Empowering Bystanders: Leveraging Generative AI to Enhance Direct Cyberbullying Intervention and Support Teen Well-Being

PEINUAN QIN, School of Computing, National University of Singapore, Singapore
JUNTI ZHANG, Institute of Data Science, National University of Singapore, Singapore
JITING CHENG, School of Computing, National University of Singapore, Singapore
JUNGUP LEE, Department of Social Work, National University of Singapore, Singapore
ZHIXING LIU, School of Computing, National University of Singapore, Singapore
YI-CHIEH LEE, Computer Science, National University of Singapore, Singapore

With the rise of social networks, cyberbullying has become pervasive among teenagers. While bystander intervention can help, direct action remains challenging. We identified key barriers—effort and lack of confidence—through a formative study with 67 participants and developed EmojiGen, a Large Language Model (LLM)-powered tool for intervention support. In an experiment with 90 participants on a simulated platform, EmojiGen significantly increased intervention frequency, boosted defending self-efficacy and perception of knowing how to help, as well as reduced anxiety and effort. This work demonstrates the potential of LLMs in designing AI-assisted interventions, fostering proactive engagement in online safety.

ACM Reference Format:
Peinuan Qin, Junti Zhang, Jiting Cheng, Jungup Lee, Zhixing Liu, and Yi-Chieh Lee. 2025. Empowering Bystanders: Leveraging Generative AI to Enhance Direct Cyberbullying Intervention and Support Teen Well-Being. 1, 1 (February 2025), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The rise of social networks has reshaped adolescent interactions but also exposed them to risks such as cyberbullying, which negatively impacts mental health and social development [47]. Cyberbullying includes harassment, impersonation, and exclusion [28, 50]. Bystanders are large groups who witness cyberbullying, but rarely intervene [22, 38]. Although prior work has made progress in encouraging indirect bystander intervention (e.g., reporting) [14], efforts to promote direct intervention—actively confronting perpetrators or supporting victims—have had limited success [15, 46]. Typically, direct intervention can deter perpetrators, provide immediate emotional support, and encourage intervention [2, 3].

To address this, we first conducted a formative study identifying key barriers: the effort needed to formulate responses and the lack of confidence in communication skills. Based on these insights, we developed EmojiGen, a Large Language Model (LLM)-powered tool that assists bystanders in cyberbullying intervention. We then evaluated its impact through

Authors' Contact Information: Peinuan Qin, e1322754@u.nus.edu, School of Computing, National University of Singapore, Singapore, Singapore; Junti Zhang, juntizhang@u.nus.edu, Institute of Data Science, National University of Singapore, Singapore, Singapore; Jiting Cheng, e1237249@u.nus.edu, School of Computing, National University of Singapore, Singapore, Singapore; Jungup Lee, swklj@nus.edu.sg, Department of Social Work, National University of Singapore, Singapore, Singapore; Zhixing Liu, e1101808@u.nus.edu, School of Computing, National University of Singapore, Singapore, Singapore; Yi-Chieh Lee, yclee@nus.edu.sg, Computer Science, National University of Singapore, Singapore, Singapore.

Manuscript submitted to ACM

a mixed-methods experiment, focusing on the following research questions: **RQ1: How does EmojiGen influence the frequency of bystander direct intervention? RQ2: How does EmojiGen impact bystanders' perceptions of cyberbullying intervention? RQ3: How do changes in bystanders' perceptions mediate the effect of EmojiGen on their direct intervention behavior in cyberbullying situations?**

The results show that EmojiGen significantly increased direct intervention, improved self-efficacy and perceived ability to help, and reduced effort and anxiety. Further analysis revealed that its impact on resisting perpetrators was primarily mediated by enhanced self-efficacy and intervention knowledge.

2 Related Work

Cyberbullying and Bystander Intervention. Cyberbullying refers to repeated online attacks against individuals struggling to defend themselves [7, 19, 29]. It poses serious risks, particularly for teenagers, leading to anxiety, depression, and low self-esteem [26, 28, 44]. Bystanders play a crucial role in mitigating cyberbullying [30], with interventions categorized as indirect (e.g., reporting content) and direct (e.g., confronting bullies or supporting victims) [16, 31]. While direct intervention effectively halts bullying and supports victims, it requires effort and emotional investment, resulting in low engagement [23, 41]. The bystander effect, where individuals assume that others will intervene, further discourages action [25]. These challenges highlight the need for strategies to empower bystanders to take direct action.

