* 23, 2 (2011), 204–225. https://doi.org/10.1080/10538720.2011.561474
[102] BBC News. 2021. *Grindr faces £8.5m fine for selling user data.* https://www.bbc.com/news/technology-55811681.
[103] Gwenn Schurgin O'Keeffe, Kathleen Clarke-Pearson, Council on Communications, and Media. 2011. The impact of social media on children, adolescents, and families. *Pediatrics* 127, 4 (2011), 800–804. https://doi.org/10.1542/peds.2011-0054
[104] Peterson K Ozili. 2023. The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native english speakers.* IGI Global, 134–143. https://doi.org/10.2139/ssrn.4128165
[105] Xinru Page, Andrew Capener, Spring Cullen, Tao Wang, Monica Garfield, and Pamela J. Wisniewski. 2022. Perceiving Affordances Differently: The Unintended Consequences When Young Autistic Adults Engage with Social Media. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–21. https://doi.org/10.1145/3491102.3517596
[106] Anastasia Powell, Adrian J Scott, and Nicola Henry. 2020. Digital harassment and abuse: Experiences of sexuality and gender minority adults. *European Journal of Criminology* 17, 2 (2020), 199–223. https://doi.org/10.1177/1477370818788006
[107] Trevor Project. 1998. *The Trevor Project.* https://www.thetrevorproject.org/.
[108] Cassidy Pyle, Lee Roosevelt, Ashley Lacombe-Duncan, and Nazanin Andalibi. 2021. LGBTQ Persons' Pregnancy Loss Disclosures to Known Ties on Social Media: Disclosure Decisions and Ideal Disclosure Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–17. https://doi.org/10.1145/3411764.3445331
[109] Emmanuel N-B Quarshie, Mitch G Waterman, and Allan O House. 2020. Prevalence of self-harm among lesbian, gay, bisexual, and transgender adolescents: a comparison of personal and social adversity with a heterosexual sample in Ghana. *BMC research notes* 13, 1 (2020), 1–6. https://doi.org/10.1186/s13104-020-05111-4
[110] Afsaneh Razi, Ashwaq AlSoubai, Seunghyun Kim, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. 2023. Sliding into My DMs: Detecting Uncomfortable or Unsafe Sexual Risk Experiences within Instagram Direct Messages Grounded in the Perspective of Youth. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–29. https://doi.org/10.1145/3579522
[111] Afsaneh Razi, Ashwaq AlSoubai, Seunghyun Kim, Nurun Naher, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. 2022. Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. (2022). https://doi.org/10.1145/3491101.3503569
[112] Afsaneh Razi, Ashwaq Alsoubai, Seunghyun Kim, Nurun Naher, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J. Wisniewski. 2022. Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 39, 9 pages. https://doi.org/10.1145/3491101.3503569
[113] Afsaneh Razi, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2020. Let's talk about sext: How adolescents seek support and advice about their online sexual experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13. https://doi.org/10.1145/3313831.3376400
[114] Afsaneh Razi, Karla Badillo-Urquiola, and Pamela J Wisniewski. 2020. Let's talk about sext: How adolescents seek support and advice about their online sexual experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.
[115] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J. Wisniewski. 2021. A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 465 (oct 2021), 38 pages. https://doi.org/10.1145/3479609
[116] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J Wisniewski. 2021. A human-centered systematic literature review of the computational approaches for online sexual risk detection. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–38. https://doi.org/10.1145/3479609
[117] Reddit. 2023. Moderator Code of Conduct. https://www.redditinc.com/policies/moderator-code-of-conduct
[118] Reddit. 2023. Reddit Content Policy. https://www.redditinc.com/policies/content-policy
[119] Erin Reed. 2023. Senator Admits "Kids Online Safety Act" Will Target Trans Content Online. https://www.erininthemorning.com/p/senator-admits-kids-online-safety
[120] Ariane Resnick. 2023. *Deadnaming—What It Is and Why It's Harmful to Mental Health.* https://www.verywellmind.com/what-is-deadnaming-and-why-is-it-harmful-5188575
[121] Diana Rieger, Anna Sophie Kümpel, Maximilian Wich, Toni Kiening, and Georg Groh. 2021. Assessing the extent and types of hate speech in fringe communities: a case study of alt-right communities on 8chan, 4chan, and Reddit. *Social Media+ Society* 7, 4 (2021), 20563051211052906. https://doi.org/10.1177/20563051211052906
[122] Stephen T Russell and Jessica N Fish. 2016. Mental health in lesbian, gay, bisexual, and transgender (LGBT) youth. *Annual review of clinical psychology* 12 (2016), 465–487. https://doi.org/10.1146/annurev-clinpsy-021815-093153
[123] Stephen T Russell and Jessica N Fish. 2019. Sexual minority youth, social change, and health: A developmental collision. *Research in Human Development* 16, 1 (2019), 5–20. https://doi.org/10.1080/15427609.2018.1537772
[124] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–22. https://doi.org/10.1145/3361108
[125] Rob Sanders. 2021. *The Great Migration from Public Social to Private Social Media.* https://www.simplilearn.com/public-social-to-private-social-media-article
[126] Rebecca Scharlach, Blake Hallinan, and Limor Shifman. 2023. Governing principles: Articulating values in social media platform policies. *new media & society* (2023), 14614448231156580. https://doi.org/10.1177/14614448231156580

