"I'm gonna KMS:" From Imminent Risk to Youth Joking about Suicide and Self-Harm via Social Media

Naima Samreen Ali
Vanderbilt University
Nashville, USA
naima.samreen.ali@vanderbilt.edu

Sarvech Qadir
Vanderbilt University
Nashville, USA
sarvech.qadir@vanderbilt.edu

Ashwaq Alsoubai
Vanderbilt University
Nashville, USA
ashwaq.alsoubai@vanderbilt.edu

Munmun De Choudhury
Georgia Institute of Technology
Atlanta, USA
mchoudhu@cc.gatech.edu

Afsaneh Razi
Drexel University
Philadelphia, USA
afsaneh.razi@drexel.edu

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

ABSTRACT

Recent increases in self-harm and suicide rates among youth have coincided with prevalent social media use; therefore, making these sensitive topics of critical importance to the HCI research community. We analyzed 1,224 direct message conversations (DMs) from 151 young Instagram users (ages 13-21), who engaged in private conversations using self-harm and suicide-related language. We found that youth discussed their personal experiences, including imminent thoughts of suicide and/or self-harm, as well as their past attempts and recovery. They gossiped about others, including complaining about triggering content and coercive threats of self-harm and suicide but also tried to intervene when a friend was in danger. Most of the conversations involved suicide or self-harm language that did not indicate the intent to harm but instead used hyperbolical language or humor. Our results shed light on youth perceptions, norms, and experiences of self-harm and suicide to inform future efforts towards risk detection and prevention.

Content Warning: This paper discusses the sensitive topics of self-harm and suicide. Reader discretion is advised.

CCS CONCEPTS

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

KEYWORDS

Youth, Social Media, Self-harm, Suicide

ACM Reference Format:
Naima Samreen Ali, Sarvech Qadir, Ashwaq Alsoubai, Munmun De Choudhury, Afsaneh Razi, and Pamela J. Wisniewski. 2024. "I'm gonna KMS:" From Imminent Risk to Youth Joking about Suicide and Self-Harm via Social Media. 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, 18 pages. https://doi.org/10.1145/3613904.3642489

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

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642489

1 INTRODUCTION

Instagram is one of the most popular social media sites among youth [89], but it has recently gained scrutiny after several high-profile cases of youth suicide were associated with self-harm content viewed on the platform [68]. Although engaging with self-harm or suicide content online provides the youth with an opportunity to connect with others having similar experiences [33], research also has shown that exposure to such content may increase the risk of engaging in these behaviors (i.e., suicide contagion [49]) or normalizing harmful behavior [61]. Given that suicide is a growing public health concern and has been ranked the second leading cause of death among youth (10-24) in the United States [24, 45], these concerning trends warrant more rigorous analyses to capture the nuances when youth discuss these sensitive topics on social media. As such, suicide and digital manifestations of self-harm have both become phenomena of interest to the Human-Computer Interaction (HCI) research community (c.f., [30, 47, 56, 63, 67]).

Many scholars have studied how social media can lead to contagion effects, increasing offline risks of suicide and self-harm [1, 18, 59]; therefore, a priority of the HCI community has been to detect online indicators of mental health issues [50, 52, 87] to mitigate harm. Although these efforts provide valuable insights, the exclusive reliance on publicly available digital trace data or self-reported survey data limits our general understanding of how youth engage in conversations around suicide and self-harm with their peers in private contexts. For instance, survey-based studies are prone to recall biases [35, 71], representing a methodological constraint in our understanding of how these mental health issues manifest in private contexts, which often serves as the arena for some of the most severe instances of online risks [90]. Prior research has pointed to substantial distinctions between the characteristics of public versus private social media interactions [79], particularly among youth [13, 58]. For instance, the posting behaviors of young people have been observed to have a notable change, with their private posts primarily reflecting their emotions and personal affairs, while their public posts adopt a more neutral tone [77]. Recently, HCI researchers have delved into understanding the nature of unsafe interactions among youth within their Instagram *private* conversations (e.g., [5, 6, 47, 70]). These studies enhance our understanding of youths' personal experiences across different risk contexts, such as sexual risk [70], risky images [5], and seeking

support [47]. While these studies lay crucial groundwork for understanding unsafe private conversations among youth to improve online safety, there is still a dire need to study how self-harm or suicide discussions unravel in private conversations of youth while also highlighting the importance of understanding their context to better understand the discussions' goals. Therefore, we pose the following high-level research questions:

- **RQ1:** *What types of conversations do youth have with others regarding their personal experiences with self-harm and/or suicide? And, how do their conversation partners typically respond?*
- **RQ2:** *How do youth discuss self-harm and/or suicide with others when talking about a third-party?*
- **RQ3:** *In what other ways do youth use self-harm and/or suicide language in their private social media conversations?*

To answer these RQs, we utilized an ecologically valid private dataset carefully collected as part of the Instagram data donation (IGDD) study by Razi et al. [71], which contained Direct Messages (i.e., DMs or private conversations) from youth (aged 13-21). We qualitatively analyzed 2,019 unique sub-conversations in 1,224 direct message conversations from 151 participants. We conducted a thematic qualitative analysis [23] to identify the prominent themes that emerged from the data. Our research is the first to study online discussions around suicide or self-harm among youth in their *private conversations* on Instagram.

Overall, we found that youth disclosed their current urges of self-harm and/or suicide that mostly posed an imminent risk while also sharing experiences of being bullied to kill themselves and threatened with suicide by close relationships (RQ1). At the same time, close friends and acquaintances reached out to check on the well-being of at-risk youth. Moreover, youth confided and reminisced about their past self-harm and suicidal experiences, at times, bearing celebratory announcements of successful recovery from such behaviors to inspire others. Youth also gossiped about others, forming judgments about their intentions and exhibiting frustration over them sharing triggering content (RQ2). Conversely, they also devised intervention plans to support their at-risk friends and prevent tragic events from happening. A huge majority of youth discussions showed casual usage of hyperbolic suicide language to convey extreme emotions ranging from joy to stress (RQ3). Furthermore, they engaged in general discussions about suicide and/or self-harm mostly to raise awareness for prevention as well as to share thoughts on media content focused on such topics. In a light-hearted manner, youth engaged in suicide and/or self-harm discourse humorously. Based on these findings, we make the following novel contributions to the literature on youth discourse around suicide and self-harm based on their Instagram DMs:

- Evidence showing that youth openly discuss self-harm and/or suicide with their peers in ways that can be both beneficial and detrimental to their mental and/or physical well-being.
- Insights on how youth rely heavily on peer-support when struggling with self-harm and suicidal thoughts, leading to evidence-based recommendations for supporting these positive support behaviors, rather than overly policing the private conversations of youth.
- Concerns around the prevalence of violent language and humor in the private conversations of youth that may lead to desensitization and unintended harm to vulnerable youth populations. These findings also raise caution for computational research to consider false positives in automated risk detection for youth suicide and self-harm prevention.
- Sheds light on leveraging youths' perceptions and experiences of self-harm and suicide to inform and support future interventions regarding early prediction and prevention of suicidal risk among youth.

Next, we cover the wealth of knowledge garnered through HCI and interdisciplinary research regarding self-harm and suicide, as it relates to youth and their social media use. Due to the sensitive nature of the topics described in this paper, we suggest the reader's discretion as well as self-care strategies for coping.

2 RELATED WORK

Over the past decade, significant interest has grown in studying youth self-harm and suicidal tendencies through their social media usage [30, 47, 56, 63, 67]. There are three primary lines of research: the role of social media in discussions of self-harm and suicide, studies that focus on understanding how young people seek support online regarding these topics, and studies that involve the prediction of self-harm and suicide for risk mitigation. In this section we synthesize the literature on these thrusts, situating the contributions made in this paper.

2.1 The Role of Social Media in Discussions of Self-Harm and Suicide

Social media has changed youth communication patterns from offline forms of discourse with its increased accessibility [89] and sheer scale [85], while also allowing for disinhibition [53] that enables youth to express themselves more openly or engage in behaviors they might avoid in face-to-face interactions. A bulk of prior research has concentrated on *describing* youth behaviors in disclosing their self-harm and suicide ideation *publicly* on social media, ranging from uncovering what youth discuss in such postings online, to how social media signals could be indicative of self-harm and suicidal behaviors [59, 65, 75, 91]. Foundational research by Patchin et al. [65] and Pater et al. [66] revealed that anonymous posting, sending, or sharing of hurtful content on the internet is frequent among adolescents and may be related to higher depressive levels and initial suicide ideation [65, 66]. Many studies [15, 34, 54, 59] also discuss language, symbolism, and conversational patterns in self-harm discussions on social media platforms. Moreover, studies have also [75, 91] explored self-harm and suicidal conversations by youth and their associated affiliations of whom they choose to converse about it and revealed that self-harm discussions are likely to occur with close friends, intimate romantic partners, and siblings and less likely to occur with acquaintances. While this literature identified the key themes, patterns, and ideas on how youth interact with suicide and self-injury in social media, these approaches provide limited insights into how youth talk about its context, particularly in private discourse, where increasingly greater numbers of youth are communicating [96]. Hence, our study bridges this gap by studying how youth talk about and

seek support regarding suicide and self-injury topics in private conversations. We further investigate various contexts of youth discussions regarding suicidal topics disclosed in private discourse.

Complementary to the above-reviewed research, various interdisciplinary efforts have attempted to look at causal relationships between social media use and self-harm/suicide. These causal investigations have centered on the assumption and the observation that social media use can have negative impacts on people's well-being (see Hancock et al. [42] for a meta-analysis of this relationship), and even has the capacity to create social contagion among its most vulnerable users [18, 59]. In this context, it is defined as the impact of social media content on self-harm and suicide on individuals [62]. Cataldo et al. suggested that exposure to suicide and NSSI content on Instagram could lead to a comparison of self-harm, with users competing to exhibit the most injuries [18]. In fact, Brown et al. [15] highlighted that adolescents with a history of NSSI are more active on social media than adolescents with no NSSI history. Apart from that, social media content such as internet suicide pacts, NSSI posts, and "extreme" communities seem to increase suicidal behavior among youth [55]. Although the above constitutes an important research direction to better inform platform design and outline policy for social media corporations' obligation in protecting youth mental health, such efforts require richer examinations of how and what youth discuss when they share their own or others' self-harm and suicide-related experiences. Causal investigations need to be informed by the myriad ways today's youth engage online, particularly on private channels and specifically on sensitive topics. This paper aims to fill this gap.

2.2 Youth's Online Support Seeking and Provisioning for Self-Harm and Suicide

Not all investigations of the role of social media on youth mental health have been to uncover the former's negative effects. There have been both qualitative and quantitative approaches to studying how social media channels cater to the support-seeking and provisioning needs of youth around the topics of self-harm and suicide [28, 29, 56]. From research efforts analyzing textual data from public online forums and social media for self-injury and suicidal ideation [54, 92], while triggering, such discourse can have benefits. Youth use social media platforms to share any emotional distress related to suicide ideation in search of receiving emotional support [41] and to have an outlet to talk about and process their struggles with self-harm [74]. Manikonda et al. [56] further employed computer vision techniques on Instagram public posts, revealing unique self-disclosures through imagery for emotional distress, calls for help, and explicit displays of vulnerability, differing from textual disclosures. Therefore, these studies collectively highlight the significant role of social media platforms for youth to share emotional distress, suicide ideation, and self-harm, emphasizing their importance in fulfilling unique self-disclosure needs.