Encouraging Bystander Intervention. The Bystander Intervention Model (BIM) [10] outlines five steps: observing, interpreting, assuming responsibility, possessing skills, and acting. Prior work has encouraged indirect intervention by enhancing responsibility (e.g., audience awareness) [14] and empathy (e.g., emotional design) [46]. Educational interventions and chatbot-based simulations have also been explored [19, 20]. However, these methods provide limited support for direct intervention, often failing to address the complex, real-time nature of cyberbullying [14, 46]. There remains a gap in equipping bystanders with immediate and adaptive intervention support.

LLMs for Direct Cyberbullying Intervention. Bystanders often struggle with how to intervene, despite recognizing the need to act [10, 37]. Lack of confidence and knowledge leads to anxiety and avoidance, whereas self-efficacy fosters proactive behavior [4, 5]. Recent LLM advancements offer potential solutions by providing real-time guidance, intervention knowledge, and anxiety reduction [17, 27, 39]. LLM-driven tools have enhanced peer-to-peer mental health support [45] and provided personalized emergency guidance [39]. However, their application in direct cyberbullying intervention remains underexplored, necessitating further empirical research.

3 Methods

To inform intervention design, we surveyed 67 social media users to identify barriers to direct intervention. Participants preferred indirect responses, citing (1) effort in composing comments and (2) lack of confidence as primary deterrents. To address these, we developed EmojiGen (see Figure 1(e)), an LLM-driven tool that supports *Emoji Selection* – Users express emotional intent via emojis, reducing textual formulation complexity [8, 18]. *Comment Generation* – The LLM (GPT-4o) generates contextually relevant, positively framed intervention comments [1, 35].

Experimental Platform and Materials. To ensure ecological validity, we developed SnapShare, a simulated social media platform, aligning with prior cyberbullying intervention studies [14]. A pilot test with 15 users assessed usability and credibility, leading to refinements. We designed nine cyberbullying posts covering three common topics: appearance, gender, and race. Each post, adapted from real cases, contained 1–6 anonymized images and one cyberbullying comment categorized as trolling, harassment or flaming. To mimic real engagement, each post also mixed 1–5 non-cyberbullying comments from actual social media discussions. To validate realism, 27 participants rated the authenticity of posts and

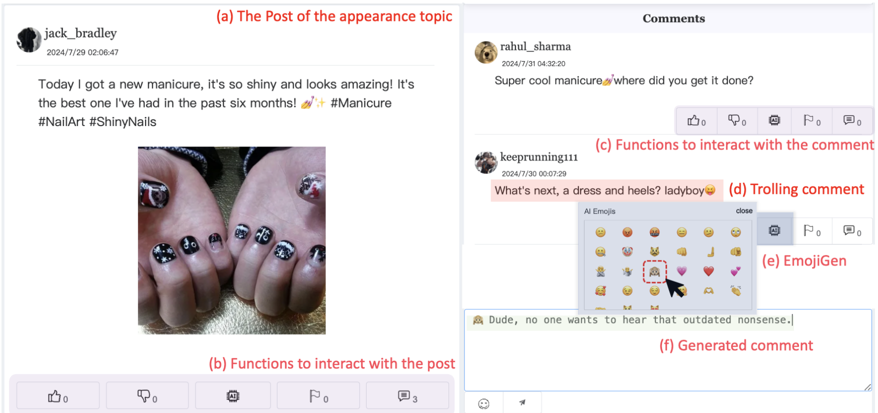

Fig. 1. Screenshot of SnapShare with EmojiGen. (a) A post on the appearance topic. (b) Interaction buttons: like, dislike, EmojiGen, flag, and comment. (c) Comment interaction buttons with the same options. (d) A trolling cyberbullying comment. (e) The participant selects 🙈 in the EmojiGen panel. (f) The system generates a resisting response: "🙈 *Dude, no one wants to hear that outdated nonsense.*"

cyberbullying comments on a 5-point Likert scale. The mean realism ratings for the post-topics were: appearance: $M = 4.148, SD = 0.602$, gender: $M = 4.259, SD = 0.507$, race: $M = 4.241, SD = 0.594$; while the realism score for cyberbullying comments were: harassment: $M = 4.047, SD = 0.442$, trolling: $M = 4.133, SD = 0.601$, flaming: $M = 4.080, SD = 0.512$. No significant differences in realism perception were found across topics ($F(2, 78) = 0.295, p = 0.745$) or categories ($F(2, 78) = 0.175, p = 0.840$).