[127] Shari Kessel Schneider, Lydia O'donnell, Ann Stueve, and Robert WS Coulter. 2012. Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American journal of public health* 102, 1 (2012), 171–177. https://doi.org/10.2105/AJPH.2011.300308

[128] Cynthia Silva. 2021. *Top social media platforms 'unsafe' for LGBTQ users, report finds.* https://www.nbcnews.com/nbc-out/out-news/top-social-media-platforms-unsafe-lgbtq-users-report-finds-rcna889

[129] Civic Space. 2017. *Egypt: Crackdown on LGBTQI people and their supporters must stop.* https://www.article19.org/resources/egypt-crackdown-on-lgbtqi-people-and-their-supporters-must-stop/.

[130] Matthias Spielkamp. 2017. *Inspecting Algorithms for Bias.* https://www.technologyreview.com/2017/06/12/105804/inspecting-algorithms-for-bias/.

[131] Jessica Steinke, Meredith Root-Bowman, Sherry Estabrook, Deborah S Levine, and Leslie M Kantor. 2017. Meeting the needs of sexual and gender minority youth: formative research on potential digital health interventions. *Journal of Adolescent Health* 60, 5 (2017), 541–548. https://doi.org/10.1016/j.jadohealth.2016.11.023

[132] Paul R Sterzing, G Allen Ratliff, Rachel E Gartner, Briana L McGeough, and Kelly C Johnson. 2017. Social ecological correlates of polyvictimization among a national sample of transgender, genderqueer, and cisgender sexual minority adolescents. *Child Abuse & Neglect* 67 (2017), 1–12. https://doi.org/10.1016/j.chiabu.2017.02.017

[133] Stela Stojisavljevic, Bosiljka Djikanovic, and Bojana Matejic. 2017. 'The Devil has entered you': A qualitative study of Men Who Have Sex With Men (MSM) and the stigma and discrimination they experience from healthcare professionals and the general community in Bosnia and Herzegovina. *PLoS One* 12, 6 (2017), e0179101. https://doi.org/10.1371/journal.pone.0179101

[134] Wendy A Suzuki, Mónica I Feliú-Mójer, Uri Hasson, Rachel Yehuda, and Jean Mary Zarate. 2018. Dialogues: The science and power of storytelling. *Journal of Neuroscience* 38, 44 (2018), 9468–9470. https://doi.org/10.1523/JNEUROSCI.1942-18.2018

[135] Ruth Tennant, Louise Hiller, Ruth Fishwick, Stephen Platt, Stephen Joseph, Scott Weich, Jane Parkinson, Jenny Secker, and Sarah Stewart-Brown. 2007. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes* 5, 1 (Nov. 2007), 63. https://doi.org/10.1186/1477-7525-5-63

[136] Fight the New Drug. 2022. *How Porn Misrepresents the LGBTQ+ Community.* https://fightthenewdrug.medium.com/how-porn-misrepresents-the-lgbtq-community-b121555b558e

[137] thinkuknow. 2023. *Welcome to CEOP Education.* https://www.thinkuknow.co.uk/.