At the same time, disclosing such sensitive topics online is uniquely challenging. Therefore, to support these positive uses of social media, researchers have argued that ensuring privacy and anonymity holds key significance in ensuring youth's disclosures of and engagement with self-injury and suicidal discussions online [83, 93]. A qualitative study by Scourfield et al. [80] learned that talking about self-harm is a private and personal association for each individual and they may struggle to communicate about it [80]. Zhang et al. [98] and Robinson et al. [75] investigated self-harm and suicide-related posts and highlighted how youth often turn to social media platforms as supportive environments to share their distress. However, prior work establishing how respondents of self-harm and suicide messages have responded is limited. A study by Marchant et al. [57] examined how youth interact with their peers concerning suicide and self-harm. Their findings revealed that youth are comfortable engaging in private conversations regarding self-harm with their peers since they can receive emotional support to help them feel less isolated. This paper builds on this body of work, examining the conversations on self-harm and suicide that youth engage in on private channels, and how they communicate their own lived experiences or others in this process.

2.3 Towards Prediction of Self-Harm and Suicide for Risk Mitigation

Taking a step further, other researchers have recognized the importance of automated *prediction* of self-harm and suicidal behaviors from social media that could inform technological interventions. Thus, significant strides have been made by the HCI scholars in addressing the pressing issue of detecting mental health problems, such as suicide and self-harm, through automated approaches [17, 26, 30, 36, 44, 63, 64, 88]; see Chancellor and De Choudhury [20] for a review. A prominent study by De Choudhury et al. [30] analyzed Reddit posts in mental health forums to identify linguistic cues linked to an increased likelihood of future discussions about suicidal ideation on the platform. Specific to Instagram – the platform of interest in this paper – Brown et al. [15] used regression analyses to examine predictors for current suicidal ideation in adolescents who had posted non-suicidal self-injury pictures on the platform, revealing that a majority had encountered active suicidal thoughts on Instagram, with 25% expressing such thoughts themselves. Huh-Yoo et al. [47] also looked into how youth seek help in their Instagram Direct Messages and found that conversations usually start casually among friends and gradually develop into discussing negative instances in topics related to everyday stress to severe mental health disclosures, including suicide. These works stand as a valuable milestone for they not only have created a viable approach to risk detection but have also illustrated the complex ways in which these behaviors manifest in social media.

However, users' language in social media especially young individuals has posed a longstanding difficulty for the detection algorithms, mainly because this language often does not adhere to typical patterns or regulations [2, 21]. Hence, further qualitative research is necessary, especially concerning youth's self-harm and suicide expressions on social media, to ensure that automated methods incorporate a nuanced human perspective, minimizing unintended harm [31, 81]. This paper will bridge this gap through an in-depth thematic analysis of youths' private conversations on Instagram, aiming to comprehensively grasp the distinctive challenges and expressions of youth regarding self-harm and suicide, encompassing both personal and referenced experiences.

3 METHODS

In this section, we describe the dataset we analyzed, our scoping process, and our qualitative analysis approach. Finally, we provide an overview of the descriptive characteristics of our sample conversations and participants.

3.1 Instagram Data Collection and Scoping Process

3.1.1 Instagram Data Donation Project. We analyzed data that was originally gathered through a youth Instagram data donation (IGDD) project by Razi et al. [71], which was approved by the original authors' Institutional Review Board (IRB). The data was collected from 2020 until mid-2022. The dataset consisted of Instagram Direct Messages (DMs) in 32,055 conversations from 195 participants ages 13-21, who were English speakers based in the U.S., had an active Instagram account for at least 3 months when they were a teen (ages 13-17), exchanged direct messages with at least 15 people, and had at least 2 direct messages conversations that made them or someone else feel uncomfortable or unsafe. The participants were asked to download their Instagram data (with parental consent), upload it to a secure web-based system, and flag their private conversations as safe or unsafe. The participants were asked to provide further details about each conversation marked as unsafe by labeling at least one unsafe message for risk types (i.e. self-injury, sexual solicitations, harassment, etc.) derived from Instagram reporting feature risk categories [1]. Also, six undergraduate research assistants (RAs) were recruited to label unsafe DMs and the risk type in those messages. Besides that Instagram is known for having a predominantly younger user base, the availability of their Instagram private conversation data from Razi et al. [71] work was pivotal for us as researchers, given the ethical and legal challenges associated with obtaining such a sensitive dataset [72].

3.1.2 Dataset Scoping and Relevancy Coding. We leveraged a multi-level scoping process. First, we identified conversations marked as self-injury by the participants themselves (n = 12) and the 394 conversations labeled as self-injury by the research assistants. Then, we leveraged previous literature [33, 69] to search for common search terms indicative of suicide or self-harm. We also looked up urban dictionaries [86] for common slang words or secretive language [2] that teens use when they discuss such topics. By initially examining these conversations, we generated a list of additional relevant keywords based on the conversations. For instance, common vernacular used by youth included "kms" (short for "kill myself" or "kill me slowly") and "kys" (short for "kill yourself"). Our initial query retrieved 4,605 conversations based on these keywords of which some terms got zero hits and others appeared irrelevant to the scope of our research; therefore, they were removed (see Table 1). Combining the results of our final query with the risk-flagged data from participants and annotators, we identified 5,001 conversations to code for relevancy. Conversations were coded for relevancy by the first two co-authors, and consequently, this process helped us scope our three high-level research questions. The conversations were deemed relevant based on the following inclusion criteria:

1. The conversation contained self-harm or suicide references.
2. The conversation involved the usage of self-harm language, including any form of utterances where members inflicted some sort of harm upon themselves or used the language metaphorically or ironically.

First, 10% of the conversations (n = 500) were reviewed by both coders, and inter-rater reliability (IRR) was calculated which indicated a significant level of agreement (Cohen's Kappa = 0.76). The two coders met to resolve conflicts and then divided the rest of the conversations among themselves. When in doubt, they consulted with one another and the last author to reach a consensus. In general, conversations were coded as irrelevant for alternative use of keywords (e.g., "*I cut open the package*") or due to word stems (e.g., die within "*buddies*"). After relevancy coding, 1,224 conversations remained for analysis.

In our study, a conversation is defined as the entire history of messages exchanged between participants (either one-on-one or group chat) on a digital platform, which may span from a few minutes to several days, or years. A conversation does not necessarily have to adhere to a coherent thread of communication and can have several discussion topics. Each conversation was parsed into sub-conversations to prepare for data analysis. A 'sub-conversation' was defined as a singular conversation, where a self-harm or suicide discussion was initiated and concluded, which was identified based on four components: 1) initiation of the self-harm and/or suicide discussion, 2) the context of the disclosure, 3) the responses of the conversation partners, and 4) the conclusion of the conversation, often indicated by a switch in the subject and/or passing of time (given timestamps of messages were available). Oftentimes, a single conversation contained multiple sub-conversations, indicating a history of these types of discussions that unfolded over time. We identified 2,019 sub-conversations, which were used as our unit of analysis for our qualitative results. Figure 1 further illustrates our data scoping and relevancy coding process.

3.2 Qualitative Data Analysis Approach

To answer our RQs, we leveraged the qualitative thematic analysis approach of Braun and Clark [14] to identify the key aspects of youths' discussions around self-harm and/or suicide. First, we familiarized ourselves with the data when coding the conversations based on our inclusion criteria for relevancy. This process allowed us to create initial codes and formulate our high-level research questions. Then, the first two authors contextualized the data by identifying whether sub-conversations involved suicide, self-harm, or both. Then, they open-coded a subset of the sub-conversations to gain initial insights and create the initial code book for analysis. They reviewed these codes with the last author and the larger research team to form a consensus on how to code the remaining sub-conversations. Next, they coded the sub-conversations based on whether the topics of conversation were centered around the youth or others: 1) the personal experiences of the participant or their conversation partners, 2) the experiences of people other than the participant and their conversation partners, or 3) any other self-harm and/or suicide-related topics. The coders convened regularly

[1] https://www.facebook.com/help/instagram/192435014247952
[2] https://socialmediavictims.org/resources-for-parents/text-slang-emoji-dictionary/

Table 1: Keywords for the Search Query

Search Results	Keywords
Relevant	die, kill, cut, burn suicide, self harm, kms, kys, slit, shooting, pain, pills, scarred, scratch, punching walls, punch yourself, punch myself
No hits	syringe, needle, pinch, pinching, substance, head banging, scarring, starve, SUE, SVV, SUE, Secretsociety123, Unalive, bonspo and thinspo, IHML, C U T, selfharn, eating disorder
Coded as irrelevant	Burn, wound, bruise, beat, blade, bite, biting, inject, pierce

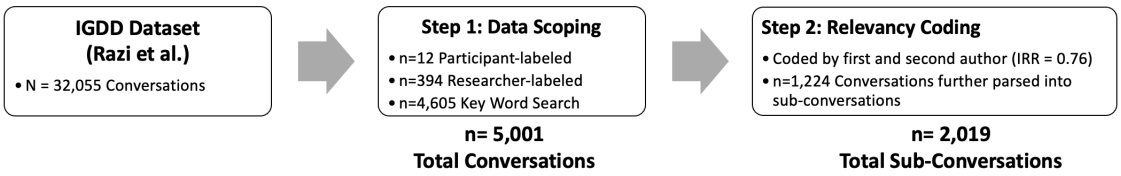

Figure 1: Scoping and Relevancy Coding Process

to iterate, refine, and finalize their codes with constant guidance from the last author. Once coding was completed, the research team worked together to group the codes conceptually into the themes presented in this paper.

For RQ1, we identified two main themes when youth engaged in discussions related to their personal experiences: 1) Disclosing current and/or future thoughts of self-harm or suicide and 2) Past experiences with self-harm or suicide. We also further examined how youth discussed experiences related to others (RQ2), which included themes related to gossip, judgment, and intervention. For RQ3, we saw suicide and self-harm language hyperbolically, and general discussions on such topics. Table 2 represents our final codebook with our main themes and codes, along with illustrative quotations. Initially, we allowed for double-coding, but as we refined how the codes were operationalized, they ended up being mutually exclusive. For two-way conversations, 'P' refers to the primary participant, whose conversation is being analyzed. 'O' denotes the other individual involved in the exchange. For group conversations, we appended numbers 'O1', 'O2', 'O3', etc., to denote the other individuals participating in the group conversation. The percentages associated with each RQ were calculated based on the total number of sub-conversations coded for a given RQ.

3.3 Conversation Characteristics and Participant Demographics

Our final dataset consisted of 1,224 conversations with 2,019 sub-conversations from 151 unique participants (compared to the 195 total participants in the dataset). There were an average of 8 conversations (min = 1, max = 84 per participant), and 2 sub-conversations per conversation (min = 1, max = 15). Of the 1,224 conversations, 870 (71%) were participants and one other conversation partner, while 354 (28.9 %) were group conversations. The majority of sub-conversations were related to suicide (91%, n = 1,831), rather than self-harm (7%, n = 140). This was often due to the use of humor (e.g., "*This homeworks makes me want to kms. LOL.*") and hyperbolic language (e.g., "*It's so hot, I'm dying!*") around death. Of all the participants, 40 users had conversations that entirely contained suicide and self-harm language hyperbole, and humor. 13 users contained conversations with suicidal and self-harming content, while 98 users contained references to both. This distribution indicates that the majority of our participants used hyperbolic language for self-harm and suicide conversations, while fewer talked about it directly. Even fewer participants had conversations indicative of imminent risk. Around 7% (n = 142) of the sub-conversations were about self-harm with cutting as the most prominent mode of engaging. There were a few sub-conversations (2%, n = 46) where youth engaged in both self-harm and suicide-related discussions. Overall, we found that 23% (n = 460) of the sub-conversations were about youth sharing their personal experiences (RQ1), while 7% (n = 142) were about them discussing suicide and/or self-harm experiences of others (RQ2). Yet, the largest number of sub-conversations (70%, n = 1,417) was about using suicide and self-harm language (RQ3).