Table 1. Experimental arrangement of posts and cyberbullying comments over three days. Each participant joined for one day. A1, A2, and A3 represent three different posts under the **A**ppearance topic, with similar representations for other topics. A1 (Trolling) means that the type of cyberbullying comment that occurs under the post-A1 is the type of Trolling.

Day (participants)	Topic Order ↓	Post Order (Comment Type) →		
Day 1 (15 C_{NE} / 15 C_{EG})	Appearance	A1 (Trolling)	A2 (Flaming)	A3 (Harassment)
	Gender	G1 (Trolling)	G2 (Flaming)	G3 (Harassment)
	Race	R1 (Trolling)	R2 (Flaming)	R3 (Harassment)
Day 2 (15 C_{NE} / 15 C_{EG})	Gender	G1 (Flaming)	G2 (Harassment)	G3 (Trolling)
	Race	R1 (Flaming)	R2 (Harassment)	R3 (Trolling)
	Appearance	A1 (Flaming)	A2 (Harassment)	A3 (Trolling)
Day 3 (15 C_{NE} / 15 C_{EG})	Race	R1 (Harassment)	R2 (Trolling)	R3 (Flaming)
	Appearance	A1 (Harassment)	A2 (Trolling)	A3 (Flaming)
	Gender	G1 (Harassment)	G2 (Trolling)	G3 (Flaming)

Experimental Setup. We conducted a between-subjects experiment with 90 participants on SnapShare. All participants were selected because they reported that they would ignore (N = 32) or intervene in cyberbullying indirectly (N = 58) in daily life. Participants engaged with the platform for 10 minutes before completing 7-Likert surveys measuring *direct intervention frequency*, perceived *knowing how to help* ($\alpha = 0.873$), *defending self-efficacy* ($\alpha = 0.795$), *communication self-efficacy* ($\alpha = 0.891$) *responsibility* ($\alpha = 0.832$), perceived *workload* ($\alpha = 0.870$) and *anxiety* ($\alpha = 0.871$). Open-ended questions are also used to collect their subjective experience for qualitative insights. To mitigate sequence effects, we controlled the order in which participants encountered posts and the associated cyberbullying comment types. Each participant was randomly assigned to a single session across three days, where topics and comment categories were systematically rotated. Table 1 summarizes the experimental arrangement.

4 Results

For clarity, we denote the EmojiGen group as C_{EG} and the non-usage group as C_{NE}. **EmojiGen Promotes Direct Interventions (RQ1).** T-tests revealed that C_{EG} participants performed significantly more support interventions ($M = 4.896, SD = 2.897$) than C_{NE} ($M = 2.116, SD = 2.342$) ($t(88) = 4.996, p < 0.001$). One-way ANOVA indicated that race-related posts elicited significantly more support ($p = 0.011$). For resisting interventions, C_{EG} participants ($M = 2.125, SD = 3.085$) intervened significantly more than C_{NE} ($M = 0.279, SD = 0.734$) ($t(52.898) = 4.021, p < 0.001$). However, intervention frequency did not significantly differ ($F(2, 87) = 0.621, p = 0.54$) across cyberbullying types (flaming, trolling, harassment). **EmojiGen's Impact on Bystander Perceptions (RQ2).** EmojiGen significantly increased bystanders' defending self-efficacy ($p = 0.002$) and perceptions of knowing how to help ($p < 0.001$). It also reduced bystanders' workload ($p = 0.047$) and anxiety ($p = 0.008$) in direct cyberbullying intervention but had no significant impact on bystanders' personal responsibility ($p = 0.131$) or communication self-efficacy ($p = 0.811$). **How EmojiGen Facilitates Direct Intervention (RQ3).** A Structural Equation Model (SEM) identified two pathways: (1) Increased perception of knowing how to help → increased defending self-efficacy → increased resisting interventions (2) Increased perception of knowing of how to help → reduced anxiety (though anxiety reduction did not significantly impact intervention frequency). The model fit was strong ($CFI = 1.000, TLI = 1.021, RMSEA = 0.000$), confirming these mediation effects. **Qualitative Insights.** Participants valued EmojiGen's supportive tone and cognitive effort reduction, increasing engagement willingness. However, some noted its generic phrasing, lack of nuance, overly positive tone, and inability to convey assertiveness when resisting bullies.