[138] Nenad Tomasev, Kevin R McKee, Jackie Kay, and Shakir Mohamed. 2021. Fairness for unobserved characteristics: Insights from technological impacts on queer communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society.* 254–265. https://doi.org/10.1145/3461702.3462540

[139] Heather A Turner, David Finkelhor, and Richard Ormrod. 2010. Polyvictimization in a national sample of children and youth. *American journal of preventive medicine* 38, 3 (2010), 323–330. https://doi.org/10.1016/j.amepre.2009.11.012

[140] School of Education University of Miami and Human Development. 2020. *Steps of conducting Confirmatory Factor Analysis (CFA) in R.* https://sites.education.miami.edu/statsu/2020/10/12/steps-of-conducting-confirmatory-factor-analysis-cfa-in-r/.

[141] Linh Gia Vu, Linh Khanh Le, Anh Vu Trong Dam, Son Hoang Nguyen, Thuc Thi Minh Vu, Trang Thu Hong Trinh, Anh Linh Do, Ngoc Minh Do, Trang Huyen Le, Carl Latkin, et al. 2022. Factor structures of patient health questionnaire-9 instruments in exploring depressive symptoms of suburban population. *Frontiers in Psychiatry* 13 (2022), 838747. https://doi.org/10.3389/fpsyt.2022.838747

[142] David Walker. 2023. *Chi Square.* https://www.cedu.niu.edu/~walker/statistics/Chi%20Square%202.pdf.

[143] Pamela Wisniewski, Haiyan Jia, Na Wang, Saijing Zheng, Heng Xu, Mary Beth Rosson, and John M Carroll. 2015. Resilience mitigates the negative effects of adolescent internet addiction and online risk exposure. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 4029–4038. https://doi.org/10.1145/2702123.2702240

[144] Michele L Ybarra, Kimberly J Mitchell, Neal A Palmer, and Sari L Reisner. 2015. Online social support as a buffer against online and offline peer and sexual victimization among US LGBT and non-LGBT youth. *Child abuse & neglect* 39 (2015), 123–136. https://doi.org/10.1016/j.chiabu.2014.08.006

[145] Horim Yi, Hyemin Lee, Jooyoung Park, Bokyoung Choi, and Seung-Sup Kim. 2017. Health disparities between lesbian, gay, and bisexual adults and the general population in South Korea: Rainbow Connection Project I. *Epidemiology and health* 39 (2017). https://doi.org/10.4178/epih.e2017046

[146] Jo Yurcaba. 2022. *Social media platforms aren't doing enough to keep LGBTQ people safe, group says.* https://www.nbcnews.com/nbc-out/out-news/social-media-platforms-arent-enough-keep-lgbtq-people-safe-group-says-rcna37319

[147] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. 2017. Fairness constraints: Mechanisms for fair classification. In *Artificial intelligence and statistics.* PMLR, 962–970. https://proceedings.mlr.press/v54/zafar17a/zafar17a.pdf

[148] Christian Zilles. 2019. *Private Messages Are The New Social Network.* https://socialmediaexplorer.com/social-media-marketing/private-messages-are-the-new-social-network/.

[149] Adam G Zimmerman and Gabriel J Ybarra. 2016. Online aggression: The influences of anonymity and social modeling. *Psychology of Popular Media Culture* 5, 2 (2016), 181. https://doi.org/10.1037/ppm0000038

[150] Izabela Zych, Rosario Ortega-Ruiz, and Rosario Del Rey. 2015. Systematic review of theoretical studies on bullying and cyberbullying: Facts, knowledge, prevention, and intervention. *Aggression and violent behavior* 23 (2015), 1–21. https://doi.org/10.1016/j.avb.2015.10.001

[151] Izabela Zych, Rosario Ortega-Ruiz, and Inmaculada Marín-López. 2016. Cyberbullying: a systematic review of research, its prevalence and assessment issues in Spanish studies. *Psicología Educativa* 22, 1 (2016), 5–18. https://doi.org/10.1016/j.pse.2016.03.002

Appendix A Multivariate regression model

Table 7: Unstandardized estimates for multivariate regression model (M4) examining the relationship between online risky messages (e.g., sexual messages, harassment, and self-injury) and individual Self-Harm and Injury Behaviors (ISAS), including Cutting (ISAS1), Severe Scratching (ISAS2), Biting (ISAS3)