Of the 151 youth, their ages ranged between 13 and 21 with a mean of 17.35 years old and standard deviation of 2.11 years. Most identified as female (71%), followed by male (20%), and non-binary youth or preferred not to self-identify (9%). Among our dataset participants, 49% (n = 74) participants identified as heterosexual, whereas the rest 51% (n=77) were LGBTQ+, out of which 47 participants identified themselves as bisexual, 14 as homosexual/gay, and 16 of them preferred to self-identify. The breakdown of participant race was as follows: 56% (n=84) identified as White/Caucasian, 25% (n=38) identified as Black/African-American, 22% (n=33) identified as Asian/Pacific Islander, 17% (n=26) identified as Hispanic/Latino, 4% (n=6) identified as American Indian/Alaska Native, 3% (n=5) preferred to self-identify.

3.4 Ethical Considerations

To ensure ethical treatment of the data entrusted to us, we received Institutional Review Board (IRB) approval from the last author's institution. All research team members were required to complete IRB CITI training [40], which consists of courses on various topics related to research ethics, including the protection of human

subjects in research. Moreover, they completed the Protection of Minors (POM) Training, as well as a comprehensive background check, before accessing the dataset. Per the General Data Protection Regulation (GDPR) [37], Instagram is required to ensure the data portability of users by giving them the right to download their data and share it with third parties without restriction. This includes direct message conversations that are co-owned by other individuals. Given this legal framing, participants are free to donate their data by laws that do not assume an expectation of privacy within social media DMs. Still, we took the utmost care when handling both participant and their conversation partners' messages. In addition to removing any personally identifiable information from quotations presented in this paper, we also at times paraphrased as an extra precaution to ensure confidentiality. Further, we performed qualitative analyses on university-approved and secured shared storage, not allowing team members to upload any data to the cloud or to their individual computers. Since the dataset included disturbing self-injury and suicidal content, the research team had frequent meetings to discuss any concerns along the way. We also developed a comprehensive risk mitigation, child abuse mandated, and imminent risk reporting protocol with our university's risk compliance and legal counsel teams. All conversations that indicated any child abuse or imminent risk to a minor were reviewed based on this protocol and necessary precautions were taken.

4 RESULTS

In this section, we summarize the themes (Table 2) that emerged during analysis for answering each of our research questions.

4.1 Youth Discussing Personal Experiences with Self-harm and/or Suicide (RQ1)

Table 2 served as the overarching structure for reporting our findings along with illustrative quotations with the youth participant's age and gender. Youth used their Instagram DMs (23%, n = 460 out of 2,019 sub-conversations) to disclose and share their personal experiences of self-harm and/or suicide, which most often included current or future thoughts (84%, n = 386 out of 460), as well as their past experiences (16%, n = 74 out of 460). Below, we describe the patterns observed in these conversations.

4.1.1 Current or Future Thoughts of Suicide or Self-Harm. Within the conversations (84%, n = 386) where youth discussed their current thoughts or actions related to suicide or self-harm, they: 1) disclosed ***imminent risks*** (53%, n = 203 out of 386) posing danger of acting upon these thoughts, 2) shared about instances of ***blackmail and coercive peer-pressure*** (36%, n = 140 out of 386), and 3) ***checked-in*** (11 %, n = 43 out of 386) with one another due to known past tendencies towards self-harm.

Most concerning, youth shared situations that posed ***imminent danger/risk*** to their safety and well-being. This included telling their peers that they recently cut themselves or saying that they planned to die by suicide. Such disclosures were often coupled with sharing their struggles that led to the crisis situation, including mental health challenges ranging from anxiety and depression to bipolar and eating disorders (e.g. *"that's my fucking life. what i have to deal with on a weekly basis. feeling like im the hottest shit ever one day then completely wanting to die the next."* - 17-year-old, female). The youth shared troubles related to relationships with their partners, friends, or family which left them with a feeling of being unloveable and worthless (e.g. *"just i got cheated on and i cut myself the night i found out which was last Saturday"* - 14-year-old, male). We also noticed some youths struggling with gender dysmorphia and mentioned wanting to cut parts of their bodies to feel more aligned with their gender identity (e.g. *"Idk*[3] *any other girl like me with period dysphoria...I've cut myself down there, so I could bleed."* - 15-year-old, female). Disclosures of suicidal thoughts often emphasized the immense emotional pain and desperation to find a way out. In these cases, conversation partners often gave emotional support by showing empathy and understanding to the best of their abilities. However, oftentimes, the peer that they confided in shared many of the same struggles:

> ***P:*** *I hallucinate from all the depression...Voices in my head telling me to kill myself...(18-year-old, female)*
> ***O:*** *Don't listen to those voices.. and I doubt I can provide much consolation, as I'm also depressed.*

Youth often explained to their conversation partners that cutting themselves was a means to calm their nerves, and they asked their confidants to be gentle and not judgmental, often begging them to not be mad at them. In turn, conversation partners often eased their concerns and provided positive and affirming support, while trying to deescalate additional harm from occurring:

> ***P****: It's not very good please don't get mad. It's how I calm myself. It helps.(15-year-old, female)*
> ***O****: Oh, I am not mad. What matters to me right now is that you're feeling better...Though, is it okay if I request that we attempt to find a safer method in the future? There are endless ways, we can find one!*

Conversation partners often begged them to stop or gave practical advice on safety measures (e.g. *"My advice is to throw away razors, I know it's kinda difficult but it's what made me eventually stop"* - 13-year-old, non-binary youth). Sometimes peers also suggested talking to parents or therapists for support. However, oftentimes, the youth did not want their parents to know about their self-harming behaviors and even sought advice for how to hide them (e.g. *"My mom wants me to try on a bra...my cuts haven't healed yet."* - 15-year-old, non-binary youth). Moreover, the youth shared seeing a therapist already and revealed that it had not been effective for long. Youth also shared their relapse with self-harm after a certain period of improvement or recovery. Mostly, they expressed a fight against their instinct to dwell back into cutting as a gateway to alleviating pain and ending their suffering. Also, as they had gotten over it before, they considered having similar urges as something they would not be able to handle anymore. Two youth, who were in a group conversation with one of our participants, shared their joint struggles:

> ***O1****: I feel like I'm going to cut myself again but I don't want to but I have the urge to ugh why does my life have to be so fucked up.*
> ***O2****: I know that cutting is really hard to pass, and I'm struggling myself so I can try my best. When you get urges maybe put some ice on it, use a rubber*

[3] Refer to table 3 for the meanings of common slangs used by youth

band, stress toy, or draw. Any coping mechanism like that is better than actually cutting even if u use ice or a rubber band, well that is in my book anyway.

Sometimes peers threatened to self-harming themselves for the sake of keeping one another strong in their recovery. For instance, a member disclosed wanting to cut themselves again due to some unfortunate life problems triggering their mental health issues and the conversation partner threatened to do the same:

***P**: now im depressed again and i might go back to cutting (15-year-old, non-binary youth)*
***O**: Bitch if you cut yourself I'm cutting myself again and I'll send you how many cuts I have on my arm*
***P**: nooooo dont do thattttt*
***O**: If you don't I won't*

We also found instances of ***bully and blackmail*** regarding self-harm and/or suicide. Such discussions were mostly related to conversation members being bullied by someone outside of the chat to kill themselves. The conversation members were body-shamed, called ugly, or detestable, and encouraged to end their lives (e.g, *"Some dude just told me n ky that we fat n ugly n we should kill ourselves so yea"* - 21-year-old, female). Sometimes, these bullying incidents were committed by people whom they had close relationships with. In these cases, youth often garnered support and empathy from their friends. For example, a conversation member shares being bullied by their ex-boyfriend and how it had a toll on their mental health:

***O**: he would call me a fat whore and a bitch, and would tell me to kill myself, and i stayed with him for six months and cried everyday*
***P**: So gross honestly, Im so sorry (15-year-old, female)*

The bullying victims also shared their experience of actually engaging in self-harm and/or suicidal behavior after receiving hurtful comments that led to them contemplating death. Such instances were often tied with cutting attempts with an intention to die (e.g. *"Lately I've cut myself and have been told that if I killed myself then no one would care and I should just do it"* - 13-year-old, non-binary youth). In such cases, the conversation partners extended support by expressing disgust for the bully, showing sympathy toward the victim, and reassuring them (e.g. *"People WILL care, People DO care"* - 13-year-old,non-binary youth).

In some cases, youth were bullied by their conversation partners. They hurled hateful comments toward youth and bullied them to kill themselves (e.g. *"pls kill urself, u remind me of a fucking rats ass"* - 16-year-old, female). The victims in turn expressed their hurt feelings and retaliated with harsh remarks. (e.g. *"that kinda hurt me tho, DIEEE"* - 14-year-old, male). Sometimes bullying occurred in a joking manner where it wasn't taken seriously and responded with playful remarks (e.g. *"only if you die with me"* - 18-year-old, female). However, we observed that mostly such utterances were taken seriously and frowned upon, in which case, the offending party would usually minimize their behavior by saying they were just kidding or did not mean offense. Similar to the conversation below, victims confronted conversation partners about their self-harm struggles as a way to tell them to back off:

***P**: I tried to cut myself with a knife, You just told me to kill myself (17-year-old, female)*
***O**: I am just saying because I was angry that time, Please don't take serious*

In group conversations, conversation partners who bullied others were often reprimanded for their bad behavior. For instance, a conversation member bullied another conversation member to die and was confronted for this wrongful behavior and reminded of the detrimental consequences of their words:

***O1**: kill yourself, oo you can rid the planet of yourself and you dont make anyone lose brain cells anymore*
***O2**: That's not nice you shouldn't tell ppl to kill themselves bec if one day someone decides to listen to you it will be all your fault and you'll nEver forgive yourself*

Sometimes self-harm and/or suicide threats were used to blackmail the conversation members. Mostly, the youth discussed being threatened with suicide for rejecting someone's romantic advances (e.g. *"I said no to him, And then he threatened me with suicide"* - 17-year-old, female). They also mentioned that sometimes their romantic partners used suicide threats to pressure them into upholding the relationship (e.g., *"I know my partner held me in a relationship by threatening to kill herself if I left..."* - 18-year-old, female). That's the main reason why conversation members were stressed over breaking up with their partners sharing their fear of them resorting to suicide as a result. The conversation partners condemned such behaviors and asked them to not pay heed to such hollow threats:

***O1**: Yo, now he says he's gonna kill himself because I broke up with him*
***O2**: Thats fucked up dont listen to him*

In some cases, conversation member blackmailed their peers with suicide to get them to pay attention to them or do what they wanted them to do. It was often done over trivial things, like asking for homework solutions (e.g. *"send me the answers or i'll kms"* - 15-year-old, male), asking to do something urgently, or intervening to prevent an unwanted action. While these threats seemed playful at times, they were mostly taken seriously, where conversation partners responded with concern and tried to comply with their wishes as best they could:

***O**: imma call u and if y'all don't pick up, imma slit my throat with no hesitation I WONT HESITATE BITCH*
***P**: I cant answer rnn, ill answrr later tho (15 year-old, female)*
***O**: oh so u want me to slit my throat huh?*
***P**: Nooo*

Youth also often initiated the discussion by ***checking-in*** with peers who experience self-harm and/or suicidal urges to provide continued support and keep them from acting upon their thoughts. Most of such discussions occurred with people they knew very well and had shared their struggles with them before. The conversation members reminded them that they were not alone, offered them a listening ear, suggested alternatives, and begged them to never do it. With great dismay but in an attempt to mitigate the situation to some extent, conversation members also often requested their

peers to educate themselves to opt for less harmful ways if they were to continue self-harming activities:

> ***O***: *Um, hi. I know this is weird and out of nowhere BUT If you're gonna keep self-harming can you please educate yourself a bit so you can stay safe? I'm begging*
> ***P***: *I love you (17-year-old, female)*

Sometimes if a peer was on medication to fight a mental illness, conversation members warned them to be wary of overdosing as it can be hazardous (e.g. *"Oh so you use it as a medicine, ok i just dont want u to die"* - 17-year-old, female). We also noticed conversation members threatening to cut ties or being brutal if their friends engaged in any harmful behaviors to avoid such events from happening (e.g. *"I swear if you cut yourself EVER, I will kill you"* - 16-year-old, female). At times when youth learned that a peer was going through a rough patch from their posts, they checked on their well-being and offered emotional support (e.g. *"Talk to me what's wrong amd why do yiu want to kill yourself?...Just remember i acctuly do care although we don't know each other much"* - 15-year-old, female). Mostly, in such cases, they barely knew the person.