5 Discussion

Encouraging direct bystander intervention in cyberbullying is crucial for safeguarding teens' mental well-being, underscoring the need for innovative solutions. EmojiGen extends traditional intervention strategies by simplifying bystander participation. Unlike nudging or educational approaches [20, 32, 46], EmojiGen is an attempt to provide flexible and just-in-time adaptive interventions (JITAI) [36]. The increase in direct interventions across different cyberbullying contexts confirms its effectiveness in lowering barriers to engagement. Moreover, by reinforcing defending self-efficacy — a key determinant of intervention behavior [9, 42] — EmojiGen sets a precedent for AI-driven interventions.

Despite its effectiveness, EmojiGen's lack of personalized responses limited its impact. Some users modified or rejected AI-generated comments, particularly when resisting perpetrators, as emotionally strong responses are often necessary for effective confrontation [6, 13]. Similarly, supportive comments must be authentic and empathetic to provide meaningful reassurance to victims [11, 43]. The generic tone of AI-generated content may weaken its persuasive power and emotional resonance. A potential solution is a retrieval-augmented generation (RAG) [49] system, which can integrate users' past responses for more personalized responses.

While EmojiGen lowers the threshold for intervention, its lack of impact on perceived responsibility raises concerns about external nudging replacing intrinsic motivation [21]. AI-assisted interventions risk diffusing responsibility [34, 48], reducing user ownership [40], and leading to disengagement when AI is absent. Even that, AI's ability to increase intervention frequency may still have a deterrent effect on perpetrators and contribute to broader online behavior regulation [3, 12, 24, 33]. Given cyberbullying can cause lasting developmental harm to adolescents and our work has primarily provided implications from the bystanders' aspect, future studies should explore how generative AI can support victims, particularly in teen populations, and investigate how they perceive AI-assisted interventions in these contexts.

References

[1] Hassan Ali, Philipp Allgeuer, and Stefan Wermter. 2024. Comparing Apples to Oranges: LLM-powered Multimodal Intention Prediction in an Object Categorization Task. *arXiv preprint arXiv:2404.08424* (2024).

[2] Kimberley R Allison and Kay Bussey. 2016. Cyber-bystanding in context: A review of the literature on witnesses' responses to cyberbullying. *Children and Youth Services Review* 65 (2016), 183–194.

[3] Jenn Anderson, Mary Bresnahan, and Catherine Musatics. 2014. Combating weight-based cyberbullying on Facebook with the dissenter effect. *Cyberpsychology, Behavior, and Social Networking* 17, 5 (2014), 281–286.

[4] Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84, 2 (1977), 191.

[5] Albert Bandura and Sebastian Wessels. 1994. Self-efficacy. (1994).

[6] K Alex Burton, Dan Florell, and Jonathan S Gore. 2013. Differences in proactive and reactive aggression in traditional bullies and cyberbullies. *Journal of Aggression, Maltreatment & Trauma* 22, 3 (2013), 316–328.

[7] Angela Busacca and Melchiorre Alberto Monaca. 2023. Deepfake: Creation, Purpose, Risks. In *Innovations and Economic and Social Changes due to Artificial Intelligence: The State of the Art*. Springer, 55–68.

[8] Charles Chiang and Diego Gomez-Zara. 2024. The Evolution of Emojis for Sharing Emotions: A Systematic Review of the HCI Literature. *arXiv preprint arXiv:2409.17322* (2024).

[9] Madeleine Clark and Kay Bussey. 2020. The role of self-efficacy in defending cyberbullying victims. *Computers in Human Behavior* 109 (2020), 106340.

[10] John M Darley and Bibb Latané. 1968. Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology* 8, 4p1 (1968), 377.

[11] Pooja Datta, Dewey Cornell, and Francis Huang. 2016. Aggressive attitudes and prevalence of bullying bystander behavior in middle school. *Psychology in the Schools* 53, 8 (2016), 804–816.