	Cutting (ISAS1)			Severe Scratching (ISAS2)			Biting (ISAS3)		
Predictors	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*
(Intercept)	1.116	0.791 - 1.441	<0.00***	1.395	1.088 - 1.702	<0.00***	1.267	0.965 - 1.568	<0.00***
Sexual Messages	0.016	-0.001 - 0.035	0.072	0.010	-0.006 - 0.027	0.235	0.002	-0.014 - 0.019	0.784
Harassment	0.004	-0.010 - 0.018	0.558	-0.004	-0.018 - 0.009	0.529	-0.000	-0.013 - 0.013	0.952
Self-Injury	0.194	-0.082 - 0.472	0.167	0.109	-0.151 - 0.371	0.408	-0.104	-0.361 - 0.153	0.425
Sexual Identity (LGBTQ+)	**0.836**	**0.345 - 1.327**	**0.000*** **	**0.591**	**0.129 - 1.054**	**0.012***	**0.622**	**1.666 - 1.077**	**0.007*** **
Sexual Messages x Sexual Identity (LGBTQ+)	-0.013	-0.037 - 0.010	0.268	-0.008	-0.031 - 0.014	0.449	-0.002	-0.024 - 0.020	0.841
Harassment x Sexual Identity (LGBTQ+)	**0.031**	**0.001 - 0.062**	**0.040***	0.018	-0.010 - 0.047	0.200	0.003	-0.024 - 0.031	0.816
Self-injury x Sexual Identity (LGBTQ+)	-0.019	-0.316 - 0.277	0.896	-0.112	-0.392 - 0.167	0.426	0.127	-0.148 - 0.402	0.363
Observations	173			173			173		
R^2 / Adjusted R^2	0.263 / 0.232			0.101 / 0.063			0.105 / 0.068		

Note. *p < .05; **p<.01; ***p<.001

Table 8: Unstandardized estimates for M4 examining the relationship between online risky messages (e.g., sexual messages, harassment, and self-injury) and individual Self-Harm and Injury Behaviors (ISAS), including Banging or Hitting Self (ISAS4), Burning (ISAS5), Interfering with Wound Healing (ISAS6)

	Banging or Hitting Self (ISAS4)			Burning (ISAS5)			Interfering w/ Wound Healing (ISAS6)		
Predictors	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*
(Intercept)	1.319	1.005 - 1.633	<0.00***	1.055	0.856 - 1.254	<0.00***	2.063	1.621 - 2.505	<0.00***
Sexual Messages	0.015	-0.002 - 0.033	0.086	0.003	-0.007 - 0.014	0.508	0.013	-0.011 - 0.038	0.282
Harassment	-0.004	-0.018 - 0.009	0.522	0.002	-0.006 - 0.011	0.588	-0.002	-0.021 - 0.017	0.832
Self-Injury	0.090	-0.177 - 0.358	0.506	-0.029	-0.199 - 0.139	0.729	-0.123	-0.500 - 0.253	0.517
Sexual Identity (LGBTQ+)	**0.697**	**0.223 - 1.170**	**0.004** **	0.201	-0.098 - 0.501	0.187	0.368	-0.298 - 1.035	0.277
Sexual Messages x Sexual Identity (LGBTQ+)	-0.016	-0.039 - 0.007	0.170	0.002	-0.012 - 0.016	0.778	-0.013	-0.045 - 0.019	0.427
Harassment x Sexual Identity (LGBTQ+)	0.021	-0.007 - 0.050	0.152	0.006	-0.011 - 0.025	0.468	0.0325	-0.008 - 0.073	0.121
Self-injury x Sexual Identity (LGBTQ+)	-0.042	-0.329 - 0.244	0.769	0.053	-0.128 - 0.234	0.563	0.275	-0.127 - 0.678	0.179
Observations	173			173			173		
R^2 / Adjusted R^2	0.129 / 0.092			0.078 / 0.039			0.090 / 0.025		

Note. *p < .05; **p<.01; ***p<.001

Table 9: Unstandardized estimates for M4 examining the relationship between online risky messages (e.g., sexual messages, harassment, and self-injury) and individual Self-Harm and Injury Behaviors (ISAS), including Carving (ISAS7), Rubbing Skin against Rough Surface (ISAS8), Pinching (ISAS9)