4.1.2 Youth confided and reminisced about past experiences with self-harm and/or suicide. In conversations where youth disclosed their past experiences (16%, n = 74 out of 460 sub-conversations) with self-harm and/or suicide, they: 1) ***confided and reminisced*** (69%, n = 51 out of 74) about their past behaviors, and 2) ***celebrated recovery*** (31%, n = 23 out of 74) from past harmful urges. We found youth **confiding or reminiscing** to discuss incidents of self-harm and/or suicide from their past. They mostly confided in friends or romantic partners whom they had not known for long, seeking to strengthen their bond without the fear of being judged. The recipients would offer emotional support by understanding and validating their feelings. In one such instance, a conversation member shared that they hesitated to open up because of unsupportive and harsh reactions:

> ***P***: *I hated everyone and everything in 7th grade. I couldn't open up to people cause people sometimes yelled at me and stuff for it (including some friends) I attempted suicide a couple of times-It was overall a mess (15-year-old, non-binary youth)*
> ***O***: *Omg please...you won't try to kill your self i know how it feels not being able to tell people things*

Youth often discussed the long-term negative consequences of engaging in self-harm and/or suicide in their past. They shared the challenge of hiding their scars from self-harming to avoid embarrassment. Similarly, youth ranted about relationship issues such as being abandoned by their ex-romantic partners because of their struggles with self-harm and/or suicidal urges (e.g. *"my last ex stopped talking to me for a while after I had a relapse in self-harm (was clean for over a year...Don't worry I'm still clean"* - 13-year-old, non-binary youth). As youth confided about their past with relatively new people to foster deeper connections, in parallel, they reminisced and revisited memories of self-harm and/or suicide with close people. They mostly discussed the challenges of forming new friendships or relationships, as sharing their history of self-harm and/or suicidal thoughts often led to others distancing themselves. The peers often recommended withholding details from their past until they have reached a point of complete comfort and trust:

> ***O***: *I've been at such a low point in my life a few years ago I actually cut myself...Every time I tell sombody about my feelings and my past they never see me the same way. I just want people to see me how I am now and not how I used to be*
> ***P***: *Then sometimes it's best not to tell the entire back-story until you feel comfortable (21-year-old, male)*

While going through the distressing narratives of youth, the instances of ***celebrating recovery*** from such harmful behaviors came out as a ray of hope. Mostly, youth announced proudly that they had overcome self-harming. Moreover, they celebrated getting past suicidal thoughts and they felt lighter and more appreciative of life. They mentioned certain projects employed to help with suicide and self-harm which proved effective in their journey of recovery (e.g. *"The semi-colon project was basically a way for people who suffer with sui # ∗ #thoughts, self $ ∗ #∗, etc. to say that "my story isn't over"*, 13-year-old, non-binary youth youth). Their peers celebrated and commended their accomplishments. Youth often posted their success stories in the support group conversations that they had been a part of which helped them through the self-harming (e.g. *"i kept going and then i met you guys and i haven't hurt myself with my blades"* - 18-year-old, female). It was often to encourage and inspire others to stay put and assure that they can overcome it too.

4.2 Self-harm and/or Suicide Discussions Involving Others (RQ2)

We observed that youth (7%, n = 142 out of 2,019 sub-conversations) engaged in the discussions surrounding self-harm and/or suicide incidents of those around them. In these conversations, they shared: 1) ***gossip and judgment*** (65%, n = 93 out of 142) on others' harmful behaviors, and 2) ***intervention*** (35 %, n = 49 out of 142) techniques to prevent them from self-harming. Below we unwind on the various prominent trends we found in such discussions.

4.2.1 Youth gossiped and passed judgments on others' experiences of self-harm and/or suicide. We noticed self-harm and/or suicide instances of others having a social impact on youth that was mostly manifested as sharing ***gossip and judgment*** on their suicide incidents. Youth often engaged in such discussions to inform about the tragic death of their friends due to suicide. The conversation partners showed curiosity in knowing all the details of those instances while also trying to speculate the connectivity of the incidents with mental health issues (e.g. *"How? Where? Was he having PTSD issues?"* - 18-year-old, female). Youth also discussed their friends being admitted to psyche wards or residential for contemplating or attempting suicide. The peers mostly showed support by offering sympathy, while it was also revealed that being in a residential proved to be effective for their friends:

> ***P***: *And she told her therapist she was gonna do it and then kill herself. But the good part is she is doing really really well in residential (15-year-old, female)*
> ***O***: *Aww that's so sad!!! I feel very bad for her. I'm glad she's doing better tho.*

Table 2: CodeBook (SH = Self-harm and S = Suicide)

Themes	Codes	Illustrative Quotations
RQ1: Discussing personal experiences about SH/S (23%, n = 460)		
Current or Future thoughts of SH/S (84%, n = 386)	**Disclosing Imminent Danger/Risk (53%, n = 203)**	*O: I cut myself* *P: Why* *O: I was depressed* *P: Idc. You shouldn't have. How many? (15-year-old, male)* *O: 9*
	Bully and Blackmail/Peer Pressure (36%, n = 140)	*O3: please stfu & go choke* *O2: Yj?* *O1: Yu go die wtf bitch*
	Checking-in (11%, n = 43)	*O: All I ask is please don't kill yourself people care about you I promise it might not seem like it but they do* *P: Thank you (15-year-old, male)*
Past experiences with SH/S (16%, n = 74)	**Confiding or Reminiscing (69%, n = 51)**	*P: Not only that do u know I used to cut myself... trust me wen I say u have no clue (15-year-old, female)*
	Celebrating Recovery (31%, n = 23)	*O1: oh shoot, it's been 3 years since i've self harmed* *O2: Oh HELL YEAH* *O3: FUCK YEAH LOV U !!!!!!!!!!*
RQ2: SH/S Discussion involving others (7%, n = 142)		
Gossip and Judgement (65%, n = 93)	*O: what happened to her?? Both of them?* *P: She tried to kill herself in residential, I don't know exactly what happened And she couldn't tell me yet because she didn't want the staff to overhear (15-year-old, female)*	
Intervention (35%, n = 49)	*P: Hi..xyz is threatening to take a whole bottle of pills, SHE TOOK SEVEN NOW* P: *NINE...IDK, SHE HELP TEN, 'O' HELP, She took twelve* O: *we can't physically stop her* P: *I KNOW* O: *there isn't much we can do..I can call the suicide prevention hotline* P: *Do it (16-year-old, female)*	
RQ3: The hyperbolic and humorous use of SH/S language (70%, n = 1,417)		
Using hyperbole for "killing oneself" as an exaggerated expression (85%, n=1,209)	**as a reaction to frustration (31%, n = 373)**	*O: ABC GOT FUCKING LICE!!!!!* *IM GONNA KILL MYSELF*
	as a response to social stressors (22%, n = 270)	*P: and i'm sitting in the back seat because i don't want to see her eye...i'm gonna throw up just kill me(18-year-old, female)*
	as a response to work stress (20%, n = 228)	*O: School gonna kill me next year...It's gonna be my toughest year*
	as a response to viewed content (15%, n = 180)	*O: Lmaoooo that video is killing me* *P: OH MY GOD HOW ARE YOU SO FINEEEEEE My chest , Thank you (21-year-old, female)*
	as a response to physical discomfort (7%, n = 90)	*O1:Can I just like die...I'm coughing up a lung over here*
	as a reaction to joyousness (5%, n = 68)	*P: I know i mean die from blushing or laughter (15-year-old, female)*
General Discussions about SH/S (11%, n = 158)	**Suicide Prevention/Awareness (61%, n = 96)**	*P: They might threaten the kids along the way to get them to keep going (empty threats sometimes, sometimes not) but it mainly works off of his idea of escalating challenges...it's worked* *O: That's awful!!* *P: It's awful so be careful (18-year-old, female)*
	Media and Culture (39%, n = 62)	*O: But 13 reasons why is good, did u watch it.* *I just hate how she's tallying up all these things and kills herself BC of them... Ppl kill themselves for deeper reasons*
Humor directly about killing or harming oneself (4%, n=50)	**Killing (96%, n=48)**	*P: ok i'll kill myself then smh my head for legal reasons that's a joke (19-year-old, male)* *O: if u wanna kill urself ull have 2 go thru me first*
	Harming (4%, n=2)	*O1:i made a joke and said 'i'd cut off my left arm to see it' and this girl replied 'i'd cut off both my legs to see it again' and i'm just*

While sharing self-harming and/or suicidal behaviors of mutual friends, youth often exchanged judgmental comments terming it as attention-seeking or causing drama (e.g. *"All she does is just cause drama... she constantly talks about suicide and shit and it's just frustrating"* - 17-year-old, female). They mentioned that if people who engage in such activities are not seeking attention only, they are often secretive about it. However, youth also advocated for their friends who are ill-perceived as attention-seekers when in reality they are not and should be given a chance to explain themselves:

> ***P****: ppl Hate him because All he talks is about suicide and stuff He's seeking for attention...He's actually a really nice and funny dude (17-year-old, male)*
> ***O****: He is but he has a bad reputation*

Youth also stressed over and complained about the secondary harm caused by the self-harming experiences shared or posted by others including explicit pictures of cuts. It mostly made them uncomfortable and led to mental distress (e.g. *"she kept posting abt her self harm which made me rlly uncomfy n then when she kept posting pics...it made me have an anxiety attack"* - 21-year-old, female). They also mentioned that such harmful content acts as a trigger for them and entices their urges which makes them block or ignore the person sharing it. The peers often supported them by expressing empathy and validating their feelings. However, sometimes peers defied the judgments and defended those who shared their self-harming experiences laying out their stance as a way to communicate their emotional pain and to seek support. They suggested unfollowing those people if it's triggering but not to blame them for one's own actions.

4.2.2 Youth discussed strategies to intervene and help others at risk of suicide and/or self-harm. We found that youth shared the self-harm and/or suicidal thoughts of their friends to discuss immediate ***intervention*** plans to prevent such incidents from happening.

Mostly, they asked their peers for direct help with handling a crisis situation as they were worried about their friend's well-being but were unsure about how best to support them. In some cases, youth delegated the duty of monitoring or engaging in a conversation with their at-risk friends to keep them from acting upon their self-harming and/or suicidal thoughts (e.g. *"We need you to check on X do whatever you can make sure he isn't attempting suicide"* - 18-year-old, male). The peers mostly suggested trying to comfort their friends with reasons to stay alive or call the cops and alert a counselor. However, peers also advised to tread carefully while contacting counselors as they might alert the parents which can worsen the situation:

> ***O**: One of my friends has been in a really depressive state. Like she's been posting saying that she is self harming...How should I go about this? Like should I talk to her first then email like Mrs X or something?*
> ***P**: Sometime u have to be careful when emailing the counselor because i think they might tell the parents if it's really bad I think it's always good to talk to them first even though it's hard to talk about it (20-year-old, female)*

Similarly, youth discussed the struggles they are facing in currently helping and dealing with such situations, pursuing emotional support for their own mental stability (e.g. *"i had to call him out in the freezing cold for twenty minutes because he said he was gonna kill him elf... so i'm feeling really like upset and like idk rn"* - 18-year-old, female). They mentioned stressing over their friends going offline and unreachable after disclosing their suicidal urges. In this situation, the conversation partners showed sympathy and suggested not to leave their friends alone and reassure their value.