[12] Anna Davidovic, Catherine Talbot, Catherine Hamilton-Giachritsis, and Adam Joinson. 2023. To intervene or not to intervene: young adults' views on when and how to intervene in online harassment. *Journal of Computer-Mediated Communication* 28, 5 (2023), zmad027.

[13] Ann DeSmet, Sara Bastiaensens, Katrien Van Cleemput, Karolien Poels, Heidi Vandebosch, and Ilse De Bourdeaudhuij. 2012. Mobilizing bystanders of cyberbullying: An exploratory study into behavioural determinants of defending the victim. *Annual Review of Cybertherapy and Telemedicine 2012* (2012), 58–63.

[14] Dominic DiFranzo, Samuel Hardman Taylor, Franccesca Kazerooni, Olivia D Wherry, and Natalya N Bazarova. 2018. Upstanding by design: Bystander intervention in cyberbullying. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[15] Kelly P Dillon and Brad J Bushman. 2015. Unresponsive or un-noticed?: Cyberbystander intervention in an experimental cyberbullying context. *Computers in Human Behavior* 45 (2015), 144–150.

[16] Sara Erreygers, Sara Pabian, Heidi Vandebosch, and Elfi Baillien. 2016. Helping behavior among adolescent bystanders of cyberbullying: The role of impulsivity. *Learning and Individual Differences* 48 (2016), 61–67.

[17] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. From Text to Self: Users' Perception of AIMC Tools on Interpersonal Communication and Self. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[18] Rebecca Godard and Susan Holtzman. 2022. The multidimensional lexicon of emojis: A new tool to assess the emotional content of emojis. *Frontiers in Psychology* 13 (2022), 921388.

[19] Michael A Hedderich, Natalie N Bazarova, Wenting Zou, Ryun Shim, Xinda Ma, and Qian Yang. 2024. A Piece of Theatre: Investigating How Teachers Design LLM Chatbots to Assist Adolescent Cyberbullying Education. *arXiv preprint arXiv:2402.17456* (2024).

[20] Rose Hennessy Garza, Young Cho, Heather Hlavka, Lance Weinhardt, Tajammal Yasin, Sara Smith, Katharine Adler, Kacie Otto, and Paul Florsheim. 2023. A multi-topic bystander intervention program for upper-level undergraduate students: outcomes in sexual violence, racism, and high-risk alcohol situations. *Journal of interpersonal violence* 38, 15-16 (2023), 9395–9422.

[21] Frederick Herzberg, Bernard Mausner, and Barbara Bloch Snyderman. 2011. *The motivation to work*. Vol. 1. Transaction publishers.

[22] Yanru Jia, Yuntena Wu, Tonglin Jin, and Lu Zhang. 2022. How are bystanders involved in cyberbullying? A latent class analysis of the Cyberbystander and their characteristics in different intervention stages. *International journal of environmental research and public health* 19, 23 (2022), 16083.

[23] Franccesca Kazerooni, Samuel Hardman Taylor, Natalya N Bazarova, and Janis Whitlock. 2018. Cyberbullying bystander intervention: The number of offenders and retweeting predict likelihood of helping a cyberbullying victim. *Journal of Computer-Mediated Communication* 23, 3 (2018), 146–162.

[24] Sara Kiesler, Robert Kraut, Paul Resnick, and Aniket Kittur. 2012. Regulating behavior in online communities. *Building successful online communities: Evidence-based social design* 1 (2012), 4–2.

[25] Bibb Latané and John M Darley. 1970. The unresponsive bystander: Why doesn't he help? *(No Title)* (1970).

[26] Jungup Lee, JongSerl Chun, Jinyung Kim, Jieun Lee, and Serim Lee. 2021. A social-ecological approach to understanding the relationship between cyberbullying victimization and suicidal ideation in South Korean adolescents: The moderating effect of school connectedness. *International Journal of Environmental Research and Public Health* 18, 20 (2021), 10623.

[27] Yen-Fen Lee, Gwo-Jen Hwang, and Pei-Ying Chen. 2022. Impacts of an AI-based cha bot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational technology research and development* 70, 5 (2022), 1843–1865.

[28] Amanda Lenhart, Mary Madden, Aaron Smith, Kristen Purcell, Kathryn Zickuhr, and Lee Rainie. 2011. Teens, Kindness and Cruelty on Social Network Sites: How American Teens Navigate the New World of" Digital Citizenship". *Pew Internet & American Life Project* (2011).