	Carving (ISAS7)			**Rubbing Skin Against Rough Surface (ISAS8)**			**Pinching (ISAS9**		
Predictors	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*
(Intercept)	1.075	0.881 - 1.269	<0.00***	1.364	1.056 - 1.671	<0.00***	1.305	1.008 - 1.602	<0.00***
Sexual Messages	**0.113**	**0.000 - 0.022**	**0.048***	0.008	-0.008 - 0.025	0.345	0.008	-0.007 - 0.025	0.304
Harassment	-0.003	-0.012 - 0.004	0.393	-0.007	-0.213 - 0.006	0.273	-0.002	-0.016 - 0.010	0.658
Self-Injury	-0.014	-0.180 - 0.150	0.860	-0.141	-0.403 - 0.120	0.289	0.009	-0.243 - 0.262	0.940
Sexual Identity (LGBTQ+)	-0.076	-0.368 - 0.216	0.608	0.055	0.408 - 0.518	0.815	0.244	-0.203 - 0.691	0.283
Sexual Messages x Sexual Identity (LGBTQ+)	-0.006	-0.020 - 0.008	0.410	-0.004	-0.027 - 0.018	0.686	0.000	-0.021 - 0.022	0.932
Harassment x Sexual Identity (LGBTQ+)	**0.023**	**0.005 - 0.041**	**0.012***	**0.034**	**0.005 - 0.062**	**0.019***	0.033	0.006 - 0.061	0.017
Self-injury x Sexual Identity (LGBTQ+)	0.080	-0.096 - 0.257	0.372	0.131	-0.149 - 0.411	0.356	-0.311	-0.311 - 0.230	0.767
Observations		173			173			173	
R^2 / Adjusted R^2		0.093 / 0.055			0.070 / 0.030			0.122 / 0.084	

Note. *p < .05; **p<.01; ***p<.001

Table 10: Unstandardized estimates for M4 examining the relationship between online risky messages (e.g., sexual messages, harassment, and self-injury) and individual Self-Harm and Injury Behaviors (ISAS), including Sticking Self with Needles (ISAS10), Pulling Hair (ISAS11), Swallowing Dangerous Substances (ISAS12)

	Sticking Self w/ Needles (ISAS10)			**Pulling Hair (ISAS11)**			**Swallowing Dangerous Substances (ISAS12)**		
Predictors	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*	*Estimates*	*Std. CI*	*p*
(Intercept)	1.055	0.827 - 1.283	<0.00***	1.366	1.038 - 1.693	<0.00***	1.023	0.874 - 1.173	<0.00***
Sexual Messages	0.011	-0.000 - 0.024	0.067	**0.020**	**0.002 - 0.039**	**0.026***	0.001	-0.007 - 0.009	0.770
Harassment	0.000	-0.009 - 0.010	0.970	0.003	-0.011 - 0.017	0.664	0.007	0.000 - 0.013	0.034*
Self-Injury	-0.039	-0.234 - 0.154	0.685	-0.066	-0.345 - 0.212	0.636	-0.019	-0.147 - 0.107	0.758
Sexual Identity (LGBTQ+)	-0.055	0.399 - 0.288	0.751	0.281	-0.211 - 0.775	0.261	0.180	-0.0448 - 0.406	0.116
Sexual Messages x Sexual Identity (LGBTQ+)	0.005	-0.011 - 0.022	0.505	-0.017	-0.041 - 0.007	0.164	0.000	-0.010 - 0.011	0.960
Harassment x Sexual Identity (LGBTQ+)	0.019	0.001 - 0.041	0.067	0.020	-0.009 - 0.051	0.182	0.001	-0.012 - 0.015	0.815
Self-injury x Sexual Identity (LGBTQ+)	0.084	-0.123 - 0.293	0.422	0.119	-0.179 - 0.417	0.432	0.048	-0.088 - 0.184	0.488
Observations		173			173			173	
R^2 / Adjusted R^2		0.137 / 0.100			0.084 / 0.039			0.094 / 0.056	

Note. *p < .05; **p<.01; ***p<.001