4.3 The Hyperbolic and Humorous Use of Self-Harm and Suicide Language (RQ3)

A large proportion (70%, n = 1,417 out of 2,019) of sub-conversations used suicide or self-harm language, without active engagement in such behaviors. Within these conversations, we saw that youth: 1) used ***hyperbole for "killing oneself" as an exaggerated expression*** (85%, n = 1209 out of 1,417) of their emotions, 2) discussed suicide and/or self-harm topics in ***general*** (11%, n = 158 out of 1,417), and 3) engaged in ***humor directly about killing or harming oneself*** (4%, n = 50 out of 1,417). Below, we elaborate on the detailed characteristics of such discussions.

4.3.1 Youth used hyperbolic language for "killing oneself" without humor. Youth used suicide hyperbolic language to express extreme 1) ***frustration*** (31%, n = 373 out of 1,209), 2) reaction to ***social stressors*** (22%, n = 270 out of 1,209), 3) ***work stress*** (20%, n = 228 out of 1,209), 4) ***reaction to viewed content*** (15%, n = 180 out of 1,209), 5) ***physical discomfort*** (7%, n = 90 out of 1,209), and 6) ***joyousness*** (5%, n = 68 out of 1,209).

Youth mostly used suicide hyperbolic language to exhibit their ***frustration*** in different situations. It often stemmed from experiencing minor inconveniences like the website taking too long to load, boredom, and finding something annoying. We also saw that the youth were aware that their hyperbolic language does not show any actual intent and decided to make it clear as well. (*"Just cause I say I wanna die in the morning. DOESNT mean I actually do"* - 18-year-old, female). Sometimes youth showed frustration because of making trivial errors e.g. a silly typing error, that ignited feelings of embarrassment and awkwardness. Similarly, a conversation member expressed their frustration on sending the wrong emoticon to someone and their peer tried to calm them down:

> ***P**: SHE IS LEGIT GOING TO THINK IM WEIRD (17-year-old, female)*
> ***P**: KMS KMS KMS*
> ***O**: She doesn't care*

Youth used hyperbolic language as a response to ***social stressors*** (i.e. attraction, infatuation, fondness, and strong dislike). Youth leveraged exaggerated language to show immense admiration, and affection for any other person and how they would "die" for them. These conversations entailed emotions of extreme fondness and infatuation by often referring to someone's beauty and personality and how it "kills" them. Youth also conveyed strong emotions about disliking a person, using exaggerated language (*"If I had to deal with that person, I'd kill myself in class"* - 15-year-old, non-binary youth). Youth also employed hyperbolic phrases to express ***work stress***. They used such phrases to emphasize their feelings of being overwhelmed by academic or professional work-related stress including school homework, courses, exams, and tiring work shifts. The conversation partners found it relatable and disclosed their own extreme emotions in the form of hyperbolic suicide language when conversing about their work stressors (e.g., *"Worked a double yesterday 14 hours I almost died"* - 21-year-old, female).

Youth used hyperbolic language as a strong ***response to viewed content*** i.e. videos, memes, and songs they found amusing or emotionally impactful (e.g "*Lmaoooo*[4] *that video is killing me"* - 21-year-old, female). The respondents acknowledged and also found it relatable. Similarly, youth utilized exaggerated statements to express extreme ***physical discomfort*** caused by health issues, sleep deprivation, or tiredness and exhaustion. They also used such language to emphasize severe body pain (e.g. *"my head has been killing me for two days lmao end me"*). Additionally, youth employed exaggerated suicide language to express extreme ***joyousness***. They used it as a way to emphasize the extent of their happiness on different occasions e.g. receiving surprises and gifts, responding to funny texts, petting puppies, and getting to watch their favorite celebrity. Youth sometimes showed joyousness when being nostalgic for specific memories and adventures from the past (e.g. *"One of the funniest times I had with that ex, I still die laughing"*). We saw that the conversation partners reciprocated with similar levels of excitement and happiness.

4.3.2 Youth talked about self-harm and/or suicide in a general context mostly to raise awareness for prevention. We observed that youth engaged in general discussions (11%, n = 158 out of 1,417) to 1) promote ***suicide prevention or awareness*** (61%, n = 96 out of 158), and 2) discuss ***media and culture*** (39%, n = 62 out of 158) based on self-harm and suicide topics. Youth often shared instructions on taking part in the ongoing suicide prevention campaign as a result of their peers liking their story where they had posted a picture with the caption "Stay Alive." These messages were shared

[4]Refer to table 3 for the meanings of common slangs used by youth

without context as to whether the participant was actually suicidal or not. The campaign intended to reassure individuals who have suicidal thoughts that they are not alone and therefore, should stay strong as their lives matter. The conversation partners were mostly appreciative of such messages and agreed to share a similar post on their story. However, sometimes, they ignored the message, refused to participate, considered it a hassle, or seemed surprised to receive the message, as they were not suicidal:

> ***O**: Since you liked my Stay Alive pic: Okay so you have to post a black and white picture of anything and put the caption as "Stay alive challenge accepted" and then tag me. This is for suicide prevention*
> ***P**: Do i look suicidal? (20-year-old, female)*

Moreover, youth made each other aware of the harmful social media trends which caused deaths by suicide e.g. the 'Blue Whale Challenge', terming them as the worst thing on the Internet. The conversation partners expressed their frustration with its continued existence and lack of prohibition owing to the tragic outcomes of such challenges (e.g. *"WTFFF AND THT GAME STILL EXISTS??"*). Sometimes, they exchanged national suicide prevention hotlines as a resource for any unfortunate circumstances. They also discussed the importance of knowing how to help people who are suicidal as oftentimes they don't actually want to act upon their thoughts and need someone to be there to stop them and console them.

Youth also discussed ***media and culture*** based on self-harm and/or suicide topics. They mentioned various TV shows, movies, and books that were centered around themes of suicide and/or self-harm, the most prominent of which was "13 Reasons Why." They applauded the show claiming it as a realistic portrayal without over romanticizing such behaviors. The peers mostly agreed with the stance and added how it made them realize that people are too complex and its difficult to decipher the main reason behind them killing themselves. However, sometimes, they considered it as glorifying suicide, which made them feel uncomfortable:

> ***P**: 13 reasons why is getting another season (18-year-old, female)*
> ***O**: ITS BASICALLY AN INSTRUCTION MANUAL ON HOW TO COMMIT SUICIDE AND NOT TAKE ANY RESPONSIBILITY FOR YOUR OWN ACTIONS, Bitch kills herself and makes everyone feel bad cause they didn't notice her*

Sometimes, youth discussed the self-harming behaviors associated with certain cultures, mainly goth and emo (e.g. *"Goth girls always be cutting themselves"* - 21-year-old, female). The peers denied such assumptions and stated that they were just stereotypes. Youth also used suicide and/or self-harm references in role-playing dialogues while assuming a fictional character. Mostly, they made their character engage in self-harm and/or suicide as part of the role-play. We observed that they used "\" when delivering dialogues and "(" or "\\" to speak normally when out of character. They also said they wanted to die or kill themselves while talking in terms of their character because they wanted to quit it and do something else (e.g. *"(I wanna die, here, see ya when I get home and can rp!"* - 19-year-old, female).

4.3.3 Humor directly about killing or harming oneself . Youth engaged in comedic or humorous discussions that involved explicit references to 1) killing oneself (96%, n = 48 out of 50) and 2) harming oneself (4%, n = 2 out of 50). Most of such instances referred to humor directly about ***killing*** oneself or joking about suicide. Youth often joked about what would happen if they died of suicide with their close friends (e.g. *"You'll be the first to read the suicide letter, LOL"* -17-year-old, female). They also jokingly talked about shooting themselves by tweaking the term and using "parashoot" and "kashoot" instead to add that fun element suggesting it was said in a light manner (e.g. *"imma jus kashoot myself then"* - 15-year-old, female). It was also evident that the respondents showed a lack of responsiveness and ignored suicidal humor as they considered it not harmful. Similarly, at times, conversation members took on the role of a person exhibiting suicidal behavior and a suicide hotline operator respectively, and exchanged humor as a dialogue:

> ***P**: Are you going to stop with the dad jokes *cocks shotgun* or what... Calls national suicide hotline* Hello it's me again (18-year-old, female)*
> ***O**: *answers on other end* Hi me again, I am [anon]*
> ***P**: I'll go shoot myself now*
> ***O**: This sounds like a cheesy comedy*

Similarly, they often joked about how they would rather choose to kill themselves or wish to die than be in a specific situation (e.g. *"I'd rather kill myself than do so much in one day LAMAOOOO"* - 14-year-old, male). They also made fun of each other for being in the habit of saying "kill myself" to anything and everything to which the conversation partners agreed. Furthermore, in only two instances, ***self-harming*** humor language was used where youth were seen making jokes about cutting body organs either to get attention from their crush or to catch their favorite musical show. In one instance, they also clarified that it was a joke (e.g. *"i made a joke and said 'i'd cut off my left arm to see it"*). The self-injury humor was taken lightly without much speculation about the context or reasoning behind the use of this language by the respondents.

5 DISCUSSION

In this section, we discuss the implications of our results and provide design recommendations for assisting youth in combating suicidal and/or self-harm urges.

5.1 From Positive Past Reflections to Imminent Risks of Self-Harm and Suicide (RQ1)

Our findings uncovered extreme cases, ranging from youth celebrating recovery from self-harm to imminent risk situations where immediate intervention was needed to prevent serious consequences. A poignant finding from our study was the clear importance of Instagram DMs in providing a needed outlet for youth to reach out to their peers to share their personal struggles with self-harm and suicide. As several youth in our study disclosed to their peers, therapy had not been effective for them, and seeking peer support provided a more empathetic and less judgmental avenue for helping them cope. As such, our participants often used social media as a "valve" to share their journeys of recovery and provide support and encouragement to those still struggling with harmful urges. In a beautifully raw way, youth in our study discussed suicide and

self-harm topics openly and authentically, and we witnessed advanced strategies being used by youth to mitigate harm. Further, we saw youth developing social norms regarding standing up against bullies who told others to kill themselves or tried to degrade their self-worth. We reflect on these positive instances of recovery, hope, and support *first*, so they are not overshadowed by fear invoked by the thought of youth having such heavy and hard discussions without the guidance of adults. Prior research [9, 27] has pointed to the importance of leveraging social media as a channel for supporting young people with self-harm and suicidal ideation. Our research brings home the importance of providing outlets for peer support through unfiltered private conversations among youth.

Meanwhile, the prevalent narrative among policy experts and youth advocates is that social media companies are at fault and have not adequately protected young users from exposure to self-harm content on their platforms [5]. Researchers and news media alike have raised concerns regarding suicide contagion effects propagated through social media platforms [6] [7]. As a response to these criticisms and lawsuits, Instagram responded by implementing sensitivity filters and banning graphic images of self-harm and suicide [7] but decided not to remove non-graphic self-harm related content. This was a nuanced decision by Instagram that other social media platforms (e.g., X, formerly known as Twitter [8]) have also made between promoting self-harm and suicide (e.g., "You should") versus support seeking or discussing self-harm or suicide topics. Still, beyond letting such content remain on the platform, while promises had been made to support people struggling with self-injury, since 2019, Instagram has only provided guidelines on how to help friends in their Help Center [9]. There is still no additional help intervention in Instagram DMs besides reporting content. Some parental control applications already employ automatic monitoring to notify parents about potential suicide or self-harm risks found in young people's private messages [25]. Yet, our results underscore the need to avoid excessive surveillance that might deter youth from participating in beneficial peer support. Therefore, an outstanding question remains open: Should platforms silence such content, while risking stigma or encourage it, while risking other unintended consequences? These difficult decisions are worth continuing to grapple with as our findings show that youth often reflect on their past self-harm behaviors and seek support, which can generally be a positive and helpful experience for them.