[29] Qing Li. 2007. Bullying in the new playground: Research into cyberbullying and cyber victimisation. *Australasian Journal of Educational Technology* 23, 4 (2007).

[30] D Lynn Hawkins, Debra J Pepler, and Wendy M Craig. 2001. Naturalistic observations of peer interventions in bullying. *Social development* 10, 4 (2001), 512–527.

[31] Robert D Lytle, Tabrina M Bratton, and Heather K Hudson. 2021. Bystander apathy and intervention in the era of social media. In *The emerald international handbook of technology-facilitated violence and abuse*. Emerald Publishing Limited, 711–728.

[32] Jiayue Mao. 2022. The Role of Nudges in Mitigating and Preventing Cyberbullying on Social Media. In *2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022)*. Atlantis Press, 1404–1408.

[33] Alice E Marwick. 2021. Morally motivated networked harassment as normative reinforcement. *Social Media+ Society* 7, 2 (2021), 20563051211021378.

[34] Mareike Möhlmann, Lior Zalmanson, Ola Henfridsson, and Robert Wayne Gregory. 2021. Algorithmic management of work on online labor platforms: When matching meets control. *MIS quarterly* 45, 4 (2021).

[35] Alberto Muñoz-Ortiz, Carlos Gómez-Rodríguez, and David Vilares. 2023. Contrasting linguistic patterns in human and llm-generated text. *arXiv preprint arXiv:2308.09067* (2023).

[36] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* (2018), 1–17.

[37] Amanda B Nickerson, Ariel M Aloe, Jennifer A Livingston, and Thomas Hugh Feeley. 2014. Measurement of the bystander intervention model for bullying and sexual harassment. *Journal of adolescence* 37, 4 (2014), 391–400.

[38] Magdalena Obermaier, Nayla Fawzi, and Thomas Koch. 2016. Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. *New media & society* 18, 8 (2016), 1491–1507.

[39] Hakan T Otal, Eric Stern, and M Abdullah Canbaz. 2024. Llm-assisted crisis management: Building advanced llm platforms for effective emergency response and public collaboration. In *2024 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 851–859.

[40] Jon L Pierce, Tatiana Kostova, and Kurt T Dirks. 2001. Toward a theory of psychological ownership in organizations. *Academy of management review* 26, 2 (2001), 298–310.

[41] Joshua R Polanin, Dorothy L Espelage, and Therese D Pigott. 2012. A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *School psychology review* 41, 1 (2012), 47–65.

[42] Virpi Pöyhönen, Jaana Juvonen, and Christina Salmivalli. 2010. What does it take to stand up for the victim of bullying? The interplay between personal and social factors. *Merrill-Palmer Quarterly (1982-)* (2010), 143–163.

[43] Jeroen Pronk, Tjeert Olthof, Frits A Goossens, and Lydia Krabbendam. 2019. Differences in adolescents' motivations for indirect, direct, and hybrid peer defending. *Social Development* 28, 2 (2019), 414–429.

[44] Anurag Sarkar, Shalabh Agarwal, Abir Ghosh, and Asoke Nath. 2015. Impacts of social networks: A comprehensive study on positive and negative effects on different age groups in a society. *International Journal* 3, 5 (2015), 177–190.

[45] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023), 46–57.

[46] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N Bazarova. 2019. Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.

[47] Elizabeth Whittaker and Robin M Kowalski. 2015. Cyberbullying via social media. *Journal of school violence* 14, 1 (2015), 11–29.

[48] Beibei Yue and Hu Li. 2023. The impact of human-AI collaboration types on consumer evaluation and usage intention: a perspective of responsibility attribution. *Frontiers in psychology* 14 (2023), 1277861.

[49] Ruichen Zhang, Hongyang Du, Yinqiu Liu, Dusit Niyato, Jiawen Kang, Sumei Sun, Xuemin Shen, and H Vincent Poor. 2024. Interactive AI with retrieval-augmented generation for next generation networking. *IEEE Network* (2024).

[50] Chengyan Zhu, Shiqing Huang, Richard Evans, and Wei Zhang. 2021. Cyberbullying among adolescents and children: a comprehensive review of the global situation, risk factors, and preventive measures. *Frontiers in public health* 9 (2021), 634909.