On the other hand, we also found instances of youth being blackmailed by friends or partners who threatened self-harm or suicide, as a means of exerting coercive control over the actions of others. We observed youth telling one another to harm or kill themselves as a form of bullying. Such toxic behaviors are potentially harmful to youth; thus, finding effective ways to mitigate such coercive and violent behaviors is a worthy endeavor. Bailey et al. [9] addressed such concerns by incorporating content moderation in a social network platform designed for youth actively experiencing suicidal ideation; however, research has typically found that content inhibited youth from posting about their emotional and mental challenges or encouraged them to alter moderated (e.g., banned) keywords to overcome the platforms' restrictions [21]. We propose an alternative solution to content moderation. Our findings provide compelling evidence for social media platforms to be more proactive in providing crisis intervention, given that they are a first line of defense that provides the private channels in which youth seek direct help from their peers when in crisis. Responsibility may manifest in the form of providing in-the-moment help resources, integrated within the private chats through contextual prompts [8], to at-risk youth and/or their peers on how to react in crisis situations or where to call to get immediate help. Such context-aware responses, however, would require accurate and timely mechanisms for risk detection that are well-designed to support, rather than censor or police, youth conversations that involve self-harm or suicide. However, such design-based interventions must be inherently trauma-informed [22, 73] and sensitive to the privacy needs and vulnerable mental state of at-risk youth. For instance, the design of these trauma-informed interventions should ensure both psychical and psychological safety, trust of youth, effective peer support, collaboration with youth to design the technology, and enabling youth to have more power over their lives [22, 73].

5.2 From Gossip to Intervention in the Self-Harm and Suicide of Others (RQ2)

While prior work has focused mainly on the personal experiences of youth with self-harm and suicide [67, 91], understanding how youth talked about third-parties and their suicidal or self-harm tendencies gave us new insights into how youth think about these topics. When youth gossiped about others, they often accused those individuals of using suicide or self-harm as a way to seek attention, trivializing their struggles. Similarly, youth shared frustration and discomfort with others who shared triggering content, especially when they struggled with these tendencies themselves. In some cases, youth knew how to set boundaries and communicate those boundaries with others to save themselves from mental distress. Yet, it is unclear whether boundary setting of this type caused additional distress to the person who made the initial disclosure. Prior research has studied the harmful effects of self-harm content on vulnerable youth [59] but has yet to examine whether being blocked from sharing such content with others could also be detrimental to those in crisis. Therefore, we call upon the need to understand how youth create norms and how they set healthy boundaries in their interactions with others while also offering support. Norms are important in online mental health support, including that around self-harm and suicide, as prior research has indicated how normative conformance can often result in more quality emotional and informational support, compared to contexts where help seekers make little normative accommodation [82].

Our study also found instances where third-parties tried to intervene and provide peer support to those in danger of self-harm and/or suicide. This builds on the prior efforts that focused on understanding ways youth seek help on online platforms [75, 98] and sheds light on how recipients of such disclosures go above and

[5] https://www.bbc.com/news/uk-47114313
[6] https://www.goodmorningamerica.com/wellness/story/suicide-contagious-teens-parent-98075448
[7] https://about.instagram.com/blog/announcements/supporting-and-protecting-vulnerable-people-on-instagram
[8] https://help.twitter.com/en/rules-and-policies/glorifying-self-harm
[9] https://help.instagram.com/388741744585878

beyond to come up with feasible plans to protect those at risk. However, we found that they were not always sure about the right action plan due to a lack of expertise. Scholars have recognized this gap in knowledge as well; thus, prompting efforts to develop algorithmic solutions (e.g., AI-based automatic suggestion generation) to assist conversational peers on online forums in support provisioning/writing [46]. Building on these ideas and efforts, as well as our findings in this paper, we posit that it is necessary to equip youth with skills to know how to react in case of hearing peer struggles through education and training. Training from organizations such as Crisis Text Line [10] could be helpful in private chats for youth to use de-escalating techniques when eminent self-injury presents, which could go hand in hand with algorithmic strategies to help to write quality responses. Also, dealing directly with such intense situations had an effect on their mental prowess. Such training needs to incorporate sessions and aspects for youth where they can discuss their own needs in addition to peers' potential struggles. For instance, training could enlighten peers on the types of social roles they can adopt when they encounter another person sharing self-harm or suicidal content. Yang et al. [97] conceptualized such roles to comprise "providers, welcomers, and storytellers"; thus youth could be provided with strategies corresponding to these varied roles to rise to the occasion when such sensitive messages appear in their private social media conversations.

5.3 From Humor to Hypberbole When Using Self-Harm and Suicide Language (RQ3)

While prior studies [16, 78, 84] documented the utilization of humor and figurative language in public posts by users, our paper extends these findings by revealing that "youth," a vulnerable population, similarly employ such language in private settings, especially on a topic as sensitive as self-harm. It is recognized that humor can be a way to communicate stigmatizing experiences and find ways to cope [10]. Given that Instagram's DM conversations are inherently private, this might make room for greater disinhibition, especially in engagements with close peers [48], thus allowing youth to fight longstanding perceptions of stigma around self-harm and suicide, and hence making the use of hyperbolic language an acceptable norm. For the same reason, participants in our data may have felt comfortable bringing up such experiences in a facetious manner, with hilarity, or even sometimes with purposeful exaggeration. Speaking of norm-setting, expressing vulnerability through humor and hyperbole may also be a mechanism through which young people experiment with boundary-setting with their peers – gauging their reactions to what could be acceptable private discourse online. Moreover, as hyperbolic language is increasingly normalized in young people's online presence [4], they may conform to this culture to gain acceptance within those online spaces.

However, our analysis unraveled the potential for desensitization to suicide-related topics through humor and normalizing language. While communicating stigmatizing experiences is an effective way to cope as mentioned above [11], it is a double-edged sword with potentially being triggering to those who struggle with suicide ideation or grief from suicide loss [95]. Since words carry power based on individual experiences, background, and perceptions, they can be re-traumatizing for victims of violence or loss [32]. The use of dark humor about suicide as a coping mechanism also has the downfall of peers not knowing the actual intention, therefore not being able to provide support when there is a real need. Peers may also misconstrue hyperbolic references to self-harm and suicide to be attention-seeking behaviors, enacted to gather sympathy from others online, which in turn can fulfill their emotional needs. We call on educators and researchers to provide understanding and training for youth to take more effective ways of coping with situations that are not triggering to the community.

Moreover, given how enormously youth have integrated suicide hyperbolic language into their conversations, it can highly affect the ML community's algorithmic risk detection and prevention efforts [20] as it poses a great likelihood of false positives. The cost of high false positives could be severe depending on the risk mitigation approaches after the risk is identified. For instance, if the risk prevention interventions involve escalating to mental health services, it could overburden the mental health system. Prior efforts [16, 78] to detect suicidal ideation from users' public posts have shared struggles caused by flippant references to suicide or self-harm. However, youths' language is inherently more complex [21], as also exhibited in our study (e.g., *kms* or *kashoot*), which intensifies the concerns of accurate prediction. For this purpose, the context of the discussion can play a great role in training models to detect hyperbole effectively and dismiss them as non-risky. Our work delved deeper and showed that participants were aware of their hyperbole and even mentioned that they did not literally mean any actual harm. The researchers can look for subtle cues within the discussions that can efficiently inform the models to identify and overlook such instances. However, using suicide language is not completely harmless as it can still contribute to desensitizing the topic, and ignoring it is not preferable. Instead, we should consider teaching youth about social pragmatics and why using violent language, while it seems harmless, could harm others.

5.4 Implications for Design

Here we propose design-based solutions to help youth navigate suicide and/or self-harm urges.

5.4.1 Create safe places where youth can discuss self-harm and suicide. While youth privately discuss self-harm and suicide, this paper uncovered concerns about misunderstanding and the inability to provide effective support that limited their engagement. Thus, in light of our recommendation for youth to choose the right social roles (refer to section 5.2), we recommend social media companies create a trusted circle feature where youth can connect with professional therapists, supportive friends, family members, and peers who've had similar experiences. This 'circle of trust' concept, known for enhancing online safety for adolescents [38], would enable young individuals to privately and anonymously express their self-harm and suicidal urges, with professionals conducting regular check-ins for timely support. Resources should also be available for at-risk youth and their peers looking to intervene, as previously mentioned in section 5.2. Within these circles of trust, a crisis intervention plan should be established in cases of imminent risks of self-harm or suicide occurring. As we observed instances of bullying

[10] https://www.crisistextline.org/become-a-volunteer

and blackmail regarding self-harm and suicide, mostly by people outside of the chat, maintaining a respectful and cyberbullying-free environment, akin to the TalkLife platform's model for peer support among youth [51], is crucial for these circles to ensure effective engagement from youth. However, obtaining consent for personal data use, as seen in Facebook's suicide AI [39], remains inadequately addressed. Users may be unaware that their online activities are analyzed by individuals without healthcare expertise but with the power to involve law enforcement in self-harm cases [19]. Therefore, for these circles of trust, online platforms are urged to prioritize robust consent mechanisms to safeguard user privacy.

5.4.2 Develop timely and accurate risk detection for imminent risks and toxic behavior. Manual moderation is inefficient due to time constraints, especially for cases requiring immediate intervention. Thus, it is crucial to develop accurate and timely detection systems for self-harm and suicide in private conversations. However, the lack of high-quality annotated datasets sensitive to context has long hindered training these models to have accurate detection outcomes [20]. To address this, future annotation efforts should consider our approach, which considers the shifts in topics within a given conversation and considers the context by differentiating suicidal hyperbolic language from high-risk situations, especially since 70% of the private sub-conversations used the self-harm and suicide language hyperbolically. As mentioned in Section 5.3. youth might have different intentions of using hyperbolic or humorous references to self-harm and suicide and a one-size-fits-all approach is not a solution. This nuanced annotation moves beyond the binary classification of self-harm or suicide for a context-aware detection model that would decrease the false-positives instances of hyperbolic or humorous incidents. Additionally, most existing systems detect self-harm and suicide retrospectively, missing the opportunity for effective prevention [20]. To address this, a proactive detection approach could be formulated by employing forecasting techniques to predict and preempt these risks ahead of time, a strategy that has demonstrated effectiveness in various contexts, including marketing and finance, when applied to social media [76].

5.4.3 Consider nudging youth towards less violent language. While addressing high-risk incidents is crucial, the frequent use of suicide humor and hyperbolic language can divert our attention and obscure more urgent issues. Although improving detection efficiency can mitigate false positives, we should also consider long-term strategies to encourage youth to adopt less violent language. To achieve this, we can employ "nudging," which involves subtle prompts aimed at influencing behaviors without infringing on users' decision-making autonomy [60]. Research has shown that nudging is a promising approach to help youth reflect on their actions and suggest alternative behaviors [3]. For example, researchers [12, 94] have attempted to prevent cyberbullying behaviors by reminding the consequences and emphasizing the harm through warning pop-ups integrated within social media platforms. Moreover, Hardin et al. [43] introduced an interactive game called "Digital Privacy Detectives" to teach teens how to manage their online privacy through plot-based challenges. Building upon these ideas, when youth use harmful language hyperbolically, nudges can serve as a valuable tool for them to reconsider the meaning behind their words and guide them toward using milder and less harmful language to express their emotions. Likewise, we can introduce educational nudges to impart an understanding of the consequences and seriousness of making jokes about such sensitive topics. However, this should be executed thoughtfully to avoid appearing as censorship. Therefore, a collaborative effort involving mental health professionals, educators, and youth can be instrumental in designing more effective nudges that align with youth's needs and language preferences.

5.5 Limitations and Future Work

Some limitations of our study inform areas of future research. First, we chose to analyze private conversations because prior research has already utilized public posts to understand youths' discussions around suicide and self-harm. We compared our findings with the existing literature; however, future research could leverage both private and public data for direct comparisons. Furthermore, the dataset in the study by Razi et al. [71] was collected solely from Instagram. Therefore, to assess the generalizability of our findings, we recommend future research to explore other popular social media platforms among youth, such as TikTok. We only had demographic information from participants who donated their data, not from their conversation partners. While in some cases we could infer the relationship between the individuals (e.g., peers, strangers), we did not have pertinent details, such as the age of these conversation partners. While they also appeared to be youth in most cases, it is possible that some of these conversations occurred with older adults. Future research should understand how age plays a role in the peer-based conversations youth have around suicide and self-harm. The Razi et al.[71] study included youth with at least two unsafe messages (not necessarily about suicide/self-harm), possibly indicating a higher range of unsafe interactions compared to a general sample of youth. Therefore, future research should involve diverse youth who have experienced suicide/self-harm for a holistic understanding of their online interactions with those issues.

It is noteworthy to point out that only a handful of the conversations were flagged by participants as risky. This suggests that conversations around suicide and self-harm may be more commonplace and accepted among youth than adults may imagine. This only increases the urgency of educating youth on how to safely manage these situations when they occur with their peers. This is particularly true given that many of the youth participants explicitly stated how they did not trust their parents or other adults to help them with these situations. As such, future work should consider applications and guidelines for peer-based support (rather than more paternalistic solutions) that are developmentally appropriate and safe for youth. Moreover, this study was conducted with a sample from the United States, and results may vary in different countries. In particular, the hyperbolic use of self-harm or suicide may differ in different languages. Future research should incorporate a more diverse, global perspective to understand the nuances of youth language across different cultures.

5.6 Conclusion

Our research is the first to analyze youths' private conversations regarding suicide and/or self-harm to provide deep insights into their struggles and triumphs. While these findings have the potential to invoke fear and paternalism, which could lead to restriction

and protectionism [13], they also have the potential to invoke hope and partnership with our younger generations to help them navigate these serious topics in ways we never before had to even imagine. Now is the time to learn from the experts (i.e., youth) to inform social media platforms on how they can proactively support the mental health and well-being of our youth, rather than debate whether and how social media impacts their mental health. Instead of pointing fingers and trying to place blame on social media for the mental health crisis of our youth, let us all work together to create viable solutions. There is no time to waste.

ACKNOWLEDGMENTS

This research was supported 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, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our sponsors.

REFERENCES

[1] Jaclyn Abraham, Rebecca Roth, Heidi Zinzow, Kapil Chalil Madathil, and Pamela Wisniewski. 2022. Applying behavioral contagion theory to examining young adults' participation in viral social media challenges. *Transactions on Social Computing* 5, 1-4 (2022), 1–34.

[2] Muhammad Abulaish, Ashraf Kamal, and Mohammed J Zaki. 2020. A survey of figurative language and its computational detection in online social networks. *ACM Transactions on the Web (TWEB)* 14, 1 (2020), 1–52.

[3] 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.

[4] Saiful Akmal, Nadia Ulfah, Nabila Fitria, et al. 2022. Here Comes the Acehnese Gen-Z! Language And Identity in Social Media Communication. *Communication Today* 23 (2022).

[5] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Joshua Gracie, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. 2022. Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 148, 14 pages. https://doi.org/10.1145/3491102.3501969

[6] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Chen Ling, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. 2023. Getting Meta: A Multimodal Approach for Detecting Unsafe Conversations within Instagram Direct Messages of Youth. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 132 (apr 2023), 30 pages. https://doi.org/10.1145/3579608

[7] Florian Arendt, Sebastian Scherr, and Daniel Romer. 2019. Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. *New Media & Society* 21, 11-12 (2019), 2422–2442.

[8] Karla Badillo-Urquiola, Diva Smriti, Brenna McNally, Evan Golub, Elizabeth Bonsignore, and Pamela J Wisniewski. 2019. Stranger danger! social media app features co-designed with children to keep them safe online. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 394–406.

[9] Eleanor Bailey, Mario Alvarez-Jimenez, Jo Robinson, Simon D'Alfonso, Maja Nedeljkovic, Christopher G Davey, Sarah Bendall, Tamsyn Gilbertson, Jessica Phillips, Lisa Bloom, et al. 2020. An enhanced social networking intervention for young people with active suicidal ideation: safety, feasibility and acceptability outcomes. *International journal of environmental research and public health* 17, 7 (2020), 2435.

[10] Amy M Bippus. 2000. Humor usage in comforting episodes: Factors predicting outcomes. *Western Journal of Communication (includes Communication Reports)* 64, 4 (2000), 359–384.

[11] Bill Borcherdt. 2002. Humor and its contributions to mental health. *Journal of rational-emotive and cognitive-behavior therapy* 20 (2002), 247–257.

[12] Leanne Bowler, Eleanor Mattern, and Cory Knobel. 2014. Developing design interventions for cyberbullying: A narrative-based participatory approach. *iConference 2014 Proceedings* (2014).

[13] Danah Boyd. 2014. *It's complicated: The social lives of networked teens.* Yale University Press.

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

[15] Rebecca C Brown, Eileen Bendig, Tin Fischer, A David Goldwich, Harald Baumeister, and Paul L Plener. 2019. Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses. *PloS one* 14, 9 (2019), e0220623.

[16] Pete Burnap, Walter Colombo, and Jonathan Scourfield. 2015. Machine classification and analysis of suicide-related communication on twitter. In *Proceedings of the 26th ACM conference on hypertext & social media*. 75–84.

[17] Elena Campillo-Ageitos, Hermenegildo Fabregat, Lourdes Araujo, and Juan Martinez-Romo. 2021. NLP-UNED at eRisk 2021: self-harm early risk detection with TF-IDF and linguistic features. *Working Notes of CLEF* (2021), 21–24.

[18] Ilaria Cataldo, Bruno Lepri, Michelle Jin Yee Neoh, and Gianluca Esposito. 2021. Social media usage and development of psychiatric disorders in childhood and adolescence: a review. *Frontiers in Psychiatry* 11 (2021), 508595.

[19] Karen L Celedonia, Marcelo Corrales Compagnucci, Timo Minssen, and Michael Lowery Wilson. 2021. Legal, ethical, and wider implications of suicide risk detection systems in social media platforms. *Journal of Law and the Biosciences* 8, 1 (2021), lsab021.

[20] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine* 3, 1 (2020), 43.

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

[22] Janet X Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-informed computing: Towards safer technology experiences for all. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–20.

[23] Victoria Clarke, Virginia Braun, and Nikki Hayfield. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* 3 (2015), 222–248.

[24] S Cohen. 2022. Suicide rate highest among teens and young adults. *UCLA Health* (2022).

[25] Bark-Advanced content monitoring for all your kid's devices. 2022. New suicidal ideation research from the CDC and Bark. https://www.bark.us/learn/suicidal-ideation-study-cdc-bark/

[26] Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights* 10 (2018), 1178222618792860.

[27] Georgina Cox and Sarah Hetrick. 2017. Psychosocial interventions for self-harm, suicidal ideation and suicide attempt in children and young people: What? How? Who? and Where? *BMJ Ment Health* 20, 2 (2017), 35–40.

[28] Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 71–80.

[29] Munmun De Choudhury and Emre Kiciman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *Proceedings of the international AAAI conference on web and social media*, Vol. 11. 32–41.

[30] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2098–2110.

[31] Chris DeBrusk. 2018. The risk of machine-learning bias (and how to prevent it). *MIT Sloan Management Review* 15 (2018), 1.

[32] Richard Delgado. 2019. *Understanding words that wound.* Routledge.

[33] Jamie M Duggan, Nancy L Heath, Stephen P Lewis, and Alyssa L Baxter. 2012. An examination of the scope and nature of non-suicidal self-injury online activities: Implications for school mental health professionals. *School Mental Health* 4 (2012), 56–67.

[34] Michele P Dyson, Lisa Hartling, Jocelyn Shulhan, Annabritt Chisholm, Andrea Milne, Purnima Sundar, Shannon D Scott, and Amanda S Newton. 2016. A systematic review of social media use to discuss and view deliberate self-harm acts. *PloS one* 11, 5 (2016), e0155813.

[35] Robert J Fisher and James E Katz. 2000. Social-desirability bias and the validity of self-reported values. *Psychology & marketing* 17, 2 (2000), 105–120.

[36] Joseph C Franklin, Jessica D Ribeiro, Kathryn R Fox, Kate H Bentley, Evan M Kleiman, Xieyining Huang, Katherine M Musacchio, Adam C Jaroszewski, Bernard P Chang, and Matthew K Nock. 2017. Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological bulletin* 143, 2 (2017), 187.

[37] General Data Protection Regulation (GDPR). 2023. Art. 20 GDPR – right to data portability. https://gdpr-info.eu/art-20-gdpr/

[38] Arup Kumar Ghosh, Charles E Hughes, and Pamela J Wisniewski. 2020. Circle of trust: a new approach to mobile online safety for families. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[39] Norberto Nuno Gomes de Andrade, Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. 2018. Ethics and artificial intelligence: suicide prevention on Facebook. *Philosophy & Technology* 31 (2018), 669–684.

[40] Kristie B Hadden, Latrina Prince, Laura James, Jennifer Holland, and Christopher R Trudeau. 2018. Readability of human subjects training materials for research. *Journal of Empirical Research on Human Research Ethics* 13, 1 (2018), 95–100.

[41] Sang-Hyuk Han, Seungyoon Lee, and Hyeoncheol Kang. 2018. A content analysis of suicide-related tweets. *Crisis* 39, 1 (2018), 55–63.

[42] Jeff Hancock, Sunny Xun Liu, Mufan Luo, and Hannah Mieczkowski. 2022. Psychological well-being and social media use: A meta-analysis of associations between social media use and depression, anxiety, loneliness, eudaimonic, hedonic and social well-being. *Anxiety, Loneliness, Eudaimonic, Hedonic and Social Well-Being (March 9, 2022)* (2022).

[43] Caroline D Hardin and Jen Dalsen. 2020. Digital Privacy Detectives: An Interactive Game for Classrooms. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 184–189.

[44] Samiha Binte Hassan, Sumaiya Binte Hassan, and Umme Zakia. 2020. Recognizing suicidal intent in depressed population using NLP: a pilot study. In *2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 0121–0128.

[45] Keith Hawton, Kate EA Saunders, and Rory C O'Connor. 2012. Self-harm and suicide in adolescents. *The Lancet* 379, 9834 (2012), 2373–2382.

[46] Shang-Ling Hsu, Raj Sanjay Shah, Prathik Senthil, Zahra Ashktorab, Casey Dugan, Werner Geyer, and Diyi Yang. 2023. Helping the Helper: Supporting Peer Counselors via AI-Empowered Practice and Feedback. *arXiv preprint arXiv:2305.08982* (2023).

[47] Jina Huh-Yoo, Afsaneh Razi, Diep N. Nguyen, Sampada Regmi, and Pamela J. Wisniewski. 2023. "Help Me:" Examining Youth's Private Pleas for Support and the Responses Received from Peers via Instagram Direct Messages. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 336, 14 pages. https://doi.org/10.1145/3544548.3581233

[48] Adam N Joinson. 2007. Disinhibition and the Internet. In *Psychology and the Internet*. Elsevier, 75–92.

[49] Amro Khasawneh, Kapil Chalil Madathil, Emma Dixon, Pamela Wiśniewski, Heidi Zinzow, and Rebecca Roth. 2020. Examining the self-harm and suicide contagion effects of the Blue Whale Challenge on YouTube and Twitter: qualitative study. *JMIR mental health* 7, 6 (2020), e15973.

[50] Meeyun Kim, Koustuv Saha, Munmun De Choudhury, and Daejin Choi. 2023. Supporters First: Understanding Online Social Support on Mental Health from a Supporter Perspective. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–28.

[51] Kaylee Payne Kruzan. 2019. *Self-Injury Support Online: Exploring Use of the Mobile Peer Support Application TalkLife*. Cornell University.

[52] Mrinal Kumar, Mark Dredze, Glen Coppersmith, and Munmun De Choudhury. 2015. Detecting changes in suicide content manifested in social media following celebrity suicides. In *Proceedings of the 26th ACM conference on Hypertext & Social Media*. 85–94.

[53] Cheng-Yu Lai and Chia-Hua Tsai. 2016. Cyberbullying in the social networking sites: An online disinhibition effect perspective. In *Proceedings of the The 3rd Multidisciplinary International Social Networks Conference on SocialInformatics 2016, Data Science 2016*. 1–6.

[54] Anna Lavis and Rachel Winter. 2020. # Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal of child psychology and psychiatry* 61, 8 (2020), 842–854.

[55] Jorge Lopez-Castroman, Bilel Moulahi, Jérôme Azé, Sandra Bringay, Julie Deninotti, Sebastien Guillaume, and Enrique Baca-Garcia. 2020. Mining social networks to improve suicide prevention: A scoping review. *Journal of neuroscience research* 98, 4 (2020), 616–625.

[56] Lydia Manikonda and Munmun De Choudhury. 2017. Modeling and understanding visual attributes of mental health disclosures in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 170–181.

[57] Amanda Marchant, Keith Hawton, Ann Stewart, Paul Montgomery, Vinod Singaravelu, Keith Lloyd, Nicola Purdy, Kate Daine, and Ann John. 2017. A systematic review of the relationship between internet use, self-harm and suicidal behavior in young people: The good, the bad and the unknown. *PloS one* 12, 8 (2017), e0181722.

[58] Alice E Marwick and Danah Boyd. 2014. Networked privacy: How teenagers negotiate context in social media. *New media & society* 16, 7 (2014), 1051–1067.

[59] Aksha M Memon, Shiva G Sharma, Satyajit S Mohite, and Shailesh Jain. 2018. The role of online social networking on deliberate self-harm and suicidality in adolescents: A systematized review of literature. *Indian journal of psychiatry* 60, 4 (2018), 384.

[60] Christian Meske and Ireti Amojo. 2020. Ethical guidelines for the construction of digital nudges. *arXiv preprint arXiv:2003.05249* (2020).

[61] Megan A Moreno, Adeline Ton, Ellen Selkie, and Yolanda Evans. 2016. Secret society 123: Understanding the language of self-harm on Instagram. *Journal of Adolescent Health* 58, 1 (2016), 78–84. https://doi.org/10.1016/j.jadohealth.2015.09.015

[62] Carla Moss, Christopher Wibberley, and Gary Witham. 2023. Assessing the impact of Instagram use and deliberate self-harm in adolescents: A scoping review. *International journal of mental health nursing* 32, 1 (2023), 14–29.

[63] Alicia L Nobles, Jeffrey J Glenn, Kamran Kowsari, Bethany A Teachman, and Laura E Barnes. 2018. Identification of imminent suicide risk among young adults using text messages. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11.

[64] Bridianne O'dea, Stephen Wan, Philip J Batterham, Alison L Calear, Cecile Paris, and Helen Christensen. 2015. Detecting suicidality on Twitter. *Internet Interventions* 2, 2 (2015), 183–188.

[65] Justin W Patchin and Sameer Hinduja. 2017. Digital self-harm among adolescents. *Journal of Adolescent Health* 61, 6 (2017), 761–766. https://doi.org/10.1016/j.jadohealth.2017.06.012

[66] Jessica Pater and Elizabeth Mynatt. 2017. Defining digital self-harm. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1501–1513.

[67] Jessica A Pater and Elizabeth D Mynatt. 2017. Digital self-harm: Understanding adolescent participation in an emerging phenomenon. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), 4031–4043.

[68] Jacobo Picardo, Sarah K McKenzie, Sunny Collings, and Gabrielle Jenkin. 2020. Suicide and self-harm content on Instagram: A systematic scoping review. *PLoS one* 15, 9 (2020), e0238603.

[69] Paul L Plener, Michael Kaess, Christian Schmahl, Stefan Pollak, Jörg M Fegert, and Rebecca C Brown. 2018. Nonsuicidal self-injury in adolescents. *Deutsches Ärzteblatt International* 115, 3 (2018), 23.

[70] 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. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 89 (apr 2023), 29 pages. https://doi.org/10.1145/3579522

[71] Afsaneh Razi, Ashwaq AlSoubai, Seunghyun Kim, Nurun Naher, Shiza Ali, Gianluca Stringhini, , Munmun De Choudhury, and Pamela J Wisniewski. 2022. Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–9.

[72] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Xavier Caddle, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela Wisniewski. 2021. Teens at the Margin: Artificially Intelligent Technology for Promoting Adolescent Online Safety. In *ACM Conference on Human Factors in Computing Systems (CHI 2021)/Artificially Intelligent Technology for the Margins: A Multidisciplinary Design Agenda Workshop*.

[73] Afsaneh Razi, John Seberger, Ashwaq Alsoubai, Nurun Naher, Munmun De Choudhury, and Pamela J. Wisniewski. 2024. Toward Trauma-Informed Research Practices with Youth in HCI: Caring for Participants and Research Assistants When Studying Sensitive Topics. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 134 (oct 2024), 31 pages. https://doi.org/10.1145/3637411

[74] Jacob M Ring, Taylor A Burke, and Kristen M Janson. 2019. "I felt like I wasn't alone": A qualitative study of adolescents who blog about self-harm. *Journal of Child and Adolescent Psychiatric Nursing* 32, 1 (2019), 22–28.

[75] Jo Robinson, Georgina Cox, Eleanor Bailey, Sarah Hetrick, Maria Rodrigues, Steve Fisher, and Helen Herrman. 2016. Social media and suicide prevention: a systematic review. *Early intervention in psychiatry* 10, 2 (2016), 103–121.

[76] Dimitrios Rousidis, Paraskevas Koukaras, and Christos Tjortjis. 2020. Social media prediction: a literature review. *Multimedia Tools and Applications* 79, 9-10 (2020), 6279–6311.

[77] María-José Rubio-Hurtado, Marc Fuertes-Alpiste, Francesc Martínez-Olmo, and Jordi Quintana. 2022. Youths' Posting Practices on Social Media for Digital Storytelling. *Journal of new approaches in educational research* 11, 1 (2022), 97–113.

[78] Ramit Sawhney, Prachi Manchanda, Puneet Mathur, Rajiv Shah, and Raj Singh. 2018. Exploring and learning suicidal ideation connotations on social media with deep learning. In *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis*. 167–175.

[79] Rachel Schwartz, Laura Curran, and Marian Diksies. 2020. Online social work education and the disinhibition effect. *Advances in Social Work and Welfare Education* 21, 2 (2020), 107–122.

[80] Jonathan Scourfield, Katrina Roen, and Elizabeth McDermott. 2011. The non-display of authentic distress: Public-private dualism in young people's discursive construction of self-harm. *Sociology of health & illness* 33, 5 (2011), 777–791.

[81] Nigam H Shah, Arnold Milstein, and Steven C Bagley. 2019. Making machine learning models clinically useful. *Jama* 322, 14 (2019), 1351–1352.

[82] Eva Sharma and Munmun De Choudhury. 2018. Mental health support and its relationship to linguistic accommodation in online communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[83] Ariana C Simone and Chloe A Hamza. 2020. Examining the disclosure of non-suicidal self-injury to informal and formal sources: A review of the literature. *Clinical psychology review* 82 (2020), 101907.

[84] Kamesha Spates, Xinyue Ye, and Ashley Johnson. 2018. "I just might kill myself": Suicide expressions on Twitter. *Death studies* (2018).

[85] Lauren A Spies Shapiro and Gayla Margolin. 2014. Growing up wired: Social networking sites and adolescent psychosocial development. *Clinical child and family psychology review* 17 (2014), 1–18.

[86] The Bark Team. 2023. 2023 teen slang meanings every parent should know. https://www.bark.us/blog/teen-text-speak-codes-every-parent-should-know/

[87] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.

[88] Qamar Un Nisa and Rafi Muhammad. 2021. Towards transfer learning using BERT for early detection of self-harm of social media users. *Proceedings of the Working Notes of CLEF* (2021), 21–24.

[89] Emily A Vogels, Risa Gelles-Watnick, and Navid Massarat. 2022. Teens, social media and technology 2022. (2022).

[90] David Wadden, Tal August, Qisheng Li, and Tim Althoff. 2021. The effect of moderation on online mental health conversations. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 751–763.

[91] Janis Whitlock, Jennifer Muehlenkamp, John Eckenrode, Amanda Purington, Gina Baral Abrams, Paul Barreira, and Victoria Kress. 2013. Nonsuicidal self-injury as a gateway to suicide in young adults. *Journal of adolescent health* 52, 4 (2013), 486–492.

[92] Janis Whitlock, Jennifer L Powers, John Eckenrode, and editors. 2006. Internet support groups for suicide survivors: a new mode for gaining bereavement assistance. *Suicide and Life-Threatening Behavior* 36, 1 (2006), 25–41.

[93] Janis L Whitlock, JL Powers, and John Eckenrode. 2006. The virtual cutting edge: the internet and adolescent self-injury. *Developmental psychology* 42, 3 (2006), 407. https://doi.org/10.1037/0012-1649.42.3.407

[94] Anne Williford, L Christian Elledge, Aaron J Boulton, Kathryn J DePaolis, Todd D Little, and Christina Salmivalli. 2013. Effects of the KiVa antibullying program on cyberbullying and cybervictimization frequency among Finnish youth. *Journal of Clinical Child & Adolescent Psychology* 42, 6 (2013), 820–833.

[95] Donna M Wilson, Michelle Knox, Gilbert Banamwana, Cary A Brown, and Begoña Errasti-Ibarrondo. 2022. Humor: A Grief Trigger and Also a Way to Manage or Live With Your Grief. *OMEGA-Journal of Death and Dying* (2022), 00302228221075276.

[96] Pamela Wisniewski, Haiyan Jia, Heng Xu, Mary Beth Rosson, and John M Carroll. 2015. " Preventative" vs." Reactive" How Parental Mediation Influences Teens' Social Media Privacy Behaviors. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 302–316.

[97] Diyi Yang, Robert E Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[98] Renwen Zhang, Natalya N. Bazarova, and Madhu Reddy. 2021. Distress disclosure across social media platforms during the COVID-19 pandemic: Untangling the effects of platforms, affordances, and audiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

A GLOSSARY OF ABBREVIATIONS AND SLANG TERMS

Table 3: Abbreviations/Slangs and their Meanings

Abbreviations/Slangs	Meaning
NSSI	Nonsuicidal self-injury
LMAO	Laughing my ass off
LOL	Laughing out loud
IDK	I don't know
KMS	Kill myself
KYS	Kill yourself
PPL	People
IDC	I don't care
STFU	Shut the fuck up
SMH	Shaking my head
PTSD	Post-traumatic stress disorder
RN	Right now
WTF	What the fuck?
RP	Role-play
OMG	Oh my God
SUE	Suicide
Secretsociety123 and SVV	Self-harm
bonspo and thinspo	Promoting extreme thinness
IHML	I hate my life